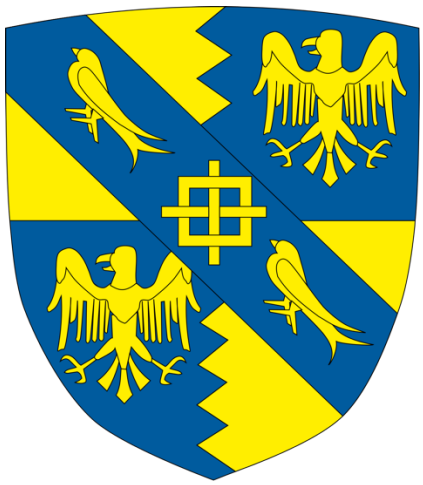
Modelling prognostic trajectories in Alzheimer's disease



Joseph Julian Giorgio

Magdalene College, University of Cambridge

25th of May 2020

This thesis is submitted for the degree of Doctor of Philosophy

Disclosures

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

Modelling prognostic trajectories in Alzheimer's disease.

Joseph Giorgio

Abstract

Progression to dementia due to Alzheimer's Disease (AD) is a long and protracted process that involves multiple pathways of disease pathophysiology. Predicting these dynamic changes has major implications for timely and effective clinical management in AD. There are two reasons why at present we lack appropriate tools to make such predictions. First, a key feature of AD is the interactive nature of the relationships between biomarkers, such as accumulation of β-amyloid -a peptide that builds plaques between nerve cells-, tau -a protein found in the axons of nerve cells- and widespread neurodegeneration. Current models fail to capture these relationships because they are unable to successfully reduce the high dimensionality of biomarkers while exploiting informative multivariate relationships. Second, current models focus on simply predicting in a binary manner whether an individual will develop dementia due to AD or not, without informing clinicians about their predicted disease trajectory. This can result in administering inefficient treatment plans and hindering appropriate stratification for clinical trials. In this thesis, we overcome these challenges by using applied machine learning to build predictive models of patient disease trajectories in the earliest stages of AD. Specifically, to exploit the multi-dimensionality of biomarker data, we used a novel feature generation methodology Partial Least Squares regression with recursive feature elimination (PLSr-RFE). This method applies a hybrid-feature selection and feature construction method that captures co-morbidities in cognition and pathophysiology, resulting in an index of Alzheimer's disease atrophy from structural MRI. We validated our choice of biomarker and the efficacy of our methodology by showing that the learnt pattern of grey matter atrophy is highly predictive of tau accumulation in an independent sample. Next, to go beyond predicting binary outcomes to deriving individualised prognostic scores of cognitive decline due to AD, we used a novel trajectory modelling approach (Generalised Metric Learning Vector Quantization – Scalar projection) that mines multimodal data from large AD research cohorts. Using this approach, we derive individualised prognostic scores of cognitive decline due to AD, revealing interactive cognitive, and biological factors that improve prediction accuracy. Next, we extended our machine learning framework to classify and stage early AD individuals

based on future pathological tau accumulation. Our results show that the characteristic spreading pattern of tau in early AD can be predicted by baseline biomarkers, particularly when stratifying groups using multimodal data. Further, we showed that our prognostic index predicts individualised rates of future tau accumulation with high accuracy and regional specificity in an independent sample of cognitively unimpaired individuals. Overall, our work used machine learning to combine continuous information from AD biomarkers predicting pathophysiological changes at different stages in the AD cascade. The approaches presented in this thesis provide an excellent framework to support personalised clinical interventions and guide effective drug discovery trials.

Acknowledgements

The work presented in this thesis was funded by the Global Alliance, Alan Turing Institute and the EU Seventh Framework Programme. Beyond these funding bodies this work would not have been possible without the large longitudinal data bases funded by a multitude of public, private and philanthropic funders.

I am very thankful for the continued support of my supervisor, Professor Zoe Kourtzi. Without Zoe's determination to support this research, provide crucial feedback and allow me the freedom to explore so many facets of computational neuroscience this body of work would not have been delivered. Along with all the members of the Adaptive Brain Lab, the environment that this work was undertook allowed for state-of-the-art application of approaches, expertise for critical appraisal, but most importantly a friendship group that fosters excellent scientific work.

The direction and driving force of this thesis is one of clinical relevance, this would not have been possible without the expertise, guidance and exceptional collaborative spirit of Professor Bill Jagust and the members of his UC Berkeley lab; Dr Susan Landau and Dr Suzanne Baker.

In parallel to the clinical aspects of this work, efforts to exploit relevant computational approaches were a key motivation. These tools would have been useless to me without the time and care that Professor Peter Tino took to explain and guide me through the mathematical principles that underpin all the insights derived from this thesis.

Finally, I would like to thank all members in this research field. From the PIs that lay the direction, the researchers who painstakingly appraise data, the technical staff who collect the data, the assistants who collate the data and most importantly the participants who donate their data, big data research is impossible without you.

Table of Contents

List of figures, tables, graphical examples and algorithms

List of abbreviations

AD: Alzheimer’s disease
APP: amyloid precursor protein
EOAD: early onset Alzheimer’s disease
LOAD: late onset Alzheimer’s disease
ApoE: apolipoprotein E
PART: primary age-related tauopathy
MCI: mild cognitive impairment
PET: positron emission tomography
FTP: flortaucipir
SUVR: standardised uptake value ratio
FDG: Fluorodeoxyglucose
MRI: magnetic resonance imaging
SNAP: suspected non-Alzheimer pathophysiology
NIA-AA: National Institute on Aging - Alzheimer's Association
A: fibrillary β-amyloid plaques
T: neurofibrillary tangles -tau-
N: neurodegeneration/neuronal injury
CSF: cerebrospinal fluid
t-tau: total tau
p-tau: phosphorylated tau
CN: cognitively normal
sMCI: stable Mild cognitive Impairment
pMCI: progressive mild cognitive impairment
LDA: linear discriminant analysis
MMSE: mini-mental state examination
ADAS-cog: Alzheimer's disease assessment scale-cognitive subscale
PLSr: partial least squares regression
PLSr-RFE: partial least squares regression – recursive feature elimination
GMLVQ: generalised matrix learning vector quantisation
GMLVQ-Scalar projection: generalised matrix learning vector quantisation-scalar projection
DOF: degrees of freedom
ADNI: Alzheimer’s disease neuroimaging initiative
ADNI-Mem: Alzheimer’s disease neuroimaging initiative-memory
ADNI EF: Alzheimer’s disease neuroimaging initiative-executive function
3D MP-RAGE: three-dimensional (3D) magnetization-prepared rapid acquisition with gradient echo
IR-SPGR: inversion recovery spoiled gradient echo
VBM: voxel based morphometry
DARTEL: diffeomorphic anatomical registration through exponentiated lie algebra
ROIs: regions of interest
GM Grey Matter
FBP: florbetapir
3T: 3 tesla
mCi: millicuries
GDS: geriatric depression scale
MAE: macro averaged error
TP: true positive

TN: true negative
CI: confidence intervals
BACS: Berkeley aging cohort study
SD: standard deviation
PiB: Pittsburgh compound B
SC: stable condition
EAD: early Alzheimer’s disease
LBNL: Lawrence Berkeley national laboratory
1.5T: 1.5 tesla
CT: computed tomography
DVRs: distribution volume ratios
PACC: preclinical Alzheimer’s cognitive composite
ACEr: Addenbrooke's cognitive examination-revised
Non-ND: non-neurodegenerative disease
DTI: diffusion tensor imaging
SNPs: single nucleotide polymorphisms
fMRI: functional magnetic resonance imaging

1. Introduction

Preamble

Alzheimer's Disease

Alois Alzheimer in early November of 1906 gave a seminal lecture describing the neuropathological and clinical presentations of one of his patients, Auguste D. Alzheimer described the death of his patient as caused by the death of a dementing illness, describing clinical presentations of progressive dementia and the pathological presence of intracellular plaques and tangles (Alzheimer, 1907). This characteristic progressive dementia in the presence of plaques and tangles, termed Alzheimer's disease (AD), now accounts for up to 70% of the 50 million people living with dementia world-wide, at a global societal cost approaching 1 trillion US dollars (Patterson, 2018). However, these numbers are only a fraction of the projected diagnosis, with an anticipated 152 million cases by 2050 (Prince et al., 2015). The current societal and financial burden coupled with the anticipated increase in AD globally presents a critical need for treatment. Alarmingly, a pharmacological intervention still eludes the pharmaceutical industry with clinical trials failing to achieve FDA approval 99.6 percent of the time (Patterson, 2018). The long development time of 13 years with an average cost of 5.6 billion US dollars (compared to on average 793.6 Million for a cancer drug) (Cummings, Reiber, & Kumar, 2018) has driven pharmaceutical giants out of the race with Pfizer pulling out of AD research in 2018. The increasing prevalence and lack of viable treatment makes AD one of the most debilitating diseases in the first world. This is evident when comparing the relative change in cause of death in the US between 2000 and 2015: 5 of the 6 leading causes of death declined(1% for breast cancer, up to 55% for HIV), while death due to AD (6th) increased by 123% (Alzheimer's Association, 2018).

Since 1906, however, the biological underpinnings of AD have been heavily studied and increasingly clear hypotheses have been proposed as to the pathological cascade that ultimately leads to dementia due to AD. This foundational neuroscientific inquiry coupled with robust biological definitions of AD provide a framework to study AD. Further, with the ingress of more computationally complex tools in neuroscience, predictive multivariate models can be built to elucidate pathophysiological interactions and prognostic trajectories in AD. These

multivariate approaches are well poised to combine multimodal biomarkers to stratify individuals at the earliest stages of AD, increasing the likelihood of effective clinical interventions. When considered together, these three facets of AD research can inform clinical trials and guide personalised interventions, ultimately alleviating the estimated 342.000 US dollar cost for a person living with dementia (Alzheimer's Association, 2018).

Machine learning in clinical decision making

Machine learning is one such multivariate approach to integrate multimodal biomarkers to guide clinical decision-making. However, as of yet machine learning is yet to see widespread proliferation in clinical practice. Recently, deep learning has been lauded as the answer to this problem (Esteva et al., 2019). Deep learning attempts to model hidden interactions between input data with parameters tuned throughout the learning process (Lecun, Bengio, & Hinton, 2015). In a typical application to neuroimaging data this parameter search space will often exceed the sample size by many orders of magnitude. The most efficient designs still have 1000 times more parameters than samples (Spasov, Passamonti, Duggento, Liò, & Toschi, 2019). While deep learning has excelled at image recognition (Krizhevsky, Sutskever, & Hinton, 2012), access to labelled training data for image recognition dwarves labelled clinical data. Therefore, there is a drastic need to accurately validate these heavily parameterised approaches when making clinical decisions (Topol, 2019). Further, given the hidden architecture of deep learning models, internal decisions made by the model are largely blind; the machine learning practitioner often sees only a final output from the 'black box' architecture.

This black box approach of deep learning seems at odds with clinical decision making. Clinical diagnosis is not blindly made, rather, informed from well categorised and interpretable biomarkers. Blind appraisal of data by deep leaning models is not the only feature at odds with a clinical decision maker. Throughout the learning process a deep learning model will multiply weights, fit non-linearities and -in some cases- model recurrent processes (Lecun et al., 2015). However, when tasked with making a clinical diagnosis the clinician is reminded *"Entities should not be multiplied unnecessarily" (*William of Occam, c. 1320*).* As a pillar of clinical science, Occam's razor drives decision making, favouring diagnostic parsimony. These incongruities between a clinician and deep learning architectures (i.e. heavy parameterisation, poor interpretability and unnecessary complexity) question their current and potential clinical

utility in diagnosis. However, machine learning need not be so discordant with clinical ethos. In 1987 Blumer and colleagues showed when hypotheses of minimum complexity are applied to data in machine learning -so called Occam's learner- a model can achieve optimal predictive performance (Blumer, Ehrenfeucht, Haussler, & Warmuth, 1987).

For machine learning to become more accepted in clinical decision support, researchers must focus on developing models that are interpretable and generalisable without needless complexity. AD research is well poised to offer a course of inquiry where interpretable and robust machine learning can be deployed. Throughout the AD cascade there are well characterised pathophysiological processes. Under the recent biological definition of AD these measurable pathological processes can be reduced down to dynamic changes in β-amyloidosis, tau deposition, neurodegeneration and cognitive decline (Jack, Bennett, et al., 2018). Therefore, machine learning approaches that provide interpretable and intuitive predictions based within these categories of biomarker changes, can provide mechanistic insights into pathological processes while providing a clear and succinct clinical support tool.

Underlying pathology: the amyloid hypothesis

The amyloid hypothesis is the most widely accepted series of pathological changes that lead to dementia due to AD. The amyloid hypothesis first put forward in the 1980's, posits that the primary cause event of AD is the accumulation of β-amyloid plaques between neural cells (Glenner & Wong, 1984a). Similar work at the time showed that β-amyloid peptide, a biproduct of the amyloid precursor protein (APP), was the core chemical component of senile plaques (Masters et al., 1985) first proposed by Alois Alzheimer in 1906. These initial investigations showed that β-amyloid plaques found in patients with trisomy 21 (Down's Syndrome) were the same as senile plaques characteristic of AD (Glenner & Wong, 1984b; Masters et al., 1985). When considering the invariable development of AD for older Down's syndrome patients, this work suggested a genetic cause of AD. Subsequent work isolated a genetic deficit on chromosome 21 that leads to the accumulation of β-amyloid plaques (Goldgaber, Lerman, McBride, Saffiotti, & Gajdusek, 1987; Kang et al., 1987; Tanzi et al., 1987).These findings provided a mechanistic understanding of how genetic mutations can alter the function of APP, resulting in AD (Goate et al., 1991; Mullan et al., 1992).

Amyloid Precursor Protein (APP)

β-amyloid is a fragment of the transmembrane amyloid precursor protein (APP). When β- then γ-secretases cleave APP in the amyloidogenic pathway, neurotoxic $A\beta_{42}$ isoforms are formed. Due to the hydrophobic nature of $A\beta_{42,}$ these protein subunits self-aggregate to form β-amyloid fibrils and subsequent β-amyloid plaques (i.e. senile plaques) (**Figure 1.1)**. While this process in itself does not cause dementia, familiarly inherited genetic mutations to the presenilin 1,2 PSEN1, PSEN2) or BACE 1 gene invariably lead to Early Onset AD (EOAD). These mutations result in the favouring of proteolytic processing of APP by γ-secretase (PSEN1, PSEN2) or the adjacent β-secretase (BACE 1) subunit, which favours the production of the more fibrilogenic $A\beta_{42}$ (over $A\beta_{1\text{-}40}$) and subsequent β-amyloid plaques. Given the functional changes these mutations drive in APP processing, leading to accumulation of β-amyloid, and development of AD, these genetic studies provide strong evidence for the primary role of β-amyloid in AD (Hardy & Higgins, 1992; Selkoe, 1991). Additional support of the amyloid hypothesis comes from genetic studies investigating rare mutations on the 21st chromosome. These studies show that when the APP gene is duplicated individuals acquire AD in their 50s (Rovelet-Lecrux et al., 2006), while Downs syndrome individuals who only have the distal part of chromosome

21 replicated (not including the APP gene locus) do not get AD (Prasher et al., 1998). Similarly, several protective mutations to the APP gene show a decrease in β-amyloid accumulation and a lower risk or AD or age-related cognitive decline (Jonsson et al., 2012; Kero et al., 2013).

ApoE 4

While mutations to the APP gene are known to lead to EOAD, the major genetic risk factor for Late Onset AD (LOAD) is the epsilon 4 allele on the apolipoprotein E (ApoE) gene (Corder et al., 1993). The ApoE4 allele was identified as the major genetic risk factor for LOAD in the early 1990's, where it was observed that individuals expressing the ApoE4 allele develop more β-amyloid deposits than those expressing only ApoE3 (Rebeck, Reiter, Strickland, & Hyman, 1993). Subsequent work on transgenic mice has highlighted the influence of ApoE in the clearing of β-amyloid. Mice expressing the APP gene that are ApoE deficient show a significant decrease in β-amyloid accumulation, highlighting the role of ApoE in β-amyloid clearance not production (Castellano et al., 2011; Kim et al., 2013). Specifically, this work in transgenic mice showed that the epsilon 4 allele hindered β-amyloid clearance the greatest, with a decreased effect by ApoE4 > E3 > E2 (Castellano et al., 2011). These studies suggest a genetic link between β-amyloid dyshomeostasis (i.e. increased production and decreased clearance) and AD.

Figure 1.1 β-amyloid cascade

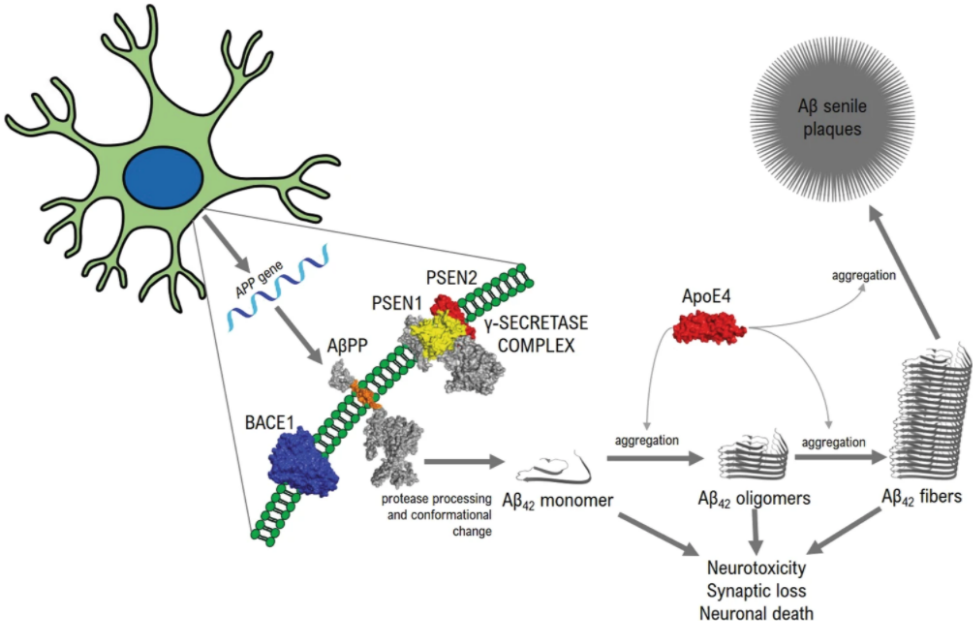


Figure 1.1 *adapted from (Castellani, Plascencia-Villa, & Perry, 2019) β-amyloid cascade hypothesis. β-amyloid precursor protein (APP) gene expression is followed by consecutive protease processing by β-secretase 1 (BACE1) and γ-secretase complex (PSEN1 (presenilin 1), PSEN2 (presenilin 2), releasing the β-amyloid peptide, including pathogenic species Aβ42. Aβ42 undergoes conformational change, assembles into oligomers and protofibrils, and ultimately deposits as senile plaques*

Tau

The second hallmark proteinopathy in AD is the formulation of neurofibrillary tangles. Tangles comprise filaments of the hyperphosphorylated forms of the microtubule binding protein tau (Grundke-Iqbal et al., 1986). These straight filaments twist around each other to form paired helical filaments initially appearing within the soma and dendritic processes of neurons. When these filaments aggregate, they combine into neuruofibrillary tangles and accumulate within dendrites, axons and cell bodies (Brion, 1998).

The amyloid hypothesis is not without contention, with critics suggesting that deposition of tau not β-amyloid is the primary event in AD (Heiko Braak & Del Tredici, 2011). Histological work investigating the deposition of cortical tau across the lifespan shows subcortical tau deposition within the first decade of life (Heiko Braak & Del Tredici, 2011). Subsequent studies show that the subcortical deposition of tau as well as deposits in the medial temporal

lobe are commonly found in cognitively healthy amyloid negative older people (Heiko Braak & Del Tredici, 2012). While studies show that medial temporal tau in the absence of β-amyloid can lead to poorer episodic memory and medial temporal atrophy, this is referred to as primary age-related tauopathy (PART) and is distinct from AD (Crary et al., 2014).

However, there are several lines of genetic evidence supporting β-amyloid as the primary driver of AD and not tau. When investigating the temporal order of events for individuals with the inherited mutations leading to EOAD and in Downs syndrome patients who acquire AD it has been shown that β-amyloid deposition precedes the accumulation of tangles (Bateman et al., 2012; C. A. Lemere et al., 1996; Cynthia A. Lemere et al., 1996). Further, when investigating genetic mutations in the gene encoding tau (FTDP-17), these mutations lead to non-AD tauopathies such as frontal-temporal dementia and parkinsonism (Hutton et al., 1998; Poorkaj et al., 1998; Spillantini, Van Swieten, & Goedert, 2000). These genetic studies provide evidence that supports the upstream role of β-amyloid relative to tau accumulation.

The AD Cascade

The hypothetical model of the AD cascade proposed by Jack and colleagues presents a schema of the dynamic pathophysiological changes that lead to AD-dementia (Jack et al., 2013; Jack, Knopman, et al., 2010) (**Figure 1.2)**. This model provides a sequence of pathological changes initiated by β-amyloid accumulation. The model extends the amyloid hypothesis, positing a series of pathological changes that are measurable in vivo. One critical aspect of the AD cascade is that while $A\beta_{42}$ dysregulation that results in β-amyloid plaques is necessary, it is not sufficient to lead to AD. This is an important distinction as β-amyloid deposition is regularly seen in cognitively normal older adults, with approximately 30% of individuals being β-amyloid positive by 80 (Jansen et al., 2015). The dynamic model of the AD pathological cascade shows upstream changes in β-amyloid and downstream neurodegeneration, with maximum rate of biomarker change moving from one pathophysiological domain to the next. The sequence of pathophysiological change has been widely supported in many international cohorts investigating in vivo changes in EOAD (Gordon et al., 2018).

Figure 1.2 The AD cascade

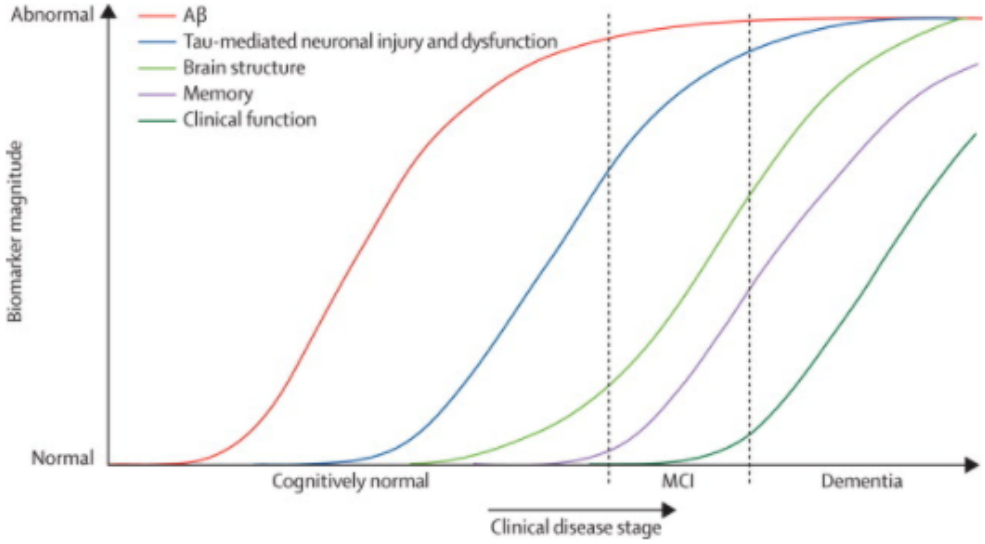


Figure 1.2 *adapted from* (Jack, Knopman, et al., 2010). *Dynamic biomarkers of the Alzheimer's pathological cascade: Aβ=β-amyloid. MCI=mild cognitive impairment.*

Dynamic Changes in β-amyloid Biomarkers Precedes Tau.

The widespread cortical deposition of β-amyloid initiates the AD cascade. This rising deposition of β-amyloid has been shown in vivo using PET, with β-amyloid levels associated with tau deposition in coming years (Leal, Lockhart, Maass, Bell, & Jagust, 2018; Tosun et al., 2017). Further, this deposition of β-amyloid precedes cognitive symptoms by up to 20 years and has a weak association with the cross-sectional severity of dementia symptoms (Furst et al., 2012; Klunk et al., 2004; Rowe et al., 2007). These studies support that β-amyloid is a prerequisite of both changes in tau and cognition in AD.

Tau Deposition.

The next key event s is tau deposition across the cortex. This deposition can be measured in vivo using flortaucipir PET and closely follows the characteristic spreading pattern of tau deposits across the cortex in AD, known as the Braak staging system (Schöll et al., 2016)(**Figure 1.3)**.

The Braak staging describes the characteristic spreading pattern of tau deposits across the cortex in AD (H. Braak & Braak, 1991). The spreading pattern shows that neurofibrillary tangles and neuropil threads spread across the cortex at 6 different stages. Stages I and II are localised to transentorhinal regions. Stages III and IV are the limbic stages and involve the

spreading from the transentorhinal regions into the entorhinal cortex with no effect on neocortical regions. Stages V and VI are the neocortical stages, this involves the spreading of tau from the limbic regions throughout the neocortex.

Figure 1.3 Braak staging system

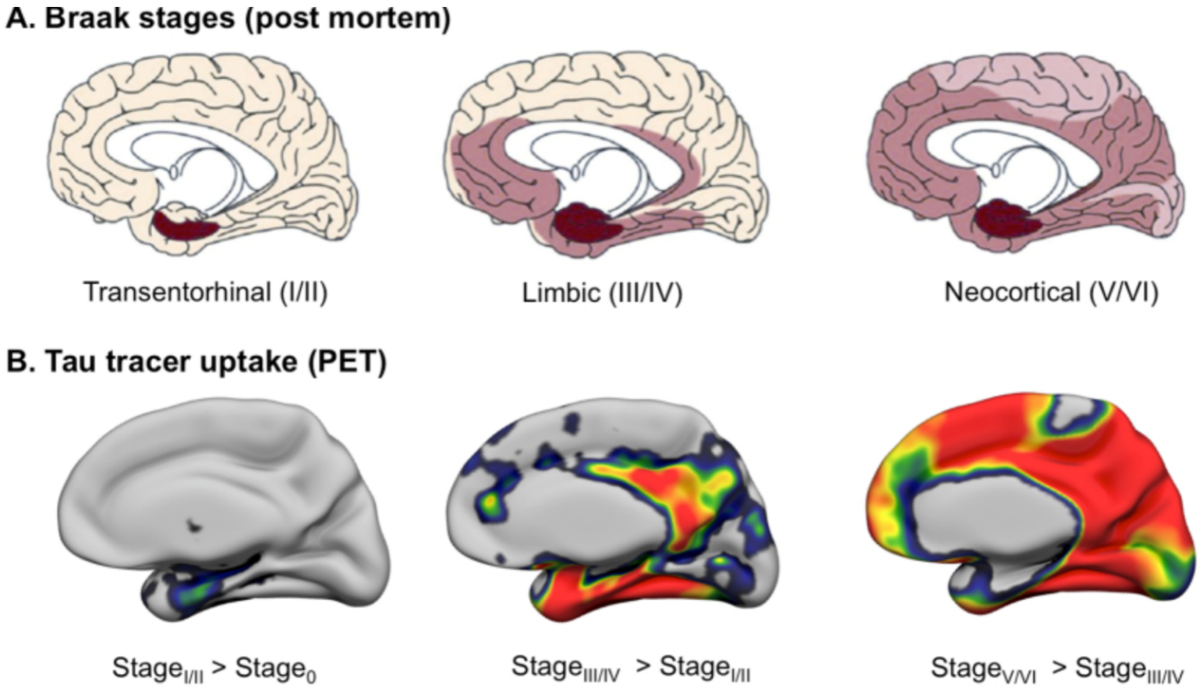


Figure 1.3 *adapted from (Schöll et al., 2019) FTP Tau tracer deposition patterns resemble post mortem Braak stages.* ***A****. post mortem Braak stages of tau deposition throughout Alzheimer's disease.* ***B****. Voxel-wise two-sample t-tests on FTP SUVR images between subjects assigned to tau-PET based Braak stages.*

This spreading of tau in the presence of amyloid has been closely linked to cross sectional and longitudinal cortical atrophy and hypometabolism (Bischof et al., 2016; L. Wang et al., 2016). Further, the deposition of tau is related to cognitive function in the presence of β-amyloid, however this association is weaker in the absence of β-amyloid (Brier et al., 2016; Cho et al., 2016; Johnson et al., 2016; Schwarz et al., 2016).

Cortical Atrophy and Hypometabolism.

Closely following the changes in tau deposition, a characteristic pattern of hypometabolism and atrophic changes occur (Bradford C. Dickerson et al., 2009; Furst et al., 2012; V. Singh et al., 2006). Hypometabolism as measured by FDG-PET follows a signature pattern initially occurring in the posteromedial and lateral temporal parietal cortex, then spreading to other

temporal regions (W. J. Jagust et al., 2009). Structural MRI measures of individuals who transition from mild cognitive impairment to dementia show a pattern of atrophy that parallels the pattern of tau deposition (Whitwell et al., 2007). Further, whole brain measures of tau are closely related to temporoparietal atrophy both longitudinally and cross-sectionally (LaPoint et al., 2017), suggesting that tau spreads prior to medial temporal atrophy. These AD typical patterns of atrophy and hypometabolism lead to an increased risk of future cognitive decline in cognitively healthy individuals, particularly in individuals with β-amyloid (De Leon et al., 2001; B. C. Dickerson et al., 2011; W. Jagust et al., 2006; Smith et al., 2007). However, similar patterns of atrophy can be observed in the absence of β-amyloid (Bakkour, Morris, Wolk, & Dickerson, 2013; Knopman et al., 2013) and this process of neurodegeneration is thought to be independent of AD, referred to as suspected non-Alzheimer pathophysiology (SNAP) (Jack, Knopman, et al., 2016). These studies highlight the lack of specificity of neurodegeneration, particularly in the medial temporal cortices when diagnosing AD.

The large body of work that supports this hypothetical model of the AD cascade makes it an ideal framework to characterise an individual with AD. That is, this series of dependent changes throughout the AD cascade provides an elegant framework to classify and stage an individual with AD.

Biological Definition of AD

Established by an international consortia of AD researchers, the 2018 NIA-AA Research Framework was developed to unify and guide AD researchers by formulating a biological definition of Alzheimer's disease (Jack, Bennett, et al., 2018).

The underlying principle of this framework is to provide a single classification scheme of AD across the disease spectrum. Given the reliance on biological markers and the relative infancy of the framework, the intent was not to guide clinical practice but to design longitudinal observational cohort studies and interventional studies. At the core of the research framework is the dissociation of syndromic definitions (i.e. defined by clinical labels) of AD that constituted the previous 2011 NIA-AA research framework (Albert et al., 2011; McKhann et al., 2011) from neuropathological changes that define a disease. However, the majority of ongoing longitudinal observational cohort studies were initiated prior to the introduction of the 2018 NIA-AA diagnostic framework with the clinical categorisation of participants following the 2011 NIA-AA syndromic definitions of AD (R. C. Petersen et al., 2010). Therefore, research approaches and techniques that are amenable to both syndromic and biological diagnostic criterion are well suited to capitalise on decades of longitudinal AD research.

Syndromic Definitions: Cognitively Normal, Mild Cognitive Impairment and Probable AD Dementia

Under the previous classification schemes the AD continuum is defined as the transition from a preclinical cognitively unimpaired state to a state of mild cognitive impairment (MCI) to dementia (Ronald C. Petersen et al., 2001). Crucially the guidelines note that there are no sharp demarcations that separate MCI from cognitively normal or demented and subjective clinical judgement is required to make the final diagnostic decision.

Mild Cognitive Impairment.

MCI due to AD is a four-stage diagnostic process (Albert et al., 2011). First, the patient should have a declining state of cognition, which can be determined by a patient, person close to the patient or a trained clinician. Second, the patient should perform below education and age matched controls in one or more cognitive domains (i.e. memory, executive function visuospatial skills). If repeated longitudinal measurements are taken, then a decline in performance should be evident. Third, the patient should have only mild problems performing

everyday tasks and is able to function independently. Finally, the patient must not be demented; that is the symptoms are mild enough to not impair social or occupational functioning.

Probable AD Dementia.

The 2011 NIA-AA diagnoses probable AD dementia as a distinct subtype of all cause dementia, having all clinical presentations of all cause dementia in addition to several specific criterion (McKhann et al., 2011). All dementias require declining cognitive symptoms that interfere with the ability to work or maintain usual activities. These symptoms must not be explained by delirium or neuropsychiatric conditions. To be diagnosed as probable AD dementia the declining cognitive symptoms must include a minimum of two of the following: Impaired ability to acquire and remember new information, impaired reasoning and handling of complex tasks, impaired visuospatial abilities, impaired language functions and changes in personality. In addition to these symptoms a patient with probable AD dementia must have had an insidious onset of these symptoms as well as longitudinal data documenting the worsening of cognition. Finally, the diagnosis should not be applied when a patient has concomitant substantial cerebrovascular disease or features of other dementia causing illnesses (i.e dementia with Lewy bodies, Frontotemporal dementia, primary progressive aphasia).

Sensitivity and Specificity of Syndromic Definitions.

While these transitions take place in AD, they are not specific to AD. Similarly, the most common syndromic definition of AD dementia, multi-domain amnestic dementia, is not sensitive to atypical AD variants such as dysexecutive, visuospatial or language dementias (Murray et al., 2011; Ossenkoppele et al., 2015). Nor is multi-domain amnestic dementia specific to AD pathological change (Nelson et al., 2011; Serrano-Pozo et al., 2014). These inherent syndromic definitions therefore are unable to provide a sensitive and specific definition of AD. In light of this, a single biological framework to define AD across the Alzheimer's continuum is preferable.

A/T/N Classification Framework: A Biological Definition of AD

The ATN classification framework was designed to be an unbiased classification scheme that codifies biomarkers commonly used in AD research. The scheme initially introduced by Jack et al (Jack, Bennett, et al., 2016) provides a framework to classify individuals based on three categories of AD biomarkers that measure the underlying AD pathophysiology: fibrillary β-

amyloid plaques (A), neurofibrillary tangles -tau- (T) and neurodegeneration/neuronal injury (N). This framework is appealing as it provides an objective classification framework across the AD spectrum.

Amyloid Beta Plaques: A.

As discussed previously, the hallmark pathological process in AD is the accumulation of β-amyloid plaques. The presence of β-amyloid deposition can be measured in-vivo either with an increased uptake of β-amyloid PET tracers, or subthreshold levels of $A\beta_{42}$ in the CSF, sampled via lumbar puncture. Biomarkers of β-amyloidosis are well targeted to the pathophysiological process in AD (Ikonomovic et al., 2008; Strozyk, Blennow, White, & Launer, 2003), although not specific to the disease. For example, following traumatic brain injuries it has been shown that there is a significant uptake in β-amyloid PET traces (Hong et al., 2014). Further, subthreshold levels of CSF $A\beta_{42}$ have been reported due to atrophic (Holmberg, Johnels, Blennow, & Rosengren, 2003) or inflammatory responses (Krut et al., 2013) un-related to AD pathophysiology. This lack of specificity can limit the diagnostic usefulness of this marker when measured in isolation (i.e. only diagnosing based on a β-amyloid positive biomarker).

Fibrillar Tau: T.

The wide-spread accumulation of fibrillar tau is the second stage of the amyloid cascade hypothesis, however tau accumulation is also a hallmark of other primary tauopathies (i.e. frontotemporal dementia, progressive supranuclear palsy or cortical-basal syndrome) (B. Hall et al., 2017; Sander et al., 2016; Schöll et al., 2019). Several tau biomarkers do, however, have reasonably high specificity to AD. Using lumbar punctures to extract CSF markers, paired increases of t-tau (total-tau) and p-tau (phosphorylated tau) are indicative of AD (K. Blennow et al., 1995). Further, these CSF markers are not present in non-AD primary tauopathies (S. Hall et al., 2012), with p-tau correlating closely with the severity of AD tau pathology postmortem (Tapiola et al., 2009). However, elevated levels of t-tau do occur as a result of neuronal damage either due to traumatic brain injuries or underlying neuronal disease (Öst et al., 2006; T. et al., 2014). When taken together the two markers give an indication of the intensity of neurodegeneration (t-tau) and the accumulation of tau (p-tau) due to AD (Kaj Blennow & Hampel, 2003). With the introduction of the flortaucipir tau-PET ligands, the uptake and topology of fibrillar tau can be measured across the AD spectrum. tau-PET ligands have a reasonably high specificity to paired helical filament tau found in AD but a much lower

binding affinity to non-AD tauopathies with straight filament tangles (Marquié et al., 2015). Finally, the spatial topography of tau accumulation follows closely the Braak staging of AD related tau at autopsy (Schöll et al., 2016, 2019). The ability to accurately measure the pathological spread of tau in-vivo makes it a good candidate biomarker for biological diagnosis, particularly if an individual also has a positive β-amyloid biomarker.

Neurodegeneration / Neuronal Injury: N.

The final, but least specific, pathophysiological process in the amyloid cascade is neurodegeneration (Bakkour et al., 2013). Neurodegeneration / neuronal injury can be measured indirectly from t-tau levels in the CSF. Neuroimaging provides a more direct way to measure neurodegeneration / neuronal injury with FDG-PET used to measure hypometabolism and structural MRI to measure atrophy. However, abnormal levels of atrophy or hypometabolism are not specific to AD, even if there are concordant measures of AD-like signatures on both FGD-PET and structural MRI (i.e. SNAP) (Burnham, Bourgeat, Doré, Savage, Brown, Laws, Maruff, Salvado, Ames, Martins, Masters, Rowe, & Villemagne, 2016; Jack, Knopman, et al., 2016; Jack et al., 2012). While these neuroimaging markers are non-specific, they both have a topology that is characteristic of typical AD (i.e. generally resulting in amnestic multidomain dementia). Atrophy measured using structural MRI shows atrophic patterns that initially occur within the medial temporal cortex extending to the basal lateral temporal cortex (Jack & Holtzman, 2013), whereas, FDG-PET hypometabolism initially occurs in the posteromedial and lateral temporal parietal cortex (W. J. Jagust et al., 2009).

Defining AD With The ATN Classification Scheme: The Alzheimer's Continuum.

The 2018 NIA-AA framework establishes a set of definitions of AD without the need of clinical diagnosis (**Table 1.1)**. As β-amyloidosis is a specific pathophysiological process to AD, an individual with a positive biomarker for β-amyloid (notated A+) is defined as *AD continuum*. While non-specific to AD, tau is the second proteinopathy that is a hallmark of AD. Therefore, an individual with a positive biomarker for amyloid (A+) and a positive biomarker for tau (notated T+) is defined as having *Alzheimer's Disease* (i.e. A+T+). An individual with a positive biomarker for β-amyloid (A+) but a negative biomarker for tau (T-), is at an earlier stage of the AD continuum compared to an individual with A+T+ and is defined as presenting *Alzheimer's pathological change* (i.e. A+T-).

As a non-specific biomarker for AD, neurodegeneration / neuronal injury is not used to define AD within the NIA-AA framework. However, the presence of a positive biomarker for

neurodegeneration / neuronal injury (notated as N+) is used to stage an individual with AD. As neurodegeneration / neuronal injury is the last stage of the amyloid cascade, an individual with the biomarker profile A+T+N+ is in a more advanced stage of AD compared to an individual with A+T+N-.

Table 1.1 ATN classification scheme

<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 1.1 *adapted from* (Jack, Bennett, et al., 2018) *describes the biomarker profiles that are associated with different diagnostic classification in the 2018 research framework.*

Biomarker cut-offs: Amyloid Positivity.

Within the ATN classification system all biomarkers exist on a continuous scale. However, under the NIA-AA classification framework a binary indication of biomarker positivity is required to make a diagnosis. The appropriate discretisation of these continuous biomarkers is crucial for informing clinical decision making or determining inclusion in a clinical trial. As β-amyloid biomarkers are one of the most heavily studied in AD, binary categorisation approaches exist for determining β-amyloid positivity (i.e. A+). Large AD cohort studies have observed that β-amyloid PET is distributed bimodally (Palmqvist et al., 2015). This has led to deriving β-amyloid PET cut off values that denote if an individual is A+ vs A- (Jack et al., 2017). Similarly, when investigating CSF $A\beta_{42}$, researchers have been able to determine the abnormal level indicating that cortical β-amyloid is not being sufficiently cleared (Palmqvist

et al., 2014). These discretisation approaches have proven robust however they do not account for individuals who are on the boundary of β-amyloid positivity.

Biomarker Cut-Offs: Tau and Neurodegeneration / Neuronal Injury.

Unlike β-amyloid biomarkers, there is no clear discretisation approach to determining a categorical classification of in vivo tau (i.e. T+) and neurodegeneration / neuronal injury (N+). The most applicable way of categorising these biomarkers is using topographical staging. The topographic accumulation of tau PET closely follows the Braak neurofibrillary tangle staging topography (Schöll et al., 2016). This classification scheme would mirror the Braak postmorten staging based on the location of tau PET uptake. A recent approach measures the accumulation of tau in meta regions with stages closely aligned with cognitive decline (Jack, Wiste, et al., 2018). A similar approach could be used for structural MRI as patterns of cortical atrophy have been shown to mirror the progressive Braak staging (Jack et al., 2017).

Combining ATN Biomarkers To Make Prognostic Decisions In AD.

Many studies on independent cohorts show that the combination of ATN biomarkers provide a fine stratification of individuals on the AD continuum based on prognostic trajectories of cognitive decline (Allison et al., 2019; Bilgel et al., 2018; Burnham, Bourgeat, Doré, Savage, Brown, Laws, Maruff, Salvado, Ames, Martins, Masters, Rowe, Villemagne, et al., 2016; Dumurgier et al., 2017; Insel et al., 2015; Jansen et al., 2018; van Maurik et al., 2019). While the NIA-AA proposed that a numeric scale could be used to define normal vs abnormal biomarkers to inform clinical decision making, there is no widely accepted approach used. The inherent continuous nature of these biomarkers and the reliable biomarker trajectories make multivariate methods, such as machine learning, optimal for combining ATN biomarkers to make prognostic decisions.

Predictive Models in AD

Machine learning techniques have become one of the most favoured means of analysing neuroimaging data. Annual publications involving both neuroimaging and machine learning have increased nearly exponentially since 2003 (Christos Davatzikos, 2019). The application of machine learning approaches in AD is prevalent and accordingly has been extensively reviewed (Arbabshirani, Plis, Sui, & Calhoun, 2017; Maroco et al., 2011; Pellegrini et al., 2018; Rathore, Habes, Iftikhar, Shacklett, & Davatzikos, 2017; Tanveer et al., 2020). In particular, machine learning approaches are commonly used for a) classifying individuals as cognitively normal (CN) vs. AD patients, achieving accuracies ranging from 60-90% or b) predicting conversion to dementia due to AD from MCI (i.e. stable MCI vs progressive MCI) achieving accuracies ranging from 50-85% (Pellegrini et al., 2018). The most common and arguably most simple classification frameworks rely on Support Vector Machines (Arbabshirani et al., 2017; Pellegrini et al., 2018; Rathore et al., 2017). However, large publicly available longitudinal research cohorts have encouraged more complex learning approaches to be used (for example; elastic nets, random forests, gaussian process regressions, multiple kernel learning, artificial neural networks and deep learning)) (Ebrahimighahnavieh, Luo, & Chiong, 2020; Jo, Nho, & Saykin, 2019; Jollans et al., 2019; Tanveer et al., 2020). Attempts have been made to compare the performance of different modelling approaches (Jollans et al., 2019; Maroco et al., 2011; Samper-González et al., 2018). However, comparison of modelling techniques is difficult as cross validation approaches, sample sizes and sample heterogeneity have a significant effect on model performance metrics (i.e. accuracy or receiver operator characteristics) (Jollans et al., 2019).

Further, when concluding that a given algorithm is best for making predictions across the AD continuum the machine learning approach must be able to capture changes in pathology, cognition and diagnosis. However, in trying to achieve optimal performance across a set of different prediction problems the machine learning practitioner should refer to the no free lunch theorem. The *no free lunch theorem* (Wolpert & Macready, 1997) is an empirically assessed theorem that states that over all possible predictions all machine learning approaches average out to have the same performance. This occurs because a machine learning approach is commonly designed to achieve optimal performance for a specific task (i.e. predicting change in diagnosis). In constraining the algorithm to perform optimally in this task the same machine learning algorithm is therefore limited to solve a new task (i.e. predicting change in cognitive scores). Therefore, even though prediction tasks may be associated, the same algorithm with

optimal performance on the first task will sacrifice prediction performance on the second task (i.e. better performance in predicting change in diagnosis will likely decrease performance in predicting change in cognitive scores). With this in mind it is perhaps better to not ask which algorithm is best for making a specific prediction related to AD, but rather, which algorithm is able to make a range of predictions related to AD. In light of this a potential way to compare modelling frameworks is to test models on the same hold out data set across a set of different problems (i.e. changes in diagnosis and change in cognitive scores). These prediction challenges offer a more unbiased and potentially more relevant approach to determining the efficacy of prediction algorithms (e.g. TADPOLE (Marinescu et al., 2018)). This is particularly relevant in classification of cognitively normal vs AD or sMCI vs pMCI, these classes are discriminated with similar accuracies using simple linear discriminant analysis (i.e. LDA) or complex and heavily optimised nonlinear models (i.e deep learning) (Pellegrini et al., 2018). Given that heavily constrained models (i.e. deep learning) will be less effective at multiple prediction tasks (i.e. no free lunch) and offer no predictive benefit, simpler and less parameterised machine learning approaches may be optimal to predict a range of AD related changes while achieving benchmark performance.

Alternatively, insights can be gained by approaches that compare performance within the same framework. That is, research that uses the same sample and modelling approach to classify patients but alters training features can provide information on the predictive nature data modalities. By doing this neuroscientists and clinicians alike can optimise their data collection to be as predictive as possible.

Data modalities used for classification

Cognitive Data

Several studies have used only cognitive variables to make predictions in AD (Belleville, Fouquet, Hudon, Zomahoun, & Croteau, 2017; Chapman et al., 2011; Donovan et al., 2014; Gallucci et al., 2018; Pereira et al., 2018, 2017; Silva et al., 2013; Tabert et al., 2006). These models have reasonable performance separating sMCI vs pMCI, achieving classification accuracies up to 84% (Chapman et al., 2011). Of these cognitive variables, those that relate to memory are often the most heavily weighted (i.e. are more informative for the classification) particularly when predicting conversion from MCI to AD (Silva et al., 2013; Tabert et al.,

2006). However, multivariate models that combine cognitive data across several domains are favourable to single domain models (Chapman et al., 2011).

Clinical Information

In addition, several studies have investigated whether clinical variables associated with affective disturbance (i.e. depression and anxiety) are predictive of conversion to dementia due to AD (Chen et al., 2008; Defrancesco et al., 2017; Donovan et al., 2014). These studies show that moderate to severe depressive symptoms are predictive of conversion to dementia, (Defrancesco et al., 2017) while mild depressive symptoms, which are frequent in individuals with MCI, do not increase the risk of conversion (Chen et al., 2008; Defrancesco et al., 2017).

Clinical Neuroimaging Data

Different approaches have been used to perform discrete classifications using neuroimaging data (for reviews: (Arbabshirani et al., 2017; Maroco et al., 2011; Pellegrini et al., 2018; Rathore et al., 2017; Tanveer et al., 2020)). However, alternate continuous approaches have also been used to reduce neuroimaging data into an AD index or indices (Christos Davatzikos et al., 2019). These studies look for latent disease markers within structural MRI scans, often observing that the heavily weighted regions in these latent structures are in the hippocampal and parietal lobes (C. Davatzikos, Xu, An, Fan, & Resnick, 2009; Y. Fan, Batmanghelich, Clark, & Davatzikos, 2008). Further, using these data reduction approaches, several AD atrophy signatures have been discovered, often relating to separable cognitive trajectories (Dong et al., 2017; Young et al., 2018).

Multimodal Data

Given the interactive pathophysiological processes that occur in AD, models have been built to combine multimodal data to make predictions in AD commonly combining; genetic, cognitive, CSF and neuroimaging data (Cui et al., 2011; Devanand et al., 2007; Frölich et al., 2017; Hinrichs, Singh, Xu, & Johnson, 2011; Mazzeo et al., 2016; K. H. Thung, Yap, Adeli, Lee, & Shen, 2018; Zhou et al., 2018; Zhu, Suk, Lee, & Shen, 2016). There is general agreement that the combination of multimodal data improves model prediction performance, with varied improvements in classification accuracies (Rathore et al., 2017).

Binary Classification Of Disease vs Predicting Disease Trajectories

Most modelling approaches thus far have focussed on binary classifications. These assign class labels based on changes in diagnosis that largely correspond to the NIA-AA 2011 diagnostic criterion. Accordingly, methodological frameworks for mining neuroimaging data (Jollans et al., 2019) and predicting progression to AD (Samper-González et al., 2018) have been introduced. These approaches generally focus on binary classifications that are based on discrete clinical labels (i.e. stable vs. progressive MCI), as determined by arbitrary criteria (e.g. within a 3 year period of clinical assessment). However, these class labels (or targets) are poor ground truths for building and testing the performance of classification models (i.e. *target uncertainty*). One issue is that by discretising patients in two broad categories (i.e. sMCI vs pMCI) the model ignores the continuous nature of AD progression. Take for example patients at either side of the class boundary with very different disease trajectories; patient who progresses to AD within 1 day from clinical assessment and a patient who converts in 3 years will both be classified as pMCI. Similarly, patients with similar disease trajectories may be classified in different groups; a patient who converts in 3 years will be classified as pMCI, while a patient who remains stable for 3 years and progresses to dementia 1 day after the clinical assessment will be classified as sMCI. As a result, crucial continuous prognostic information is ignored, and models can easily misclassify patients based on arbitrary decision boundaries. To overcome this limitation and make meaningful predictions in AD, modelling approaches need to capture continuous information in prognostic trajectories.

Some models have achieved prediction of individual variability in disease progression (for review (Tang et al., 2015; Woo, Chang, Lindquist, & Wager, 2017)). A common approach to account for individual variability is to generate predictions using time-to-event models. These approaches assign a probability that a person will convert from MCI to dementia due to AD within a certain period of time (Alsaedi, Abdel-Qader, Mohammad, & Fong, 2018; Casanova et al., 2013; Desikan et al., 2010; Jack, Jr, et al., 2010; S. M. Landau et al., 2010; K. Liu, Chen, Yao, & Guo, 2017; Michaud, Su, Siahpush, & Murman, 2017; Oulhaj, Wilcock, Smith, & De Jager, 2009; Young et al., 2014), while others aim to predict exact time to conversion (Dukart, Sambataro, & Bertolino, 2015; K.-H. Thung, Yap, Adeli, Lee, & Shen, 2018; Vogel et al., 2018).

Although time-to-event models (e.g. survival analysis models predicting time to conversion) capture continuous information in patient trajectories, they are limited by target uncertainty. One source of uncertainty is the limited and variable frequency of clinical follow-ups, with

many studies interpolating the exact 'time to conversion' between clinical diagnoses. While this approach is common it has been shown to adversely affect model reliability (Li, Iddi, Aisen, Thompson, & Donohue, 2019). A further source of target uncertainty in models that rely on syndromic definitions is the poor inter-rater reliability in AD diagnosis (i.e. diagnoses may differ across clinicians). This issue is not limited to just AD prediction but most classification tasks in clinical neuroscience using diagnosis as class labels (for review (Woo et al., 2017)).

While these are clear limitations to using diagnostic labels as training targets, these types of predictions do fit well under the 2011 NIA-AA diagnostic recommendations for preclinical, mild cognitive impairment or dementia stages in AD. Accordingly, most large research cohorts focus heavily on changes in syndromic labelling as outcome variables (R. C. Petersen et al., 2010). However, the 2018 NIA-AA research framework has transitioned to defining AD as a continuum; that is, in this framework cognitive staging is achieved using continuous measures. Therefore, for predictive models to be of use within the current AD research framework, predictions need to move beyond discrete classifications based on noisy labels and predict continuous trajectories of more objective measures (i.e. cognitive scores).

Modelling Trajectories in MCI:

Several approaches predict future changes in cognitive scores by grouping longitudinal cognitive scores into discrete trajectories (Bhagwat, Viviano, Voineskos, Chakravarty, & Initiative, 2018; Hochstetler et al., 2015; Y. Wang et al., 2019; Wilkosz et al., 2010). These approaches have used various data driven ways to define prototypical patterns of cognitive change based on future changes of MMSE or ADAS-Cog. (Bhagwat et al., 2018) used a novel learning approach to fuse baseline and follow up imaging and clinical scores to predict if an individual will decline fast or slow (MMSE) or fast, moderate or slow (ADAS-cog) on cognitive scores. (Hochstetler et al., 2015) used classification and regression trees to determine if baseline variables can predict from three latent classes who had similar growth patterns of cognitive and functional changes. (Wilkosz et al., 2010) derived a much larger set of outcome labels discovering six different trajectories for cognitive decline due to AD. Recently, the generalisability and interoperability of this type of approach have been questioned (Y. Wang et al., 2019). Here, Wang et al. investigated the interoperability of defining broad latent trajectories of cognitive decline in two separate longitudinal studies. While they found that both cohorts had trajectories of slow and rapid dementia progression,

the number of latent classes and relative class size could not be replicated across the two datasets when identical statistical approaches were used. Moving away from discrete trajectories approaches have been proposed to predict exact changes in MMSE or ADAS-Cog scores (Y. Fan et al., 2008). (Zhang, Shen, & Initiative, 2012) used baseline and follow-up data to predict future scores on cognitive tests at different time intervals, while (Y. Fan et al., 2008) used structural MRI to predict the rate of change in MMSE score. While these two studies don't assume latent classes, their predictions are based on measures with low resolution and sensitivity to change (i.e. MMSE).

Modelling trajectories in preclinical AD

The work discussed so far primarily focused on predicting changes from a baseline state of MCI. However, determining biological predictors of future pathological change for cognitively normal individuals is of crucial importance particularly as clinical trials move to less severely AD affected individuals. Several studies using primarily mixed effects models have studied the association of cortical β-amyloid and cognitive trajectories in cognitively normal populations (Allison et al., 2019; Bilgel et al., 2018; Burnham, Bourgeat, Doré, Savage, Brown, Laws, Maruff, Salvado, Ames, Martins, Masters, Rowe, & Villemagne, 2016; Dumurgier et al., 2017; Insel et al., 2015; Jansen et al., 2018). These studies have shown that when considering both atrophy and cortical amyloid burden separable patterns of cognitive decline are observed (Bilgel et al., 2018; Burnham, Bourgeat, Doré, Savage, Brown, Laws, Maruff, Salvado, Ames, Martins, Masters, Rowe, & Villemagne, 2016; Dumurgier et al., 2017; Insel et al., 2015), concluding that for optimal predictive performance both amyloid and atrophy should be considered. Further, it has been shown that there is a dose dependent relationship of amyloid and cognitive decline (Bilgel et al., 2018). As more longitudinal data from preclinical research cohorts become available, machine learning approaches that accurately predict prognostic trajectories in preclinical AD will have the most utility in clinical trial design.

Thesis Motivation and summary

The work presented in this thesis contains novel adaptations to machine learning tools for prediction in AD. However, these tools were designed to be fit for purpose, that is, biological and clinical characteristics of the data necessitated computational innovation and not the inverse. Accordingly, the motivation of the thesis is not methodological but demonstrates an integration of multivariate prediction techniques to make clinically relevant predictions in a well-defined biological problem. The methods section introduces two approaches to capitalise on the characteristics of the data used for prediction in AD. The first section, Partial Least Squares regression – Recursive Feature Elimination (PLSr-RFE), introduces a new technique to reduce high dimensional collinear neuroimaging data into a single disease related index. The second section, Generalised Matrix Learning Vector Quantisation -Scalar Projection (GMLVQ-Scalar projection), introduces a trajectory modelling approach that derives a continuous prognostic index from a model trained on diagnostic labels with high target uncertainty. We then deploy these tools in three separate experiments, described in respective chapters. The first derives a single index of grey matter atrophy that is predictive of AD biomarkers across modalities. The second generates prognostic trajectories of cognitive decline for individuals who are MCI at baseline. The third generates prognostic trajectories of tau accumulation for individuals who are cognitively normal at baseline. For all models we validate using independent data, not only testing the performance on the task the model was trained on, but also making predictions of other markers in line with the model of the AD pathological cascade. For example, we validate our atrophic maker by relating it to tau burden and cognition, which are immediately upstream and downstream, respectively, to atrophy. We generate prognostic trajectories and associate these to cognitive decline for MCI individuals, as this measure is hypothesised to have the maximum rate of change for this population. Similarly, for cognitively normal individuals, we generate prognostic trajectories and associate these with tau accumulation, which is hypothesised to have maximum rate of change prior to cognitive deficits.

2. Methods

Introduction

Machine learning has emerged as one of the most common approaches to analyse neuroimaging data (Christos Davatzikos, 2019). Within the context of using neuroimaging data to make predictions in AD many machine learning approaches have been proposed (Ebrahimighahnavieh et al., 2020; Jo et al., 2019; Jollans et al., 2019; Tanveer et al., 2020). Broadly, machine learning can be separated into two categories, a.) machine learning approaches that require independent feature creation or feature selection or b.) approaches that do not require separate feature creation and prediction stages (i.e. deep learning). Approaches like deep learning operate in a way where a subspace within the entire image is learnt and within this hidden subspace a prediction is made. However, in neuroimaging as the number of features (e.g. grey matter voxels) is far greater than the number of observations (e.g. number of voxels >300,000, number of observations <1000), these approaches are heavily susceptible to overfitting. Further, as the learnt subspace is within hidden layers, it can be difficult to determine how the algorithm makes a prediction and what are the important signatures within the image. Machine learning in contrast to deep learning approaches often requires a separate feature selection or feature creation stage and model prediction stage. Successful feature selection is a critical determinant of model success in making accurate and robust predictions (Chu, Hsu, Chou, Bandettini, & Lin, 2012). Common approaches for selecting neuroimaging features involve exhaustively searching through a reduced set of possible predictors from structural MRI. Alternatively, unsupervised feature creation approaches (i.e. PCA) are often deployed to reduce dimensionality and remove collinearity between features (Rathore et al., 2017).

In this work we performed supervised feature set construction using PLS and feature reduction using recursive feature elimination. PLSr determines multivariate relationships between predictor variables to best describe response variables. These multivariate relationships decompose the neuroimaging into a subset of orthogonal latent variables with maximum covariance with the response variable(s) (Krishnan, Williams, McIntosh, & Abdi, 2011; McIntosh & Lobaugh, 2004). Further, we use recursive feature elimination to remove voxels with poor predictive utility so as to reduce the likelihood of the resultant feature being overfit. The result of this process is in an easily interpretable sparse set of voxels that can be used to

make robust predictions. The PLS regression algorithm -plsregress.m- used is part of the MATLAB statistics and machine learning toolbox. All additional analysis steps were built using custom code in MATLAB and developed for the purpose of the work presented in this thesis.

Following the feature selection process predictions within the machine learning framework are then made. The most common machine learning frameworks are classification, with the model assigning each individual a specific class label based on the multivariate relationship between predictor features. Alternatively, a machine learning approach can be developed to predict an outcome measure which may be discrete or continuous within a regression framework. The most common approach to performing binary classifications in AD is the support vector machine (SVM) (Rathore et al., 2017). While this approach has been successful at classifying cognitively normal vs AD and sMCI vs pMCI the approach is limited in several aspects. The first, SVMs are not well suited to handling correlated features. Correlated features may hinder interpretability of model outputs as predictor weights within the SVM may oscillate between the correlated features. In addition, these pairwise interactions within the feature space (e.g. feature covariance) may be useful in the prediction and may not be accounted for in model formulation. In this work we use the GMLVQ classification framework to perform binary classifications (Schneider, Biehl, & Hammer, 2009). The GMLVQ classification model learns the multivariate relationship between the baseline predictors (metric tensor) and the location in multidimensional space (prototype) that best discriminates between the two classes. GMLVQ is well suited to classification in AD as the model induces univariate and multivariate feature scaling via a highly interpretable metric tensor to improve classification performance.
Despite the high prediction accuracies achieved by machine learning algorithms, binary classification approaches are poorly constrained, as they are based on clinical labels rather than capturing information in longitudinal patient trajectories. To derive a continuous prediction for each individual we introduce the GMLVQ-scalar projection methodology. This approach extends the GMLVQ classification framework to derive a distance to a given class prototype along the predictive axis for each individual. The GMLVQ algorithm described by Schneider et al (Schneider et al., 2009) was implemented in MATLAB by collaborators in the school of computer science at the University of Birmingham. The extension of the GMLVQ framework was derived and implemented using custom code in MATLAB and was developed for the purpose of the work presented in this thesis.

Partial Least Square Regression

To extract a predictive AD index from grey matter voxels we used PLS regression (PLSr). PLSr is a technique that determines multivariate relationships between predictor variables to best describe response variables. PLSr applies a decomposition on a set of predictors to create latent variables that show the maximum covariance with the response variables (Krishnan et al., 2011; McIntosh & Lobaugh, 2004).

These latent variables are derived by performing a decomposition of both the predictor variables and the response variables. In our case, we only use a single response variable therefore only the predictor variables are decomposed into latent variables. This decomposition process iteratively computes a vector of weights so that the projections of the predictors (i.e. latent variables) have maximum covariance with the response variable. After the first weights vector is computed, the reconstructed predictor data is subtracted from the original predictor data resulting in a deflated predictor matrix. This process is then repeated on the deflated predictor data to generate a second weights vector and latent variable and it can be further repeated to derive as many dimensions as the user specifies.

For our case the PLSr model is set up as such:

For each individual, the predictor variables are values of grey matter density from a structural MRI scan and the response variable is a cognitive score (**Figure 2.1**).

The 3-dimensional structural scan is vectorized so that the predictor matrix (denoted as X) is subjects (rows) by voxels (columns) and the response variable (denoted as Y) is subjects (rows) by 1 (cognitive score).

Figure 2.1 PLSr model

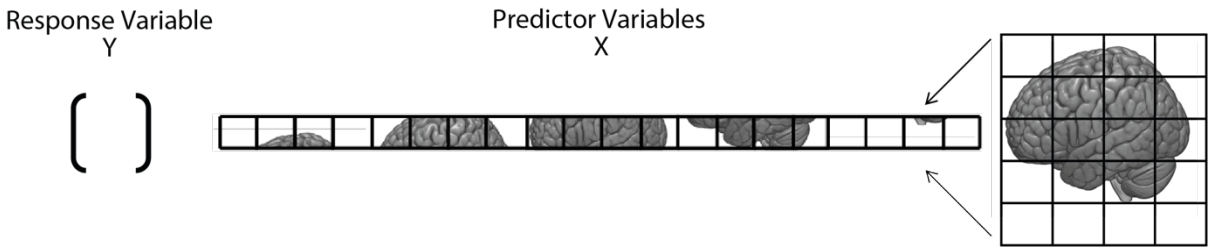


Figure 2.1 *shows a visual representation of the PLSr model, the column on the left is the response variable Y, the row on the right is the vectorized structural MRI for a single subject.*

The PLSr method will decompose the predictor variables X into T latent projections.

Such that:

T =XW, so that the cov(T,Y)=max

Where X is the matrix of grey matter voxels, T is a subjects by L (number of dimensions) matrix of projections, W is a voxels by L matrix of voxel weights and Y is the response variable.

Voxel weight stability

Given that neuroimaging data often has many more predictors (i.e. voxels) than samples (i.e. patients), it is highly likely that the PLSr approach will overfit the data rendering interpretation of the voxel weights matrix unreliable. Therefore, it is ideal to assess the predictive nature of a given voxel. One way of doing this is to determine the stability of the weight assigned to each voxel. A common approach is to use bootstrapping to assess this, where stability is assessed by generating n sets of bootstraped samples from the data set. Then for each bootstrapped sample a PLSr is performed resulting in a distribution of weights per voxel. Across bootstraps axis rotation and reflection can occur, this is corrected for using Procrustes rotation (Milan & Whittaker, 1995). To summarise the resultant distributions for each weights vector, we calculated a variance-normalised weight for each voxel $Zw_{vox,dim}$ akin to a z-statistic.

$Zw_{vox,dim} = \frac{\mu(bootstrap)}{\sigma(bootstrap)}$ where the voxel location is indexed by $vox \in \{1, \dots, V\}$, the PLSr dimension is indexed by $dim \in \{1, \dots, D\}$ and $bootstrap$ is the distribution of weights for the given voxel and dimension across bootstrapped samples. i.e. The mean weight across all bootstrap samples was normalised by the corresponding standard deviation.

PLSr Recursive Feature Elimination (PLSr-RFE)

While PLSr is robust so that the number of predictors can be more than the number of observations, the high variance explained by the complete model can show that merely having many predictors can explain “erroneous” variance in the outcome variable. To account for this, we performed recursive feature elimination, whereby we iteratively removed voxels (original input features) that have weak predictive value. While recursive feature elimination is generally used for feature selection, we performed a simultaneous feature selection and feature construction method using the PLSr recursive elimination framework.

Using the variance normalised weight (i.e. $Zw_{vox,dim}$) we determined how stable the predictive power of each voxel is by summing the variance normalised weight across PLSr dimensions.

As the weights can be either negative or positive, we calculated the stability Sw_{vox} of a voxel by the sum of the absolute z values across the PLSr dimensions.

$$Sw_{vox} = \sum_{dim=1}^{D} |Zw_{vox,dim}|$$

In order to determine the least stable voxels to be eliminated in each iteration, we ranked each voxel by its stability from lowest to highest and removed the lowest 25th percentile. We then repeated this process iteratively removing the least stable 25th percentile of voxels. After each reduction, we tested how well the model using the remaining voxels predicts the response variable. This value was generated using a nested cross validation framework. To determine the optimal number of voxels, we implemented an early stopping paradigm, where we tested after each iteration of the recursive feature elimination loop the generalisation performance of the learnt PLSr dimensions generated from the voxel subset. We simultaneously assessed the performance of each PLSr dimension to generalise out-of-sample and the "optimal" number of voxels required to generate these dimensions.

k-Fold Nested Cross Validation

Standard methods to test for model acceptance involve permutation testing to generate a null hypothesis (McIntosh & Lobaugh, 2004). Here, we use k-fold cross validation to asses model performance (**Figure 2.2a**). Within each of the cross folds we first need to determine the optimal number of voxels and PLSr dimensions to retain. Here, we accepted a model, specifically a given PLSr dimension and the number of voxels retained, using nested cross validation. Nested cross validation is an approach of determining hyperparameters (i.e. PLSr dimension and the number of voxels retained) by assessing the generalisation performance across a set of hyperparameters (**Figure 2.2b**).

Figure 2.2 Data splitting for cross validation framework.

Figure 2.2a K-fold cross validation example, where k=5.

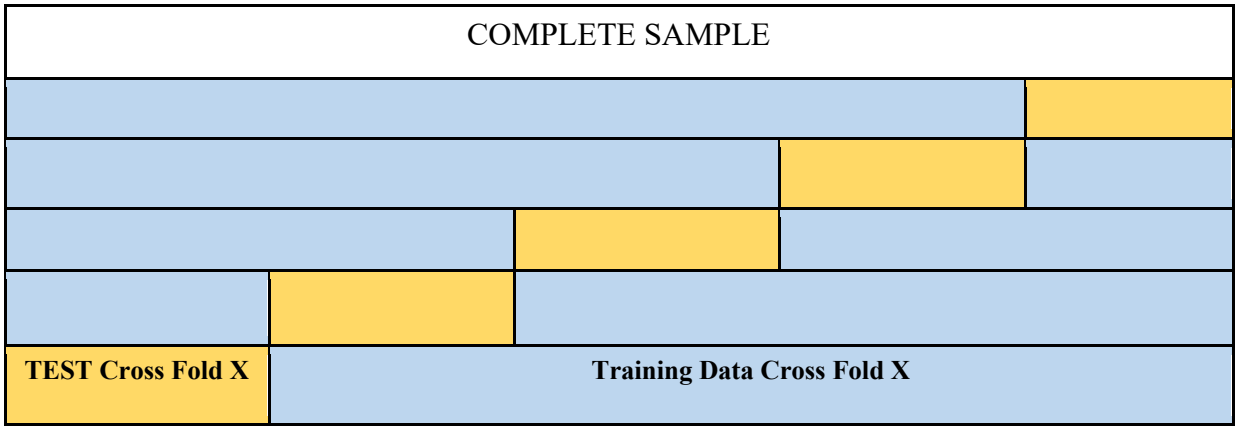


Figure 2.2a *shows an example of splitting a data set for 5-fold cross validation. The data is split into 5 non overlapping sets of test data (yellow box) and 5 sets of training data.*

Figure 2.2b Nested k fold cross validation, where number of nested loops nk = 5.

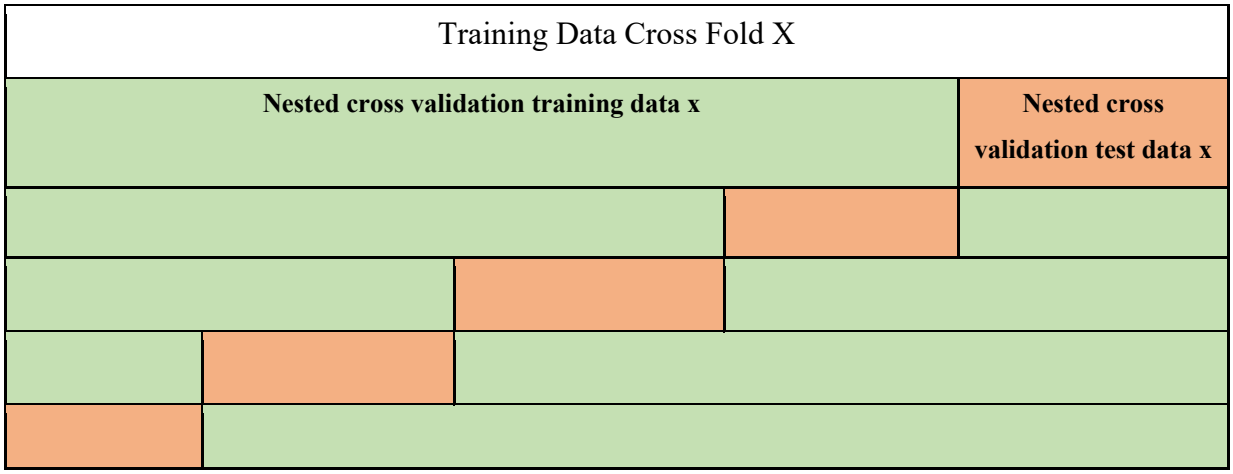


Figure 2.2b *shows an example of splitting one training set (Training Set Data Cross Fold X) from figure 1a into 5 nested cross validation folds. The training set from figure 1a is split into 5 non overlapping sets of nested cross validation test data (orange box) and 5 sets of nested cross validation training data (green box).*

Recursive feature elimination within nested cross folds.

We assessed the generalisation performance of different hyperparameters by learning a voxel weights matrix on a nested training data set that is iteratively reduced (i.e. voxels removed). Within each recursive loop, we learned the weights matrix from the remaining cross validation training data and then generated the latent dimensions for a cross validation test set by multiplying the voxel weights matrix by the cross-validation test data. We then tested the variance that can be explained by each of these latent dimensions in our predictor variable. Following this we assessed and ranked the stability of how predictive each voxel is, removing the least stable 25%. We then repeated this process, iteratively removing voxels from the cross-validation training set. We determined the optimal number of voxels by assessing when prediction performance starts to decrease on the cross-validation test set (i.e. too few voxels remaining). Simultaneously, we determined how many PLSr dimensions to retain by assessing the difference between cross validation training variance and test variance explained in the response variable.

Algorithm 2.1. PLSr-RFE implementation

PLSr-RFE algorithm

For each 1:k cross-fold resampling do

  For each 1:nk nested cross fold resampling do

    For each recursive loop

      generate PLSr weights for a set of latent dimensions for nested training data.

      $T_{(Nested\ Training\ x)} = X_{(Nested\ Training\ x)}\ W_{(Nested\ Training\ x)}$,

      Where, $Cov(T_{(Nested\ Training\ x)}, Y_{(Nested\ Training\ x)}) = max$

      project the nested test data onto the training weights matrix to derive a set of latent variables for the nested test data.

      $T_{(Nested\ Test\ x)} = X_{(Nested\ Test\ x)}\ W_{(Nested\ Training\ x)}$

      calculate the shared variance (R^2) between the nested test latent variables and the nested test set response variable

      determine predictive stability of each voxel

$$Sw_{vox} = \sum_{dim=1}^{D} |Zw_{vox,dim}|$$

      rank and remove the least stable 25% of voxels

    end

  end

  Set number of voxels retained(***nLoops***) and PLSr dimensions (***nPLSdim***)

  For 1:***nLoops***

    generate PLSr weights for a set of latent dimensions for cross-fold training data

    $T_{(Training\ Data\ Cross\ Fold\ X)} = X_{(Training\ Data\ Cross\ Fold\ X)}\ W_{(Training\ Data\ Cross\ Fold\ X)}$,

    Where, $Cov(T_{(Training\ Data\ Cross\ Fold\ X)}, Y_{(Training\ Data\ Cross\ Fold\ X)}) = max$

    determine predictive stability of each voxel

$$Sw_{vox} = \sum_{dim=1}^{D} |Zw_{vox,dim}|$$

    rank and remove the least stable 25% of voxels

  end

  project the cross-fold test data onto the training weights matrix to derive a set of latent variables for the cross-fold test data. $T_{(TEST\ Cross\ Fold\ X)} = X_{(TEST\ Cross\ Fold\ X)}\ W_{(Training\ Data\ Cross\ Fold\ X)}$

  calculate the R^2 between the cross-fold test data ***nPLSdim*** latent dimensions and the cross-fold test data response variable.

  end

PLSr-RFE: Summary/Conclusion

The PLSr-RFE process was designed to alleviate two issues with PLSr when applied to neuroimaging data. The first is, by determining the optimal model by nested cross validation, we increase the likelihood of the model to generalise out of sample. That is, the resultant sparse voxel weights matrix is likely to extract a predictive latent score in a new representative population. This is critical for neuroimaging data as the number of voxels in a scan (in the order of 10^5-10^6) will often be far greater than the number of observations, increasing the chance of overfitting the data. The second benefit of this approach is that the resultant sparse weights matrix is easily interpretable, giving a strong indication of the regionally specific multivariate relationships between grey matter density and the response variable (i.e. cognitive score).

Generalised Matrix Learning Vector Quantisation

Learning Vector Quantisation (LVQ)

Learning Vector Quantisation (LVQ) comprises classifiers that operate in a supervised manner to iteratively modify class-specific prototypes to find boundaries of discrete classes. In particular, LVQ classifiers are defined by a set of vectors (prototypes) that represent classes within the input space. These prototypes are updated iteratively throughout the training phase, resulting in changes in class boundaries. For each training example, the closest prototype for each class is determined. These prototypes are then updated so that the closest prototype representing the same class as the input example is moved towards the input example and those representing different classes are moved further away.

Training data sample of size *n* is denoted by $(x_i, y_i) \in \mathbb{R}^m \times \{1, \dots, K\}, i = 1,2, \dots, n$, where *m* denotes data dimensionality and *K* the number of different classes. In this instance the LVQ network comprises *L* prototypes $w_q \in \mathbb{R}^m, q = 1,2, \dots, L$ defined by their location in the input space and their class label $c(w_q) \in \{1, \dots, K\}$.

Prototypes are adapted automatically throughout training so that distances between points of class $c \in \{1, \dots, K\}$ and their corresponding prototypes sharing the same label *c* is minimised, while keeping the prototypes of other classes as far as possible. Classification is then performed on test data based on a winner-takes-all process, whereby a previously unseen input vector $x_i \in \mathbb{R}^m$ is assigned the class label $c(w_q)$ of the closest prototype (w_q).

The Generalised Matrix LVQ (GMLVQ)

The Generalised Matrix LVQ (GMLVQ) (Schneider et al., 2009) extends the LVQ utilising a full metric-tensor for a more robust (with respect to the classification task) distance measure in the input space. To do this, the metric tensor induces feature scaling in its diagonal elements, while accounting for task conditional interactions between pairs of features (co-ordinates of the input space).

Given a positive definite matrix Λ, $\Lambda \succ 0$, the generalised form of the squared distance is calculated as

$d_\Lambda(x, w) = (x - w)^T \Lambda (x - w)$.

Positive definiteness of Λ can be ensured by defining $\Lambda = \Omega^T\Omega$ where $\Omega \in \mathbb{R}^{m \times m}$ is a full rank matrix. Note that only relative distances of input points to the prototypes are important. Hence, the metric tensor can be multiplied by any positive real number without effecting the classifier performance in any way. To account for this inherently ill-posed nature of the model fitting and hence to ensure the stability of the algorithm, Λ must be normalised following each learning step e.g. by making sure that $\sum_i \Lambda_{i,i} = 1$, fixing the trace throughout learning.

Using the steepest descent method, the cost function to be minimised through online learning is

$$f_{\mathrm{GMLVQ}} = \sum_{i=1}^{n} \varphi(\mu_\Lambda(x_i))$$

where

$$\mu_\Lambda(x_i) = \frac{d_\Lambda(x_i, w^+) - d_\Lambda(x_i, w^-)}{d_\Lambda(x_i, w^+) + d_\Lambda(x_i, w^-)}$$

In this instance φ is a monotonic identity function $\varphi(\ell) = \ell$, $d_\Lambda(x_i, w^+)$ is the distance between the sample vector x_i from the closest prototype with the same class label $c(w)^+ = c(x_i)$ and $d_\Lambda(x_i, w^-)$ is the distance from x_i to the closest prototype of a different class. We assessed model performance by classification accuracy, true positive rate, true negative rate and averaged error (i.e. average error of false positive and false negative).

Interrogating the Metric Tensor Matrix:

The metric tensor provides a way to determine how predictive different data is in the classification task. Below are toy examples highlighting how the elements of the metric tensor transform the feature space by weighting univariate and pairwise features. The examples show a hypothetical binary classification task. The data and metric tensor are shown in the original space and in the learnt space (i.e. after feature scaling by the metric tensor). The metric tensor is learnt to ensure that all points have a neighbour of the same class, that is, the shortest pairwise distance for all points is with a point from the same class.

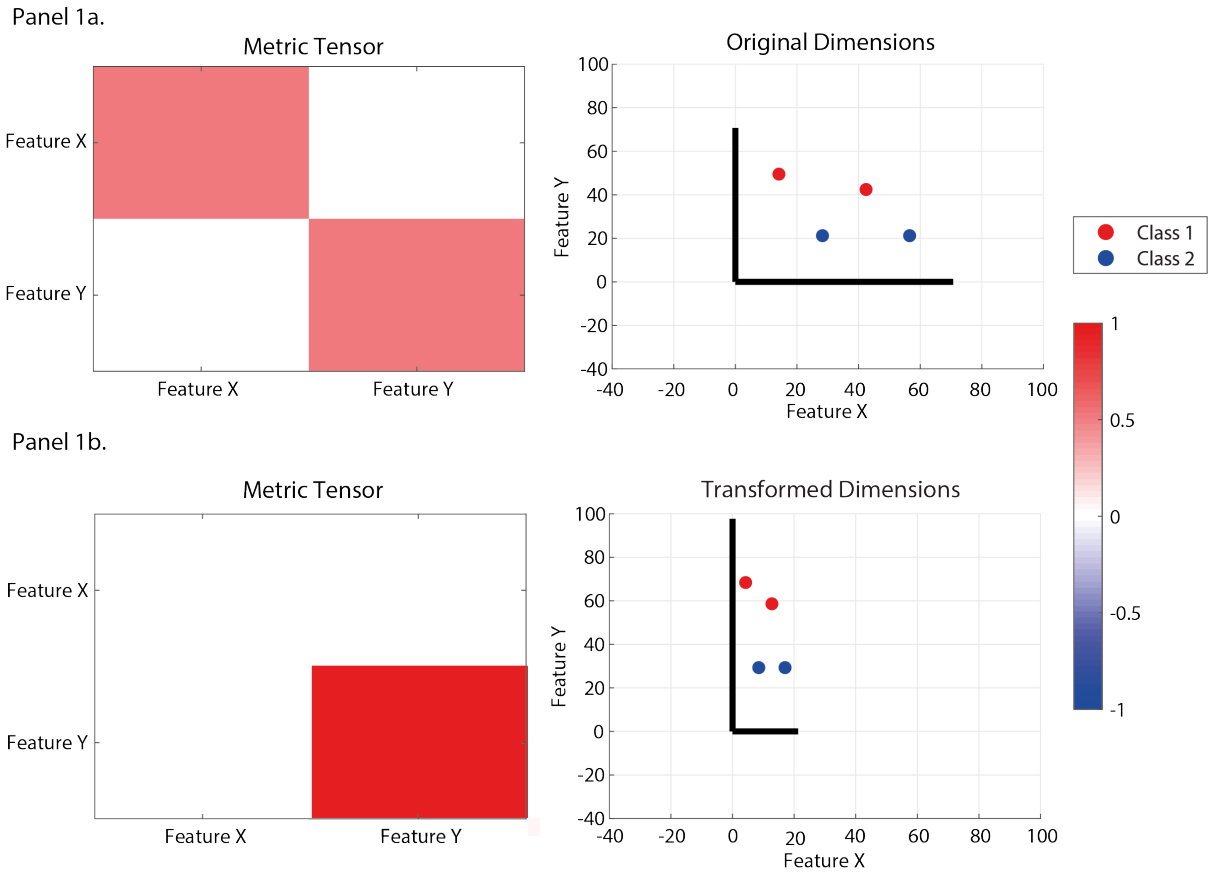
Here, we provide 3 examples of how the elements of the metric tensor scales the data in the input space and the interpretation of the metric tensor.

Panel a: All examples show the initialised metric tensor on the left panel, where the diagonal values are both equal to 0.5 and the off diagonals are equal to 0. The panel on the right gives an example of the data distribution in the original space, where red circles are training data of

class 1 and blue dots are of class 2. The solid black line represents the X and Y axis in the original space.

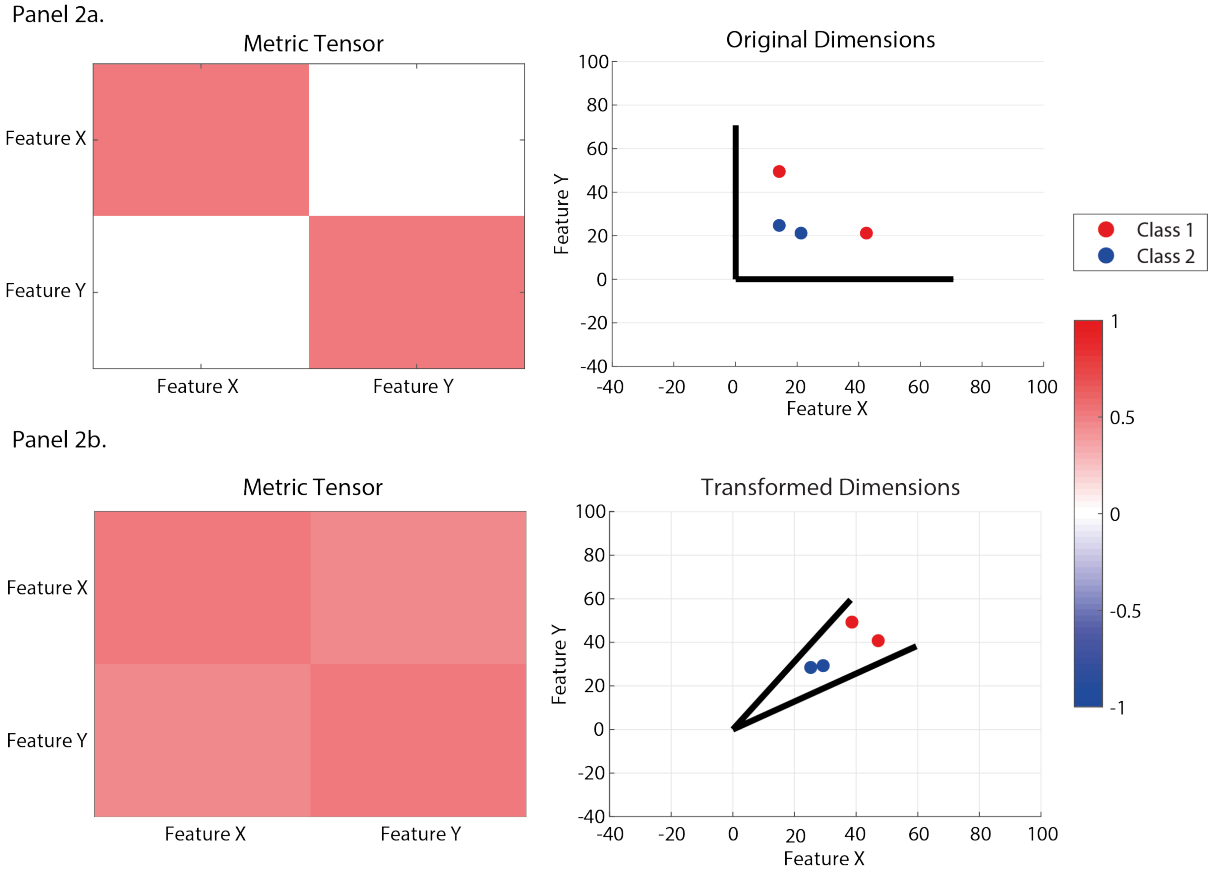
Panel b: All examples show the learnt metric tensor following the GMLVQ process on the left panel, note the diagonal terms sum to 1 and the off diagonals can be either negative of positive. The right panel shows the transformed space when scaled by the learnt metric tensor: the red circles and the blue dots are the same individuals from **Panel a** in the transformed space following the GMLVQ process and the solid black lines show the transformation of the X and Y axis into the learnt space.

Example 2.1: Univariate weights



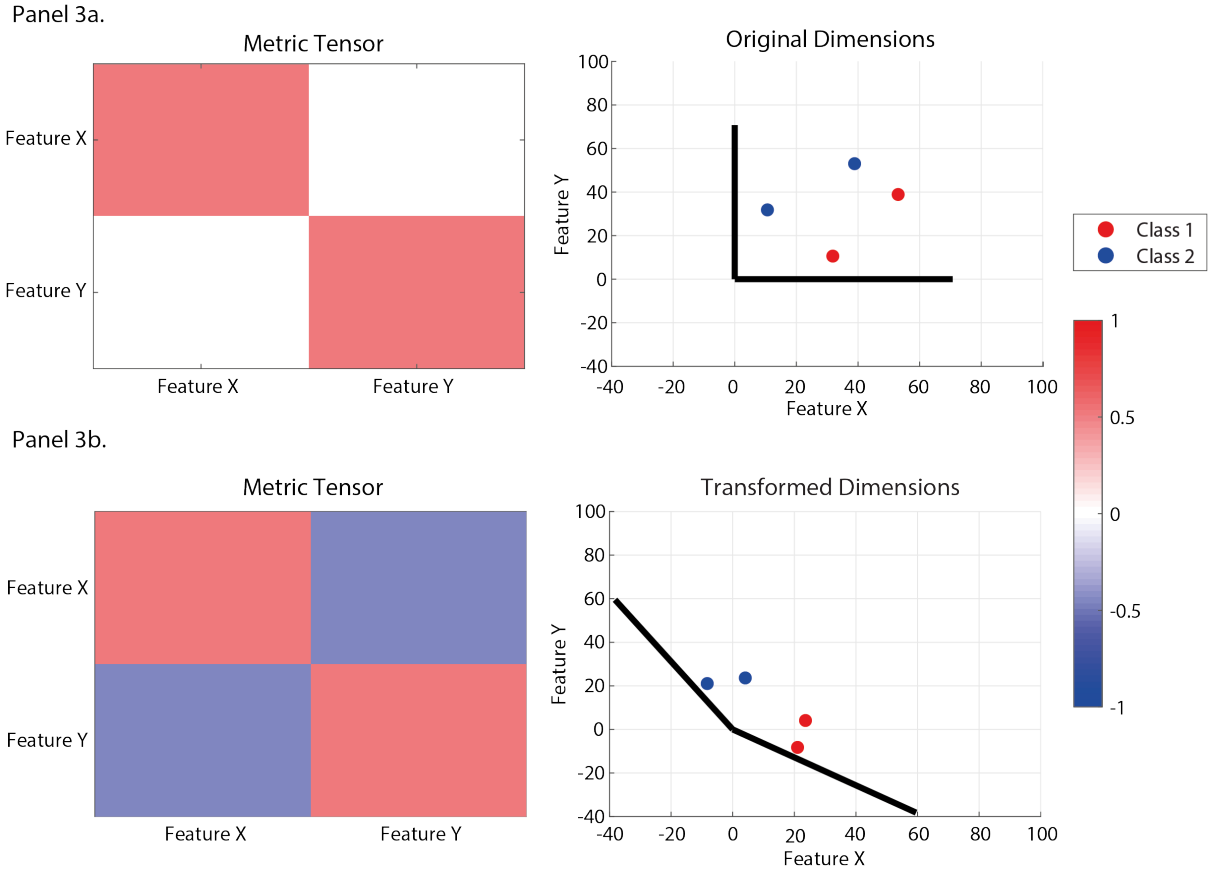
Example 2.1 shows that for all data have a neighbour of the same class the diagonals need to be scaled so that the term for Feature Y has to be increased relative to Feature X The conclusion that is drawn from the metric tensor is that Feature Y is the most relevant feature.

Example 2.2 Multivariate weights (positive)



Example 2.2, the interaction between Feature X and Feature Y is highly relevant to group training samples of the same class (Note that the off-diagonal terms are positive). This interaction term collapses the angle between the two dimensions causing data of the same class to group together.

Example 2.3 Multivariate weights (negative)



Example 2.3 shows that for all data to have a neighbour of the same class the interaction between Feature X and Feature Y is relevant; however, as this term is negative it increases the angle between the two dimensions. Thus, we conclude that these two features separate points from different classes.

To interpret these examples mathematically, note that the squared distance can be written as $d_{\Lambda}(x,w) = (x-w)^T \Lambda (x-w) = (X-W)^T (X-W) = d_2(X,W)$,
where $d_2(.,.)$ denotes the squared Euclidean distance and $X = \Lambda^{1/2}x, W = \Lambda^{1/2}w$. Hence, the original vectors (including the unit directional vectors of the standard axis) are transformed by the linear operator equal to the square root of the symmetric positive-definite metric tensor Λ into a space where the new distance (defined by Λ) can be re-interpreted as the usual Euclidean distance.

GMLVQ – Scalar Projection

Let's assume that our class labels are defined by a longitudinal change in diagnostic category, where someone who has a change of diagnosis attributed to a deteriorating condition is classed *progressive*, whereas someone with a stable diagnosis and condition is classed *stable*.

The GMLVQ- Scalar Projection method extends the GMLVQ framework to extract specific distance information from the sample vector x_i and the learnt prototypes $w_{(stable,progressive)}$. Specifically, we determine the distance in the learnt space (i.e. after applying the learnt metric tensor) between an individual with sample vector x_i and the learnt prototype w_{stable} along the vector separating w_{stable} and $w_{progressive}$

Following the learning process in GMLVQ we transformed the sample vector x_i and prototypes $w_{(stable,progressive)}$ into the learnt space via the metric tensor Λ.

$$X_i = \Lambda^{1/2}\, x_i$$

$$W_{(stable,progressive)} = \Lambda^{1/2} w_{(stable,progressive)}$$

As the metric tensor Λ is learnt in the non-Euclidean space:$d_\Lambda(x,w) = (x-w)^T\Lambda(x-w)$, we applied the square root of this tensor to re-represent the data so that the squared norm of a vector in the non-Euclidian space is equal to the squared Euclidean norm of the transformed space.

We centred the coordinate system on $W_{(stable)}$ and calculated the orthogonal projection of each vector X_i onto the vector $W_{(progressive)}$, in this co-ordinate system.

$$Projection = \frac{\overrightarrow{X_\iota W_{stable}} \cdot \overrightarrow{W_{progressive} W_{stable}}}{\left|\overrightarrow{W_{progressive} W_{stable}}\right|}$$

To normalise the projections with respect to the position of the prototype $W_{(progressive)}$, we divided the projection by the norm of $W_{(progressive)}$:

$$Scalar\ Projection = \frac{\overrightarrow{X_\iota W_{stable}} \cdot \overrightarrow{W_{progressive} W_{stable}}}{\left|\overrightarrow{W_{progressive} W_{stable}}\right|^2}$$

The resultant value indicates the separation of a test point from prototype W_{stable} along the direction of $\overrightarrow{W_{progressive} W_{stable}}$. To determine the relative separation from the stable prototype, we normalised the projection by the distance between each prototype $W_{progressive}, W_{stable}$, as indicated by squaring the norm of vector $\overrightarrow{W_{progressive} W_{stable}}$. A large positive projection indicates a large separation from W_{stable} in direction $\overrightarrow{W_{progressive} W_{stable}}$

and a large negative projection indicates a large separation from W_{stable} in the opposite direction i.e. $\overrightarrow{W_{stable}W_{progressive}}$. In particular, a value of 1 indicates that a sample is incident to prototype $W_{progressive}$, whereas a value of 0 indicates that a sample is incident to prototype W_{stable}. A value of 0.5 is the decision boundary separating the two classes within the binary classification framework. That is, the scalar projection has a large positive value for progressive individuals and zero or negative value for stable individuals.

Deriving the GMLVQ – Scalar Projection: Prototype Vector

The figure below uses the data from Example 2 above to graphically derive the Prototype Vector ($\overrightarrow{W_qW_j}$) that is generated following the GMLVQ learning process. Note: Here we return to general notation, where class 1 is *progressive* and class 2 is *stable* in the description above.

Figure 2.3 Prototype vector

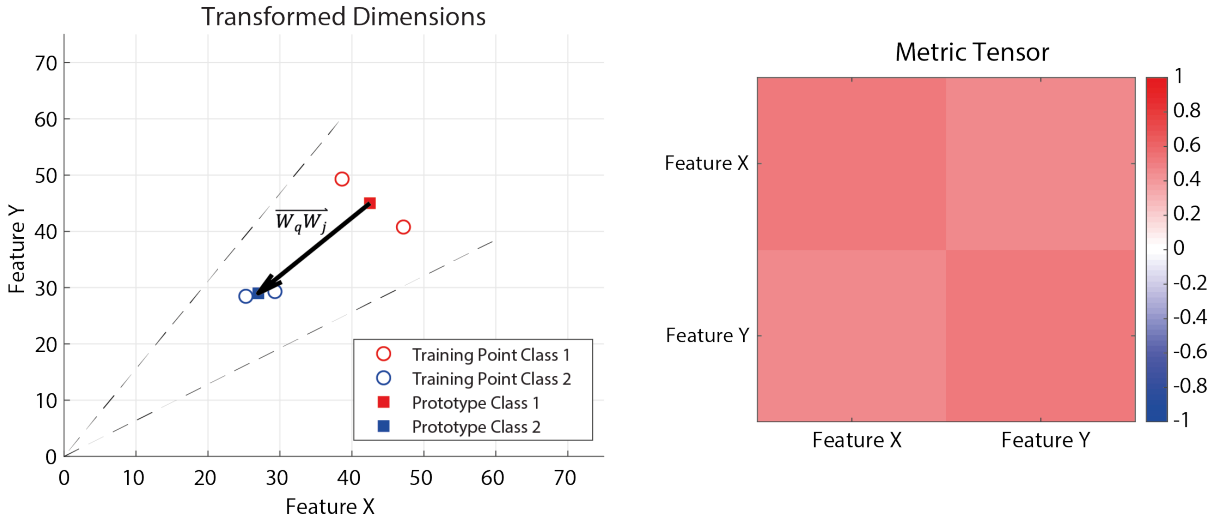


Figure 2.3 *The left panel shows the vector $\overrightarrow{W_qW_j}$ represented by the black arrow connecting prototypes for class 1 and class 2. The dashed lines show the transformed axis from the original space. The right panel is the learnt metric tensor used to transform the data from the original space into the learnt space.*

Deriving the GMLVQ – Scalar Projection: Test Projections

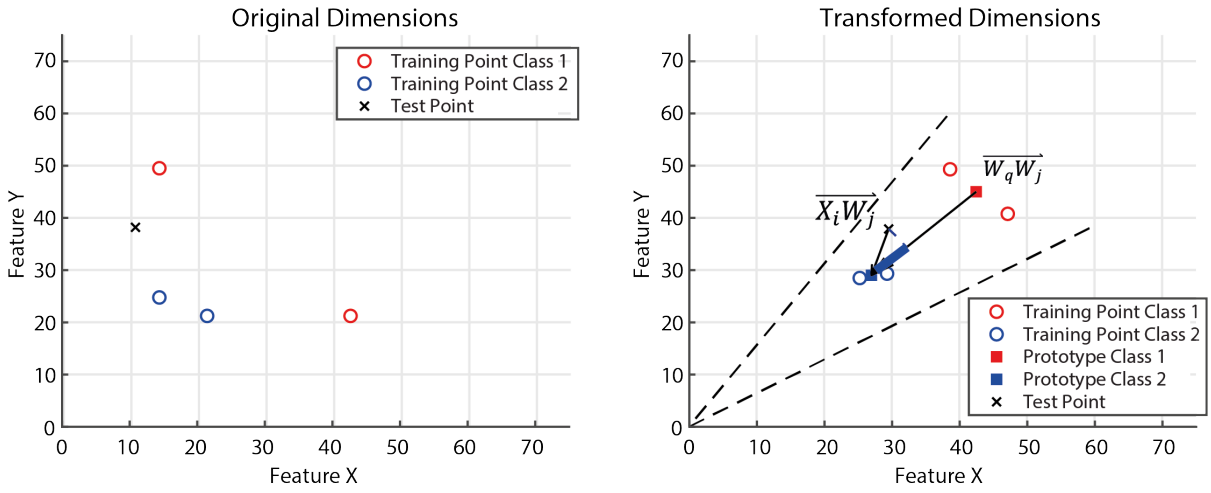
Here, we use the data and learnt metric tensor from **Example 2.2** in **Interrogating the Metric Tensor Matrix** to show three scenarios where values for the scalar projection are calculated and interpreted in a general sense.

The left panel shows a possible test point in the original data space of **Example 2.2**. The red dots are the training sample for class 1, the blue dots are the training sample for class 2 and the black X is the test point in the original space.

The right panel shows the graphical derivation of the scalar projection in the learnt space for the test data point X. Here, the red dots are the training sample for class 1 and the blue dots are the training sample for class 2, the black X is the test point in the learnt space (i.e. after the GMLVQ learning process) and the dashed black lines show the transformation of the X and Y axis into the learnt space.

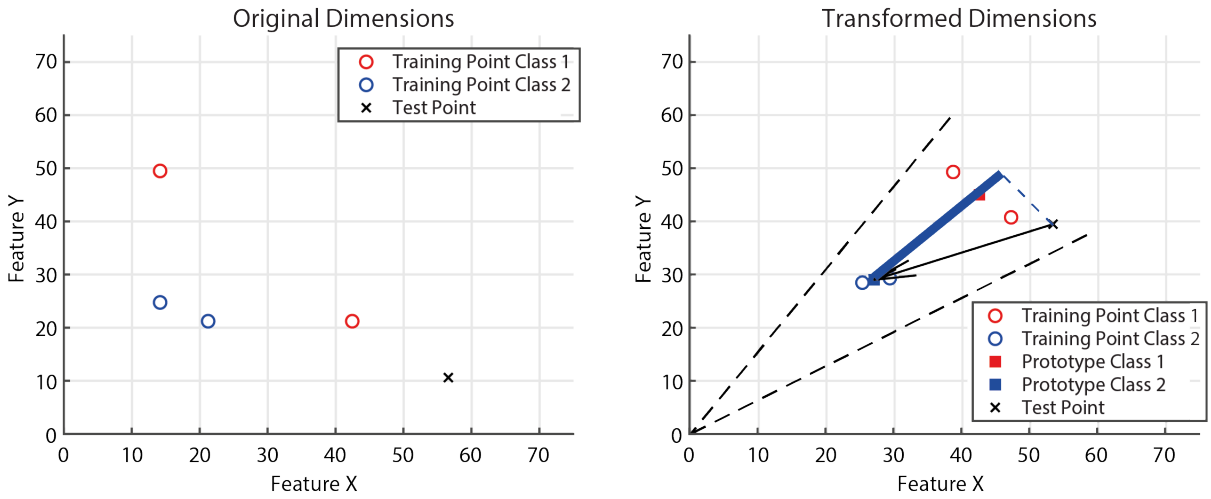
The red square is the prototype for class 1 in the learnt space, the blue square is the prototype for class 2 in the learnt space. The vector $\overrightarrow{W_q W_j}$ connects the prototypes for class 1 and class 2, and vector $\overrightarrow{X_1 W_j}$ connects the test point in the learnt space with the prototype for class 2. The dashed blue line is the orthogonal projection from the test point in the learnt space to vector $\overrightarrow{W_q W_j}$. The solid blue line represents the scalar projection from the test point in the learnt space to the class 2 prototype along the vector $\overrightarrow{W_q W_j}$.

Example 2.4 Between two classes



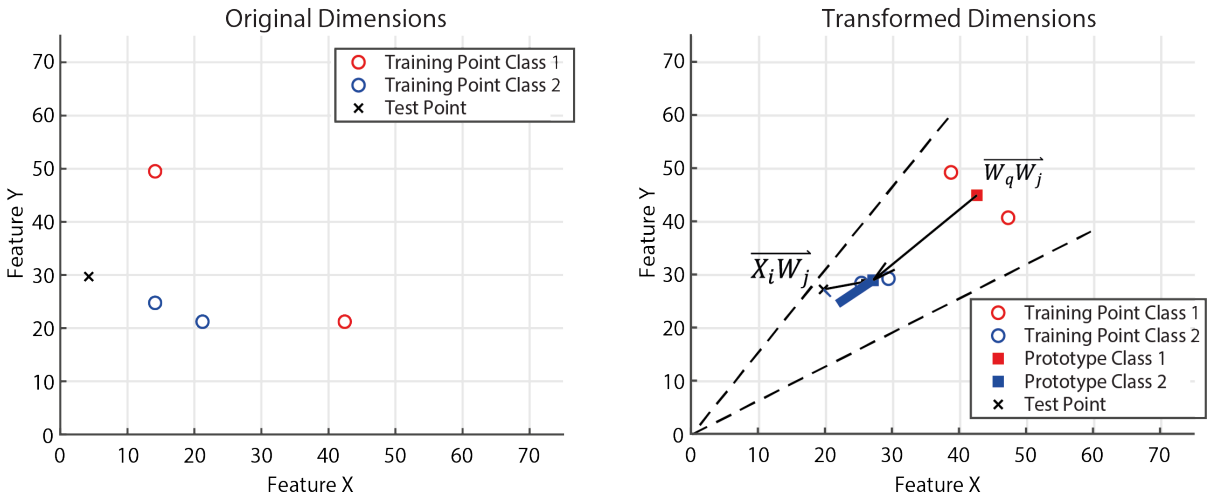
Example 2.4 shows an example of a scalar projection that has a length between 0 and 1. Within the binary classification framework a decision boundary exists at 0.5, where a value below this will cause the point to be labelled as class 2 and a value beyond this will cause the point to be labelled class 1. This is interpreted as an individual that is between the two classes.

Example 2.5 Positive value greater than one



Example 2.5 is given where the length of the scalar projection is larger than 1, in this example the individual is further away from class 2 prototype along the prototype vector than the class 1 prototype. This is interpreted as being further away from the class 2 prototype than the majority of class 1 individuals.

Example 2.6 Negative value



Example 2.6 shows an example where the length of the scalar projection is negative. In this example the individual is further away from the class 1 prototype than the class 2 prototype along the prototype vector. This is interpreted as being further away from the class 1 prototype than the majority of class 2 individuals.

GMLVQ – Scalar Projection: Summary/Conclusion

The GMLVQ-scalar projection approach is a method for deriving a continuous predictive variable from the GMLVQ classification framework. Our methodology defines a novel way of learning a continuous prognostic metric while training using 'noisy' diagnostic labels. The development of this approach was motivated to capitalise on broad class labels to train a model and derive an individual continuous score. By building a model to separate two continuous target values into two general categories (i.e. stable or progressive), we have forced the model to concentrate on the main structures in the data that lie between distinguishing the higher and lower target values. As our model has relatively - compared to deep learning approaches- few degrees of freedom (DOF), the model is unlikely to overfit the training data.

Where DOF for the model is DOF for prototypes + DOF for metric tensor.

DOF for the prototypes = D*C*P, and, metric tensor = D^2- (D^2-D)/2

D = number of original input features (input dimensions)

C = number of classes

P = number of prototypes per class

The GMLVQ-scalar projection approach addresses three inherent issues with prognostic models in AD research. 1.) The GMLVQ-scalar projection approach is able to account for target uncertainty. This is achieved by having a model that learns a low-parameter task-dependent scaling matrix (metric tensor), and only two locations in hyperdimensional space (prototypes). These univariate (diagonal) and multivariate (off diagonal) relationships are learnt to separate the two classes (stable vs progressive) as best possible from a global perspective (vs local metric tensors) without over constraining the predictor data. Similarly, by defining only one position in this learnt space that best determines if a person is stable or progressive, the model must ignore subtle differences for any given target, learning a broad location that best describes stable / progressive populations. By not over constraining the data, this type of model will not be sensitive enough to overfit based on subtle difference in diagnostic criteria or by patients on the class boarder. 2.) The GMLVQ-scalar projection approach is able to derive a continuous metric of how far an individual is from the stable prototype along the dimension that best separates stable and progressive training samples. This allows the model to learn implicitly a continuous prognostic trajectory for an individual that may be predictive of underlying pathophysiological change that leads to a deteriorating condition (i.e. progressive pathophysiology). 3.) The GMLVQ-scalar projection approach is perfectly suited to harmonise data that was collected using diagnostic criteria based on syndromic definitions with inherently poor sensitivity and specificity to extract a highly sensitive continuous prognostic index.

3. Experimental Chapters Overview.

Related publications and presentations

Publications

Giorgio, J., Landau, S. M., Jagust, W. J., Tino, P., Kourtzi, Z., & Alzheimer's Disease Neuroimaging Initiative. (2020). Modelling prognostic trajectories of cognitive decline due to Alzheimer's disease. *NeuroImage: Clinical*, *26*, 102199 (Joseph Giorgio, Landau, Jagust, Tino, & Kourtzi, 2020).

Giorgio, J., Jagust, W. J., Baker, S. L, Landau, S. M., Tino, P., Kourtzi, Z., & Alzheimer's Disease Neuroimaging Initiative. (*in review*). Predicting future tau accumulation in asymptomatic and early Alzheimer's disease (Joseph Giorgio, Jagust, et al., 2020).

Presentations

Alzheimer's Imaging Consortium 2020

Giorgio, J., Jagust, W. J., Baker, S. L, Landau, S. M., Tino, P., Kourtzi, Z. (2020) (POSTER): Predicting future regional tau accumulation in preclinical Alzheimer's disease

Alzheimer's Association International Conference 2020

Giorgio, J., Jagust, W. J., Baker, S. L, Landau, S. M., Tino, P., Kourtzi, Z. (2020) (ORAL): Predicting future regional tau accumulation in preclinical Alzheimer's disease

Alzheimer's Association International Conference 2019

Giorgio, J., Landau, S. M., Jagust, W. J., Tino, P., & Kourtzi, Z. (2019). (POSTER): PREDICTING FUTURE RATE OF COGNITIVE DECLINE DUE TO ALZHEIMER'S DISEASE. *Alzheimer's & Dementia*, *15*, P228-P228.

Experiment Chapter 1. Deriving an interpretable and interoperable score of Alzheimer's related atrophy

Here we use the novel feature generation methodology procedure introduced in **Methods: PLSr Recursive Feature Elimination (PLSr-RFE)** to generate an index of Alzheimer's disease atrophy from structural MRI. PLSr-RFE is a hybrid-feature selection and feature construction method that captures co-morbidities in cognition and pathology. Using predictive, study specific information (cognitive composites) we generate a highly interpretable disease related biomarker (i.e medial temporal atrophy due to AD). This biomarker can then be used across studies, transferring information rich study specific information between cohorts and modalities (i.e. interoperation). Here, we confirm the interoperability of our methodology and biomarker showing that the learnt pattern of grey matter atrophy is highly predictive of tau burden in an independent sample.

Experimental Chapter 2. Modelling prognostic trajectories of cognitive decline due to Alzheimer's disease

Here, we use the novel trajectory modelling approach presented in **Methods: GMLVQ-Scalar Projection** to mine multimodal data from MCI patients in the Alzheimer's disease Neuroimaging Initiative (ADNI) cohort, deriving individualised prognostic scores of cognitive decline due to AD. We integrate the biomarker generated from the PLSr-RFE procedure (**Experiment Chapter 1: Deriving an interpretable and interoperable score of Alzheimer's related atrophy**) into the machine learning methodology that extends beyond binary patient classification into discrete subgroups (i.e. stable vs. progressive MCI), determining individual profiles from baseline (i.e. cognitive or biological) data that are predictive of individual cognitive trajectories (i.e. change in memory scores from baseline).

Experimental Chapter 3. Predicting future regional tau accumulation in early Alzheimer's disease

Here, we extended previous work modelling cognitive trajectories for MCI individuals (**Experimental Chapter 2**. **Modelling prognostic trajectories of cognitive decline due to Alzheimer's disease)** to preclinical AD samples. We use our machine learning framework (**Methods: GMLVQ-Scalar Projection)** to test whether a prognostic index derived from baseline data can classify and stage preclinical AD individuals based on future pathological tau accumulation. Our results show that the characteristic spreading pattern of tau in preclinical AD is predictable using baseline biomarkers, particularly when stratifying groups using multimodal data. Finally, we are able to predict spatially specific individualised changes in tau accumulation in an out-of-sample group of cognitively normal people, highlighting the strong predictive ability of this metric and prediction framework.

4. Experimental Chapter 1: Deriving an interpretable and interoperable score of Alzheimer's related atrophy

Introduction

With the introduction of MRI in clinical research, atrophic patterns characteristic of AD can be reliably measured in-situ (W. Jagust, 2018). However, there is no single measure that quantifies the degree of AD related atrophy. Nor, is there a single measure of cognitive dysfunction due to AD. Therefore, it is critical to generate scores for AD related atrophy and cognitive decline that can be used within the ATN diagnostic framework to support effective stratification. Through the use of longitudinal databases such as the Alzheimer's Disease Neuroimaging Initiative (ADNI) researchers are able to derive informative biological and cognitive composite scores. These composites are generated from many variables and crucially avoid the so called "curse of dimensionality" when building models of disease progression. Cognitive composites have rapidly become the norm for generating summative measures when investigating AD (Ayutyanont et al., 2014; Donohue et al., 2014; Jutten et al., 2017, 2018; Langbaum et al., 2014), with measures of memory (i.e., ADNI- Mem (Crane et al., 2012))) and executive function (i.e., ADNI EF (Gibbons et al., 2012)) proving highly effective for diagnosing ADNI individuals with MCI. However, these cognitive composite scores may not generalise across studies as scores are often within-sample normalised and generated using cohort specific cognitive batteries.

Here, we aim to create biomarkers from data that is readily transferable across studies (i.e. structural MRI) that can predict a cohort specific cognitive composite measure (i.e. ADNI-mem). Our aim is to move beyond correlational analysis of ADNI-Mem (i.e. (Nho et al., 2012)), in doing so we will derive a score that allows us to transfer information from study to study via structural MRI (i.e. interoperation). While many studies in AD use structural MRI to separate sMCI vs. pMCI within a binary classification framework (for reviews: (Pellegrini et al., 2018; Rathore et al., 2017)) we aim to capture continuous information that directly predicts cognition. Further, given the close link between memory, medial temporal lobe atrophy, and regional deposition of tau (Cho et al., 2016; Harrison et al., 2019; Johnson et al.,

2016; Knopman et al., 2019; Schöll et al., 2016) this continuous structural MRI score should relate to both memory dysfunction and cortical tau deposition.

Using the novel, hybrid feature selection and feature construction method presented in (**Methods: PLSr Recursive Feature Elimination (PLSr-RFE)**) we generate an interpretable feature that can transfer pathological information from behavioural composite measures through structural MRI (i.e grey matter atrophy due to Alzheimer's Disease). We highlight the efficacy of our methodology and validate our biomarker using two approaches. First, we generate and test predictions of ADNI Mem using structural MRI in an independent ADNI sample. Next, we show that our approach is able to transfer information from one domain (cognitive) to another (cortical tau burden) via structural MRI in an independent sample.

Methods and materials

ADNI Participants

Data were obtained from the Alzheimer's disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. A major goal of ADNI has been to examine biomarkers including serial magnetic resonance imaging (MRI), and positron emission tomography (PET), with clinical and neuropsychological assessment to predict outcomes in mild cognitive impairment (MCI) and early Alzheimer's disease (AD). Data samples are defined as: (1) Development data samples used for model formulation and within-sample validation, (2) independent validation data samples used for out-of-sample validation. Below we provide details for each data sample:

1. Development Sample:

Data from 589 individuals (baseline diagnoses: Normal =317, MCI=272) from ADNI-GO and ADNI-2 were used for model formulation and within-sample validation. All individuals had cognitive measurements and 3T structural MRI. All individuals were included in this cross-sectional sample independent of their future diagnosis (i.e. whether following baseline an individual's diagnosis changed from cognitively normal to MCI/AD).

2. Cross-modal associations validation sample:

To out-of-sample-validate the model that predicted cross-modality associations (e.g. predict ADNI-Mem scores from grey-matter), we drew an independent validation sample comprised of 446 individuals (Normal=263, MCI=172, AD=11) from ADNI-3. These individuals have a 3T structural MRI, and cognitive measures in addition to a flortaucipir (FTP) PET scan for measuring cortical tau.

2. Cross-modal associations validation sample I:

We selected 219 individuals from the Cross-modal associations validation sample (Normal=122, MCI=89, AD=8), excluding individuals with an FTP-PET scan who were part of the Development Sample. Individuals in the Cross-modal associations validation sample I were newly recruited into ADNI-3, that is they had not been enrolled in ADNI-GO or ADNI-2 prior to enrolling in ADNI-3. This independent sample was used to validate cross-modal

associations of grey matter and ADNI-Mem scores. All data from the Cross-modal associations validation sample were taken from assessments closest in time to the FTP-PET scan.

Brain Imaging data

MRI Acquisition

Structural MRIs were acquired at ADNI-GO, ADNI-2 and ADNI-3 sites equipped with 3 T MRI scanners using a 3D MP-RAGE or IR-SPGR T1-weighted sequences, as described online (http://adni.loni.usc.edu/methods/documents/mri-protocols).

PET Acquisition

The FTP-PET protocol entailed the injection of 10 mCi of tracer followed by acquisition of 30 min of emission data from 75-105 min post injection. For each participant FTP-PET scans were acquired in close proximity to the structural MRI scan (mean =0.21 +- std=0.51 years).

Image Analysis: FTP (Flortaucipir PET) Tau

FTP data were realigned, and the mean of all frames was used to coregister FTP to each participant's MRI acquired closest to the time of the FTP-PET. FTP standardised uptake value ratio (SUVR) images were normalised to inferior cerebellar grey matter (Baker, Maass, & Jagust, 2017). MR images were segmented and parcellated using Freesurfer (V5.3) and regions of interest were used to extract cerebellar-normalised regional SUVR data. SUVR data was summarised for three Braak staging regions 12 (medial temporal), 34 (inferolateral temporal) and 56 (extra-temporal neocortical) by averaging uptake across individual Freesurfer region of interests (ROIs) comprising each Braak region (Maass et al., 2017). Finally, we assigned individuals as tau positive for each Braak stage if their SUVR value was greater than the 90th percentile of amyloid-negative, cognitively normal individuals.

Image Analysis Voxel Based Morphometry (VBM)

Structural scans were segmented into grey matter, white matter and CSF (Cerebrospinal Fluid). The DARTEL toolbox (Ashburner, 2007) was then used to generate a study specific template to which all scans were normalised. Following this, individual grey matter segmentation volumes were normalised to MNI space without modulation. The unmodulated values for each voxel represent grey matter density at the voxel location. We chose to use the

unmodulated grey matter data as it has been shown that there is a marked decrease in sensitivity to detecting abnormal regions within grey matter when the data is modulated (Radua, Canales-Rodríguez, Pomarol-Clotet, & Salvador, 2014). We repeated the PLSr-RFE procedure with modulated data to confirm that unmodulated data is preferred when generating AD related biomarkers. All images were then smoothed using a 3mm3 isotropic kernel and resliced to MNI resolution 1.5x1.5x1.5 mm voxel size. We used a small kernel size, as topographically complex and relatively small cortical regions are likely to be affected in AD (i.e. structures within the medial temporal cortex; e.g. hippocampus, entorhinal cortex). It has been suggested that smoothing beyond a 3mm kernel may artificially link small but discrete clusters of voxels, reducing topographic sensitivity (Radua et al., 2014). Further, our analysis applies a spatial decomposition across voxels. By sampling the spatial covariance structure across voxels, disease related non-parametric variations at the voxel level (that are mitigated using larger smoothing kernels in parametric statistical tests across participants) are preserved when using smaller kernel sizes, improving the efficacy of the analysis method. All structural MRI pre-processing was performed using Statistical Parametric Mapping 12 (http://www.fil.ion.ucl.ac.uk/spm/).

Partial Least Squares Regression with Recursive Feature Elimination (PLSr-RFE)

We implemented Partial Least Squares Regression with Recursive Feature Elimination **(**PLSr-RFE) (**Methods: PLSr Recursive Feature Elimination (PLSr-RFE))** to generate a grey matter density feature based on data from the Development sample (cognitively normal and MCI individuals)**.** In particular, we used grey matter density measured by structural MRI as a predictor variable to determine multivariate relationships between grey matter voxels that best predict ADNI-Mem. We performed feature set construction using PLSr and feature reduction using recursive feature elimination. PLSr determines multivariate relationships between predictor variables to best describe response variables. In particular, PLSr applies a decomposition on a set of predictors to create orthogonal latent variables that show the maximum covariance with the response variables (Krishnan, Williams, McIntosh, & Abdi, 2011b; McIntosh & Lobaugh, 2004). In our study, we used PLSr to generate a set of latent predictor variables from structural MRI data, where a) the number of features (i.e. grey matter voxels) is far greater than the number of observations (e.g. number of voxels >300,000, number of observations <1000), b) there is high degree of multi-collinearity

between voxels. PLSr reduces redundant information and maximises the amount of variance that the latent variables predict in the response variable. Further, we performed recursive feature elimination by iteratively removing voxels that have weak predictive value to reduce the chance of over fitting.

Methods Cross Validation Framework

Within-sample validation: k-fold cross-validation (PLSr-RFE)

Within-sample generalisation for PLSr-RFE (Development Sample) frameworks was assessed using k -fold cross validation (k=5). In brief, data was split into a series of training and test sets, whereby each sample point is present once in a test set. To select hyper-parameters for a given training sample, we used nested cross-validation. To assess model generalisation performance, we averaged variance explained in ADNI-Mem from the test set across all cross folds.

Resampling PLSr-RFE

To generate our test cross-folds and validation cross-folds, we performed a stratified permutation. As we have a priori knowledge of the distribution of the ADNI-Mem score within our complete sample, we respected this distribution within each fold of the data. To do this we ordered our sample by ADNI-Mem and resampled the data to have population representative distributions relative to this metric.

Out-of-Sample validation: Cross-modal associations (PLSr-RFE)

To test the out-of-sample association of the PLS derived grey matter feature (represented by the voxel weight matrix) with memory and cortical tau we used an independent sample of individuals from ADNI 3 (Cross modal associations validation sample I). We pre-processed the structural MRI scans independent of the Development sample used for model formulation following the same VBM methodology. We derived a grey matter score for the validation sample by multiplying the weight matrix generated from the PLSr-RFE trained on Development sample by the grey matter voxel values of the Cross-modal associations validation sample. We then correlated the grey matter score from the Cross-modal associations validation sample I with the ADNI-Mem composite score to test the generalisability of the grey matter score across data samples. Further, we tested the cross-modal association of the PLS derived grey matter score with cortical tau burden (measured with FTP-PET) (Cross-modal

associations validation sample). Using independent sample t-tests, we tested whether there is a significant difference in PLS derived grey matter score between tau positive and tau negative individuals across 3 different Braak staging regions: 12, 34 and 56. We then tested whether individual variability in the PLS derived grey matter score relates to individual variability in cortical tau burden across each of the Braak staging regions (Pearson's Correlation).

Results

Nested Cross Validation Results Modulated vs Un-modulated structural MRI

Optimising Number of Remaining Voxels

To determine the optimal number of voxels to retain in the PLSr-RFE model, we used a nested cross validation framework. Our acceptance criterion is the amount of variance explained in the ADNI-Mem score from the cross validation hold out set. We folded our training set 5 times to tune the parameter of number of retained voxels**.** Further, we investigated how the PLSr-RFE model performs when the same grey matter data is modulated in the VBM process.

PLS 1 on un-modulated grey matter:

For the first PLS dimension across cross validation runs we observed a decline in test variance explained after the 21st recursive loop iteration corresponding to 902 remaining voxels. **Figure 4.1** shows the converging performance of the training and test sets throughout the recursive elimination process. Here, we show that generalisation performance approximately doubles when eliminating variant (un-stable) voxels.

Figure 4.1 PLS 1 on un-modulated grey matter early stopping generalisation

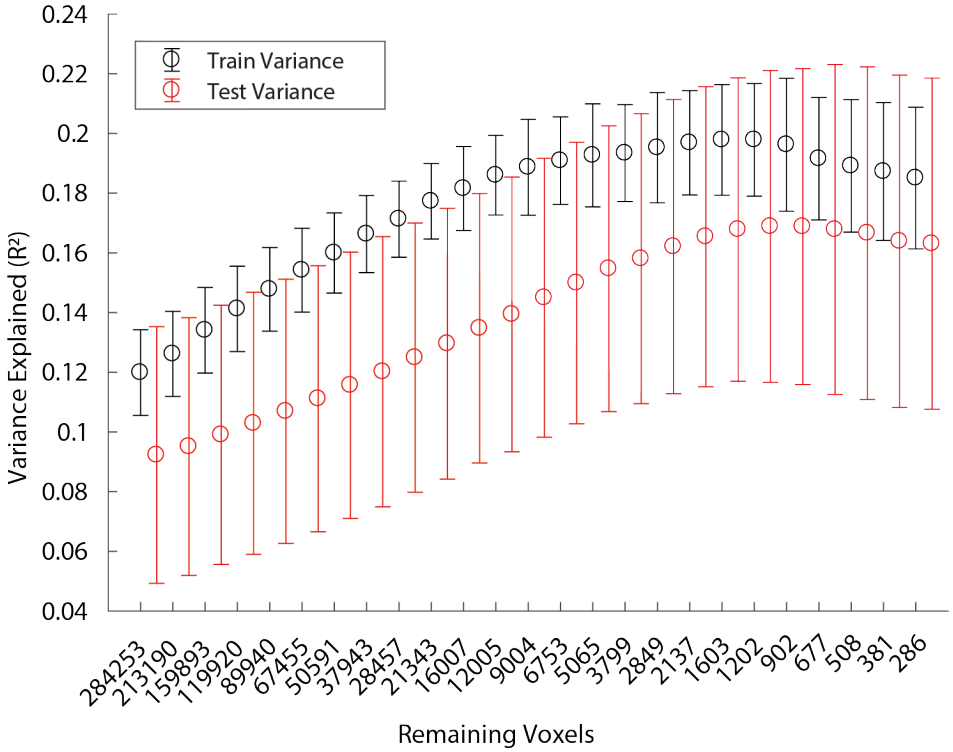


Figure 4.1 *shows the training and test performance of PLS 1 derived from un-modulated data across each recursive elimination iteration. The black dots and error bars show the mean and standard deviation of the variance explained in the ADNI-Mem score by PLS1 for the training sample across all nested validation loops. The red dots and error bars show the mean and standard deviation of the variance explained in the ADNI-Mem score by PLS1 for the test sample across all nested validation loops.*

PLS 2-5 on un-modulated grey matter:

For the remaining PLS features we observe that the training variance explained is far greater than the cross-validation test variance as seen in **Figure 4.2**. Thus, we conclude that all remaining dimensions are fitting noise / over fitting the training sample. We therefore reject these components from further analysis.

Figure 4.2 PLS 2-5 on un-modulated grey matter early stopping generalisation performance

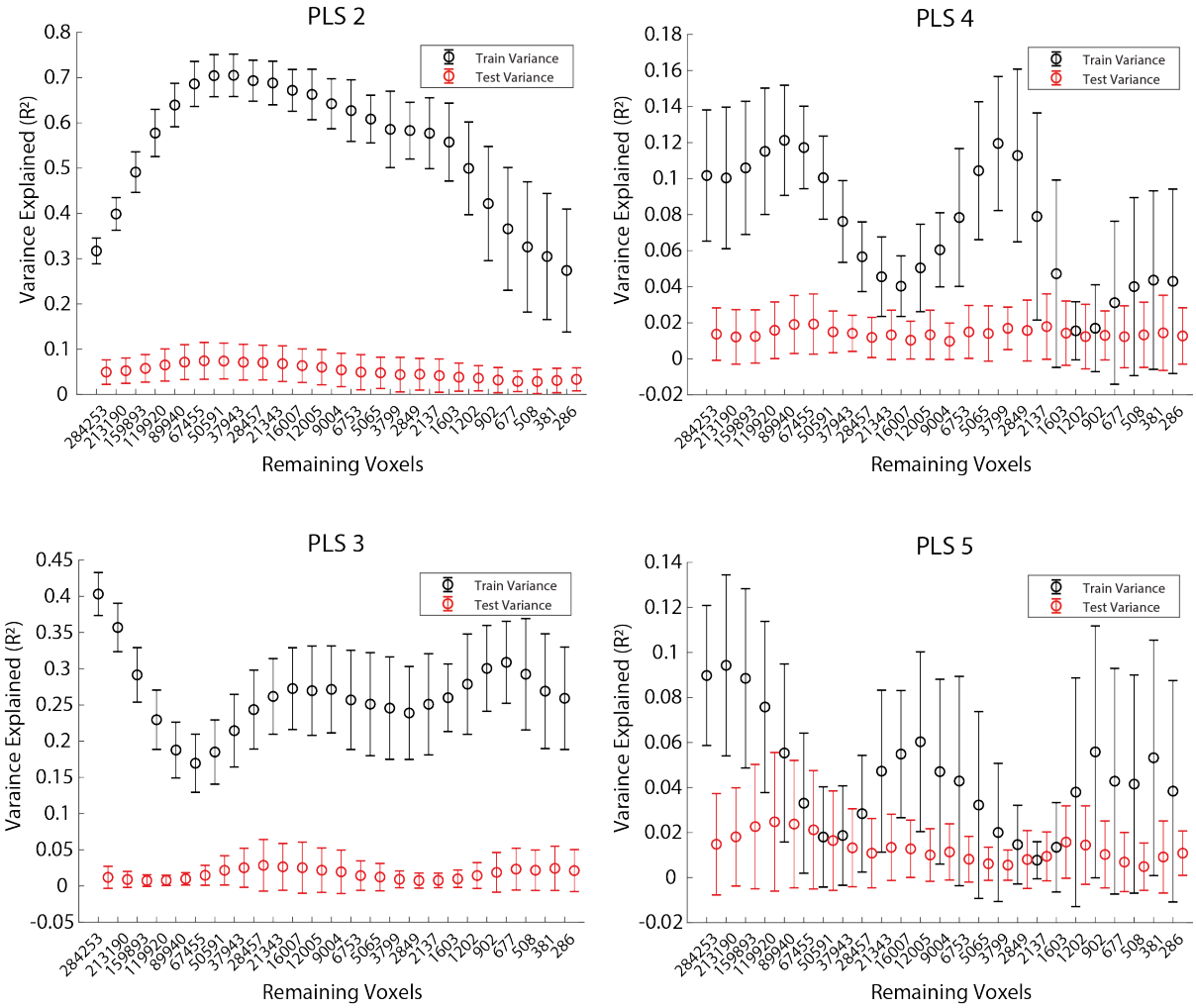


Figure 4.2 *shows the training and test performance for the remaining PLS dimensions (2-5) for unmodulated data across each recursive elimination iteration. The black dots and error bars show the mean and standard deviation of the variance explained in the ADNI-Mem score by PLS dimensions for the training sample across all nested validation loops. The red dots and error bars show the mean and standard deviation of the variance explained in the ADNI-Mem score by the PLS dimensions for the test sample across all nested validation loops. Given the last few dimensions are capturing noise that describe small amounts of variance in ADNI-Mem, their order across recursive loops may vary. This can be seen by the out of phase oscillatory behaviour of PLS 4 and PLS 5.*

Generalisation Performance un-modulated data:

To determine the efficacy of the PLS derived voxel weights matrix to generalise, we performed the matrix multiplication of the hold-out grey matter and the voxel weights matrix within each cross fold following the recursive elimination process. **Table 4.1** shows the hold out performance for each of the five cross folds using un-modulated data. The high correspondence between the training and test variance across the cross folds suggests that the PLS derived voxel weights matrix generated is robust.

Table 4.1: PLS 1 un-modulated data cross fold performance

	train variance	**test variance**
cross fold	**ADNI-Mem**	**ADNI-Mem**
1	0.16	0.28
2	0.18	0.18
3	0.18	0.18
4	0.21	0.09
5	0.20	0.14
MEAN	0.19	0.18

Table 4.1 *shows the training and test variance explained in the ADNI Mem score by PLS derived grey matter score for the un-modulated data from the individual cross folds and the mean performance across all cross folds.*

PLS 1 on modulated grey matter:

For the first PLS dimension across cross validation runs we observed a decline in test variance explained after the 16th recursive loop iteration corresponding to 3799 remaining voxels. Figure 3 shows a different profile to that of the un-modulated data (**Figure 4.2**) on the training and test sets throughout the recursive elimination process. Here, we show that generalisation performance peaks with approximately 4 times the amount of voxels retained than the un-modulated data.

Figure 4.3: PLS 1 on modulated grey matter early stopping generalisation performance

Figure 4.3 *shows the training and test performance of PLS 1 derived from modulated data across each recursive elimination iteration. The black dots and error bars show the mean and standard deviation of the variance explained in the ADNI-Mem score by PLS1 for the training sample across all nested validation loops. The red dots and error bars show the mean and standard deviation of the variance explained in the ADNI-Mem score by PLS1 for the test sample across all nested validation loops.*

Generalisation Performance, modulated data:

To determine the efficacy of the PLS derived voxel weights matrix to generalise to another sample, we performed the matrix multiplication of the hold-out grey matter and the voxel weights matrix within each cross fold following the recursive elimination process. **Table 4.2** shows the hold out performance for each of the five cross folds using modulated data. The low correspondence between the training and test variance across the cross folds suggests that the PLS derived voxel weights matrix generated is not robust to the cognitive measure ADNI-Mem. Suggesting that when investigating the spatial correlates of AD related atrophy, un-modulated data may be more appropriate. Given the lack of within sample generalisability when using modulated grey matter we do not continue to use these data in further analysis.

Table 4.2: PLS 1 modulated data cross fold performance

	train variance	test variance
cross fold	**ADNI-Mem**	**ADNI-Mem**
1	0.15	0.07
2	0.13	0.05
3	0.13	0.05
4	0.13	0.06
5	0.12	0.04
MEAN	0.13	0.05

Table 4.2 *shows the training and test variance explained in the ADNI Mem score by PLS derived grey matter score for the modulated data from the individual cross folds and the mean performance across all cross folds.*

Spatial Organisation of Voxel Weights Matrix: Figure 4.4 and **Table 4.3** show the 'z-value' for each cluster contained within a given anatomical region within the average voxel weights matrix across cross folds using unmodulated data. We show that the voxels cluster within the medial temporal cortex covering regions in the bi-lateral hippocampus and amygdala. Reference regions of interest (ROIs) are extracted from the Brainnetome Atlas (L. Fan et al., 2016).

Figure 4.4 PLS voxel weights matrix

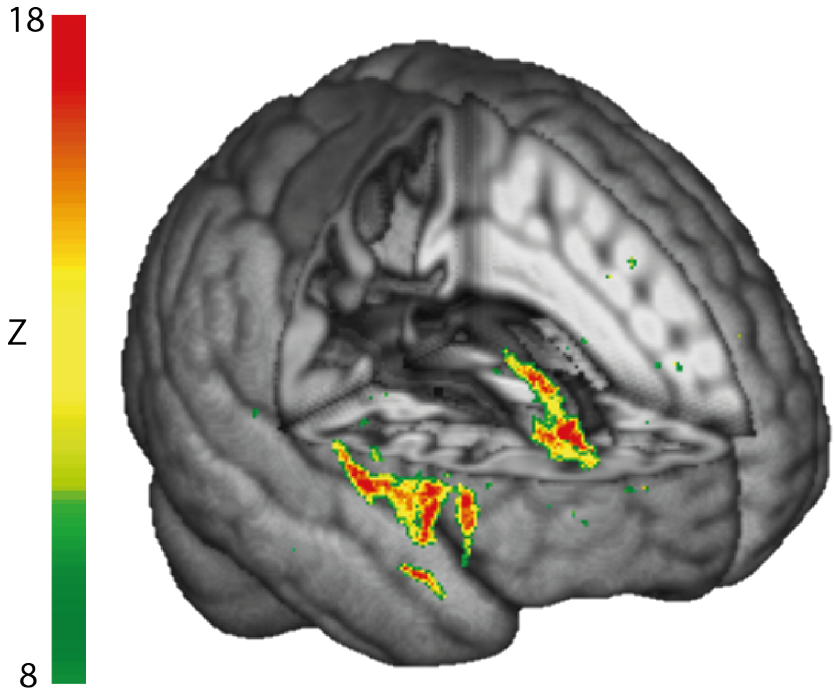


Table 4.3: PLS 1 voxel weights

%Overlap	Max Z	Mean Z	nVoxels	ROI (Left/Right)	Anatomical Description (BA Area)
0.55	9.8	9.08	13	FuG L	*rostroventral (area 20)*
6.2	10.29	9.67	24	PhG L	*rostral (area 35/36)*
17.2	15.22	11.76	66	Amyg L	*medial amygdala*
18.3	15.2	11.89	90	Amyg R	*medial amygdala*
21.29	16.44	12.71	43	Amyg L	*lateral amygdala*
14.29	15.7	13.09	45	Amyg R	*lateral amygdala*
18.28	17.69	12.49	247	Hipp L	*rostral hippocampus*
6.59	13.87	11.59	73	Hipp R	*rostral hippocampus*
6	14.64	11.88	82	Hipp L	*caudal hippocampus*
10.09	12.81	10.88	144	Hipp R	*caudal hippocampus*

Figure 4.3 *and* ***Table 4.3*** *reports brain regions within the mean PLS 1 voxel weights matrix across cross folds. The table shows the percentage overlap of each ROI with the voxel weights matrix, the peak Z value within each region, the average Z value within each region, the number of voxels and the gross and fine anatomical description of the regions. FuG: Fussiform Gyrus, PhG: Parrahippocampal Gyrus, Amyg: Amygdala, Hipp: Hippocampus, L:Left, R:Right*

Validating the PLS derived grey matter score

Next, we derived the PLS derived grey matter score for a validation sample that did not include individuals that were used in the model development (Cross-modal associations validation sample I). This value represents the weighted linear sum of grey matter voxels that best described the ADNI-Mem score in the Development sample. We showed that this score accounts significantly for variance in ADNI-Mem for the Cross-modal associations validation sample I (n= 219) that was not previously used in the PLSr-RFE feature generation ([$r^2(217) = 0.33$, $P < 0.0001$]). This relationship remained significant when we controlled for (Age; $r^2(216) = 16\%$, P<0.0001, Gender; $r^2(216) = 24\%$ P<0.0001, or Education; $r^2(216) = 37\%$ P<0.0001). It is likely that the higher variance explained by the PLS derived grey matter score for the validation sample I relative to the development sample is due to the significantly higher degree of atrophy (lower PLS grey matter score) in the validation sample I (Wilcoxon Signed Rank;

z=-3.42, p<0.0001). That is, a greater amount of variance in ADNI-Mem is likely explained by the greater amount of AD related atrophy in the validation sample.

Further, we observed significant differences (independent sample t-tests) in the PLS derived grey matter score between three sub-groups (normal cognition, sMCI -retain diagnosis of MCI for 3+ years- and pMCI -progress to dementia within 3 years-) within the Development sample used in the PLSr-RFE analysis. In particular, the normal cognition group showed significantly higher scores than the pMCI group (t(170)=9.13, P<0.0001, Cohens D =1.5) and the sMCI showed significantly higher score than pMCI group (t(165)=5.7, P<0.0001, Cohens D =0.94). However, when comparing normal cognition vs. sMCI individuals we observed only a small effect [t(230)=3.7, P=0.00072 (FWE Corrected), Cohens D =0.48] (**Figure 4.4**).

Taken together these results suggest that the PLS derived grey matter score captures variance that relates to memory dysfunction (i.e. poor ADNI-Mem scores) due to AD.

Figure 4.4 distribution of PLS derived grey matter scores

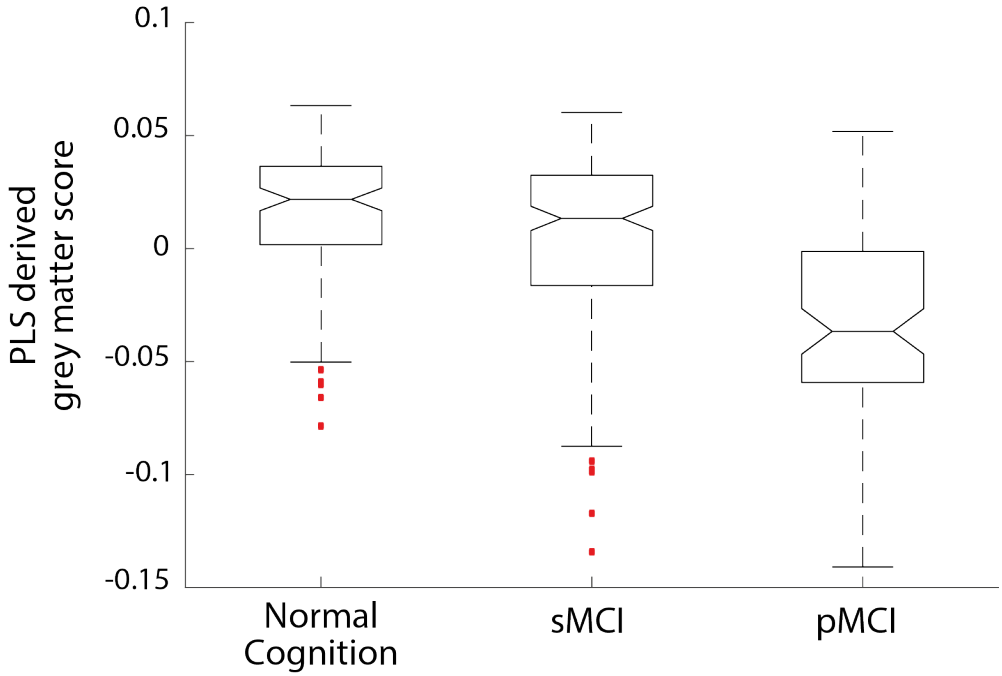


Figure 4.4. *PLS derived grey matter scores for normal cognition, sMCI and pMCI groups: Boxplots of the PLS derived grey matter scores for cognitively normal, sMCI and pMCI groups. The centre line represents the median and the edges of the boxes represent the 25th and 75th percentiles of each sample. The medians of two samples are significantly different at the p<0.05 if the edge of the intervals around each notch do not overlap. Red points denote outliers*

We next compared the variance explained in ADNI-Mem by the PLS derived grey matter score to the variance explained by the average grey matter density in medial temporal regions (i.e. amygdala, hippocampus) known to be related to ADNI-Mem (Nho et al., 2012). For each test set within the nested cross-validation framework, we extracted mean grey matter density from regions in the amygdala and hippocampus, as defined using the Brainnetome atlas (L. Fan et al., 2016). We then compared the variance explained in ADNI-Mem for these a-priori selected regions with the variance explained by the PLS derived grey matter feature.

PLS derived grey matter score vs. A-Priori Regions

We compared the variance explained in ADNI-Mem by the PLS derived grey matter score with the mean grey matter density from a set of corresponding regions known to be related to ADNI-Mem (Nho et al., 2012). For each of the test sets in our nested cross validation framework, we derived the PLS grey matter score and extracted the mean grey matter density from 8 anatomically defined ROIs (left/right medial Amygdala, left/right lateral Amygdala, left/right rostral hippocampus, left/right caudal hippocampus) taken from the Brainnetome Atlas (L. Fan et al., 2016). Further, we extracted the mean grey matter density across all 8 medial temporal regions (**Table 4.4)**. We found that the variance explained in the hold out test set for the PLSr-RFE framework (i.e. generalisation variance) was significantly higher than the variance explained by grey mater density from the a-priori selected regions.

Table 4.4: Generalisation variance explained: A-priori regions vs. PLS derived grey matter score

Region	***Anatomical Description***	**L**			**R**		
		p	**Cohen's D**	**R^2**	**p**	**Cohen's D**	**R^2**
Amygdala	*mAmyg, medial amygdala*	**<.0001**	**-1.14**	**0.15**	**0.01**	**-0.55**	**0.12**
Amygdala	*lAmyg, lateral amygdala*	**<.0001**	**-1.06**	**0.07**	**<.0001**	**-1.05**	**0.1**
Hippocampus	*rHipp, rostral hippocampus*	**<.0001**	**-1.31**	**0.13**	**<.0001**	**-1.21**	**0.11**
Hippocampus	*cHipp, caudal hippocampus*	**<.0001**	**-2.02**	**0.05**	**<.0001**	**-1.09**	**0.06**
Aggregate Region	**Aggregate Description**	**p**			**Cohen's D**		**R^2**
Mean	Average Medial Temporal Regions	**<.0001**			**-1.12**		**0.11**

Table 4.4 *shows the statistics and effect sizes when comparing the amount of variance explained in the ADNI Mem score by the a-priori selected areas and the PLS derived grey matter score across all nested cross folds.*

This finding suggests that the multivariate relationship between grey matter voxels captured by the PLS accounts for higher variability in individual ADNI-Mem scores than the average grey matter density in brain regions defined by coarser parcellations.

Finally, we tested whether the PLS derived grey matter score differs across individuals that vary in cortical tau pathology, as measured by FTP-PET (**Table 4.5**). Comparing individuals from an independent sample (Cross-modal associations validation sample) with tau positive vs. tau negative scores (independent samples t-test) showed the strongest effect within Braak stage 12 [$t(444)=9.6$, $P<0.0001$ Cohens D =1.9]. Further, the PLS derived grey matter score correlated (Pearson's correlation) significantly with cortical tau burden across all individuals, with the strongest effect for Braak stage 12 [$r^2(444) = 0.32$, $P< 0.0001$]. These results are consistent with previous studies showing a strong relationship between memory, medial temporal lobe atrophy, and regional (or Braak 12 stage) deposition of tau (Cho et al., 2016; Harrison et al., 2019; Johnson et al., 2016; Knopman et al., 2019; Schöll et al., 2016)

.

Table 4.5 PLS derived grey matter score relationship with flortaucipir PET Tau

Braak Stage	**Threshold**	**Tau Positive vs Tau Negative**				**GM Score vs Tau**
		p	**t**	**Cohen d**	**Pos/Neg**	**r^2**
tau Braak 12	1.95	<.0001	9.7	1.9	27/419	0.32
tau Braak 34	1.89	<.0001	8.3	1.5	33/413	0.15
tau Braak 56	1.93	<.0001	5.7	1.3	21/425	0.08

Table 4.5. (PLS derived grey matter score relationship with flortaucipir PET Tau) relationship of the PLS grey matter score with flortaucipir Tau measures. The table shows the threshold for tau positivity for each of the Braak stages, the statistical differences between the grey matter scores for tau positive vs. tau negative individuals, and the correlation of the PLS grey matter score with flortaucipir tau across all individuals.

Discussion

Here, we develop a supervised feature generation method (PLSr-RFE) that allows us to derive predictive and interpretable biomarkers based on structural brain imaging data. We demonstrate that grey matter density in the medial temporal lobe predicts variability in memory scores (i.e. ADNI-Mem score). In particular, this grey matter score is shown to be predictive for the classification of sMCI vs. pMCI individuals, consistent with previous studies showing that grey matter density in the medial temporal lobe (MTL) is associated with AD (C. Davatzikos et al., 2009; Hedden & Gabrieli, 2004; Mak et al., 2017; Matsuda, 2016; Rathore et al., 2017) and ADNI-Mem scores (Nho et al., 2012). Previous work using a comparable PLS methodology (sparse PLS) showed a similar spatial pattern of grey matter voxels that are predictive of MMSE scores (Monteiro, Rao, Shawe-Taylor, & Mourão-Miranda, 2016). Extending beyond this work, we generate a biomarker based on a projection (PLS-derived grey matter score) that is shown to explain more variance in ADNI-Mem scores than the grey matter density estimated from the corresponding atlas-defined MTL region. Importantly, we show that this PLS-derived biomarker predicts cortical tau pathology as measured by PET, providing a strong link between regional brain atrophy, memory decline, and tau pathology (Maass et al., 2018). Thus, our PLSr-RFE methodology has the potential to enhance interoperability across cohorts that typically include grey matter measurements (i.e. structural MRI scans) but may vary in the inclusion of other variables (e.g. cognitive or tau measurements). Further, our results show that when structural MRI is un-modulated in the VBM process the ability to detect structural abnormalities due to AD is enhanced over modulated data. This largely supports the results of (Radua et al., 2014), who suggest that grey matter density (un-modulated data) is better suited to reflect mesoscopic grey matter thinning that is evident in AD (W. Jagust, 2018). The analysis approach is in-line with previous studies which reduced structural neuroimaging data into AD indices (C. Davatzikos et al., 2009; Christos Davatzikos et al., 2019; Dong et al., 2017; Y. Fan et al., 2008; Young et al., 2018). Our approach extends these previous works as we have designed a method with the specific goal of interoperation. That is, the outcome of the PLSr-RFE methodology is a voxel weights matrix that can be applied to minimally pre-processed structural MRI to generate a single AD index that relates to cross modal AD biomarkers (i.e. cognition and tau accumulation). Future work using the same PLSr-RFE methodology can be extended to a wider range of measures derived from structural MRI scans (e.g. variation in cortical volume, shape and texture) that have been shown to be predictive of AD (for reviews: (Leandrou,

Petroudi, Kyriacou, Reyes-Aldasoro, & Pattichis, 2018; Mateos-Pérez et al., 2018; Matsuda, 2016)). In addition, the PLSr-RFE methodology can be used on alternative cognitive measures. For example, different imaging phenotypes related to online learning may also have predictive power in AD and offer a means to probe cognition without the use of lab based cognitive testing (J. Giorgio et al., 2018). A potential limitation of this work is that we did not extensively test and control for effects due to poor scan quality or scanner differences. All images were downloaded from the ADNI website following the standard pre-processing pipeline in an attempt to increase harmonisation across scanning sites. This included corrections due to gradient nonlinearity, non-uniform image intensity due to non-uniform receiver coil sensitivity and correction of non-uniformity due to wave effects (Jack et al., 2008). However, no additional control analysis was performed to ensure that there was no carry over differences due to scanner site. In addition, no additional quality control was run to account for image artefacts due to subject movement. Using the novel methodology presented in (**Methods: PLSr Recursive Feature Elimination (PLSr-RFE)**) we have successfully generated a voxel scaling matrix that can easily be applied to generate a single index of Alzheimer's related medial temporal atrophy. We show this grey matter score is predictive of memory dysfunction in an independent cohort. Further, the relationship between the grey matter score and cortical tau burden provides evidence the PLSr-RFE procedure is capable of transferring information from one cohort to another through a common medium (i.e. structural MRI). Finally, we introduce a single score of AD related medial temporal atrophy that can be used in subsequent classification models targeting AD patient populations.

5. Experimental Chapter 2: Modelling prognostic trajectories of cognitive decline due to Alzheimer's disease

Introduction

Progression to dementia due to Alzheimer's Disease (AD) involves multiple pathways of disease pathophysiology that impact cognition (Jack, Bennett, et al., 2018; Jack et al., 2013; Jack, Knopman, et al., 2010; W. Jagust, 2018). Individuals who develop dementia follow a trajectory from a stage of normal cognition to Mild Cognitive Impairment (MCI) and subsequent dementia (McKhann et al., 2011; Ronald C. Petersen et al., 2001; Sperling et al., 2011). Predicting early onset of neurocognitive decline due to AD has major implications for timely clinical management and patient outcomes. Yet, diagnosis at early stages of disease is impeded by heterogeneity in patient populations due to comorbidities (e.g. affective or cerebrovascular disorders) that may lead to MCI diagnosis without further progression to AD (Ronald C Petersen, 2009). Determining disease trajectories for individuals diagnosed with MCI has major implications for prognosis and personalised interventions.

Recent advances in machine learning allow us to develop predictive models of neurodegenerative disease by mining multimodal datasets that include measurements of cognition and neuropathology from large patient cohorts (Woo et al., 2017). In line with the 2011 NIA-AA diagnostic framework for mild cognitive impairment or dementia stages in AD (Albert et al., 2011; McKhann et al., 2011), most machine learning models in AD have focused on binary classifications. For example, machine learning models have been shown to predict with high accuracy whether individuals diagnosed with MCI will decline (i.e. progressive MCI; pMCI) or remain stable (i.e. stable MCI; sMCI) (Rathore et al., 2017). Fewer models have achieved prediction of individual variability in disease progression (Tang et al., 2015; Woo et al., 2017) focusing primarily on probabilistic estimates of time to conversion to AD (Alsaedi et al., 2018; Casanova et al., 2013; Desikan et al., 2010; Jack, Jr, et al., 2010; K. Liu et al., 2017; Michaud et al., 2017; Oulhaj et al., 2009; Young et al., 2014), with some models estimating exact time to conversion (Dukart et al., 2015; K. H. Thung et al., 2018; Vogel et al., 2018).

Despite the high prediction accuracies achieved by machine learning algorithms, binary classification approaches are poorly constrained, as they are based on clinical labels rather than capturing information in longitudinal patient trajectories. As a result, individual patients at the class boundary that differ only slightly in their profile may be misclassified. Further, the validity and statistical power (Li et al., 2019) of these approaches is limited by the frequency of clinical follow-ups (i.e. the point of conversion may occur between clinical assessments) and inter-rater reliability (i.e. clinicians may differ in their assessment). Extending machine learning modelling to predict measures determined by diagnostic labelling (i.e. time to conversion) suffers from the same limitations, introducing bias and limiting the interpretability and interoperability of machine learning algorithms (Janssen, Mourão-Miranda, & Schnack, 2018). Thus, novel modelling approaches that predict individualised trajectories of cognitive decline based on continuous measures need to be developed to enhance clinical validity and guide effective clinical interventions and drug discovery trials.

Here, we develop and implement a trajectory modelling approach that extends beyond binary classification. We use machine learning to stratify patients at early stages of impairment (i.e. MCI) based on baseline cognitive or biological data and determine individual prognostic trajectories based on continuous measures of cognitive decline (i.e. change in memory scores over time). Our trajectory modelling approach allows us to extract continuous information about progression to AD, in line with the current 2018 NIA-AA research framework that has transitioned to defining AD as a continuum (Jack, Bennett, et al., 2018).

In particular, we used large-scale data from the Alzheimer's disease Neuroimaging Initiative (ADNI) database. Cognitive data comprise composite scores across tasks; that is, summative measures of memory (i.e., ADNI- Mem (Crane et al., 2012)), executive function (i.e., ADNI – EF (Gibbons et al., 2012)), and depression (Yesavage, 1988). Similar composite measures have been shown to be effective for diagnosing cognitive dysfunction (Ayutyanont et al., 2014; Donohue et al., 2014; Jutten et al., 2017, 2018; Langbaum et al., 2014). In addition, we used well-studied biomarkers of AD (W. Jagust, 2018; Resnick & Sojkova, 2011); that is, grey matter density derived from structural MRI scans, β-amyloid burden from PET scans and APOE 4 status.

We adopted the metric learning framework (Generalised Metric Learning Vector Quantization, GMLVQ) and extended our approach beyond binary classification (i.e. sMCI vs. pMCI) to modelling of continuous measurements (i.e. change in ADNI-Mem scores). In particular, we first tested a low-parameter, interpretable model on a binary classification task (sMCI vs.

pMCI) and interrogated the key cognitive predictors that separate sMCI vs. pMCI individuals. We then trained our metric learning model on biological data– including the PLS-derived grey-matter feature (**Experimental Chapter 1: Deriving an interpretable and interoperable score of Alzheimer's related atrophy**), mean cortical β-amyloid burden, and APOE 4– and compared the classification accuracy across models trained with either cognitive or biological data.

To extend our modelling approach beyond binary classification (i.e. sMCI vs. pMCI), we used the novel GMLVQ- scalar projection approach (**Methods: GMLVQ-Scalar Projection**) that generates a continuous metric of disease progression. We demonstrate that this metric relates to rate of future cognitive decline (i.e. change in ADNI-Mem scores following baseline), providing evidence that our methodology delivers a continuous prognostic score of individual cognitive decline due to AD. Further, our trajectory modelling approach determines predictive trajectories using only cognitive markers; yet, predicting disease trajectories improves when including non-invasively measured and interpretable biomarkers (i.e. grey matter density and/or APOE 4).

Methods and materials

ADNI Participants

Data were obtained from the Alzheimer's disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. A major goal of ADNI has been to examine biomarkers including serial magnetic resonance imaging (MRI), and positron emission tomography (PET), with clinical and neuropsychological assessment to predict outcomes in mild cognitive impairment (MCI) and early Alzheimer's disease (AD). Data samples are defined as: (1) Development data samples used for model formulation and within-sample validation, (2) independent validation data samples used for out-of-sample validation. Below we provide details for each data sample:

1. Development Sample I:

253 MCI individuals from ADNI-GO and ADNI-2 were used for model formulation and within-sample validation. These individuals had at least 3 longitudinal cognitive testing sessions where the baseline assessments (MRI and cognitive) were those closest to the time of

the first florbetapir (FBP) PET scan. All individuals had baseline cognitive measurements, 3T structural MRI, FBP-PET scan for measuring β-amyloid, and APOE genotyping.

1. Development Sample II:

167 MCI individuals (of the 253 MCI individuals in Development Sample I) have 3 years of clinical diagnostic assessments. Data from these individuals were used as dichotomous outcomes (stable vs. progressive MCI) for longitudinal predictions.

2. Longitudinal prediction validation sample:

To out-of-sample-validate the model that generated longitudinal predictions, we drew an independent validation sample comprising 126 MCI individuals (ADNI-GO, ADNI-2). These individuals have baseline cognitive, 3T structural MRI, FBP-PET measurements and APOE 4 genotyping. As for the data used for model formulation, baseline was defined as the assessment closest in time to an individual's first FBP-PET scan acquired in ADNI. These individuals also have at least 3 longitudinal cognitive testing sessions that were used to validate the outcome measures for longitudinal predictions.

Brain Imaging data

MRI Acquisition

Structural MRIs were acquired at ADNI-GO and ADNI-2 sites equipped with 3 T MRI scanners using a 3D MP-RAGE or IR-SPGR T1-weighted sequences, as described online (http://adni.loni.usc.edu/methods/documents/mri-protocols).

PET Acquisition

PET imaging was performed at each ADNI site according to standardised protocols. The FBP-PET protocol entailed the injection of 10 mCi with acquisition of 20 min of emission data at 50-70 min post injection.

Image Analysis: FBP (Florbetapir PET) β-Amyloid

FBP data were realigned, and the mean of all frames was used to co-register FBP data to each participant's structural MRI. Cortical SUVRs were generated by averaging FBP retention in a standard group of ROIs (lateral and medial frontal, anterior and posterior cingulate, lateral parietal, and lateral temporal cortical grey matter) and dividing by the average uptake from a

composite reference region (including the whole cerebellum, pons/brainstem, and eroded subcortical white matter regions) to create an index of global cortical FBP burden for each subject (Susan M Landau et al., 2015).

Imaging Analysis-MRI: Medial Temporal Grey Matter Atrophy

Structural scans were segmented into grey matter, white matter and Cerebrospinal Fluid (CSF). The DARTEL toolbox (Ashburner, 2007) was then used to generate a study specific template to which all scans were normalised. Following this, individual grey matter segmentation volumes were normalised to MNI space without modulation. The unmodulated values for each voxel represent grey matter density at the voxel location.

All images were then smoothed using a 3mm3 isotropic kernel and resliced to MNI resolution 1.5x1.5x1.5 mm voxel size. All structural MRI pre-processing was performed using Statistical Parametric Mapping 12 (http://www.fil.ion.ucl.ac.uk/spm/).

To generate a single index of medial temporal atrophy we used a voxel weights matrix that was previously derived to generate an interpretable and interoperable disease-specific biomarker (**Experimental Chapter 1: Deriving an interpretable and interoperable score of Alzheimer's related atrophy)**. In brief, a feature generation methodology (partial least squares regression with recursive feature elimination (PLSr-RFE)) was used to apply a decomposition on a set of predictors (T1-weighted MRI voxels) to create orthogonal latent variables that show the maximum covariance with the response variable (memory score). Further, we performed recursive feature elimination by iteratively removing predictors (voxels) that have weak predictive value. The PLSr-RFE procedure results in a voxel weights matrix that is used to calculate a single score of AD related medial temporal atrophy. To generate an individual's score of medial temporal atrophy we performed a matrix multiplication of the previously derived voxel weights matrix and each subject's pre-processed T1 weighted MRI scans.

Cognitive Scores

We used three baseline cognitive scores as predictors for longitudinal models: a) composite scores of memory function (ADNI-Mem) derived from the Rey Auditory Verbal Learning, AD Assessment Schedule-Cognition, Mini-Mental State Examination and Logical Memory tests (Crane et al., 2012). b) composite scores of executive function (ADNI-EF) derived from the WAIS-R Digit Symbol Substitution, Digit Span Backwards, Trails A and B, Category Fluency and Clock Drawing tests (Gibbons et al., 2012). c) the sum of all elements from the geriatric

depression scale (GDS) (Yesavage, 1988). As individuals are excluded from ADNI with a GDS >5 we investigate affective disturbance at subthreshold levels of clinical depression.

Generalised Metric Learning Vector Quantization (GMLVQ)

We used the Generalised Metric Learning Vector Quantization (GMLVQ) framework (Schneider, Biehl, & Hammer, 2009) to generate and test binary classification models (**Methods: GMLVQ-Scalar Projection)** that classify sMCI vs. pMCI individuals (Development Sample II). Individuals were characterised as sMCI if they repeatedly received an MCI diagnosis for more than three years of clinical observation. Individuals who progressed from MCI to AD within a window of 3 years of clinical observation were characterised as pMCI. Individuals who progressed from MCI to AD after 3 years were excluded from the Development Sample II.

GMLVQ belongs to the class of classifiers referred to as Learning Vector Quantization (LVQ). These classifiers operate in a supervised manner to iteratively modify class-specific prototypes and learn boundaries between classes. For each training example, the closest prototype of each class is determined, these prototypes are then updated so that the prototype defining the same class is moved towards the training example and other prototype(s) representing different class(es) are moved further away. The Generalised Metric LVQ (GMLVQ) extends the LVQ utilising a full metric-tensor for a more robust distance measure. By applying the metric-tensor, specific feature scaling can occur while also accounting for different feature scales and pairwise task-conditional dependencies in the input space. Interrogating the diagonal terms allow us to determine the key univariate predictors for separating sMCI vs. pMCI patients. Further, interrogating the off diagonal terms of the metric tensor allow us to investigate the multivariate predictors that contribute to this classification task.

GMLVQ Cognitive model

We used the (GMLVQ) framework to generate and test binary classification models that classify sMCI vs. pMCI individuals (Development Sample II) based on cognitive measures (GDS, ADNI-Mem and ADNI-EF).

GMLVQ Biological model

We followed the same methodology as for the **GMLVQ Cognitive model** (Development Sample II) to test the GMLVQ model on biological data. That is, we generated and tested

binary classification models based on metric learning that discriminate between the same sMCI vs. pMCI individuals based on biological data (PLS derived grey matter score, β-amyloid and APOE 4). Note that this sample includes 3 pMCI individuals who are β-amyloid negative (i.e. SUVR<1.11) at baseline. We did not restrict our measure of β-amyloid to a binary value but rather used continuous SUVR values to avoid model bias near the ADNI threshold of amyloid positivity.

GMLVQ – Scalar Projection

We next generated a continuous prediction using either baseline cognitive data (GDS, ADNI-Mem, ADNI-EF) or baseline biological data (PLS Derived Grey matter score, β-amyloid, APOE 4) for MCI individuals (Development sample I). The GMLVQ- Scalar Projection method extends the GMLVQ framework to extract specific distance information from the sample vector x_i and the learnt prototypes $w_{(stable,progressive)}$. Specifically, we determine the distance in the learnt space (i.e. after applying the learnt metric tensor) between an individual with sample vector x_i and the learnt prototype w_{stable} along the vector separating w_{stable} and $w_{progressive}$ (**Methods: GMLVQ-Scalar Projection**).

A value of 1 indicates that a sample vector is incident to the pMCI prototype whereas a value of 0 indicates that a sample vector is incident to the sMCI prototype, and a value of 0.5 is the decision boundary separating the two classes within the binary classification framework. The scalar projection has a large positive value for pMCI individuals and zero or negative value for sMCI individuals.

Relating the scalar projection to individual rates of future cognitive decline

We used the GMLVQ-Scalar projection method for 253 MCI individuals (Development Sample I) to generate a cognitive scalar projection from baseline cognitive variables (GDS, ADNI-Mem, ADNI-EF), and a biological scalar projection from baseline biological variables (PLS Derived Grey matter score, β-amyloid, APOE 4). To test whether individual scalar projections relate to individual rates of future cognitive decline, we correlated (Pearson's correlation) the scalar projection (generated using baseline predictors) with the rate of future change in ADNI-Mem scores. We computed the rate of future cognitive change by fitting a linear model to the ADNI-Mem scores across multiple measurements (Development Sample I: mean=5.7, std=1 time points; mean=4, std=1.7 years, Longitudinal prediction validation sample: mean=5, std=1.7 time points; mean=4.4, std=1.5 years). The slope of the linear model

represents the rate of change in ADNI-Mem score. Individual scores higher than 2 standard deviations from the sMCI mean score or less than 2 standard deviations from the pMCI mean score were determined as outliers and excluded from further analysis.

Statistical Validation

Within-sample validation: k-fold cross-validation (GMLVQ)

Within-sample generalisation for the GMLVQ (Development Sample (b)) was assessed using k -fold cross validation (k=10). In brief, data was split into a series of training and test sets, whereby each sample point is present once in a test set. To select hyper-parameters for a given training sample, we used nested cross-validation. Within each fold, we used grid search in the parameter space; that is, we folded our training data into hyper parameter cross-validation test and training sets and searched across all possible combinations of hyper parameters. We selected the optimal set of hyper parameters for a given cross fold based on mean performance across hyper parameter cross-validation test sets. We then fixed the hyper parameters within each cross fold and train the models. To assess model generalisation performance, we averaged metrics (GMLVQ: Accuracy, Macro Averaged Error (MAE), True Positive (TP), True Negative (TN)) from the test set across all cross folds.

Resampling GMLVQ

We performed a resampling throughout the GMLVQ learning procedure to address class imbalance for the binary classification tasks. That is, within each cross fold, we down-sampled the majority class to the same number of samples as the minority class. We performed a random down-sampling of the data set 100 times respecting the class split defined, thus creating 100 classifiers forming an ensemble. To determine the class of a given test point we performed majority voting among the 100 classifiers in the ensemble. Similarly, we interrogated the efficacy of our selected variables to perform the classification task via the average metric tensor across all resamples and cross folds.

Within-sample validation: Random resampling (GMLVQ-Scalar Projection)

Within-sample generalisation for the GMLVQ-scalar projection framework (Development Sample I) was assessed using random resampling. In brief, we tested the relationship between the scalar projection and rates of future cognitive decline by randomly splitting our sample into test and training data 1000 times. To avoid biasing the model in the training phase due to class

imbalance in the data (majority class: sMCI= 113 vs. minority class: pMCI = 54), we resampled the data to generate balanced classes (i.e. number of sMCI equals number of pMCI individuals). This resampling process randomly selects half of the individuals in the minority class and the same number of individuals from the majority class as training data; with the remaining sample used as test data. The training data within a resampling was used in the GMLVQ framework to learn the metric tensor and prototype locations for each training set. We then calculated the GMLVQ-Scalar Projection for the corresponding test set and correlated these values (Pearson's correlation) with the rate of future ADNI-Mem change for the same individuals. We assessed within-sample generalisation based on the median of the correlation coefficients generated from the test sets across resampling using 95% confidence intervals.

Out-of-Sample validation: Longitudinal Predictions (GMLVQ-Scalar Projection)

To test the out-of-sample generalisability of the GMLVQ-Scalar Projection in predicting individual rates of future cognitive decline, we drew a validation sample with longitudinal information (i.e. cognitive data from three or more measurements) (Longitudinal prediction validation sample). As described above, we derived the grey matter score for this validation sample from the voxel weight matrix generated from the PLSr-RFE (**Experimental Chapter 1. Deriving an interpretable and interoperable score of Alzheimer's related atrophy**). To generate the scalar projections for the longitudinal validation sample, we used the metric tensor and prototype locations generated from the trained model with the median test performance on Development sample II. Using the metric tensor and prototype locations from the cognitive and biological models, we then generated a cognitive scalar projection and a biological scalar projection for each individual in the longitudinal validation sample. We then correlated these scalar projections with the rate of change in future ADNI-Mem scores (i.e. following baseline). Correlation coefficients were computed using skipped Pearson Correlation (Robust Correlation Toolbox; (Pernet, Wilcox, & Rousselet, 2013). This method accounts for potential outliers and determines significance using bootstrapped confidence intervals (CI) from 1000 permutations.

Comparing Correlations between samples:

To test if the relationship between the GMLVQ-Scalar Projection and rate of future cognitive decline is significantly different between Development sample II and the Longitudinal prediction validation samples we used Fisher's r to Z transformation. To compare if the relationship of the GMLVQ-Scalar Projection and rate of future cognitive decline is

significantly different between models using biological or cognitive data we generate a Steiger Z statistic (Steiger, 1980).

Results

Cognitive Classification Models for predicting sMCI vs pMCI

We tested whether a classification model that is based on the Generalised Metric Learning Vector Quantization (GMLVQ) framework and trained and tested on baseline cognitive data predicts progression from MCI to AD. In particular, we trained and tested both a linear and non-linear classifier to discriminate between sMCI and pMCI using cognitive data (Geriatric Depression Scale (GDS), ADNI Memory (ADNI-Mem) and ADNI Executive Function (ADNI-EF) from a sample of 167 MCI individuals (Development Sample II). We tuned the model with 2 hyper-parameters using nested cross validation and assessed its performance using 10-fold cross validation. The model successfully classified stable (sMCI; n=113) vs. progressive (pMCI; n=54) individuals [Accuracy: 81.4%, MAE: 17.6%, TP: 84.9%, TN 79.8%]. We obtained identical performance by increasing model complexity to a non-linear classifier by increasing the number of prototypes per class to two [Accuracy:81.4%, MAE:17.6%, TP:84.9%, TN 79.8%], and therefore selected the linear model for further analysis. Interrogating the average metric tensor **(Figure 5.1)** showed that the most predictive feature was ADNI-Mem (mean:0.55, std:+-0.12), while ADNI-EF (mean:0.35, std:+-0.09) and GDS (mean:0.1, std:+-0.05) had moderate and minor contributions to the classification task, respectively. These results suggest that the baseline ADNI-Mem score is the most discriminative cognitive feature for classifying sMCI vs. pMCI, as indicated by the diagonal terms in the metric tensor that are scaled to sum to one. Further, learning a metric in the input space of the classifier enables us to extend beyond the weighting of individual input features (such as ADNI-Mem score) and study the higher-order interplay between pairs of features with respect to the classification task. Interrogating the off diagonal terms of the metric tensor indicates that the interaction of GDS with ADNI-Mem or ADNI-EF is important for classifying sMCI vs. pMCI individuals. The positive off-diagonal terms indicate a positive interaction between the ADNI-Mem and ADNI-EF scores that group individuals from the same class. In contrast, the negative off diagonal terms indicate that the GDS score has a negative interaction with the ADNI-Mem and ADNI-EF scores and separate individuals into different classes. For example, individuals with similar baseline ADNI-Mem and ADNI-EF scores may be classified in different groups depending on their baseline GDS score, with higher scores likely reflecting affective disturbance and MCI comorbidity.

Figure 5.1 Cognitive model metric tensor

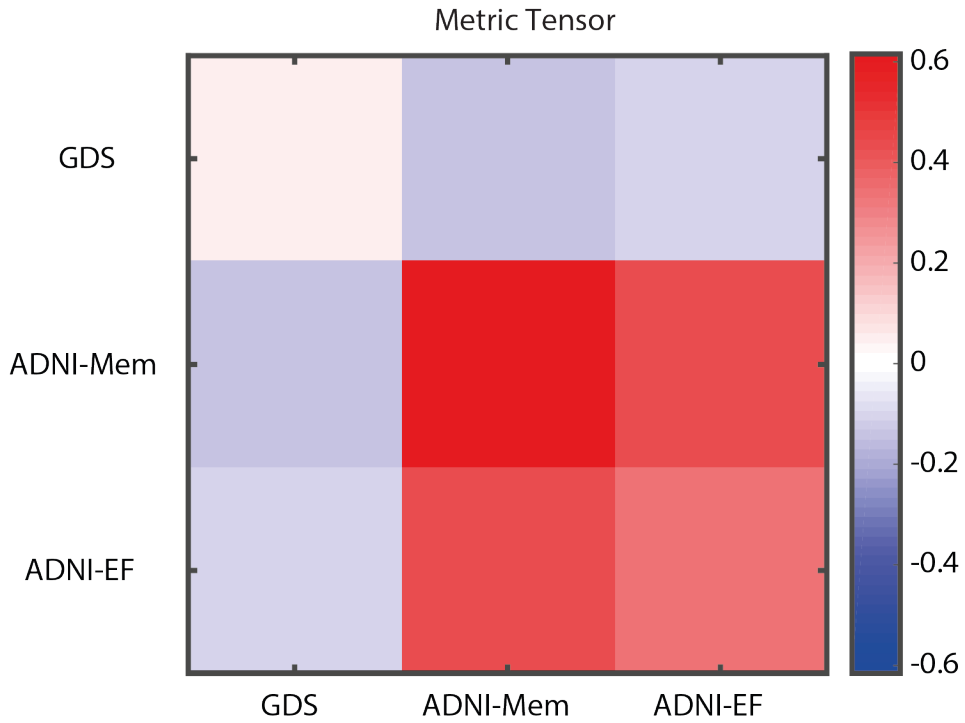


Figure 5.1*. Cognitive classification Model - Metric Tensor Metric tensor for the classification model (sMCI vs pMCI) generated using cognitive data (GDS, ADNI-Mem, ADNI-EF). The colour scale indicates the predictive value for each cell in the metric tensor, where diagonal terms sum to 1. The diagonal terms show strong contribution of the ADNI-Mem score. The positive off diagonals terms indicate a positive interaction between the ADNI- Mem and ADNI-EF scores. The negative off diagonals terms indicate the negative interaction of the GDS score with the ADNI-Mem and ANDNI-EF scores. Refer to* ***Methods: GMLVQ*** *for examples of GMLVQ and possible interpretations.*

Comparing the Performance of Biological vs. Cognitive Models

We next tested whether a classification model trained and tested on baseline biological data discriminates sMCI vs. pMCI. We developed a biological classification model of similar complexity to the cognitive model (i.e. linear classifier (1 prototype per class), 3 features, 2 hyper parameters) based on the same data sample (Development Sample II, n=167) using as predictors: PLS derived grey matter score, β-amyloid burden (measured by FBP-PET) and APOE 4 status (positive: presence of 1 or 2 APOE4 alleles, negative: no APOE4 alleles). The model successfully discriminated between sMCI vs. pMCI individuals [Accuracy: 81.9%,

MAE: 18.3%, True Positive: 81.1%, True Negative 82.3%]. We observed comparable classification performance when we increased the complexity of the biological model to a non-linear classifier (2 prototypes per class) [Accuracy: 80.7%, MAE: 19.2%, True Positive: 81.1%, True Negative: 80.5%]. The metric tensor of the model **(Figure 5.2)** indicates that the feature with the highest predictive value is baseline β-amyloid burden (mean:0.48, std:+-0.16), with similar contributions from baseline PLS derived grey matter (mean:0.28, std: +- 0.14) and APOE4 status (mean:0.24, std: +-0.10). Further, interrogating the off diagonal terms of the metric tensor indicated a positive interaction between baseline β-amyloid burden and APOE 4 status; that is baseline β-amyloid burden and APOE 4 status groups individuals from the same class. In contrast, we observed a negative interaction between baseline β-amyloid burden and the baseline PLS derived grey matter score; that is, the combination of these features separates sMCI from pMCI individuals. For example, individuals with high baseline β-amyloid burden and low baseline PLS derived grey matter score (i.e. low grey matter density in medial temporal areas) are grouped in separate classes (sMCI vs. pMCI) from individuals with high baseline PLS derived grey matter score (i.e. high grey matter density) and low baseline β-amyloid burden. Finally, we observed no significant differences (t-tests across cross folds) in classification performance between the cognitive and biological models (Accuracy: [t(9)=-0.13, P=0.90], MAE: [t(9)=0.17, P=0.87], True Positive: [t(9)=0.54, P=0.60], True Negative: [t(9)=-0.32, P=0.75]), suggesting that baseline cognitive and biological features contribute similarly to the binary classification of sMCI vs. pMCI individuals.

Figure 5.2 Biological model metric tensor

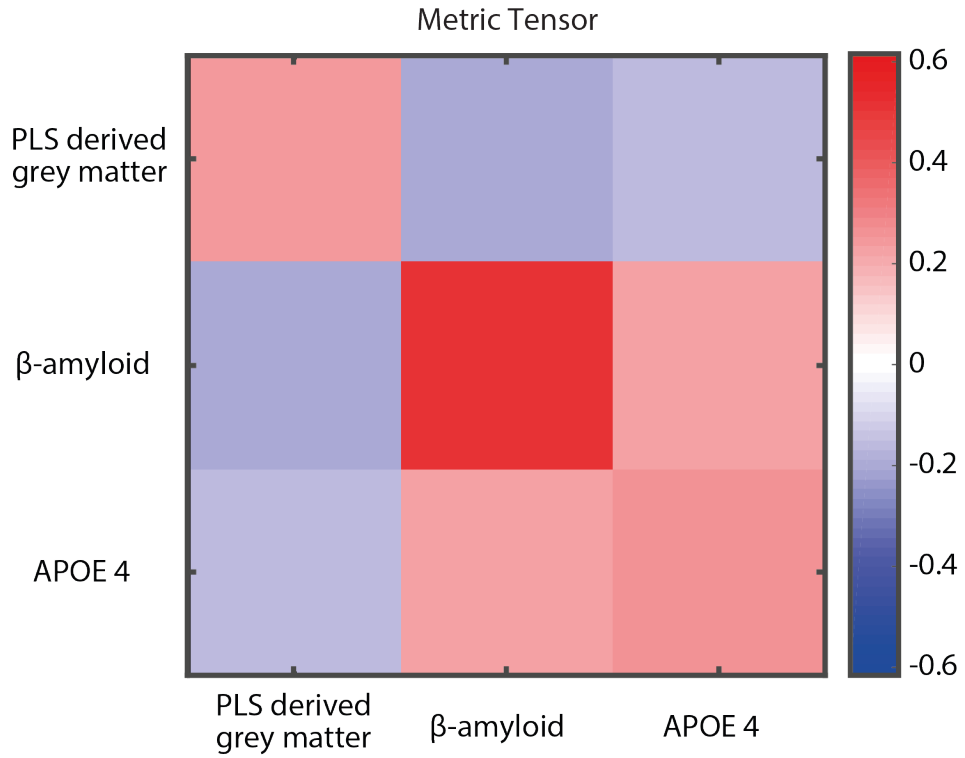


Figure 5.2. *Metric tensor for the classification model (sMCI vs pMCI) generated using biological data (PLS derived grey matter score, β-amyloid, APOE 4). The colour scale indicates the predictive value for each cell in the metric tensor, where diagonal terms sum to 1. The diagonal terms show strong contribution of β- amyloid. The positive off diagonals terms indicate a positive interaction between β-amyloid and APOE 4. The negative off diagonals terms indicate the negative interaction of the PLS derived grey matter score with both β-amyloid and APOE 4. Refer to* ***Methods: GMLVQ*** *for examples of GMLVQ and possible interpretations.*

Trajectory modelling: Predicting Individual Variability in the Rate of Future Cognitive Decline

Our analyses so far have focused on binary classifications (i.e. sMCI vs. pMCI). However, this approach is limited, as it assumes distinct patient classes and does not capture dynamic changes in disease progression over time. To extend beyond this binary framework, we developed a trajectory modelling approach by deriving a continuous metric based on a GMLVQ-scalar projection (i.e. distance of each MCI patient from the sMCI prototype) and using only baseline data. We then confirmed that this projection relates to individual variability in the rate of future

cognitive decline. In particular, we defined the rate of future cognitive decline as the rate of change in the ADNI-Mem scores across measurements following baseline, where baseline is defined as the date of the FBP-PET scan used as a predictor for deriving the GMLVQ-scalar projection. We focussed on change in memory performance as measured by ADNI-Mem, as a) memory decline has been shown to occur prior to decline in other cognitive domains in sporadic AD, b) our metric leaning model showed that ADNI-Mem was the most discriminative cognitive feature for the sMCI vs. pMCI classification compared to the other cognitive variables tested (GDS, ADNI-EF). For the same sample used in the binary classifications (Development Sample II) we observed that scalar projections derived from either the cognitive or the biological model account significantly for variance in the rate of future memory decline (**Figure 5.3**) (i.e. Cognitive: [r(165) = -0.41 (95% CI: [-0.51 -0.32]), $P < 0.0001$], Biological: [r(165) = -0.55 (95% CI: [-0.62 -0.47]), $P < 0.0001$]).

Figure 5.3 Correlation of GMLVQ-Scalar projections with rate of memory change

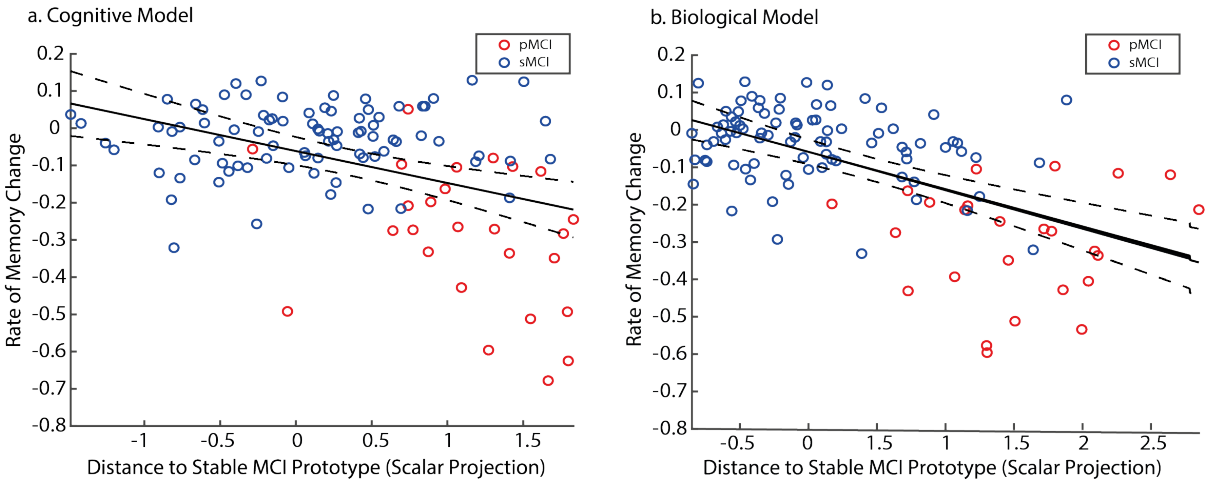


Figure 5.3. *Correlation of the GMLVQ-scalar projections derived from the a) cognitive model, b) biological model with the rate of ADNI-Mem change for Development Sample I. Red dots indicate pMCI individuals, blue dots indicate sMCI individuals. The central black line is the regression line for the fit of the GMLVQ- scalar projection to the rate of ADNI-Mem change; the dashed lines represent the 95% confidence intervals for this regression line. Data used to train the model (n = 52) were not used to test the relationship between the scalar projection and rates of future cognitive decline and are not shown here*

Validation of scalar projection: Future rate of ADNI-Mem change with incomplete (missing) data.

We next asked whether the GMLVQ-scalar projection approach could be implemented to predict future rate of cognitive decline for data with less than 3 years of clinical diagnosis. To incorporate these incomplete data from the model trained on Development Sample II, we chose the metric tensor and prototype positions from the model with the median test performance. We observed that scalar projections from both models (cognitive and biological) account for significant variance in the rate of future memory decline when the incomplete data was included in the test set (Note that data used to train the model with median test performance (n=52) were not used to test the relationship between the scalar projection and rates of future cognitive decline) (**Figure 5.4**). For the test sample used for the Cognitive model (Development Sample I) we observed a significant correlation between the cognitive scalar projection and the rate of future cognitive decline: [r(196) = -0.4 (95% CI: [-0.51 -0.29]), P < 0.0001] that remained significant when including only incomplete data ([r(85) = -0.31 (95% CI: [-0.48 -0.13]), P < 0.0001]). For the test sample used for the Biological model (Development Sample I) we observed a significant correlation between the biological scalar projection and the rates of future cognitive decline [r(196) =-0.51 (95% CI:[-0.6 -0.4]), P < 0.0001], that remained significant when including only incomplete data ([r(85) = (-0.45 (95% CI: [-0.59 -0.3]), P < 0.0001]). The observed r values for models incorporating incomplete data were comparable to the correlation values between the scalar projection and the rate of cognitive decline for individuals with complete clinical longitudinal information (Development Sample II: 3 years of clinical assessment) (Fisher's r to Z, Cognitive: [Z=-0.86, P=0.39], Biological: [Z=-1, P=0.32]). These findings suggest that the GMLVQ-scalar projection approach allows us to make longitudinal predictions based on the rate of future cognitive decline from a limited number of longitudinal clinical assessments (i.e. less than 3 years).

Figure 5.4: Correlating GMLVQ-Scalar projections with rate of memory change with incomplete data

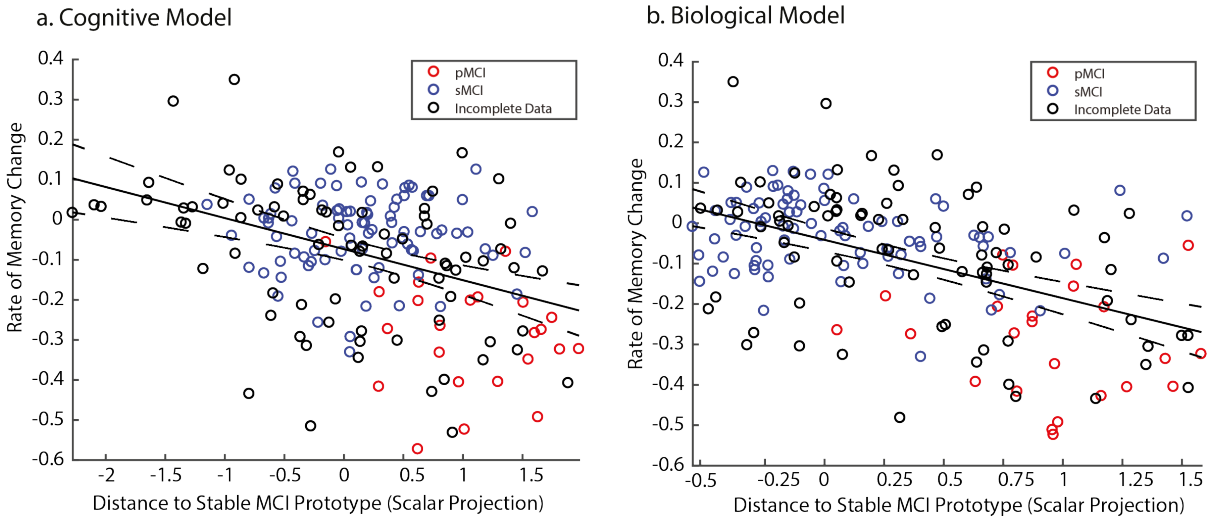


Figure 5.4 *shows the correlation of the GMLVQ-scalar projections derived from the a) cognitive model, b) biological model with the rate of ADNI-Mem change for Development Sample I. Red dots indicate pMCI individuals, blue dots indicate sMCI individuals and black dots indicate individuals with incomplete data. The central black line is the regression line for the fit of the GMLVQ-scalar projection to the rate of ADNI-Mem change; the dashed lines represent the 95% confidence intervals for this regression line. Data used to train the model (n=52) were not used to test the relationship between the scalar projection and rates of future cognitive decline and are not shown. Outliers identified by the Robust Correlation toolbox (cognitive: 11, biological=11) are not shown.*

Validation of scalar projection: Future rate of MMSE change.

To further validate the clinical relevance of our scalar projection approach, we tested whether the scalar projections derived from either the cognitive or biological model relate to the rate of future MMSE change, as MMSE is typically used in clinical diagnosis. We observed that the scalar projections correlated significantly with the future rate of change in the MMSE score (Cognitive: [$r(214) = -0.53$, $P < 0.0001$], Biological: [$r(213) = -0.43$, $P < 0.0001$] (**Figure 5.5**).

Figure 5.5: Correlating the scalar projections from cognitive and biological models with rate of MMSE change

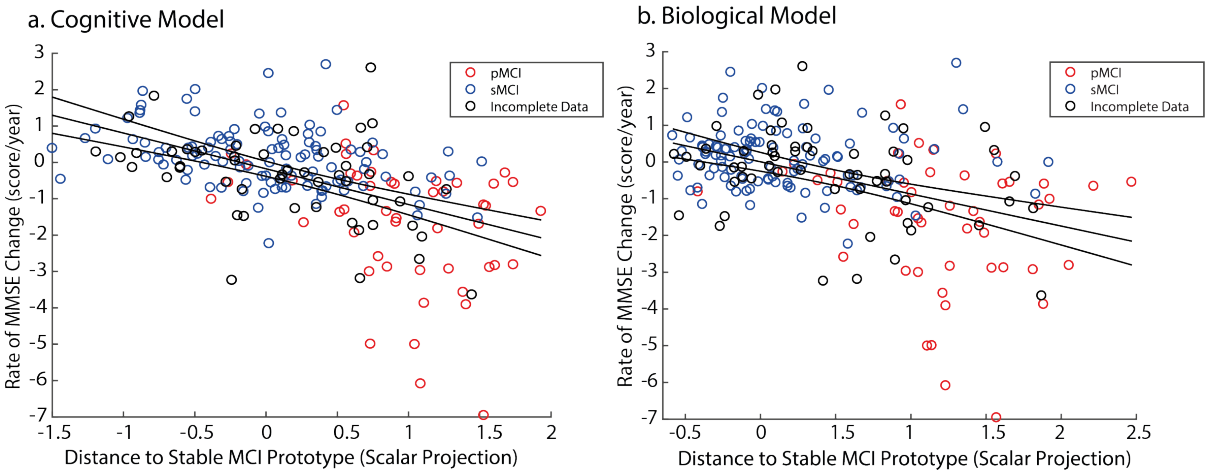


Figure 5.5 *shows the correlation of scalar projection derived from the a) cognitive model, b) biological model with the rate of MMSE change for the original data sample. The central black line is the regression line for the fit of the GMLVQ-scalar projection to the rate of MMSE change; the dashed lines represent the 95% confidence intervals for this regression line. Red circles indicate pMCI individuals, blue circles sMCI individuals and black dots for the incomplete data (i.e. less than three years of clinical assessments). Note: 23 Individuals from the previous analysis had fewer than 3 MMSE tests from baseline.*

Validating the scalar projection: Out of sample.

Next, we validated the relationship of the GMLVQ-scalar projection with cognitive decline for individuals with MCI against a new independent validation data sample (Longitudinal prediction validation sample). To calculate the scalar projections, we chose the metric tensor and prototype positions from the cognitive or biological models with the median test performance across resampling using the Development Sample II. To generate a baseline PLS derived grey matter score for the Longitudinal prediction validation sample, we multiplied the voxel weights matrix determined by PLSr-RFE on the Development sample (**Experimental Chapter 1: Deriving an interpretable and interoperable score of Alzheimer's related atrophy**) with grey matter density from baseline structural scans for the longitudinal prediction validation sample (i.e. data not used for the PLSr-RFE feature generation). We observed a significant correlation of the scalar projection with the rate of future ADNI-Mem change for cognitive data [r(124) = -0.4, (95% CI: [-0.55 -0.25]), $P <$ 0.0001]. This relationship remained significant when we controlled for: Age; [r(123) = -0.32,

(95% CI: [-0.47 -0.14]), P= 0.0003], Gender; [r(123) = -0.4, (95% CI: [-0.55 -0.22]), P< 0.0001], or Education; [r(123) = -0.4, (95% CI: [-0.54 -0.23]), P< 0.0001]). Further, we observed a significant correlation of the scalar projection with the rate of future ADNI-Mem change for biological data [r(124) = -0.68, (95% CI: [-0.76 -0.58]), $P < 0.0001$] (**Figure 5.6)**. This relationship remained significant when we controlled for: Age; [r(123) = -0.57, (95% CI: [-0.67 -0.46]), P< 0.0001], Gender; [r(123) = -0.63, (95% CI: [-0.71 -0.53]), P< 0.0001], Education; [r(123) = -0.63, (95% CI: [-0.73 -0.53]), P< 0.0001]. This relationship was not significantly different between Development Sample I vs. Longitudinal prediction validation samples (Fisher's r to Z, Cognitive model: [Z=-0.1, P=0.92], Biological model: [Z=-1.76, P=0.08]). Further, correlations between the scalar projection and the rate of future memory decline were significantly stronger for biological compared to cognitive models (Steiger's Z, [Z=-3.86, P<0.0001]). This difference between models remained significant when we controlled for Age; (Steiger's Z, [Z=-3.33, P=0.0004]), Gender; (Steiger's Z, [Z=-3.57, P=0.0002]), or Education; (Steiger's Z, [Z=-3.57, P=0.0002]). Taken together these findings suggest that the biological model explains significantly larger variance in the rate of future memory decline than the cognitive model.

Figure 5.6 Correlating scalar projections from cognitive and biological models with rate of memory change in an independent sample

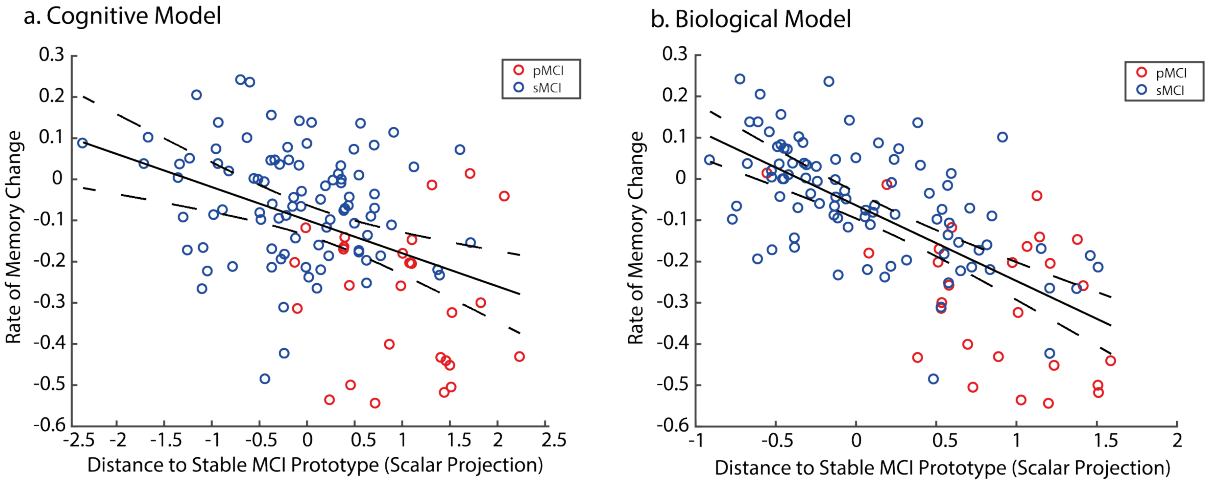


Figure 5.6. *Out-of-sample Validation Correlation of scalar projection derived from the a) cognitive model, b) biological model with the rate of ADNI-Mem change for the longitudinal validation data set. Red dots indicate pMCI individuals, blue dots sMCI individuals. The central black line is the regression line for the fit of the GMLVQ-scalar pro- jection to the rate of ADNI-Mem change; the dashed lines represent the 95% confidence intervals for this regression line. Outliers identified by the Robust Correlation toolbox (cognitive n = 7, Biological n = 8) are not shown for illustrative purposes. Note that this validation sample includes data from 3 β- amyloid negative pMCI individuals who had a scalar projection of less than 0.25 (i.e. very close to the sMCI prototype). Investigating the relationship of the scalar projection to future cognitive decline for these individuals showed dissociable cognitive trajectories from most pMCI individuals.*

Finally, we tested whether our trajectory modelling approach delivers stronger predictions when including non-invasively measured data modalities to the basic baseline cognitive model. Adding the baseline PLS derived grey matter feature and APOE 4 status to the cognitive model showed a substantial increase in the variance in the rate of future memory change explained by the scalar projection (**Table 5.1**). These results suggest that predicting the rate of future cognitive decline is enhanced by adding non-invasively measured baseline biological features to baseline cognitive data.

Table 5.1 Correlations of scalar projections with the rate of ADNI-Mem change for models based on cognitive and / or biological data

Data Type	Pearson's r [95% C.I] Cross Validation	Pearson's r Out-of-Sample Validation
Biological: (GM+APOE4+ β-Amyloid)	-0.55 [-0.66 -0.53]	-0.68 [-0.76 -0.58]
Cognitive: (GDS+ADNI-Mem+ADNI-EF)	-0.41 [-0.5 -0.30]	-0.4 [-0.55 -0.25]
Cognitive+GM	-0.46 [-0.52 -0.34]	-0.49 [-0.61 -0.35]
Cognitive+APOE 4	-0.47 [-0.55 -0.38]	-0.48 [-0.61 -0.33]
Cognitive+GM+APOE 4	-0.5 [-0.57 -0.42]	-0.53 [-0.64 -0.4]
ALL Variables: Cognitive + Biological	-0.57 [-0.63 -0.49]	-0.64 [-0.73 -0.52]

Table 5.1. *Correlations of scalar projections with the rate of ADNI-Mem change for models based on cognitive and / or biological data. Pearson's correlation coefficients are shown for Development Sample (b) based on cross-validation and the independent data used for out of sample validation (longitudinal validation sample).*

Discussion

Despite the importance of early diagnosis of AD for clinical practice and treatment, we still lack robust tools for predicting individual progression to dementia. The multimodal longitudinal measurements across large-scale samples available in ADNI provide a testbed for machine learning approaches that generate predictive features and discriminate between patient groups (Weiner et al., 2015, 2017). Here, we propose a novel trajectory modelling approach based on an integrated feature generation (PLSr-RFE) and classification (GMLVQ-scalar projection) methodology that predicts individual disease trajectories based on continuous measures of cognitive decline. Our modelling approach is in line with the current 2018 NIA-AA research framework that defines AD as a continuum and advances the state-of-the-art and clinical validity of machine learning applications to the prediction of dementia due to AD in the following main respects.

First, we successfully predict whether individuals will progress from MCI to dementia due to AD, employing a transparent machine learning approach (i.e. prototype based classifier with metric learning and linear decision boundary) trained on informative and interpretable baseline cognitive data. We show that baseline composite scores related to memory and executive function (ADNI-Mem, ADNI-EF composite score) are highly predictive of disease progression. The high cross-validated classification performance of our model is in line with previous studies showing that similar neuropsychological data are predictive of MCI progression to dementia due to AD (Belleville et al., 2017; Chapman et al., 2011; Pereira et al., 2018, 2017; Silva et al., 2013; Tabert et al., 2006). Further, we demonstrate a negative interaction between baseline cognitive (memory, executive function) and affective scores that separates individuals into different classes, with higher baseline affective scores potentially reflecting MCI comorbidity. Previous studies have shown that moderate to severe depressive symptoms (i.e. GDS> 15) are predictive of MCI conversion to AD (Defrancesco et al., 2017), while mild depressive symptoms do not increase conversion risk (Chen et al., 2008; Defrancesco et al., 2017). Here, we show that the interaction between scores that are indicative of mild depression (i.e. GDS<10) and memory dysfunction discriminates stable from progressive MCI individuals. Thus, our metric learning approach on multi domain data (i.e. cognitive and affective measurements) may provide a means of reducing MCI patient misclassification due to comorbidity (e.g. affective disturbance).

Second, our trajectory modelling approach (GMLVQ-Scalar Projection) extends beyond binary patient classification approaches (Rathore et al., 2017) that are poorly constrained.

Recent methodological frameworks for mining neuroimaging data (Jollans et al., 2019) and predicting progression to AD (Samper-González et al., 2018) have focused on binary classifications that are based on discrete clinical labels (i.e. stable vs. progressive MCI), as determined by arbitrary criteria (e.g. within a 3 year period of clinical assessment). As a result, these approaches are limited by risk of patient misclassification. That is, patients at the class boundary with different disease trajectories may be classed in the same MCI group (e.g. a patient who progresses to AD within 1 day from clinical assessment and a patient who converts in 3 years will be classified as pMCI). Similarly, patients with similar disease trajectories may be classified in different MCI groups (e.g. a patient who converts in 3 years will be classified as pMCI, while a patient who remains stable for 3 years and progresses to dementia 1 day after the clinical assessment will be classified as sMCI). To overcome this limitation and make meaningful predictions in AD, modelling approaches need to capture continuous information in prognostic trajectories and consider target uncertainty (i.e. the future clinical diagnosis) (for review (Janssen et al., 2018)). Although recent time-to-event models (e.g. survival analysis models predicting time to conversion) (Alsaedi et al., 2018; Casanova et al., 2013; Desikan et al., 2010; Jack, Jr, et al., 2010; S. M. Landau et al., 2010; K. Liu et al., 2017; Michaud et al., 2017; Oulhaj et al., 2009; Young et al., 2014) capture continuous information in patient trajectories they are limited by target uncertainty; that is, estimating the exact time to conversion is limited by the frequency of clinical follow-ups and poor inter-rater reliability (i.e. diagnoses may differ across clinicians).

Our trajectory modelling approach predicts future ADNI-Mem scores based on baseline data, allowing us to capture individual disease trajectories and reducing the risk of patient misclassification. In particular, we derive continuous prognostic scores of individual cognitive decline (i.e. scalar projection) by training the model based on 'noisy' diagnostic labels (i.e. patient classes that are poorly defined e.g. sMCI vs pMCI). As our metric learning model has limited freedom (linear low-parameter model), separating continuous target values (i.e. individualised cognitive trajectories) into two broad classes (sMCI vs. pMCI) forces the model to extract key underlying structures in the data that distinguish between target values, ignoring subtle differences in target values. Further, employing separate feature generation (i.e. PLSr-RFE) and classification (GMLVQ scalar projection) stages allows us to interrogate interpretable predictive features of progression to AD and derive predictions that generalise to patient data from independent samples from the model development sample. This is in contrast to deep learning methods that require large training samples and are shown to be difficult to

interpret and generalise (Christos Davatzikos, 2019), raising questions about the clinical utility of these approaches (for review (Topol, 2019)).

Comparing our trajectory modelling methodology to binary classifications on the same data (i.e. cognitive vs. biological) shows dissociable results. A binary metric learning algorithm shows similar performance in the binary classification of MCI subgroups (sMCI vs. pMCI) when trained on baseline cognitive vs. biological data. In contrast, the scalar projection derived from biological data explains significantly higher individual variability in the rate of future cognitive decline than the scalar projection derived from cognitive data. Further, we demonstrate that the predictive power of our trajectory modelling methodology is enhanced when including non-invasively measured baseline biological data in addition to baseline cognitive data. Although our model shows high accuracy of cognitive decline when trained on cognitive data, there is a substantial gain in predictive efficacy when adding baseline data on APOE 4 status or grey matter density (PLS derived grey mater scores). This is consistent with previous studies (Dukart et al., 2015; Vogel et al., 2018; Zhou et al., 2018) showing enhanced prediction of time to AD conversion when including biological compared to neuropsychological data alone. However, when investigating the additive effect of adding cognitive variables to the biological model of grey matter, APOE4 and β-Amyloid we did not see any improvement in modelling future cognitive change.

Previous work on trajectory modelling has focused on discretising continuous values (i.e. future change in cognitive scores) into latent classes that are then used as outcome measures in classification models (Bhagwat et al., 2018; Hochstetler et al., 2015; Y. Wang et al., 2019; Wilkosz et al., 2010). For example, previous studies (Bhagwat et al., 2018) used machine learning (i.e. longitudinal Siamese neural-network) to fuse baseline and follow up imaging and clinical scores to predict whether individuals will decline fast or slow (based on MMSE scores) or fast, moderate or slow (based on ADAS-cog). Further studies (Hochstetler et al., 2015) used classification and regression trees on baseline demographic, lifestyle, cognitive and biological data to classify individuals in three latent classes (fast, medium or slow decline) with similar growth patterns of cognitive and functional changes, while others (Wilkosz et al., 2010) used latent class trajectory models to derive six different trajectories for cognitive and behavioural decline due to AD. However, the generalisability and interoperability of these approaches have been recently questioned (Y. Wang et al., 2019). Our trajectory modelling approach differs from this previous work, as it avoids assumptions related to discretising continuous values. In particular, we derive a continuous metric (i.e.

scalar projection) from a discrete classification model (i.e. metric learning) that predicts individual rates of future cognitive change (i.e. change in ADNI-Mem). Finally, our approach is in line with previous work predicting exact changes in MMSE or ADAS-Cog scores (Y. Fan et al., 2008; Zhang et al., 2012). In particular, previous studies used baseline and follow-up structural MRI and FDG-PET data to predict future scores on cognitive tests at different time intervals (Zhang et al., 2012), or structural MRI to predict the rate of change in MMSE scores (Y. Fan et al., 2008). Our modelling approach differs from this previous work in fusing baseline multimodal data into a single metric (i.e. scalar projection) to predict future rates of cognitive change (i.e. change in ADNI-Mem, or MMSE scores). In this work we derived the rate of cognitive change using the linear least squares fit of ADNI-Mem over time. However, as the proposed AD cascade suggests that cognitive change associated with AD follows a sigmoidal shape over time, the hypothesised rate of change will follow an inverted U shape (Jack, Knopman, et al., 2010). In this work we only investigated the relationship of the scalar projection with rate of cognitive change for individuals who are MCI and therefore cognitive change over time is more likely to be linear than sigmoidal (Jack et al., 2013; Jack, Knopman, et al., 2010). However, a more complex fit may be preferred when attempting to predict cognitive change for a population including cognitively normal and demented individuals. Further, although we show no difference between the sample with incomplete (i.e less than three years of follow-up) and complete (i.e. three or more years of follow up) longitudinal follow-up, we did not control for the variability due to different duration of follow-ups. In future work we can account for this variability in follow-up length in a more principled way using linear mixed effects models.

In sum, we propose a robust methodology based on modelling multimodal data that determines predictive and interpretable markers of individual variability in progression to dementia due to AD. Although our investigations have focused on amnestic MCI, our methodology has the potential to be extended to predict individual disease trajectories specific to AD subtypes, following recent work modelling neuroimaging data (Dong et al., 2016; Young et al., 2018). Further, previous work on preclinical populations has investigated the role of grey matter atrophy and cortical amyloid burden in future cognitive decline (Bilgel et al., 2018; Burnham, Bourgeat, Doré, Savage, Brown, Laws, Maruff, Salvado, Ames, Martins, Masters, Rowe, & Villemagne, 2016; Dumurgier et al., 2017; Insel et al., 2015). Extending our trajectory modelling approach to preclinical populations using multimodal data has high clinical relevance, especially as clinical trials are moving towards less severely affected individuals who are unlikely to progress over the short time scales of clinical trials (Bilgel et al., 2014,

2017; Grober et al., 2008; Mormino et al., 2014). Thus, our approach has strong potential to deliver tools of high clinical relevance that reduce patient misclassification and facilitate effective stratification of individuals to prognostic or treatment pathways and clinical trials based on individualised rates of cognitive decline.

6. Experimental Chapter 3: Predicting future regional tau accumulation in early Alzheimer's disease

Introduction

Preclinical Alzheimer's Disease (AD) is a protracted process with multiple pathophysiological events occurring well before clinical manifestations (Dubois et al., 2016). Early phases involve molecular interactions between the β-amyloid and tau proteins, providing an opportunity to establish a more mechanistic and precise understanding of events that lead to the development of AD over the long term. One approach to tracking these events is the utilisation of machine learning to build predictive models of pathophysiological change in the earliest stages of AD. With the availability of PET scanning to image both β-amyloid and tau pathology in the brain, associations between AD related biomarkers can be examined in-vivo (for reviews: (B. Hall et al., 2017; W. Jagust, 2018; Schöll et al., 2019). These studies have shown that AD related cognitive impairment is dependent on cortical β-amyloid accumulation that facilitates the spread of tau into the neocortex resulting in neurodegeneration (L. Wang et al., 2016). In the absence of β-amyloid the presence of tau in the medial temporal lobe is insufficient to initiate widespread pathological neurodegeneration (Hanseeuw et al., 2019; L. Wang et al., 2016). In addition, recent evidence linking the apolipoprotein E gene and tau suggests that the presence of the APOE4 allele worsens tau related neurodegeneration independent of β-amyloid (Shi & Holtzman, 2018; Shi et al., 2017; van der Kant, Goldstein, & Ossenkoppele, 2020). Given this model of the AD pathological cascade (Jack et al., 2013; Jack, Knopman, et al., 2010), the staging of AD has shifted from a clinical syndromic diagnosis (Albert et al., 2011; McKhann et al., 2011) to a continuum of biomarker characteristics (Jack, Bennett, et al., 2018). In this framework, cognitively unimpaired individuals with evidence of β-amyloid and pathological tau accumulation are defined as having preclinical AD, and those with mild cognitive impairment (MCI) and similar biomarkers have AD. As a result, there is now a greater need for models that predict longitudinal change in pathological biomarkers rather than simply a change in clinical labels based on syndromic diagnosis.

Here, we utilised machine learning to quantify the multivariate relationships between the major factors underlying the pathogenesis of AD: APOE 4 genotype, β-amyloid, tau and neurodegeneration. Recently we developed a machine learning approach that derives a single prognostic index from APOE 4 genotype with measures of β-amyloid derived from either [^{18}F]-florbetapir (FBP) or [^{11}C]-PiB PET scans and a continuous measure of medial temporal atrophy derived from structural MRI (**Methods: GMLVQ-Scalar projection**). Previously we used this approach to show that the model derived prognostic index (scalar projection) is predictive of individualised rates of future memory change in MCI patients (**Experimental Chapter 2: Modelling prognostic trajectories of cognitive decline due to Alzheimer's disease**). Here, we extended this to preclinical AD samples by using our machine learning framework to test whether the scalar projection derived from baseline data can classify and stage such individuals based on future pathological tau accumulation.

Recent converging evidence highlights that patterns of tau spread (measured in-situ by longitudinal FTP-PET) are robust across preclinical AD cohorts (Jack, Wiste, et al., 2018; Pontecorvo et al., 2019; Schultz et al., 2018). This stereotypical spreading pattern for preclinical AD (i.e. β-amyloid positive individuals who are cognitively unimpaired) shows that tau initially accumulates within the temporal cortex then spreads to the superior and medial regions of the parietal cortex prior to cognitive impairment (Jack, Wiste, et al., 2018; Schultz et al., 2018).

We hypothesised that individuals classified as early AD using our machine learning approach would have the characteristic pattern tau accumulation (measured by longitudinal FTP-PET) (Jack, Wiste, et al., 2018; Pontecorvo et al., 2019; Schultz et al., 2018). Further, we anticipated that our modelling approach using multimodal biomarkers (i.e. APOE4, β-amyloid PET and MRI) will require fewer patients than β-amyloid alone to detect a clinically meaningful change in future tau accumulation. Finally, based on our previous work we believed that individual variability in the scalar projection relates to individual variability in regional future tau accumulation.

Our results show that the characteristic spreading pattern of tau in preclinical AD is predictable using baseline non-tau biomarkers, particularly when stratifying groups using multimodal data. Further, we show in an independent sample of cognitively unimpaired community dwelling individuals that the scalar projection predicts individualised rates of future tau accumulation with striking accuracy and regional specificity.

Materials and Methods

Study Design and Participants:

Three separate cohorts were used to generate and test predictive models of regional future tau accumulation. Two cohorts were drawn from the ADNI database: ADNI2/GO and ADNI 3 (adni.loni.usc.edu). ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. A major goal of ADNI has been to examine biomarkers including serial magnetic resonance imaging (MRI), and positron emission tomography (PET), with clinical and neuropsychological assessment to predict outcomes in mild cognitive impairment (MCI) and early Alzheimer's disease (AD).

A third validation cohort was taken from the Berkeley Aging Cohort Study (BACS). This cohort is comprised of community-dwelling cognitively intact elderly individuals with a Geriatric depression scale (GDS) (Yesavage, 1988) score ≤10, Mini mental status examination (MMSE) (Folstein, Folstein, & McHugh, 1975) score ≥25, no current neurological and psychiatric illness, normal functions on verbal and visual memory tests (all scores ≥−1.5 SD of age-adjusted, gender-adjusted, and education-adjusted norms) and age of 60–90 (inclusive) years. All subjects underwent a detailed standardised neuropsychological test session and neuroimaging measurements, all of which were obtained in close temporal proximity with follow up every 1 to 2 years.

Data from 488 individuals from ADNI 2/GO were used to train the machine learning model. Individuals were placed into three categories based on their baseline and longitudinal clinical diagnosis, with baseline defined as the evaluation closest to the first florbetapir (FBP) PET scan acquired in ADNI: **Demented** (n=181, 158 amyloid positive at baseline): individuals have a baseline diagnosis of demented; **Stable Condition (SC) (n=145,** 34 amyloid positive at baseline**):** individuals have a baseline diagnosis of cognitively normal and retain this diagnosis at follow up for 3 or more years; **Early Alzheimer's Disease (EAD)** (n=162, 135 amyloid positive at baseline): individuals have un unstable diagnosis of cognitively normal (n=18) or MCI (n=144) and have either reverted from or progressed to dementia throughout their enrolment in ADNI (i.e. either reverted (n=81) or progressed to dementia (n=81)). We included individuals in the EAD group who were MCI at baseline but have received a diagnosis of demented prior to baseline in this group as we anticipate they are likely affected by AD pathology but are at an earlier stage of AD than the Demented group. Further, as our machine

learning model is designed with limited freedom (only a few parameters) when training using noisy diagnostic labels it is optimised to account for target uncertainty without leading to over-fitting.

Data from 115 individuals from ADNI 3 were used to test the relationship between the scalar projection and regional future tau accumulation. These individuals were either cognitively normal (n=72) or MCI (n=43) at baseline (defined as the diagnosis closest to the first flortaucipir (FTP) PET scan acquired in ADNI 3) and have at least one follow-up FTP PET scan.

Data from 56 community dwelling individuals from BACS were used to test the accuracy of predictions of regional future tau accumulation. These individuals were cognitively normal (n=56) at baseline (defined as the diagnosis closest to the first FTP PET scan acquired in BACS) and have at least one follow-up FTP PET scan.

Brain Imaging data

MRI Acquisition

Structural MRIs for the ADNI samples were acquired at ADNI-GO, ADNI-2 and ADNI-3 sites equipped with 3 T MRI scanners using a 3D MP-RAGE or IR-SPGR T1-weighted sequences, as described online (http://adni.loni.usc.edu/methods/documents/mri-protocols). Structural MRIs for the BACS sample were collected on either a 1.5T or 3T MRI scanner at the Lawrence Berkeley National Laboratory (LBNL) or UC Berkeley using 3D MP-RAGE T1- weighted sequences. All ADNI and BACS scans were acquired with voxel sizes of approximately 1mm X 1mm X 1mm. MRI data were used for quantitation of the PET data.

PET Acquisition

PET imaging was performed at each ADNI site according to standardised protocols. The FBP-PET protocol entailed the injection of 10 mCi with acquisition of 20 min of emission data at 50-70 min post injection. The FTP-PET protocol entailed the injection of 10 mCi of tracer followed by acquisition of 30 min of emission data from 75-105 min post injection.

For the BACS, PIB PET scans were collected at LBNL. After ~15 mCi tracer injection into an antecubital vein, dynamic acquisition frames were obtained in 3D acquisition mode over a 90 min measurement interval (4 × 15 s frames, 8 × 30 s frames, 9 × 60 s frames, 2 × 180 s frames, 8 × 300 s frames, and 3 × 600 s frames) after x-ray CT. FTP PET scans were collected following

an injection of ~10 mCi of tracer in a protocol identical to that used for ADNI. All BACS participants were studied on a Siemens Biograph PET/CT.

Imaging Analysis-MRI: Medial Temporal Grey Matter Atrophy

Structural scans were segmented into grey matter, white matter and Cerebrospinal Fluid (CSF). The DARTEL toolbox (Ashburner, 2007) was then used to generate a study specific template to which all scans were normalised. Following this, individual grey matter segmentation volumes were normalised to MNI space without modulation. The unmodulated values for each voxel represent grey matter density at the voxel location. All images were then smoothed using a 3mm3 isotropic kernel and resliced to MNI resolution 1.5x1.5x1.5 mm voxel size. All structural MRI pre-processing was performed using Statistical Parametric Mapping 12 (http://www.fil.ion.ucl.ac.uk/spm/).

To generate a single index of medial temporal atrophy we used a voxel weights matrix that was previously derived to generate an interpretable and interoperable disease-specific biomarker (**Experimental Chapter 1: Deriving an interpretable and interoperable score of Alzheimer's related atrophy)**. In brief, a feature generation methodology (partial least squares regression with recursive feature elimination (**Methods: PLSr Recursive Feature Elimination (PLSr-RFE)**) was used to apply a decomposition on a set of predictors (T1-weighted MRI voxels) to create orthogonal latent variables that show the maximum covariance with the response variable (memory score). Further, we performed recursive feature elimination by iteratively removing predictors (voxels) that have weak predictive value. The PLSr-RFE procedure results in a voxel weights matrix that is used to calculate a single score of AD related medial temporal atrophy. This index of medial temporal atrophy has been shown to predict memory deficits, relate to individual tau burden and discriminates stable MCI and progressive MCI individuals. To generate an individual's score of medial temporal atrophy we performed a matrix multiplication of the previously derived voxel weights matrix and each subject's pre-processed T1 weighted MRI scans.

Imaging Analysis (ADNI)-PET: FBP (Florbetapir PET) β-Amyloid

FBP data were realigned, and the mean of all frames was used to co-register FBP data to each participant's structural MRI. Cortical Standardised Uptake Value Ratios (SUVR)s were generated by averaging FBP retention in a standard group of ROIs generated using FreeSurfer (v5.1) parcellations (lateral and medial frontal, anterior and posterior cingulate, lateral parietal, and lateral temporal cortical grey matter) and dividing by the average uptake from a composite

reference region (including the whole cerebellum, pons/brainstem, and eroded subcortical white matter regions) to create an index of global cortical FBP burden (β-amyloid) for each subject (Susan M Landau et al., 2015).

Imaging Analysis (BACS)-PET: PiB (Pittsburgh Compound B) β-Amyloid

Distribution volume ratios (DVRs) were generated with Logan graphical analysis on the aligned PIB frames using the native-space grey matter cerebellum as a reference region, fitting 35–90 min after injection.

For each subject, a global cortical PIB index was derived from the native-space DVR image coregistered to the MRI using FreeSurfer (5.1) parcellations using the Desikan-Killiany atlas (Desikan et al., 2006) to define frontal (cortical regions anterior to the precentral sulcus), temporal (middle and superior temporal regions), parietal (supramarginal gyrus, inferior/superior parietal lobules, and precuneus), and anterior/posterior cingulate regions-ROIs combined as a weighted average. There was no partial volume correction performed.

Image analysis-PET: FTP (Flortaucipir PET) tau:

FTP data were realigned and the mean of all frames used to coregister FTP to each participant's MRI acquired closest to the time of the FTP-PET. FTP SUVR images were generated by dividing voxel wise FTP uptake values by the average value within a mask of eroded subcortical white matter regions (Harrison et al., 2019). MR images were segmented and parcellated into 72 ROIs taken from the Desikan-Killany atlas using Freesurfer (V5.3). These ROIs were then used to extract from the normalised FTP-PET images regional SUVR data. Left and right hemisphere ROIs were averaged to generate 36 ROIs for further analysis. We calculated the future annualised rate of tau accumulation for each of the 36 ROIs either by taking the difference between the follow-up and baseline FTP-PET scans divided by the time interval in years from baseline (when only 2 FTP scans were taken), or fitting a linear least squares fit to 3 or more FTP-PET scans and extracting the parameter estimate for the slope of the ROI SUVR vs. time in years from baseline (when 3 or more FTP scans were taken). In the ADNI 3 sample the average time between FTP-PET scans is 1.22 +- std: 0.38 years with the number of follow-up FTP-PET scans n (2 FTP-PET scans) =93, n (3 FTP-PET scans) =17, n (4 FTP-PET scans) =5. In the BACS cohort the average time between FTP-PET scans is 1.8 +- std:0.65 years with the number of follow-up FTP-PET scans n (2 FTP-PET scans) =37, n (3 FTP-PET scans) =19.

Model Predictors and outcomes

Three baseline biological markers related to AD were used as predictors to generate the scalar projection from the machine learning model: a) Cortical amyloid burden (β-amyloid) measured using either FBP (ADNI) or PiB (BACS) PET, b) medial temporal atrophy derived from the T1 weighted structural MRI and c) APOE 4 genotype. Previously, we have shown when trained on these baseline data our machine learning approach can predict future changes in cognition for individuals diagnosed as MCI (**Experimental Chapter 2: Modelling prognostic trajectories of cognitive decline due to Alzheimer's disease)**. Here, we use the same baseline variables to generate predictions for early AD populations (i.e. individuals who are cognitively normal or MCI at baseline).

The primary outcome measure for the predictive models is regional future annualised rate of tau accumulation (SUVR/year). A secondary outcome measure is changes in future cognition over the same time scale as the longitudinal FTP scans, as measured by the Preclinical Alzheimer's Cognitive Composite (PACC).

To test changes in future cognition for individuals from ADNI 3 we used the previously derived ADNI-PACC measure (adni.loni.usc.edu) (Donohue et al., 2014). Of the 115 ADNI 3 individuals with multiple FTP-PET scans 102 individuals had multiple measures of the PACC over a similar time period (within 6 months of the baseline FTP-PET and the final FTP-PET scan). Future annualised change in PACC is calculated by either taking the difference between the follow-up and baseline PACC scores divided by the time interval in years from baseline (when only 2 PACC scores are available), or fitting a linear least squares fit to 3 or more PACC scores and extracting the parameter estimate for the slope of the PACC vs time in years from baseline (when 3 or more PACC scores are available). The average time between PACC testing sessions scans is 1.04 +- std: 0.44 years with the number of follow-up PACC testing sessions n (2 PACC sessions) =82, n (3 PACC sessions) =18, n (4 PACC sessions) =2.

Prediction Models:

Generalised Metric Learning Vector Quantization (GMLVQ)-Scalar Projection: We previously developed a machine learning approach based on the GMLVQ classification framework: GMLVQ-Scalar Projection (**Methods: GMLVQ-Scalar Projection)**. This approach allows us to derive a continuous prognostic metric by training a model based on diagnostic labels. In brief, we trained a metric learning model (GMLVQ) with baseline

multimodal data (medial temporal grey matter atrophy, β-amyloid, APOE 4 genotype). GMLVQ operates in a supervised manner to iteratively modify class-specific prototypes and learn boundaries between classes. For each training example, the closest prototype of each class is determined, these prototypes are then updated so that the prototype defining the same class is moved towards the training example and other prototype(s) representing different class(es) are moved further away. We have previously shown that this GMLVQ modelling approach classifies MCI patients into subgroups (progressive vs. stable) with high specificity and sensitivity. Extending the binary model, we derived a single prognostic distance measure (scalar projection) that separates individuals based on their future diagnosis. Previously we calculated the scalar projection to separate individuals who are stable MCI from progressive MCI showing that the continuous value predicts individualised future cognitive decline (**Experimental Chapter 2: Modelling prognostic trajectories of cognitive decline due to Alzheimer's disease**). Here, we apply the same framework on a new sample, deriving the scalar projection and making individualised predictions of future regional tau accumulation in early AD populations.

GMLVQ- scalar projection implementation:

From the training sample the model learns the multivariate relationship between β-amyloid, medial temporal atrophy and APOE 4 (metric tensor Λ) and the location in multidimensional space that best classifies SC vs EAD individuals (prototype locations: $w_{(SC,EAD)}$). For any new subject with β-amyloid, medial temporal atrophy and APOE 4 (sample vector: x_i) the scalar projection can be calculated by a series of simple linear equations.

1. Transform the sample vector x_i and prototypes $w_{(SC,EAD)}$ into the learnt space via the metric tensor Λ. Note as the metric tensor is learnt in the squared Euclidean space we transform using the square root of the metric tensor (i.e. $\Lambda^{1/2}$)

   1.a $X_i = \Lambda^{1/2}\, x_i$

   1.b $W_{(SC,EAD)} = \Lambda^{1/2} w_{(SC,EAD)}$
2. Centre the coordinate system on $W_{(SC)}$ and calculate the orthogonal projection of each vector X_i onto the vector $W_{(EAD)}$, in this co-ordinate system.

   2.a $Projection = \frac{\overrightarrow{X_i W_{SC}} \cdot \overrightarrow{W_{EAD} W_{SC}}}{|\overrightarrow{W_{EAD} W_{SC}}|}$
3. To normalise the projections with respect to the position of the prototype $W_{(SC)}$, the we divided the projection by the norm of $\overrightarrow{W_{EAD} W_{SC}}$:

4. $a\ Scalar\ Projection = \frac{\overrightarrow{X_i W_{SC}} \cdot \overrightarrow{W_{EAD} W_{SC}}}{\left|\overrightarrow{W_{EAD} W_{SC}}\right|^2}$

To determine a meaningful threshold of the scalar projection for separating individuals who are SC from EAD individuals we use logistic regression for the ADNI 2/GO sample labelled as SC, EAD, and Demented. This results in a probabilistic boundary based on scalar projection.

Determining Regions of significant AD related tau accumulation:

We first classified individuals from ADNI 3 as either SC or EAD based on each individual's scalar projection. For the individuals who are classified as EAD (based on the probabilistic threshold value) we performed a subsequent first level analysis to determine which of the 36 selected ROIs will accumulate tau in the future (i.e. regions with a future annualised rate of accumulation statistically greater than 0).

Predicting individual variability in Regional Tau Accumulation:

Finally, for regions that pass first level significance (i.e. within regions that significantly accumulating tau) we trained a series of regression models using ADNI 3 individuals classified as EAD to test if the scalar projection relates to individual variability in regional future rate of tau accumulation (dependent variable: regional future tau accumulation, independent variable: scalar projection). We then tested these models by making individualised predictions –out of sample- for individuals classified as EAD from the BACS sample. To test the accuracy of the regional predictions we calculated the shared variance between the observed future accumulation of tau and the model generated prediction using baseline biological data (i.e. scalar projection).

Statistical Analysis:

We used logistic regression to define a probabilistic boundary that separates individuals who are SC from EAD. Using the ADNI 2/GO sample we fit a three class (SC, EAD, Demented) logistic regression to determine the threshold value of the scalar projection. We set the threshold as the probability an individual is less than 50% likely to be SC. To determine the regions that will significantly accumulate tau for individuals classified as EAD we used one tailed one sample t-tests. As we are testing if regions are accumulating tau we use right tail t-tests to determine if the future rate of tau accumulation is significantly greater than 0 per ROI

for individuals classified as EAD. To compare required sample sizes for different models derived using ADNI 3 data we calculated the sample size needed for an arm of a hypothetical clinical trial designed to detect a 25% reduction in annual change (rate of tau accumulation, rate of PACC decline) with a significance of 0.05 and a power of a=0.8. For each comparison, we defined the null hypothesis as the mean and standard deviation of the rate of change calculated from the observed sample, where the alternate hypothesis is a 25% reduction of the mean of the observed sample. For each of the regions that showed significant tau accumulation we fit a robust linear regression (robustfit MATLAB) to predict future tau accumulation using the scalar projection. Setting the dependent variable as future regional tau accumulation and the independent variable as scalar projection we learnt a series of ROI regression equations.

$$Rate\ of\ Tau\ Accumulation\ (ADNI\ 3)_{\ ROI} = \beta(ADNI\ 3)_{ROI} * EAD\ Scalar\ Projection(ADNI\ 3)$$

Finally, using the fits derived from ADNI 3 data we generated predictions of tau accumulation for individuals classified as EAD in BACS.

$$Predicted\ Rate\ of\ Tau\ Accumulation\ (BACS)_{\ ROI} = \beta(ADNI\ 3)_{ROI} * EAD\ Scalar\ Projection(BACS)$$

We tested the accuracy of these predictions in the BACS sample by calculating the shared variance between the predicted future rate of tau accumulation and the observed future rate of tau accumulation after treating for outliers (robust correlation (Pernet et al., 2013)).

Results

Biological classification model for predicting stable condition vs early AD

We used a recent adaptation to the Generalised Metric Learning Vector Quantisation (GMLVQ) machine learning framework (**Methods: GMLVQ-scalar projection**) to generate a prognostic index (scalar projection) from three biological markers measured at baseline: cortical β-amyloid measured using PET, medial temporal atrophy measured using T1 weighted MRI and APOE 4 genotype.

The scalar projection is derived to discriminate stable cognitively normal individuals (SC) vs. early AD individuals with an unstable diagnosis (EAD). After the model was trained to discriminate SC vs. EAD we then determined the distance an individual is from the prototypical position of a SC. This distance is calculated along the learned axis that is predictive of future diagnosis (i.e. SC towards EAD) (i.e. scalar projection). We applied this approach training the model to discriminate SC vs EAD on data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) 2/GO cohort.

First, the model classifies SC (n=145) vs. EAD (n=162) individuals with cross-validated class-balanced accuracy of 86.4% (**Figure 6.1a)**. Further, using logistic regression we show that an individual is more than 50% likely to be classified EAD if they have a scalar projection greater than 0.4**.** Next, using the model trained on ADNI2/GO data we derived the scalar projection for individuals from two independent cohorts: a) ADNI 3 (CN=72, MCI=43) b) Berkeley Aging Cohort Study (BACS) (CN=56). The scalar projection classified 61 ADNI3 participants and 33 BACS participants as SC, with 54 ADNI3 and 23 BACS participants classified as EAD. (i.e. scalar projection greater than 0.4) (**Figure 6.1b**). We next tested if the scalar projection was significantly correlated with potential confounds using Pearson correlations. We observed that the scalar projection was weakly related to baseline age in ADNI 3 ($r(113)=0.28$, $p=0.003$) but not in BACS ($r(54)=0.187$, $p=0.168$) and it was unrelated to education (BACS: $r(54)=-0.011$, $p=0.94$, ADNI 3: $r(113)=-0.014$, $p=0.88$) or gender (BACS: $t(54)=-1.52$, $p=0.136$, ADNI 3: $t(113)=-0.582$, $p=0.56$).

Figure 6.1: Deriving the scalar projection in three cohorts

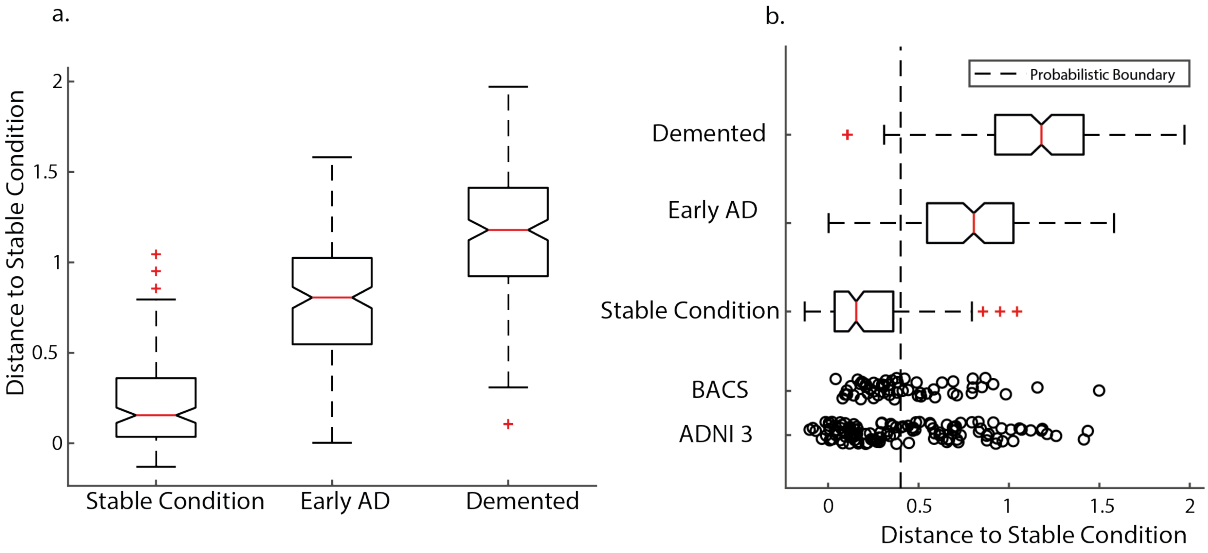


Figure 6.1a. *The distribution of scalar projection for individuals from ADNI2/GO who are SC (Stable Condition), EAD (Early AD) and Demented.* ***Figure 6.1b****. The distribution of scalar projection for BACS and ADNI 3 individuals. The dashed black line represents the probabilistic boundary used to classify SC vs EAD, all individuals to the right of the line are classified as EAD.*

Comparing rate of tau accumulation for stable condition vs early AD

For individuals from ADNI 3, we used longitudinal FTP-PET to compare regional future tau accumulation for individuals classified as SC vs. EAD. We extracted regional SUVR values from 36 Desikan-Kilany ROIs and calculated annualised rate of change per ROI. ROIs were grouped together in order to approximate the topographical distribution of tau in the Braak Staging scheme, as previously described. For individuals with 2 FTP-PET scans we used the difference between the follow-up and baseline FTP-PET scans divided by the time interval in years from baseline. For individuals with 3 or more FTP-PET scans we used the least squares fit of time from baseline vs. ROI SUVR.

We observed that EAD individuals accumulated global cortical tau (mean rate of tau accumulation within the 36 Desikan-Kilany ROIs) 2.8 times faster than SC individuals (**Table 6.1**). Further, testing which regions significantly accumulated tau (i.e. rate of accumulation significantly greater than 0 by one tail t-tests within each ROI) showed that EAD individuals accumulate tau primarily in Braak stages 4 and 5 ROIs (**Table 6.1**, **Figure 6.2**). In contrast,

those classified as SC did not show clear tau accumulation across cortical regions, with only a weak but significant effect in the middle temporal ROI t(60) =1.71, p=0.047) **Figure 6.2d**. There was no difference in cognitive change in (as measured by future annualised change in PACC) over the same time period between individuals classified as SC(mean=-0.161/year) vs. EAD(mean=-0.6/year) (t(100)=-1.04, p=0.30). However, individuals classified as EAD showed significant worsening (i.e. rate of PACC change significantly less than 0) in future cognitive ability (one tail t-test t(47)=-1.87, p=0.034).

Figure 6.2 Regional future rate of tau accumulation

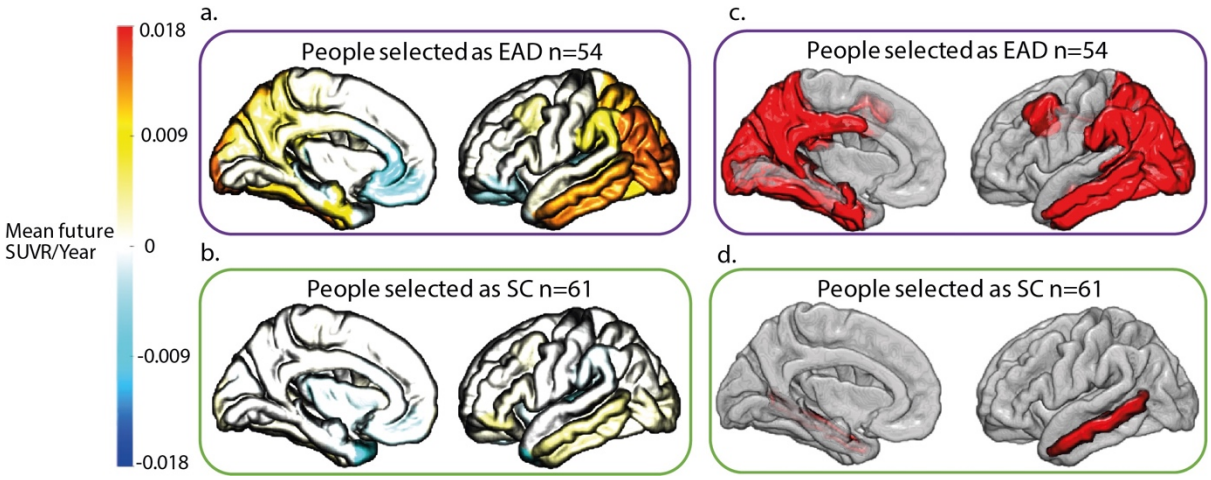


Figure 6.2. *The future annualised rate of tau accumulation for ADNI 3 individuals across the 36 Desikan Kilany ROIs.* ***Figure 6.2a****. Mean future annualised rate of tau accumulation for individuals classified as the Early AD (EAD).* ***Figure 6.2b****. Mean future annualised rate of tau accumulation for individuals classified as Stable Condition (SC).* ***Figure 6.2c*** *The regions in red are significantly accumulating tau for individuals classified as Early AD (EAD).* ***Figure 6.2d*** *The regions in red are significantly accumulating tau for individuals classified as Stable Condition (SC).*

Table 6.1. Regional future annualised rate of tau accumulation (over page)

Table 6.1. *Shows measures of future regional annualised rate of tau accumulation taken from the Desikan Killany atlas for ADNI 3 individuals within the 6 Braak stages. The mean future annualised rate of tau accumulation and test statistics describing whether a region is significantly accumulating tau for individuals classified as Stable Condition (SC) (Left block) and for individuals classified as Early AD (EAD) (Right block)*

Braak Stage	Region	Stable Condition (n=61)			Early AD (n=54)		
			Significantly Accumulating Tau			Significantly Accumulating Tau	
		mean accumulation (SUVR/Year)	t-stat	p-val	mean accumulation (SUVR/Year)	t-stat	p-val
1	ENTORHINAL	0.0006	0.130	0.448	**0.0120**	**2.259**	**0.014**
2	HIPPOCAMPUS	0.0009	0.233	0.408	-0.0011	-0.285	0.612
3	PARAHIPPOCAMPAL	0.0021	0.574	0.284	0.0049	1.149	0.128
	FUSIFORM	0.0032	0.941	0.175	**0.0135**	**2.920**	**0.003**
	LINGUAL	0.0018	0.507	0.307	0.0033	0.899	0.186
	AMYGDALA	-0.0001	-0.025	0.510	**0.0099**	**2.235**	**0.015**
4	MIDDLETEMPORAL	**0.0066**	**1.713**	**0.046**	**0.0164**	**3.332**	**<0.001**
	CAUDALANTERIORCINGULATE	0.0005	0.129	0.449	-0.0012	-0.541	0.705
	ROSTRALANTERIORCINGULATE	0.0009	0.270	0.394	-0.0070	-2.842	0.997
	POSTERIORCINGULATE	0.0025	0.796	0.215	**0.0056**	**1.741**	**0.044**
	ISTHMUSCINGULATE	0.0022	0.697	0.244	**0.0070**	**2.331**	**0.012**
	INSULA	-0.0003	-0.103	0.541	0.0022	0.809	0.211
	INFERIORTEMPORAL	0.0059	1.484	0.072	**0.0171**	**3.477**	**<0.001**
	TEMPORALPOLE	-0.0073	-1.302	0.901	-0.0001	-0.041	0.516
5	SUPERIORFRONTAL	0.0020	0.652	0.259	0.0021	0.852	0.199
	LATERALORBITOFRONTAL	0.0032	1.032	0.153	-0.0023	-0.870	0.806
	MEDIALORBITOFRONTAL	-0.0011	-0.309	0.621	-0.0041	-1.439	0.922
	FRONTALPOLE	0.0038	0.818	0.208	0.0008	0.167	0.434
	CAUDALMIDDLEFRONTAL	0.0033	1.096	0.139	**0.0069**	**1.967**	**0.027**
	ROSTRALMIDDLEFRONTAL	0.0022	0.641	0.262	0.0026	0.836	0.203
	PARSOPERCULARIS	0.0032	1.010	0.158	0.0037	1.476	0.073
	PARSORBITALIS	0.0067	1.348	0.091	-0.0021	-0.418	0.661
	PARSTRIANGULARIS	0.0035	0.920	0.181	0.0018	0.549	0.293
	LATERALOCCIPITAL	0.0045	1.007	0.159	**0.0180**	**2.887**	**0.003**
	SUPRAMARGINAL	0.0000	0.010	0.496	**0.0103**	**2.654**	**0.005**
	INFERIORPARIETAL	0.0024	0.721	0.237	**0.0184**	**3.725**	**<0.001**
	SUPERIORTEMPORAL	0.0016	0.464	0.322	0.0033	0.973	0.168
	SUPERIORPARIETAL	0.0023	0.589	0.279	**0.0150**	**3.213**	**0.001**
	PRECUNEUS	0.0016	0.587	0.280	**0.0088**	**2.694**	**0.005**
	BANKSSTS	0.0039	1.200	0.117	**0.0108**	**2.358**	**0.011**
	TRANSVERSETEMPORAL	-0.0042	-1.259	0.893	-0.0034	-0.964	0.830
6	PERICALCARINE	0.0025	0.793	0.216	0.0021	0.549	0.293
	POSTCENTRAL	0.0013	0.400	0.345	0.0029	0.856	0.198
	CUNEUS	0.0006	0.187	0.426	**0.0082**	**1.890**	**0.032**
	PRECENTRAL	0.0017	0.570	0.285	0.0025	0.910	0.184
	PARACENTRAL	0.0027	0.816	0.209	0.0051	1.499	0.070
	Mean	**0.0019**			**0.0054**		

Comparing required sample size to detect change in tau accumulation vs cognitive decline

To examine the clinical utility of regional future rate of tau accumulation as an outcome measure we contrasted the sample size (at the same power) required to observe cognitive decline vs regional future tau accumulation for individuals classified as EAD. We defined a clinically meaningful change as a 25% reduction in rate of change of either regional tau accumulation or PACC change.

For individuals classified as EAD we calculated that the required sample size (for a significance level of p=0.05 at a power of a=0.8) to detect a 25% reduction in rate of PACC change is n=1730. However, to detect a 25% reduction in regional future tau accumulation in the selected areas (**Figure 6.2c**) at the same power level an average sample size of n=1127 is required. Thus, using future rate of tau accumulation as a clinical outcome measure instead of future rate of cognitive decline delivers a reduction in required sample size of 35%.

Comparing required sample size to detect change in tau accumulation for multimodal vs unimodal classification

Next we compared prediction of regional future tau accumulation for individuals classified as EAD (n=54) based on the scalar projection vs. those classified based only on being β-amyloid positive (n=61). We observed that individuals classified as EAD using multimodal data (i.e. scalar projection) accumulated global cortical tau 1.5 times faster than individuals who are defined only by β-amyloid positivity (**Table 6.2)**.

We next compared required sample sizes to detect a 25% decrease in rate of future tau accumulation (given significance level of p=0.05 at power of a=0.8) for EAD vs β-amyloid only individuals within regions significantly accumulating tau for individuals classified as EAD (**Figure 6.2c**). We observe on average a 47% reduction in sample size when stratifying based on a classification of EAD (n=1127) vs β-amyloid positive only (n=2146). By using the tau-accumulating regions from the EAD subjects we may have biased the estimates, so we repeated the calculations using regions in which β-amyloid positive individuals are significantly accumulating tau. This showed an average a 30% reduction in sample size when stratifying based on a classification of EAD (n=831) vs β-amyloid positive only (n=1190) (**Figure 6.3**).

These results support the clinical utility of performing group stratification using a classification of EAD based on multimodal data in contrast to β-amyloid status alone.

Figure 6.3 reduction in sample size to observe change when stratifying using multimodal data vs. β-amyloid positive only

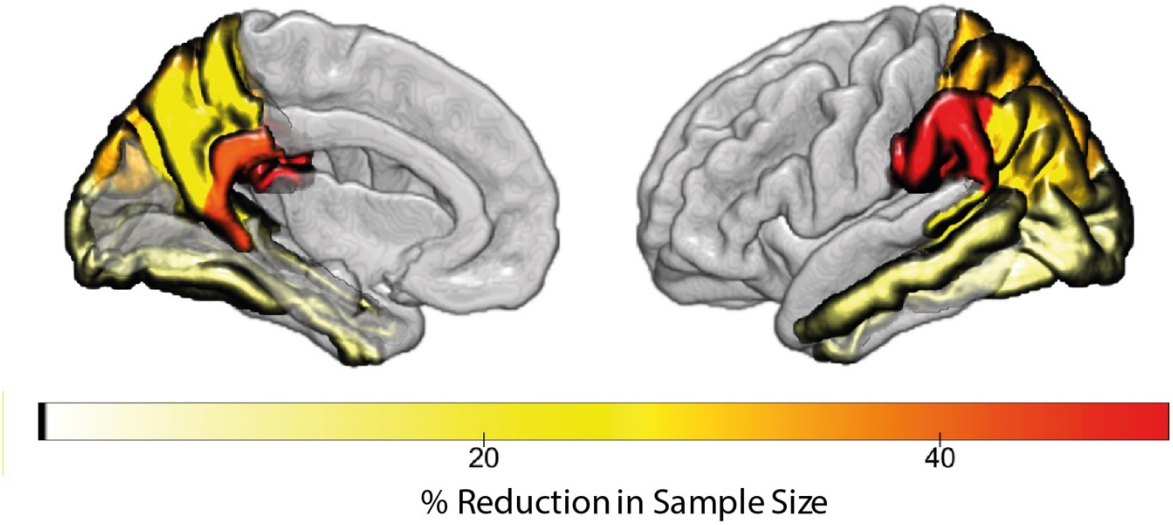


Figure 6.3 *shows the percentage reduction of sample size to observe a 25% reduction in tau accumulation per region for groups defined my multimodal data (EAD) or β-amyloid status only.*

Table 6.2. Regional future annualised rate of tau accumulation early AD vs β-amyloid positive (over page)

Table 6.2*. Shows measures of future annualised rate of tau accumulation for regions taken from the Desikan Killany atlas within the 6 Braak stages. The mean future annualised rate of tau accumulation and test statistics describing whether a region is significantly accumulating tau for individuals classified as Early AD (EAD) (Left block) and for individuals who are β-amyloid positive only (Right block). The Sample size reported is for a hypothetical clinical trial looking to observe a 25% reduction in rate of tau accumulation per region. The right most column compares the required sample size to observe the same effect for groups defined my multimodal data (EAD) or amyloid status only.*

Braak Stage	Region	Early AD (EAD)				β-amyloid Positive				EAD vs β-amyloid Positive
			Significantly Accumulating Tau				Significantly Accumulating Tau			
		mean accumulation (SUVR/Year)	t-stat	p-val	sample size	mean accumulation (SUVR/Year)	t-stat	p-val	sample size	relative difference in sample size
1	ENTORHINAL	**0.0120**	**2.259**	**0.014**	**1332**	0.007	1.386	0.085	3990	0.33
2	HIPPOCAMPUS	-0.0011	-0.285	0.611	NaN	-0.002	-0.437	0.668	NaN	NaN
3	PARAHIPPOCAMPAL	0.0049	1.149	0.127	5141	0.004	0.845	0.201	10719	0.48
	FUSIFORM	**0.0135**	**2.920**	**0.002**	**798**	**0.013**	**2.899**	**0.003**	**914**	**0.87**
	LINGUAL	0.0033	0.899	0.186	8384	0.005	1.374	0.087	4060	2.07
	AMYGDALA	**0.0099**	**2.235**	**0.014**	**1360**	0.007	1.630	0.054	2886	0.47
4	MIDDLETEMPORAL	**0.0164**	**3.332**	**<0.001**	**613**	**0.014**	**3.284**	**0.001**	**713**	**0.86**
	CAUDALANTERIORCINGULATE	-0.0012	-0.541	0.704	NaN	-0.004	-1.278	0.897	NaN	NaN
	ROSTRALANTERIORCINGULATE	-0.0070	-2.842	0.996	NaN	-0.007	-2.950	0.998	NaN	NaN
	POSTERIORCINGULATE	**0.0056**	**1.741**	**0.043**	**2240**	0.005	1.480	0.072	3501	0.64
	ISTHMUSCINGULATE	**0.0070**	**2.331**	**0.011**	**1250**	**0.006**	**1.852**	**0.034**	**2236**	**0.56**
	INSULA	0.0022	0.809	0.211	10366	0.001	0.505	0.308	30012	0.35
	INFERIORTEMPORAL	**0.0171**	**3.477**	**<0.001**	**563**	**0.017**	**3.728**	**<0.001**	**553**	**1.02**
	TEMPORALPOLE	-0.0001	-0.041	0.516	NaN	-0.004	-0.936	0.823	NaN	NaN
5	SUPERIORFRONTAL	0.0021	0.852	0.199	9354	0.000	0.163	0.435	286961	0.03
	LATERALORBITOFRONTAL	-0.0023	-0.870	0.805	NaN	-0.002	-0.834	0.796	NaN	NaN
	MEDIALORBITOFRONTAL	-0.0041	-1.439	0.922	NaN	-0.005	-1.874	0.967	NaN	NaN
	FRONTALPOLE	0.0008	0.167	0.434	242813	-0.002	-0.404	0.656	NaN	NaN
	CAUDALMIDDLEFRONTAL	**0.0069**	**1.967**	**0.027**	**1755**	0.004	1.142	0.129	5879	0.30
	ROSTRALMIDDLEFRONTAL	0.0026	0.836	0.203	9701	0.000	0.090	0.464	940607	0.01
	PARSOPERCULARIS	0.0037	1.476	0.073	3117	0.003	0.965	0.169	8226	0.38
	PARSORBITALIS	-0.0021	-0.418	0.661	NaN	-0.003	-0.613	0.729	NaN	NaN
	PARSTRIANGULARIS	0.0018	0.549	0.292	22520	0.000	0.040	0.484	4761379	0.00
	LATERALOCCIPITAL	**0.0180**	**2.887**	**0.002**	**816**	**0.016**	**2.887**	**0.003**	**921**	**0.89**
	SUPRAMARGINAL	**0.0103**	**2.654**	**0.005**	**965**	**0.007**	**1.972**	**0.027**	**1972**	**0.49**
	INFERIORPARIETAL	**0.0184**	**3.725**	**<0.001**	**491**	**0.015**	**3.322**	**0.001**	**697**	**0.70**
	SUPERIORTEMPORAL	0.0033	0.973	0.167	7172	0.001	0.435	0.333	40574	0.18
	SUPERIORPARIETAL	**0.0150**	**3.213**	**0.001**	**659**	**0.012**	**2.784**	**0.004**	**990**	**0.67**
	PRECUNEUS	**0.0088**	**2.694**	**0.004**	**937**	**0.008**	**2.454**	**0.009**	**1274**	**0.74**
	BANKSSTS	**0.0108**	**2.358**	**0.011**	**1222**	**0.009**	**2.173**	**0.017**	**1625**	**0.75**
	TRANSVERSETEMPORAL	-0.0034	-0.964	0.830	NaN	-0.004	-0.993	0.838	NaN	NaN
6	PERICALCARINE	0.0021	0.549	0.292	22521	0.003	0.703	0.242	15505	1.45
	POSTCENTRAL	0.0029	0.856	0.197	9250	0.001	0.326	0.373	71983	0.13
	CUNEUS	**0.0082**	**1.890**	**0.032**	**1900**	0.005	1.378	0.087	4035	0.47
	PRECENTRAL	0.0025	0.910	0.183	8200	0.001	0.210	0.417	174022	0.05
	PARACENTRAL	0.0051	1.499	0.070	3022	0.004	1.311	0.097	4456	0.68
	Mean	**0.0054**				**0.0037**				

Predicting individual variability in tau accumulation

Using the ADNI 3 sample we fit linear regression equations to test whether the scalar projection (derived from the model trained on ADNI2/GO) predicts individual variability in future tau accumulation (**Table 6.3)**. Of the 15 regions that were significantly accumulating tau (**Figure 6.2c**) 13 showed a significant relationship of the scalar projection and individual rates of future tau accumulation (**Figure 6.4a)**, explaining up to 30% of variance in the temporal cortex and 20% in superior and medial regions of the posterior parietal cortex (**Figure 6.4b**).

Figure 6.4 predicting individual variability in future tau accumulation in ADNI 3 early AD

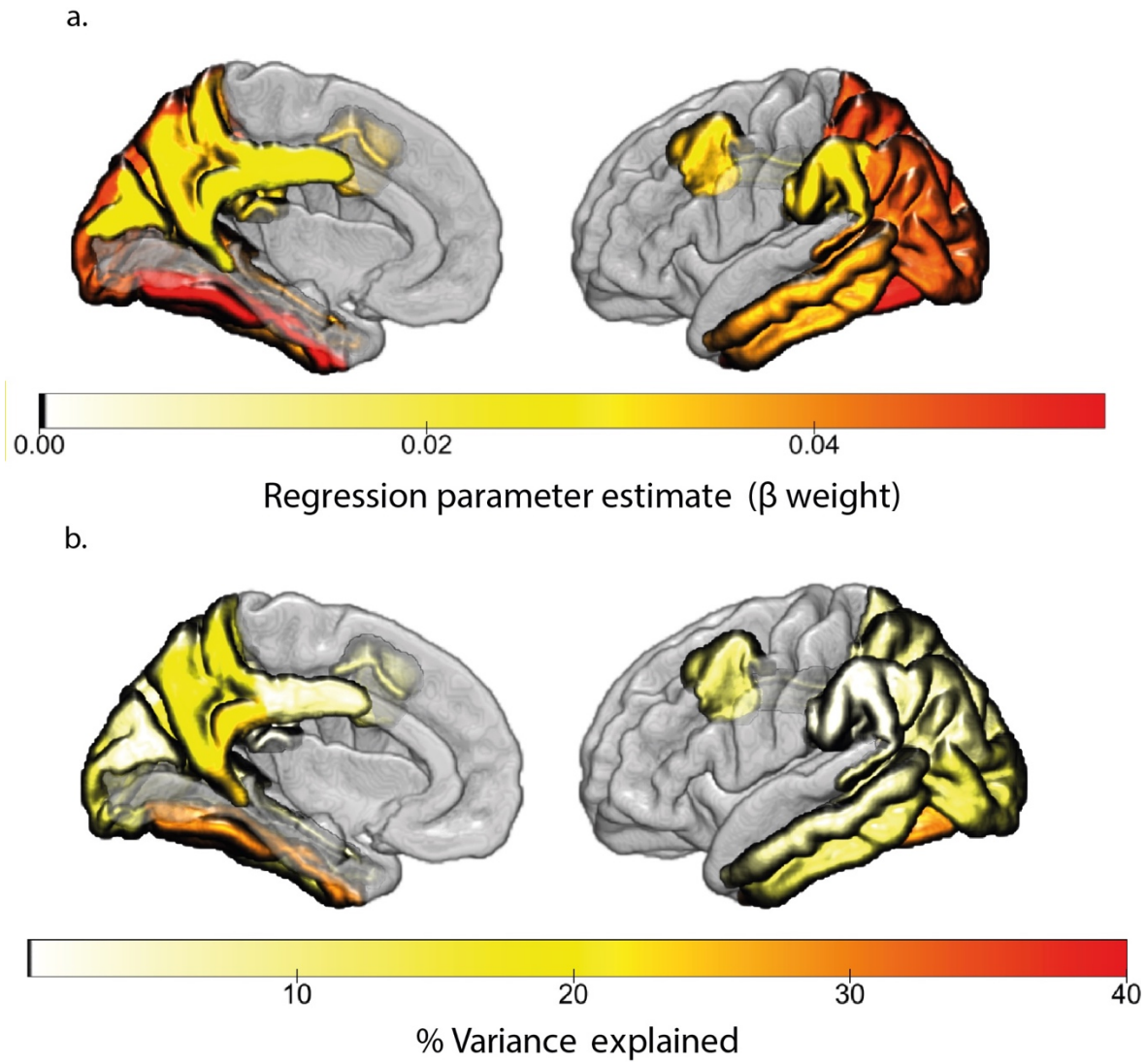


Figure 6.4a. *The significant ($p<0.05$) regional parameter estimates from linear regressions to predict future rate of tau accumulation for individuals classified as Early AD (EAD). The colour scale indicates the parameter estimate for the slope of the regression fit (i.e. relative annualised rate of future change of regional tau accumulation measured in SUVR/Year).*

Figure 6.4b. *Shows the percentage of variance explained when using scalar projection to predict future rate of tau accumulation for individuals classified as Early AD (EAD) from the ADNI 3 cohort.*

Table 6.3. Fitting individual variability in regional future annualised rate of tau accumulation

Braak Stage	Region	Beta Estimate (SUVR/Year)	t-stat	p-val	%Variance Explained
1	ENTORHINAL	-0.019	-1.327	0.190	0.69
3	FUSIFORM	**0.053**	**3.526**	**<0.001**	**30.0**
	AMYGDALA	0.015	0.919	0.362	2.04
4	MIDDLETEMPORAL	**0.040**	**2.069**	**0.044**	**7.94**
	POSTERIORCINGULATE	**0.027**	**2.460**	**0.017**	**9.27**
	ISTHMUSCINGULATE	**0.028**	**2.762**	**0.008**	**24.18**
	INFERIORTEMPORAL	**0.040**	**2.098**	**0.041**	**13.62**
5	CAUDALMIDDLEFRONTAL	**0.034**	**2.826**	**0.007**	**12.25**
	LATERALOCCIPITAL	**0.046**	**2.207**	**0.032**	**14.89**
	SUPRAMARGINAL	**0.032**	**3.041**	**0.004**	**3.89**
	INFERIORPARIETAL	**0.046**	**2.853**	**0.006**	**8.67**
	SUPERIORPARIETAL	**0.049**	**3.522**	**<0.001**	**8.67**
	PRECUNEUS	**0.034**	**3.034**	**0.004**	**19.5**
	BANKSSTS	**0.041**	**2.687**	**0.010**	**9.35**
6	CUNEUS	**0.030**	**2.129**	**0.038**	**7.84**

Table 6.3. *Shows the parameter estimates and associated statistics for the robust regression equations using the scalar projection to predict regional future tau accumulation for individuals from ADNI 3 classified as Early AD (EAD).*

Predicting individual variability in tau accumulation for an independent cognitively normal sample

To test the robustness of these regional fits we generated individualised predictions of future tau accumulation for CN individuals classified as EAD from the BACS sample. Using the model trained on ADNI2/GO individuals we derived the scalar projection from baseline multimodal biological data in BACS participants (**figure 6.1b**). We used the linear equations relating the scalar projection to rate of future tau accumulation derived from the 13 regions with significant correlations in the ADNI 3 sample to predict the future rate of tau accumulation in these ROIs in the BACS participants. Individualised predictions of future tau accumulation explain up to 39% of the observed variance in the temporal cortex and 32% in superior and

medial regions of the posterior parietal cortex (**Table 6.4, Figure 6.5).** Finally, we showed that for individuals who are classified as SC from the BACS sample no regional future tau accumulation can be predicted **(Table 6.4**). Therefore, our predictions are robust and specific for staging individuals who are transitioning from non-pathological aging to Alzheimer's Disease.

Table 6.4. Predicting regional future annualised tau accumulation in BACS

			SC (n=33)	EAD (n=23)
Braak Stage	**Region**	**ADNI 3 Beta Estimate (SUVR/Year)**	**Predicted Variance Explained %**	**Predicted Variance Explained %**
3	FUSIFORM	0.053	3.02	26.11
	MIDDLETEMPORAL	0.040	0.48	21.82
	POSTERIORCINGULATE	0.027	0.09	12.18
	ISTHMUSCINGULATE	0.028	0.00	2.79
4	INFERIORTEMPORAL	0.040	1.45	22.47
	CAUDALMIDDLEFRONTAL	0.034	0.89	0.89
	LATERALOCCIPITAL	0.046	2.38	6.79
	SUPRAMARGINAL	0.032	0.28	32.19
	INFERIORPARIETAL	0.046	0.02	2.31
	SUPERIORPARIETAL	0.049	0.59	20.95
	PRECUNEUS	0.034	2.47	26.74
5	BANKSSTS	0.041	0.51	38.46
6	CUNEUS	0.030	2.51	0.01

Table 6.4. *Shows the shared variance of the predicted regional future tau accumulation and the observed future tau accumulation for individuals from the BACS cohort. The right most column shows the shared variance for individuals classified as Early AD (EAD). The column second from the right shows the shared variance for individuals classified as Stable Condition (SC)*

Figure 6.5 prediction accuracy of future tau accumulation in BACS early AD

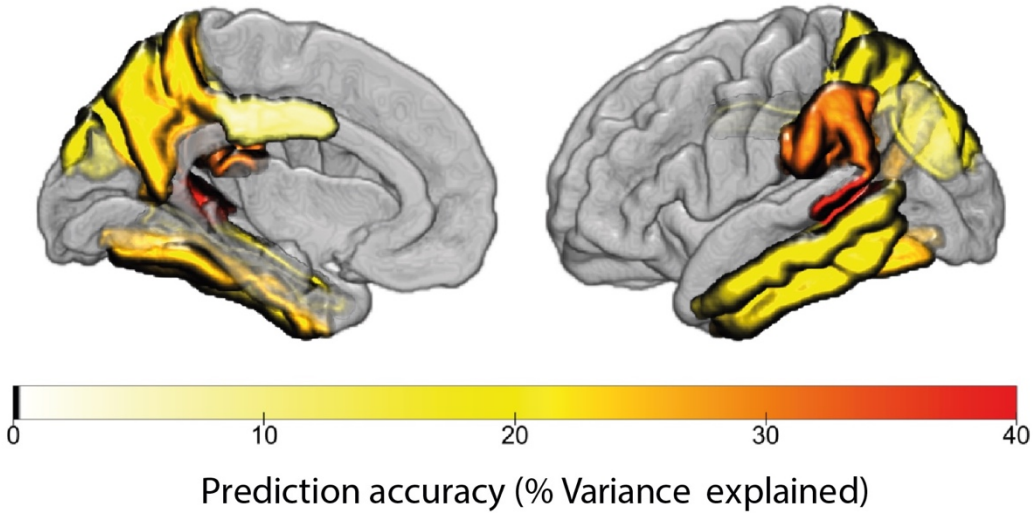


Figure 6.5. *The percentage of shared variance of the predicted regional future tau accumulation and the observed future tau accumulation for individuals classified as Early AD (EAD) from the BACS cohort.*

Discussion

These results describe the use of machine learning to combine continuous information from AD biomarkers to predict pathological changes in tau accumulation in early and asymptomatic stages of AD. We use well characterised AD biomarkers (β-amyloid, medial temporal atrophy, APOE 4) to generate a prognostic index for stratification. Using this multimodal index derived from baseline data, we predict future tau accumulation, a known pathological driver of AD progression. Further, we highlight that when stratifying by multimodal measures future tau accumulation is more sensitive to change than the gold standard cognitive instrument. Finally, we are able to predict spatially specific individualised changes in tau accumulation in an out-of-sample group of cognitively normal people, highlighting the strong predictive ability of this metric in preclinical AD populations.

Utilising a recently developed machine learning approach (**Methods: GMLVQ-Scalar Projection**) we derived a prognostic index (scalar projection) from baseline biomarkers that accurately classified individuals in the early stages of AD. Using this model derived index we showed that individuals classified as early AD will accumulate tau in a topography stereotypical of preclinical AD (Jack, Wiste, et al., 2018; Pontecorvo et al., 2019; Schultz et al., 2018). We established the clinical relevance of our approach showing that, for individuals classified as early AD, the rate of tau accumulation provides a 35% reduction in sample size to detect a clinically meaningful change relative to the gold standard cognitive instrument (PACC (Donohue et al., 2014)). Similar to previous work we have been able to successfully stratify individuals based on the pattern of tau accumulation in early AD and accurately reproduce the topography seen in several independent cohorts. The regions which we predicted would show tau accumulation are biologically plausible, corresponding strongly to regions of tau deposition demonstrated by numerous other laboratories and largely corresponding to a proposed "meta-ROI" for tau quantitation (Jack, Wiste, et al., 2018; Pontecorvo et al., 2019; Schultz et al., 2018). Use of this ROI similarly demonstrates a smaller sample size to detect clinically meaningful change than using a cognitive measure (Jack, Wiste, et al., 2018).

Extending previous work, we show that a multimodal measure is superior to a unimodal (i.e. β-amyloid positive only) classification for determining if an individual will accumulate pathological tau. Individuals who are classified as early AD using multimodal data are accumulating tau at 1.5 times the rate as β-amyloid positive only individuals. Further, when

using multimodal data vs. β-amyloid status only to stratify populations there is a greater than 30% reduction in the sample size required to observe a clinically meaningful change in the stereotypical pattern of pathological tau accumulation. The benefit of combining multimodal data for stratification in preclinical AD has been discussed in the context of future changes in cognition (Allison et al., 2019; Bilgel et al., 2018; Burnham, Bourgeat, Doré, Savage, Brown, Laws, Maruff, Salvado, Ames, Martins, Masters, Rowe, Villemagne, et al., 2016; Dumurgier et al., 2017; Insel et al., 2015; Jansen et al., 2018). These studies show that grey matter atrophy and cortical β-amyloid burden relate to separable patterns of future cognitive decline (Bilgel et al., 2018; Burnham, Bourgeat, Doré, Savage, Brown, Laws, Maruff, Salvado, Ames, Martins, Masters, Rowe, Villemagne, et al., 2016; Dumurgier et al., 2017; Insel et al., 2015). Given the association between longitudinal changes in tau and cognitive decline in the early AD phase (Hanseeuw et al., 2019), our results reiterate the benefit of combining continuous values of β-amyloid and grey matter atrophy for prognostication in early AD. Taking these findings into consideration, as well as the anticipated benefit in using tau accumulation as a primary endpoint, we show that our multimodal prognostic index can support clinical trial design, particularly in the early AD stage.

Moving beyond binary classifications we showed that the scalar projection predicts individual variability in future regional tau accumulation within regions known to be affected in early AD. We show that these individualised predictions generalise to an independent sample of cognitively normal community dwelling individuals. Further, we show within this community dwelling population these predictions are specific to individuals who are classified as early AD. Several studies have investigated individual variability in future tau accumulation, focussing primarily on associations between simultaneous changes in pathophysiology (Franzmeier et al., 2020; Hanseeuw et al., 2019; Harrison et al., 2019). Unlike these previous approaches we have used a multimodal approach to make explicit individualised predictions specific to early AD, validating these predictions in a community dwelling population.

There are several limitations of the current framework. First, our model is trained on grey matter atrophy in the medial temporal lobe. Previously we established this grey matter value relates to memory deficits, baseline tau and separates individuals who are stable MCI from progressive MCI (**Experimental Chapter 1: Deriving an interpretable and interoperable score of Alzheimer's related atrophy)**. However, as our previous research only focussed on ADNI participants this measure of atrophy likely captures information specific to typical amnestic AD populations. Given that dissociable patterns of tau spreading have been

observed for atypical AD variants (i.e. posterior cortical atrophy and logopenic progressive aphasia) (Sintini et al., 2019) additional measures of atrophy might be necessary for a larger scale predictive model. In addition, while FTP PET measures largely correspond to pathological spreading of AD related tau (Schöll et al., 2016), the off target binding of FTP limits it's sensitivity to classify other tauopathies. Specifically, a model attempting to predict tauopathies affecting the midbrain and frontotriatal regions (i.e. progressive supranuclear palsy, corticobasal syndrome) may not be possible using FTP (Lowe et al., 2016). However, by applying a similar framework using a different tracer differential diagnosis of tauopathies may be possible.

A potential limitation of this work is that we did not extensively test and control for effects due to poor scan quality or scanner differences. Images downloaded from the ADNI website following the standard pre-processing pipeline in an attempt to increase harmonisation across scanning sites. This included corrections due to gradient nonlinearity, non-uniform image intensity due to non-uniform receiver coil sensitivity and correction of non-uniformity due to wave effects (Jack et al., 2008). However, no additional control analysis was performed to ensure that there was no carry over differences due to scanner site. Further, there was no additional investigation into the potential scanner effects between data from ADNI and data from either the 1.5T or 3T scans from BACS. In addition, no additional quality control was run to account for image artefacts due to subject movement.

In this work we derived the rate of future tau accumulation using the linear least squares fit of regional FTP-PET over time. However, as the proposed AD cascade suggests that tau accumulation follows a sigmoidal shape over time, the hypothesised rate of change will follow an inverted U shape (Jack, Knopman, et al., 2010). In this work we only investigated the relationship of the scalar projection with rate of tau accumulation for individuals who are cognitively normal or MCI and therefore tau accumulation over time (i.e. rate of tau accumulation) is more likely to be linear than sigmoidal (Jack et al., 2013; Jack, Knopman, et al., 2010). However, a more complex fit may be preferred when attempting to predict future tau accumulation for a population including cognitively normal, MCI and demented individuals. Further, we did not control for the variability due to different duration of follow-ups. In future work we can account for this variability in follow-up length in a more principled way using linear mixed effects models.

In sum, we have shown that the pathological spreading pattern of tau in early AD is predictable using baseline biomarkers. We highlighted that when designing clinical trials not only is pathological tau accumulation an optimal outcome measure, but a stratification based on

multimodal data is more specific to early AD. Further, we showed that a prognostic index derived from baseline multimodal biomarkers predicts individualised rates of future pathological tau accumulation in early AD. Here, we successfully translated rich AD specific information from a large research cohort to make individualised predictions of pathophysiological change in a cognitively normal community sample.

7. Discussion

Undeniably, the proliferation of machine learning approaches in AD research is an exciting integration of computer science with neuroscience and has real-world potential. However, a fine balance must be struck in computational neuroscience ensuring the field does not favour computational innovations that lack neuroscientific insights. This deviation away from biologically constrained models and the acceptance of 'black box' approaches has questionable clinical utility and is evident with the paucity of these tools in clinical decision making (Topol, 2019). In light of this, the motivation of this thesis was to introduce a predictive framework that is easily interpretable but also not heavily constrained. That is, the driving force in the methodology presented is not one of improve accuracy at all costs by increasing model complexity or number of predictors. Rather, it is to generate and combine robust features that are widely accepted into a single framework. In doing so we limit the constraints that our model places on the data to make predictions reducing the chance of overfitting. More importantly we can make predictions that are related to AD as a clinical category, which are not specifically constrained to the task used to generate them (i.e. interoperation).

Our integrated framework (**Figure 7.1)** relies on three key components:

1. Selecting neuroimaging and cognitive data that will be linked due to disease pathology.
2. Using the PLSr-RFE framework to generate a feature optimised for generalisation performance.
3. Using the GMLVQ-scalar projection approach to model prognostic trajectories.

Figure 7.1 Integrated framework

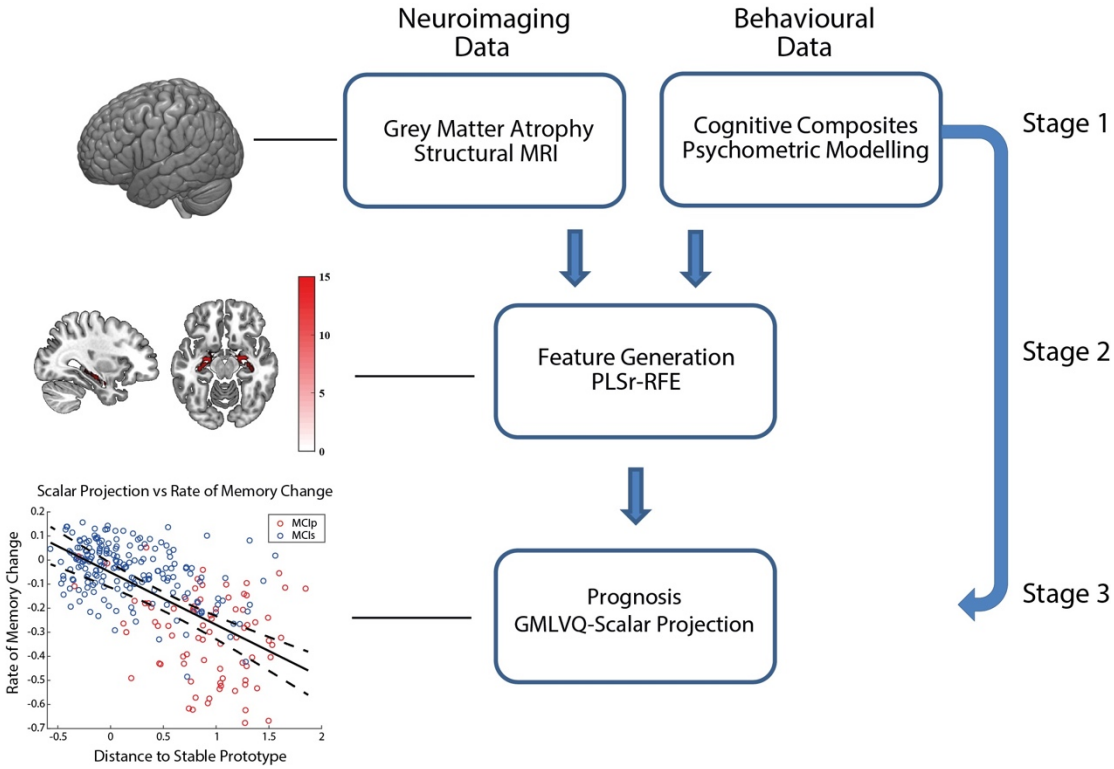


Figure 7.1 *shows the integrated framework presented in this thesis. Stage 1. Involves the selection of associated cognitive and neuroimaging markers. Stage 2. Is the feature generation process using the PLSr-RFE framework. Stage 3. Is the modelling of prognostic trajectories.*

At each stage of our modelling framework we can investigate the interactions between predictors and outcomes and gain key neuroscientific insights.

Our approach offers several specific advantages over previous modelling:

1. Our novel integrated feature generation (PLSr-RFE) method allows us to reduce high dimensional collinear data (structural MRI) into a single robust and interpretable feature (PLS-derived grey matter score).

2. Interrogating the metric tensor within the GMLVQ framework allows us to interpret the weights of univariate (diagonal) and multivariate (off diagonal) features that are predictive of progression to AD.

3. Extending the binary GMLVQ classification task to our trajectory modelling approach (GMLVQ-Scalar Projection) allows us to generate highly informative individualised disease trajectories. This approach is in line with the 2018 NIA-AA research framework that has

transitioned to defining AD as a continuum based on continuous measures rather than discrete clinical labelling.

4. The GMLVQ-Scalar Projection approach uses broad classes based on ‘noisy’ diagnostic labels to derive a highly sensitive continuous prognostic score. This allows us to capitalise on data that was collected and labelled based on the 2011 NIA-AA syndromic classification scheme.

5. The prognostic index that is derived from the GMLVQ-Scalar Projection approach is not constrained to a single predictive task. That is, the relationship between the scalar projection and pathological changes used to validate is clinical utility are implicitly learnt and therefore capable of interoperation.

Feature Generation: PLSr-RFE

Highlighting the utility of the approach, several studies have demonstrated methods for reducing structural neuroimaging data into AD indices (C. Davatzikos et al., 2009; Christos Davatzikos et al., 2019; Y. Fan et al., 2008; Young et al., 2018). Our approach is a significant extension to this work, as we have designed a method with the specific goal of interoperation. That is, the outcome of the PLSr-RFE methodology is a voxel weights matrix that can be applied to minimally pre-processed structural MRI to generate a single AD index that relates to cross modal AD biomarkers (i.e. cognition and tau accumulation).

One increasingly prevalent problem in the field has been the failure of models to generalise to new populations. There are several potential reasons that models do not generalise. The first is that approaches using the same training data to generate features and perform classification tasks are prone to overfitting. The second, is that approaches to generate indices based on a neural state or clinical symptoms may be too heavily constrained to the feature generation task and may not transfer well to the subsequent classification task.

An example of overfitting data can occur when researchers favour a feature generation approach that is specific to the subsequent classification task performed (i.e. the feature is built to maximise the classification accuracy of separating sMCI vs pMCI). Applied to the problem of AD prediction, we would require longitudinal data to

1. Generate our feature, 2. Train our classifier, 3 cross-validate our model. In our case this would limit our sample to 167 people (sMCI=113, pMCI=54). Although we could achieve high discriminability between groups, with a relatively small sample there is no guarantee that the same feature will work equally as well with a new group of patients.

Therefore, to increase the sample size available to generate our feature we favour a transfer learning approach. We reasoned that as ADNI Mem is discriminative of sMCI vs pMCI and atrophy is linked to poor memory performance, a score of atrophy that is predictive of ADNI Mem should also discriminate sMCI vs pMCI. In taking this approach we no longer require longitudinal data to generate our feature, utilising a much larger cross-sectional sample of 589 people. In using this transfer learning approach, we increased our feature generation sample by 352% greatly reducing our chance of overfitting our model development sample.

A second advantage of our approach is that we obtain greater specificity to clinical sub-groups than the outcome measure used to generate the feature (i.e. ADNI Mem). Specifically, we derive a score of atrophy that is predictive of ADNI Mem and is also sensitive to the heterogeneity of MCI pathologies. As we show in our second experimental chapter ADNI

Mem is a sensitive measure to discriminate sMCI vs pMCI. However, poor ADNI Mem scores are not specific to underlying AD pathology. That is, sMCI (false positives for AD) have a lower ADNI Mem score than healthy controls (normal cognition: NC). The heterogeneity in the underlying pathology of amnestic MCI and it's lack of specificity to AD was one of the driving motivations of the 2018 research framework transitioning from syndromic to a biological definition of AD (Jack, Bennett, et al., 2018). Using our approach, we can outperform the predictive value of raw ADNI Mem scores by capturing the variance in ADNI Mem that is both sensitive and specific to pMCI (**Figure 7.2)**.

Figure 7.2 Sensitivity and specificity of ADNI Mem and PLS derived grey matter score to AD

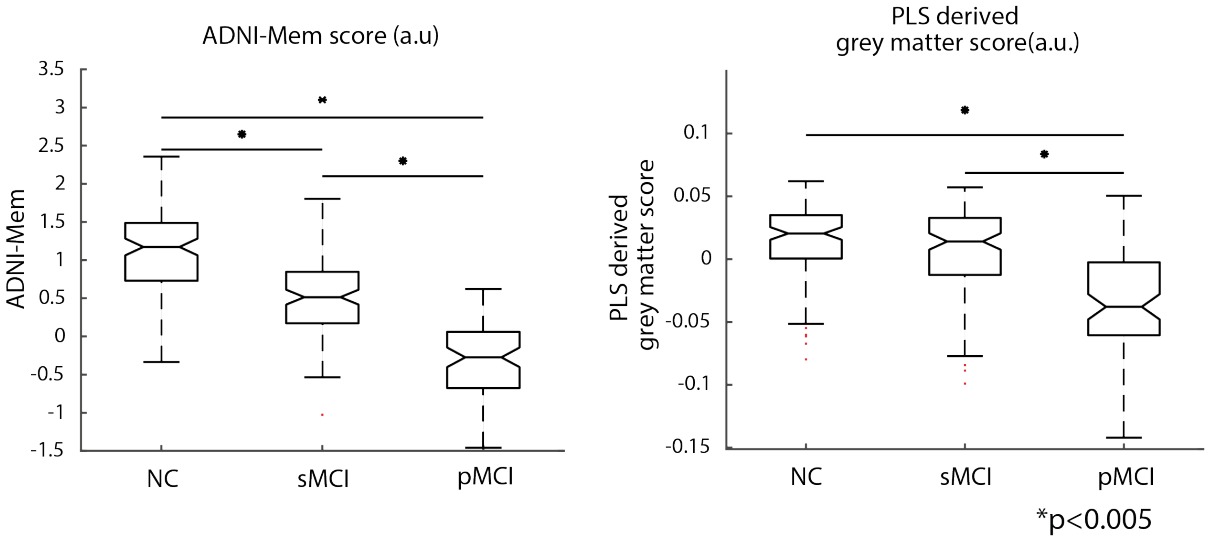


Figure 7.2 *shows that while ADNI Mem is sensitive to pMCI it is not specific (i.e. sMCI is lower than NC), however the atrophy biomarker is both sensitive and specific to pMCI. Note: Amyloid positive sMCI were removed from sample to account for potential slow converters.*

We achieved this by constraining our model to learn the association between two coupled process (atrophy and memory dysfunction). Given the sampling criteria of the MCI cohort in ADNI (multi-domain amnestic MCI), we reasoned that the most common cause of memory deficits in ADNI will likely be due to AD pathology. Thus, our biomarker of medial temporal atrophy has high face validity. Here, we leveraged the sample characteristics of ADNI to unveil a multivariate score of medial temporal atrophy as a predictor of cognitive dysfunction. However, the lack of heterogeneity in ADNI limited our ability to extract biomarkers related to atypical AD variants. Future work will apply a similar approach to a

more heterogeneous sample looking to uncover atrophic patterns that relate to non-amnestic related cognitive decline.

The outcome of the PLSr-RFE process is an easily deployable grey matter mask that can be used for quick and effective appraisal of MRI taken from the clinic. This is evident when we use the grey matter mask derived in experimental chapter 1 on a sample of patients scanned from the Cambridge Memory Clinic. We show a similar relationship for this sample, with lower grey matter atrophy scores related to lower ACEr Memory score $R^2 = 20\%$ (**Figure 7.3)**.

Figure 7.3 generalisation performance of PLS derived grey matter score in independent sample

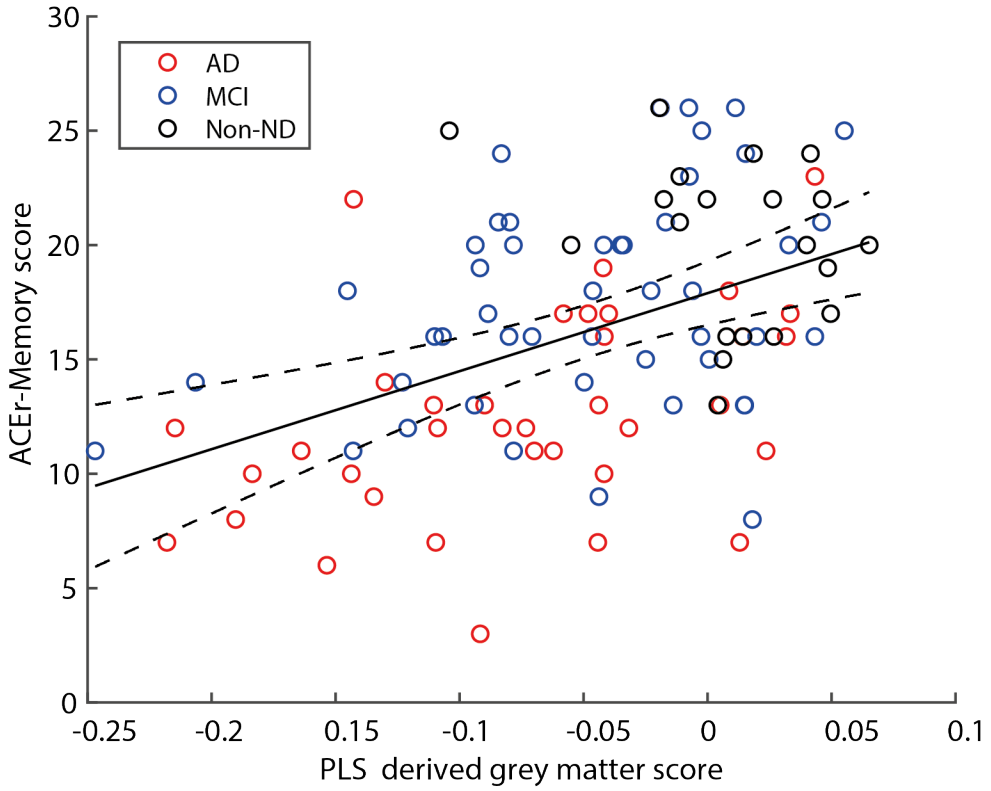


Figure 7.3 *shows the relationship of pls grey matter score derived on ADNI data in experimental chapter 1 and ACEr-Memory score from a sample recruited from the Cambridge memory clinic (r(101)=0.45, p<0.0001). Red dots are patients diagnosed as AD, blue dots are patients diagnosed as MCI and black dots are diagnosed as non-neurodegenerative disease.*

Using our approach, we observe the same relationship between the grey matter score and memory dysfunction in a heterogeneous sample taken from routine clinical patients. Further, we see that the atrophy biomarker captures memory dysfunction that can be validated to any sensitive memory score not just specific to ADNI Mem. Therefore, we have demonstrated that our method is both more sensitive and produces outputs that generalise to new patient groups.

We designed the PLSr-RFE model to handle structural MRI, however the approach can easily be deployed on other neuroimaging data (i.e. functional MRI, PET, DTI) to combine multimodal neuroimaging predictors to make predictions (Karlaftis et al., 2019). Looking beyond neuroimaging predictors, an interesting avenue of future research would utilise the PLSr-RFE approach to generate polygenic risk scores. There have been several attempts at deriving a robust polygenic risk score for late onset AD, however these have had limited success (Chasioti, Yan, Nho, & Saykin, 2019). To date the most common approaches to combine genetic information from Single Nucleotide Polymorphisms (SNPs) have been based on approximations of multivariate relationships. That is, mass univariate tests have been used to determine frequent SNPs in a sample and these prevalent SNPs are then summed to form a single score. Prior to performing this sum (either linear or weighted) SNPs are pre-processed to try and remove linkage disequilibrium (i.e. nearby SNPs have similar associations) in an attempt to remove redundancies in SNP information. This is achieved by pruning similar SNPs resulting in a set of largely uncorrelated SNPs. Both the summing of mass univariate statistics and the exclusion of correlated information are distinct weaknesses when trying to derive an AD genetic composite score. The PLSr-RFE is well suited to alleviate these issues as the PLSr model is well equipped to handle correlated predictors. Further, the PLSr-RFE approach will remove redundant information in a principled way, to optimise generalisation performance. Taken together the PLSr-RFE approach has strong potential to uncover highly predictive genetic features in AD.

Unveiling interactive predictors of AD progression: GMLVQ

Here, we used the GMLVQ classification framework to extract interactive cognitive factors that separate stable MCI from progressive MCI. Using this transparent machine learning approach, trained on informative and interpretable baseline cognitive data, we can extract key insights into cognitive data that is prognostic of cognitive decline. Our model could discriminate the two broad classes (sMCI vs pMCI) with accuracy above 80% giving us confidence that the univariate and multivariate relationships uncovered by the metric tensor are meaningful. We show that baseline composite scores related to memory and executive function (ADNI-Mem, ADNI-EF composite score) are highly predictive of disease progression. Further, we demonstrate a negative interaction between baseline cognitive (memory, executive function) and affective scores that separates individuals into different classes. Previous studies have shown that moderate to severe depressive symptoms (i.e. GDS> 15) are predictive of MCI conversion to AD (Defrancesco et al., 2017), while mild depressive symptoms do not increase conversion risk (Chen et al., 2008; Defrancesco et al., 2017). Here, similar to previous work (Chen et al., 2008; Defrancesco et al., 2017) we do not show a univariate relationship with mild depressive symptoms on MCI conversion (i.e. diagonal term of metric tensor). However, when interrogating the multivariate relationships between predictors (i.e. off diagonal term of metric tensor) we discover that the interaction between scores that are indicative of mild depression (i.e. GDS<10) and cognitive dysfunction is predictive of MCI progression. Thus, our metric learning approach on multi domain data (i.e. cognitive and affective measurements) has unveiled subtle interactions between baseline cognition and affective disturbance related to AD. Looking beyond the work presented here we can investigate this multivariate association further, elucidating neural mechanisms that underpin this relationship, potentially reducing MCI patient misclassification due to comorbidity (e.g. affective disturbance).

Modelling trajectories: GMLVQ-Scalar Projection.

Our trajectory modelling approach (GMLVQ-Scalar Projection) extends beyond binary patient classification frameworks that are poorly constrained. Here, we aimed to make meaningful predictions in AD by deriving an approach that captures continuous prognostic trajectories. Our machine learning approach is particularly well suited to tasks when training labels have a high degree of target uncertainty (e.g. future clinical diagnosis). In particular, we derived continuous prognostic scores (i.e. scalar projection) by training the model based on 'noisy' diagnostic labels. As our metric learning model has limited freedom (linear low-parameter model), separating continuous target values (i.e. individualised trajectories) into two broad classes (stable vs. progress) forces the model to extract key multivariate relationships in the data that distinguish between target populations, ignoring subtle differences in target values which can lead to overfitting. Using the GMLVQ-scalar projection model we learn the multivariate relationship between predictors (metric tensor) and the location in multidimensional space that best discriminates between stable and progressive individuals (prototype locations). Within this learnt space we then determine the distance an individual is from the prototypical position of stable individuals along the learned axis that is predictive of future diagnosis (i.e. stable towards progressive). That is, we derive a scalar value from the projection of any sample point along the prognostic axis.

Modelling trajectories of Cognitive Decline

Despite the importance of early diagnosis of Alzheimer's disease for clinical practice and treatment, we still lack robust tools for predicting individual progression to dementia. The multimodal longitudinal measurements across large-scale samples available in ADNI provide a testbed for machine learning approaches that generate predictive features and discriminate between patient groups (Weiner et al., 2015, 2017). Here, we tested whether prognostic trajectories generated from baseline cognitive data are predictive of future memory decline for MCI patients. We found that a prognostic measure derived from cognitive data is reasonably predictive of future decline (R^2 =16%), and with the inclusion of non-invasive biological predictors predictive performance improves (R^2 =23%). However, when deriving trajectories to predict cognitive decline the best performing model is built from well validated AD biomarkers (R^2 =46%).

Modelling trajectories from validated biomarkers

Numerous studies throughout the last decade have provided evidence for the hypothetical model of the AD pathological cascade (Jack et al., 2013; Jack, Knopman, et al., 2010). This widely accepted model has shifted the staging of AD from a clinical syndromic diagnosis (Albert et al., 2011; McKhann et al., 2011) to a continuum of biomarker characteristics (Jack, Bennett, et al., 2018). We show that our multimodal prognostic index (scalar projection) is successful at predicting cognitive decline in MCI individuals. Our approach has strong potential to deliver fine stratification of patients for personalised treatment plans and enrolment into clinical trials. We decided to train on available biomarkers from ADNI 2/GO cohort (i.e. structural MRI, amyloid PET, APOE 4). However, our machine learning approach can be deployed readily if different biological data is available. For example, if in a different cohort a predictor is not available (i.e. genetic markers) we can readily retrain the model on ADNI data omitting the missing feature. This lends a robustness to our models that is critical in dealing with the variability in clinical approach or available diagnostic tests between individuals. Further, as the relationship between the scalar projection and cognitive decline is implicitly learnt (i.e. change in ADNI Mem is not used in model formulation) we can validate our model using different cognitive measures (i.e. change in ACEr).

Using grey matter atrophy and β-amyloid as predictors we retrained the GMLVQ-scalar projection model to differentiate sMCI vs pMCI patients in ADNI. Below we show that we can use this model to predict cognitive decline in an independent sample of MCI patients recruited from the Cambridge Memory Clinic (**Figure 7.4)** ($R^2(16)=28.1\%$, $p=0.02$).

Figure 7.4 generalisation performance of GMLVQ-scalar projection in an independent cohort

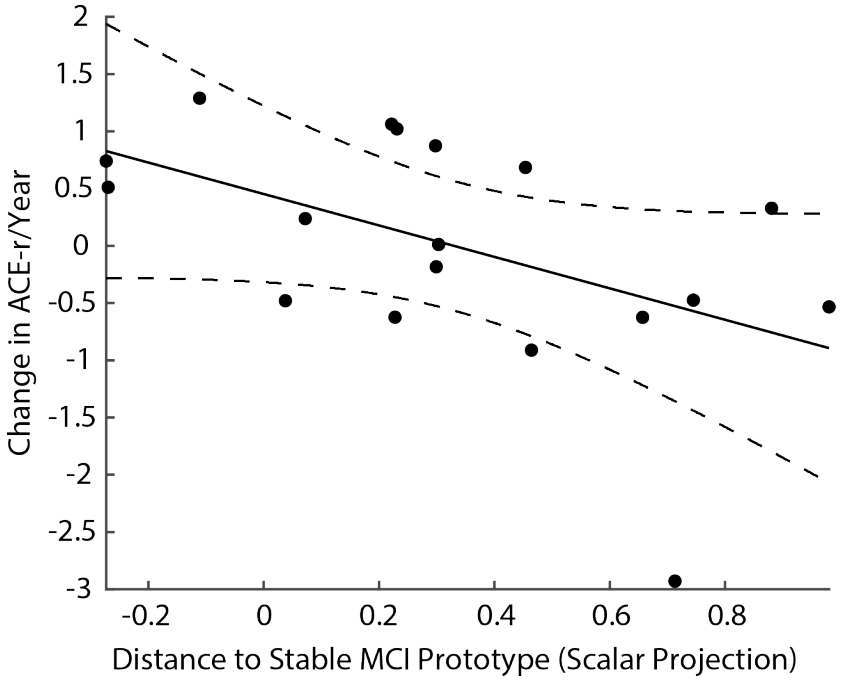


Figure 7.4 *shows the generalisation of the model trained on ADNI data to predict in an independent UK sample.*

This good generalisation performance highlights that our approach is robust and malleable to use different predictors to generate prognostic trajectories for MCI patients. However, β-amyloid PET is not routinely collected in the clinical setting. Looking forward we aim to use this approach to generate predictions on more readily available neuroimaging and cognitive data.

Modelling trajectories of biomarker change

Given the transition to defining AD in terms of biomarkers (i.e. ATN framework), machine learning models must now be designed to predict pathological biomarker changes or face becoming clinically irrelevant. Here, we use machine learning to combine continuous information from AD biomarkers to predict pathological changes in tau accumulation in the preclinical stages of AD. Our approach directly addresses the 2018 FDA framework for designing clinical trials in the stage 1 (preclinical) phase of AD (Rafii & Aisen, 2019). We use well characterised AD biomarkers (β-amyloid, medial temporal atrophy, APOE 4) to generate

a prognostic index for stratification. Using this multimodal index derived from baseline data, we predict future tau accumulation, a known pathological driver of AD progression.

Deploying our machine learning approach, we were able to accurately predict which individuals will accumulate pathological tau. In both the ADNI and BACS samples we uncovered regions that largely replicates findings from several other studies. Using baseline FTP as a proxy for disease severity, Pontecorvo et al. show that in earlier stages of AD increases in FTP were seen predominantly in the inferior lateral temporal cortex and in the posterior cingulate (Pontecorvo et al., 2019). In addition, Schultz et al. show that preclinical AD individuals (β-amyloid positive cognitively unimpaired) accumulate tau in regions extending from the amygdala, banks of the superior temporal sulcus, entorhinal, fusiform, inferior parietal, inferior temporal cortex parahippocampal gyrus and precuneus (Schultz et al., 2018). This largely supports the definition comprising the 'meta regions' of tau accumulation described in Jack et al (Jack, Wiste, et al., 2018) (**Figure 7.5)**.

Figure 7.5 correspondence of predication accuracy within tau susceptible regions reported in independent cohorts

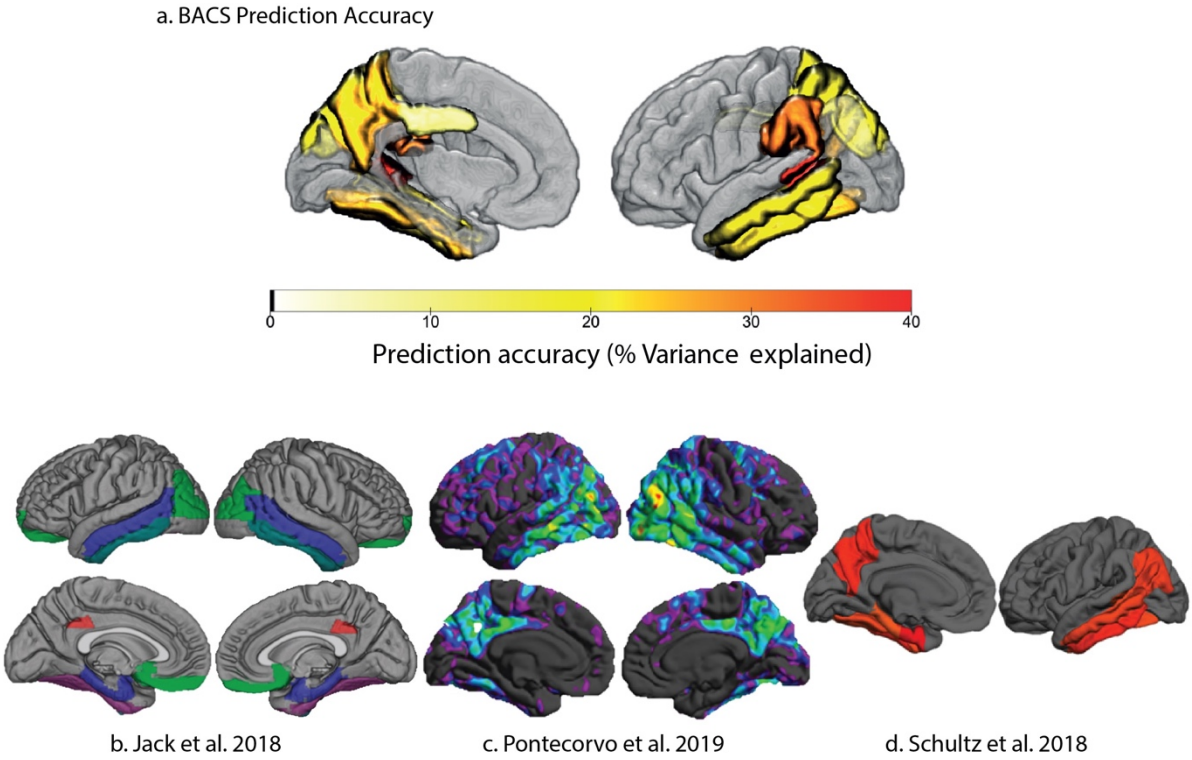


Figure 7.5 *shows the regions that are significantly accumulating tau in early AD across different cohorts.* ***Figure 7.5a****. shows the accuracy of predictions from the work presented in experimental chapter 3.* ***Figure 7.5b*** *shows the meta regions defined by Jack et al 2018, The regions comprising the early Alzheimer's disease change and temporal meta-region of interest are indicated in red, blue, aqua, magenta; regions in green are the late AD change meta regions.* ***Figure 7.5c****. shows longitudinal tau accumulation patterns from Pontecorvo et al. 2019, the heat map is the mean voxel-wise change from baseline to 18 months for β-amyloid+ subjects with intermediate tau deposition at baseline.* ***Figure 7.5d*** *taken from Schultz et al. 2018, shows the difference in tauopathy in cognitively unimpaired β-amyloid+ cohort compared to β-amyloid-.*

Extending previous work, we show that a multimodal measure is superior to a unimodal (i.e. β-amyloid positive only) classification for determining if an individual will accumulate pathological tau. We show that using multimodal data vs. β-amyloid status only to stratify populations results in a 46% reduction in the sample size required to observe a clinically

meaningful change in pathological tau accumulation. Here, we have shown that we can robustly predict the pathological spreading pattern of tau in the earliest stages of AD. Further, we have been able to empirically assess the predictive quality of different data types; a key finding being that a stratification based on multimodal data is more sensitive than amyloid status alone. Using our approach, we successfully translated rich AD information from the ADNI research cohort to make individualised predictions of pathophysiological change in an independent community-based population.

In sum, our GMLVQ-scalar projection methodology is a robust approach to combine baseline multimodal biomarkers into a prognostic metric across the AD continuum. Looking beyond the work presented in this thesis we aim to incorporate newly derived biomarkers from neuroimaging, genetic and cognitive composite scores to refine predictions of pathological change in AD.

Final Remarks

This thesis describes several key developments in the application of machine learning algorithms to clinically relevant questions of diagnosis, prognosis and disease mechanisms. Our key insight is that algorithms should not be optimised for highly specific prediction tasks, but rather attempt to make predictions that capture a disease state. That is, predictions should capture dynamic changes in multiple associated pathological processes of the same disease and not be too heavily constrained to any given task. In achieving this goal, we can capitalise on readily available data to 1) predict disease state 2) identify meaningful sub-groups 3) generate individualised predictions of disease trajectory. The developed methodology has potential to be deployed in a wide range of clinical domains. To date the success of machine learning in this field has been limited by its reliance on often hard to acquire predictors and a view of between-group discriminability as the measure to be optimised. Here, we present an alternative view and highlight a multitude of potential new research avenues with direct clinical applications.

8. Conclusions and Future Directions:

Benchmarks of Machine learning in AD research

Throughout the last decade machine learning has become a favoured approach to model multivariate interactions to diagnose and prognosticate in AD. Machine learning has been applied using a wide range of data from neuroimaging and cognitive scores, to biosensors and blood markers. To derive a benchmark of model performance in AD prediction we summarised accuracies of published papers diagnosing AD from cognitively normal and prognosticating at the MCI stage (i.e. sMCI vs pMCI) (**Figure 8.1**). These accuracies are derived from papers listed in three separate reviews (Ebrahimighahnavieh et al., 2020; Rathore et al., 2017; Tanveer et al., 2020). From the papers presented within these reviews we have separated the machine learning accuracies into either basic machine learning or deep and ensemble prediction methods.

In **Figure 8.1a** we show the averages of published accuracies in separating cognitively normal vs AD. We present an on average view of benchmark classification performances using a range of data modalities and how these accuracies have evolved over time. We see using this meta-analysis approach that classification performance of cognitively normal vs AD has remained fairly consistent over the last decade on average exceeding 90% accuracy. Further, when we separate published results into basic machine learning approaches (i.e. support vector machines, linear discriminate analysis or logistic regression) and more complex deep and ensemble approaches there doesn't appear to be any systematic benefit of one approach over the other. Taken together this shows that a classification accuracy of over 90% is required to be at benchmarked performance when classifying cognitively normal vs AD. Further, we see that although there has been methodological innovation over the past decade this benchmark accuracy has remained fairly constant.

Figure 8.1b shows the accuracies of classifiers separating sMCI from pMCI over the last 6 years. Similar to the classification of cognitively normal vs AD **Figure 8.1b** gives an average view of classification performance and encompasses predictors from different data modalities. Here, we see over the last 6 years that benchmark performances are between 70 and 80% accuracy. Unlike the classifiers built to discriminate between cognitively normal and AD there does appear to be a gradual increase in classification performance over the last 6 years.

However, as with the CN and AD classification there does not appear to be a systematic benefit of one learning approach over another.

Figure 8.1 Machine learning benchmark accuracies

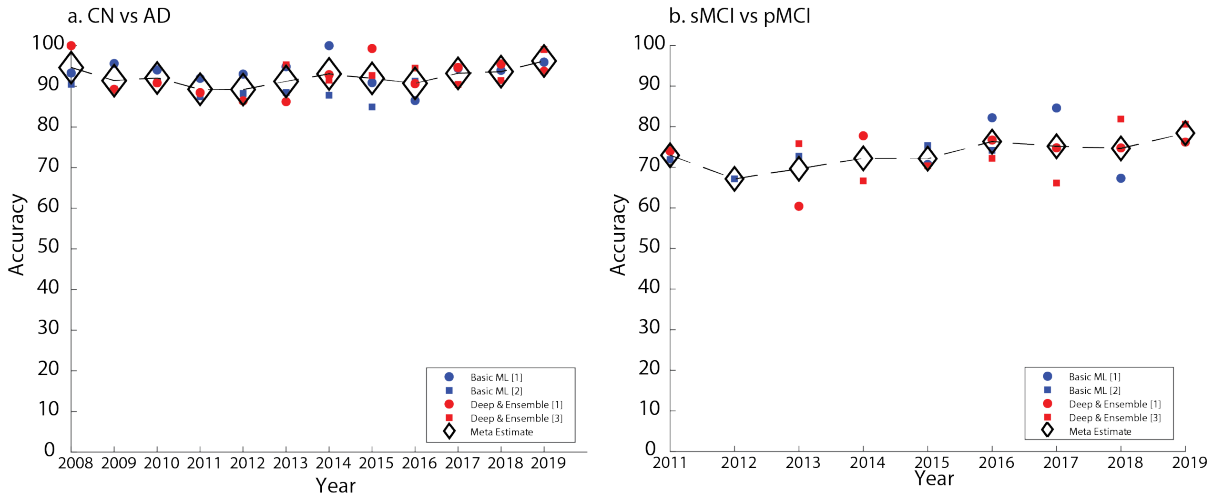


Figure 8.1a *shows the average classification accuracy in diagnosing cognitive normal (CN) vs AD each year from 2008 to 2019.* ***Figure 8.1b*** *shows the average classification accuracies in classifying individuals who are stable MCI (sMCI) vs individuals who are progressive MCI (pMCI) for each year from 2011 to 2019. Blue dots are the average classification accuracy per year for papers using basic machine learning approaches presented in [1] (Tanveer et al., 2020) (CN vs AD no. publications=47 2008:2019; sMCI vs pMCI no. publications =8 2015:2018). Blue squares are the average classification accuracies each year for papers using basic machine learning approaches presented in [2]* (Rathore et al., 2017) *(CN vs AD no. publications =43 2008:2016; sMCI vs pMCI no. publications =25 2011:2016). Red circles are the average classification accuracy per year for papers using deep or ensemble machine learning approaches presented in [1] (Tanveer et al., 2020) (CN vs AD no. publications =42 2008:2019; sMCI vs pMCI no. publications =21 2011:2019). Red squares are the average classification accuracy per year for papers using deep or ensemble machine learning approaches presented in [3] (Ebrahimighahnavieh et al., 2020) (CN vs AD no. publications =83 2012:2019; sMCI vs pMCI no. publications =27 2013:2019). Black diamonds are the meta averages across all machine learning approaches presented in the three reviews.*

Comparison of current results with benchmarks

Based on the meta-analysis shown on **figure 8.1b** the most recent benchmark performance in the sMCI vs pMCI classification is an accuracy of 78.4%. This meta average is derived from only deep learning and ensemble approaches and therefore may not represent the current benchmark accuracies across all classifier types (i.e. basic machine learning and more computationally complex approaches). In the work presented in this thesis using the GMLVQ machine learning approach we achieved a classification accuracy of 81.5% for the model built using cognitive data only and an accuracy of 81.9% for the model built using biological data. Therefore, our interpretable linear approach performs better than the average deep learning application to this problem. We are currently unable to compare the performance of the scalar projection machine learning approach to other models to predict individualised trajectories as consistent outcome measures are not as common (i.e. variability in cognitive outcome measures). However, in the future data prediction challenges offer an objective way to compare machine learning approaches that aim to capture individualised trajectories. For example, the TADPOLE challenge is designed to use standardised ADNI data sets to test predictions across three different tasks. 1.) prediction of diagnosis, 2.) change in ventricle volume and 3.) change in ADAS-cog scores (Marinescu et al., 2018). In the future we will apply our machine learning approach to make predictions in this data and compare the prediction performance of the scalar projection in modelling individualised trajectories from other methodologies submitted to the challenge.

Limitations to current benchmarks

Figure 8.1 gives an overview of how well machine learning has performed over the last decade, however this does not represent a systematic review of the literature. Rather, we provide an overall picture of how the field is evolving and whether there appears to be any perceived benefit in using more complex deep methods over basic machine learning approaches. The combined accuracies cover a wide range of data modalities, feature extraction and cross validation approaches and these are discussed in more detail in the original review articles. However, across each of these reviews it is apparent that a lack of sample heterogeneity is a major limitation of these modelling papers. For example, 90% of the papers reviewed by Ebrahimighahnavieh et.al use ADNI data to build their models (Ebrahimighahnavieh et al.,

2020). The use of ADNI in most machine learning papers poses a major concern for future developments and clinical utility. Given the exclusion criterion and sampling characteristics of ADNI, models are built on highly homogeneous and likely non representative samples. That is, the exclusion criterions of ADNI remove individuals with co-morbid conditions, including those with existing vascular conditions. Given that mixed dementia phenotypes of vascular dementia and AD are common, this limits the generalisability of models built in ADNI to a clinical setting. Further, the samples within ADNI are generally highly educated and Caucasian dominant. These two factors therefore skew samples to higher socioeconomic brackets. Given that socioeconomic factors have a considerable effect on general health (Adler & Ostrove, 1999) as well as cognitive health (Greenfield & Moorman, 2019) this greatly limits any models ability to be deployable in a real world setting.

Looking into the future, considerable work needs to be done to a.) combine data from many different cohorts to generate a more heterogeneous sample and to b.) harmonise these data so that predictions are not limited by nuances in the data (i.e. scanner differences or cognitive outcome measures). Given the diverse disease aetiologies of patients who present to a memory clinic this harmonisation is a critical interstitial step for developing machine learning tools with real world impact.

Future directions

Title of Project and Brief Description

Title: Predicting cognitive trajectories in a harmonised global cohort.

Harmonising cognitive data to generate robust trajectories: The primary goal of large global studies investigating Alzheimer's Disease (AD) is to make meaningful and robust conclusions on cognitive outcomes. However, this is challenging due to heterogeneous cognitive test batteries used across studies and non-representative sampling characteristics. Despite these limitations cognitive composites have become the favoured method for tracking and predicting cognitive trajectories. Given the variability in cognitive test batteries and sampling characteristics between studies these cognitive composites suffer from poor interoperability, that is they are study specific and cannot be generalised to the wider population or be built

leveraging multiple large data sets. The proposed project will address this issue by aggregating cognitive data across a wide range of studies and generating a cross study cognitive composite.

*Mathematical Approaches to harmonise data***:** There is significant variability across tests used in cognitive test batteries in large longitudinal data sets. However, between these studies there are many overlapping cognitive tests (i.e. shared tests across cognitive batteries). Specifically, this project will involve the aggregation of these data into an incomplete matrix and utilise data imputation to solve for missing values (i.e. the predicted score on a test that was not administered within a study). Using the imputed matrix, we will then generate a set of cross study cognitive composites leveraging modern psychometric modelling techniques. We will then test to see if the resultant cognitive composite is both sensitive to AD pathology (i.e. amyloid positivity) and robust across cohorts.

Neuroimaging biomarkers: Using ADNI we have developed and validated the PLSr-RFE biomarker generation methodology. Previously we used this methodology to capture latent structures in T1 structural MRI that relate to AD symptomology from the ADNI data base, which is a typical amnestic AD research sample. This method has generated robust, reliable and interpretable biomarkers in structural MRI. However, due to the sampling characteristics of ADNI we were able to extract only the typical medial temporal atrophy phenotype. Using the harmonised heterogeneous sample from several global cohorts we hope to uncover different neuroimaging phenotypes related to cognitive dysfunction. Specifically, by combining ADNI with samples that have different inclusion criteria (i.e. including co-morbid vascular conditions) we hope to find neuroimaging markers related to cognitive dysfunction in AD, vascular dementia and mixed AD and vascular dementia. Further, we will use the PLSr-RFE procedure to derive predictive features in different neuroimaging modalities (e.g. DTI, fMRI). Given that we will be using several imaging modalities taken from many different sites data reliability will be investigated thoroughly. To achieve this, we will generate a series of artificial neuroimaging markers (e.g. cortical thinning and white matter legions) prior to pre-processing and compare the ability to detect these artificial markers across the different scanner sites.

Prognostic modelling validation: Using the derived neuroimaging markers we will then use the GMLVQ-scalar projection approach to generate predictions of cognitive trajectories. Using the harmonised cross cohort composite we will look to determine how sensitive this measure is to AD related change (i.e. amyloid positivity) and how robust predictions of cognitive

trajectories are using the scalar projection approach. Using samples that are recruited in a memory clinic setting we will then test our predictions out of sample. Finally, we will run a cost benefit analysis of using different neuroimaging markers so we can determine the most effective and least expensive measures to accurately model individual disease trajectories.

Machine learning in clinical practice

Within the field of clinical medicine there has been a rapid expansion of data generation, curation and storage. Not only have the total numbers of data available increased but also the digitisation of large data, such as; medical imaging, genome sequencing, real time biosensors and health records. Naturally with such large quantities of data becoming available; high throughput, rapid, and autonomous analysis is an exciting prospect. This desire for an all in one kind of approach lends itself to deep learning. Accordingly, deep learning is one of the most heavily publicised routes to achieving rapid and precise diagnosis in clinical medicine.

Deep learning vs. traditional machine learning

Traditional machine learning is a two-stage process that relies on machine intelligence as well as specific expertise from a human operator to select or construct informative representations (i.e. features) of raw data. In contrast, deep learning is the desired all in one package as the learning algorithm is fed raw or minimally pre-processed data and autonomously learns a set of lower dimensional representations allowing samples to be distinguishable from one another (i.e. cognitively normal or AD). The deep learning architecture is a series of layers of basic non-linear functions, as the raw data is passed from layer to layer the input space becomes iteratively transformed so that in the final layers of the model a judgement can be made. Deep learning is not a recent concept but until recently neither the digitised data nor computational resources were available for deep learning to truly be effective. Recently, however, thanks to Moore's law and the internet, deep learning is now one of the most favoured types of learning algorithm in big data. The field in which deep learning has come into its own is computer vision, where it has excelled at image recognition (Krizhevsky et al., 2012). Given the success in computer vision deep learning is well suited to analysis of medical imaging. Accordingly,

deep learning has shown encouraging results when applied to detect abnormalities across several medical disciplines such as dermatology, radiology, ophthalmology, and pathology (for reviews (Esteva et al., 2019; Topol, 2019) . Within the field of clinical medicine deep learning models are now approaching clinician level accuracy in a variety of diagnostic tasks that use medical imaging (for review (X. Liu et al., 2019)).

Specifically, within the field of radiology deep learning is now entering the clinical pathway, with a deep learning model deployed in Indian hospitals producing interpretations of chest X-rays with accuracies that that are of equal or better than radiologists (R. Singh et al., 2018).

This applied appraisal of radiography by deep learning is also helping with rapid diagnosis of fractures in difficult regions and can improve sensitivity to small fractures over trained ER clinicians while drastically reducing to make an accurate diagnosis (Lindsey et al., 2018). Clearly deep learning is capable of performing the high throughput, rapid and precise appraisal of imaging data particularly in tasks that are time sensitive. However, although there is substantial research in deep learning for AD prediction (Ebrahimighahnavieh et al., 2020) there are no currently available clinical AD deep learning tools.

Why deep learning is not found in clinical practice in AD?

The most common applications of deep learning in AD research are for a.) Diagnosis of cognitively normal vs AD and b.) prognostication of sMCI vs pMCI. There are two fundamental barriers to these types of predictive algorithms ever becoming clinically deployable. The first regards the utility -or lack thereof- of an algorithm that separates cognitively individuals from individuals with dementia. Given that there are clearly observable cognitive differences between a cognitively normal and demented individual and that generally clear and distinctive markers can be seen on neuroimaging (i.e. large ventricles, medial temporal and global atrophy), rapid inspection by a radiographer and clinician is already feasible. With this in mind a more useful deployment in clinic is the use of deep learning to separate individuals with dementia but different disease aetiologies (i.e. AD vs PD). These types of deep learning algorithms have been developed and are successful at discerning underlying disease (Joshi et al., 2010).

In the case of deep learning approaches that have been developed for prognostication (i.e. sMCI vs pMCI) the issue is less of practicality but more of fundamental design. As discussed at multiple points in this thesis there are several inherent issues with designing an algorithm to separate sMCI vs pMCI. For example, these models define class labels based on arbitrary

follow up duration (i.e. stable MCI for how long? Progress within what time frame?) and heavily biased clinical diagnosis. Previous work interrogating the sensitivity and specificity of a diagnosis of probable AD ante mortem, and, the presence of pathological biomarkers post mortem shows that diagnoses of probable AD have a sensitivity to pathology between 70.9% and 87.3% and a specificity between 44.3% and 70.8% (Beach, Monsell, Phillips, & Kukull, 2012).

When defining the ground truth (i.e. the clinician's diagnosis) for model formulation on a diagnosis that in its own right has poor sensitivity and specificity to AD the model will always be limited by the reliability of the clinical diagnosis. Given that a deep learning algorithm is heavily parameterised to minimise error in training, the model will use variability in the neuroimaging in an attempt to describe the variance within the clinician's appraisal. Thus, the model will overfit the data to replicate a clinician and not diagnose a disease. In doing so there is a lower chance of this model generalising to a new sample, because although the variance of the clinical diagnosis is not random (i.e. patient independence and carer appraisal) it is unlikely that there will be markers within the neuroimaging that can explain this variation.

With this in mind, any deep learning that is built with the sole purpose of prognostication of sMCI vs pMCI seems invalid. Therefore, future mathematical development of deep learning models to increase performance on the heavily researched classification of sMCI vs pMCI has almost no clinical utility. However, that is not to say that all deep learning does not serve a purpose in AD prediction.

Taking Deep learning into the clinic

Previous successes of applied deep learning within a clinical setting centre around the rapid and sensitive appraisal of diagnostic imaging. Within AD research a potential application could be the rapid count and grading of cerebral micro bleeds. For example, using deep learning to combine information from a set of MRI sequences (e.g. multiparametric mapping) a rapid account of the underlying vasculature for a given subject could be made available to the clinician. Armed with this snapshot of the cerebral vascular health the unique or additive effect of vascular dementia on AD could be elucidated in a clinical setting, resulting in more targeted treatment and care plans.

Given the success of deep learning in segmentation, another potential application in AD research is the rapid and accurate segmentation of fine anatomical structures. This would allow several different contrasts (i.e. T1 T2*) to be applied simultaneously for a given subject and

fine scale volumetrics of medial temporal structures be performed. Performed longitudinally, these assessments not only give the clinician a fine scale physiological assessment of the neurodegeneration process in AD but could also provide a sensitive diagnostic tool to isolate non typical (i.e. non amnestic dominant) AD variants.

Looking into the future, deep learning offers an interesting and clinically relevant route of innovation by generating synthetic data. As the current diagnostic framework of AD can require up to three separate imaging modalities, the ability to synthesise some imaging markers (i.e. generate PET from MRI) is of critical significance both from a safety and economic perspective. A subtype of deep learning Generative Adversarial Networks (GANs) provides the perfect framework for synthesising these data. GANs are two neural networks that aim to fool each other, where the generator network aims to build a representation of the data (i.e. neuroimaging scan) that confuses the discriminator as to the real of synthetic data. This type of modelling approach has already seen success in synthesising PiB PET data to determine demyelination in multiple sclerosis, where a GAN was successfully able to combine imaging data from MRI (structural and DTI) to generate synthetic PiB PET data (Wei et al., 2020).

Bridging the gap to clinic with traditional machine learning

As discussed previously the most common classification tasks in AD research have either limited clinical utility (i.e. cognitively normal vs AD) or have questionable validity (i.e. sMCI vs pMCI), therefore the use of poorly interpretable and heavily constrained models (i.e. deep learning) may not bridge the gap from research to the clinic. Machine learning however offers a different route. As features need to be carefully constructed with expertise knowledge, a predictive tool may have more clinical utility to operate a two-stage process. First, following expert research, highly predictive features are extracted from the raw patient data. Following this these features should then be conferred to the clinician along with relevant parametric statistics (e.g. 80% of amyloid positive individuals have a value of X or greater). Following this, a multivariate prediction using the presented features is made (i.e. machine learning) and similarly this multivariate prediction is conferred to the clinician using a meaningful metric. For example, survival statistics are commonplace in clinical medicine and therefore may provide a smoother adoption into clinic.

There are several motivations behind this type of entry to the clinic. First, by presenting the clinician with the features used to make the prediction the initial stage of the model operates

as a rapid data aggregator and consolidator, allowing for near immediate interpretation of high dimensional data (i.e. neuroimaging). Second, by displaying how the algorithm makes predictions including the physiological origin of the predictors a greater transparency is achieved and will likely confer greater trust in the resultant prediction. Finally, by having an open and transparent system the machine learning researcher and the clinician can work synergistically to determine the most effective and meaningful predictions to make.

Going forward, machine learning approaches need to be looked at with healthy scepticism as models will always be limited by a.) the test-retest reliability of the outcome measure and b.) the resolution of the outcome measure. In light of this, it may be of more use to try and predict real world outcomes rather than changes in proprietary lab administered tests. For example, it is likely more useful for clinician and patient alike to accurately predict a time when the patient is no longer able to remain independent, rather than a decrease in the MMSE. This disconnect between predicted outcomes and clinically useful prognosis represents a significant barrier to machine learning entering the clinical treatment pathway for AD. Therefore, the road to implementing these highly useful and powerful tools into clinic requires a constant interaction between clinicians, patients and researchers into what is a meaningful prediction.

9. Summary

Since the first presentation on Auguste D in 1906 the scientific inquiry regarding the driving forces in AD have been well documented. Diligent work across the biological and clinical sciences has elucidated genetic, cellular and system level interactions that lead to AD. However, a treatment, and reliable prognosis of AD still eludes clinical neuroscience. Recently, the global AD research community has united to coordinate the collection, curation and sharing of expanses of predictive data. In doing so they have encouraged researchers to use data intensive computational approaches to try and integrate multi domain data into a holistic picture of AD. However, the computationally minded researcher must carefully approach this task, ensuring that due diligence is taken to respect the years of fastidious collection and collation of data by biologists. Therefore, it is imperative that computational approaches be readily distilled to biologists. Further, it is clear, for a computational approach to truly impact clinical practice the underlying decision made by the machine must be interpretable and tractable by the clinical practitioner.

Here we have endeavoured to achieve these goals. We have introduced a series of computational approaches that derive and utilise interpretable predictors to model prognostic trajectories in AD. In keeping with the state-of-the-art biological definitions of AD we have also made our models approachable to biologists; incorporating interpretable, reliable and well-defined measures of the biomarkers that define AD. In doing so we have introduced a computational approach that makes accessible but clinically meaningful predictions.

References

Adler, N., & Ostrove, J. (1999). Socioeconomic Status and Health: What We Know and What We Don't. https://doi.org/10.1111/j.1749-6632.1999.tb08101.x

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., … Phelps, C. H. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, *7*(3), 270–279. https://doi.org/10.1016/J.JALZ.2011.03.008

Allison, S. L., Koscik, R. L., Cary, R. P., Jonaitis, E. M., Rowley, H. A., Chin, N. A., … Johnson, S. C. (2019). Comparison of different MRI-based morphometric estimates for defining neurodegeneration across the Alzheimer's disease continuum. *NeuroImage: Clinical*, *23*, 101895. https://doi.org/10.1016/j.nicl.2019.101895

Alsaedi, A., Abdel-Qader, I., Mohammad, N., & Fong, A. C. (2018). Extended cox proportional hazard model to analyze and predict conversion from mild cognitive impairment to alzheimer's disease. In *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)* (pp. 131–136). IEEE. https://doi.org/10.1109/CCWC.2018.8301669

Alzheimer's Association. (2018). 2018 Alzheimer's disease facts and figures. *Alzheimer's and Dementia*, *14*(3), 367–429. https://doi.org/10.1016/j.jalz.2018.02.001

Alzheimer, A. (1907). Uber eine eigenartige Erkrankung der Hirnrinde. *Zentralbl. Nervenh. Psych.*, *18*, 177–179.

Arbabshirani, M. R., Plis, S., Sui, J., & Calhoun, V. D. (2017). Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *NeuroImage*, *145*, 137–165. https://doi.org/10.1016/j.neuroimage.2016.02.079

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage*, *38*(1), 95–113. https://doi.org/10.1016/j.neuroimage.2007.07.007

Ayutyanont, N., Langbaum, J. B. S., Hendrix, S. B., Chen, K., Fleisher, A. S., Friesenhahn, M., … Reiman, E. M. (2014). The Alzheimer's prevention initiative composite cognitive test score: sample size estimates for the evaluation of preclinical Alzheimer's disease treatments in presenilin 1 E280A mutation carriers. *The Journal of Clinical Psychiatry*, *75*(6), 652–660. https://doi.org/10.4088/JCP.13m08927

Baker, S. L., Maass, A., & Jagust, W. J. (2017). Considerations and code for partial volume

correcting [18F]-AV-1451 tau PET data. *Data in Brief*, *15*, 648–657. https://doi.org/10.1016/J.DIB.2017.10.024

Bakkour, A., Morris, J. C., Wolk, D. A., & Dickerson, B. C. (2013). The effects of aging and Alzheimer's disease on cerebral cortical anatomy: Specificity and differential relationships with cognition. *NeuroImage*, *76*, 332–344. https://doi.org/10.1016/j.neuroimage.2013.02.059

Bateman, R. J., Xiong, C., Benzinger, T. L. S., Fagan, A. M., Goate, A., Fox, N. C., … Morris, J. C. (2012). Clinical and biomarker changes in dominantly inherited Alzheimer's disease. *New England Journal of Medicine*, *367*(9), 795–804. https://doi.org/10.1056/NEJMoa1202753

Beach, T. G., Monsell, S. E., Phillips, L. E., & Kukull, W. (2012). Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005-2010. *Journal of Neuropathology and Experimental Neurology*, *71*(4), 266–273. https://doi.org/10.1097/NEN.0b013e31824b211b

Belleville, S., Fouquet, C., Hudon, C., Zomahoun, H. T. V., & Croteau, J. (2017). Neuropsychological Measures that Predict Progression from Mild Cognitive Impairment to Alzheimer's type dementia in Older Adults: a Systematic Review and Meta-Analysis. *Neuropsychology Review*, *27*(4), 328–353. https://doi.org/10.1007/s11065-017-9361-5

Bhagwat, N., Viviano, J. D., Voineskos, A. N., Chakravarty, M. M., & Initiative, A. D. N. (2018). Modeling and prediction of clinical symptom trajectories in Alzheimer's disease using longitudinal data. *PLOS Computational Biology*, *14*(9), e1006376. https://doi.org/10.1371/journal.pcbi.1006376

Bilgel, M., An, Y., Helphrey, J., Elkins, W., Gomez, G., Wong, D. F., … Resnick, S. M. (2018). Effects of amyloid pathology and neurodegeneration on cognitive change in cognitively normal adults. *Brain*, *141*(8), 2475–2485. https://doi.org/10.1093/brain/awy150

Bilgel, M., An, Y., Lang, A., Prince, J., Ferrucci, L., Jedynak, B., & Resnick, S. M. (2014). Trajectories of Alzheimer disease-related cognitive measures in a longitudinal sample. *Alzheimer's & Dementia*, *10*(6), 735-742.e4. https://doi.org/10.1016/J.JALZ.2014.04.520

Bilgel, M., Koscik, R. L., An, Y., Prince, J. L., Resnick, S. M., Johnson, S. C., & Jedynak, B. M. (2017). Temporal order of Alzheimer's disease-related cognitive marker changes in BLSA and WRAP longitudinal studies. *Journal of Alzheimer's Disease*, *59*(4), 1335–1347. https://doi.org/10.3233/JAD-170448

Bischof, G. N., Jessen, F., Fliessbach, K., Dronse, J., Hammes, J., Neumaier, B., … van Eimeren, T. (2016). Impact of tau and amyloid burden on glucose metabolism in Alzheimer's disease. *Annals of Clinical and Translational Neurology*, *3*(12), 934–939. https://doi.org/10.1002/acn3.339

Blennow, K., Wallin, A., Ågren, H., Spenger, C., Siegfried, J., & Vanmechelen, E. (1995). tau protein in cerebrospinal fluid - A biochemical marker for axonal degeneration in Alzheimer disease? *Molecular and Chemical Neuropathology*, *26*(3), 231–245. https://doi.org/10.1007/BF02815140

Blennow, Kaj, & Hampel, H. (2003, October 1). CSF markers for incipient Alzheimer's disease. *Lancet Neurology*. Lancet Publishing Group. https://doi.org/10.1016/S1474-4422(03)00530-1

Blumer, A., Ehrenfeucht, A., Haussler, D., & Warmuth, M. K. (1987). Occam's Razor. *Information Processing Letters*, *24*(6), 377–380. https://doi.org/10.1016/0020-0190(87)90114-1

Braak, H., & Braak, E. (1991, September). Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathologica*. Springer-Verlag. https://doi.org/10.1007/BF00308809

Braak, Heiko, & Del Tredici, K. (2011). The pathological process underlying Alzheimer's disease in individuals under thirty. *Acta Neuropathologica*, *121*(2), 171–181. https://doi.org/10.1007/s00401-010-0789-4

Braak, Heiko, & Del Tredici, K. (2012, December). Where, when, and in what form does sporadic Alzheimer's disease begin? *Current Opinion in Neurology*. https://doi.org/10.1097/WCO.0b013e32835a3432

Brier, M. R., Gordon, B., Friedrichsen, K., McCarthy, J., Stern, A., Christensen, J., … Ances, B. M. (2016). Tau and Ab imaging, CSF measures, and cognition in Alzheimer's disease. *Science Translational Medicine*, *8*(338), 338ra66. https://doi.org/10.1126/scitranslmed.aaf2362

Brion, J.-P. (1998). Neurofibrillary Tangles and Alzheimer's Disease. *European Neurology*, *40*(3), 130–140. https://doi.org/10.1159/000007969

Burnham, S. C., Bourgeat, P., Doré, V., Savage, G., Brown, B., Laws, S., … AIBL Research Group. (2016). Clinical and cognitive trajectories in cognitively healthy elderly individuals with suspected non-Alzheimer's disease pathophysiology (SNAP) or Alzheimer's disease pathology: a longitudinal study. *The Lancet Neurology*, *15*(10), 1044–1053. https://doi.org/10.1016/S1474-4422(16)30125-9

Burnham, S. C., Bourgeat, P., Doré, V., Savage, G., Brown, B., Laws, S., … Villemagne, V.

L. (2016). Clinical and cognitive trajectories in cognitively healthy elderly individuals with suspected non-Alzheimer's disease pathophysiology (SNAP) or Alzheimer's disease pathology: a longitudinal study. *The Lancet Neurology*, *15*(10), 1044–1053. https://doi.org/10.1016/S1474-4422(16)30125-9

Casanova, R., Hsu, F.-C., Sink, K. M., Rapp, S. R., Williamson, J. D., Resnick, S. M., … Initiative, for the A. D. N. (2013). Alzheimer's Disease Risk Assessment Using Large-Scale Machine Learning Methods. *PLoS ONE*, *8*(11), e77949. https://doi.org/10.1371/journal.pone.0077949

Castellani, R. J., Plascencia-Villa, G., & Perry, G. (2019, July 1). The amyloid cascade and Alzheimer's disease therapeutics: theory versus observation. *Laboratory Investigation*. Nature Publishing Group. https://doi.org/10.1038/s41374-019-0231-z

Castellano, J. M., Kim, J., Stewart, F. R., Jiang, H., DeMattos, R. B., Patterson, B. W., … Holtzman, D. M. (2011). Human apoE isoforms differentially regulate brain amyloid-β peptide clearance. *Science Translational Medicine*, *3*(89). https://doi.org/10.1126/scitranslmed.3002156

Chapman, R. M., Mapstone, M., McCrary, J. W., Gardner, M. N., Porsteinsson, A., Sandoval, T. C., … Reilly, L. A. (2011). Predicting conversion from mild cognitive impairment to Alzheimer's disease using neuropsychological tests and multivariate methods. *Journal of Clinical and Experimental Neuropsychology*, *33*(2), 187–199. https://doi.org/10.1080/13803395.2010.499356

Chasioti, D., Yan, J., Nho, K., & Saykin, A. J. (2019, May 1). Progress in Polygenic Composite Scores in Alzheimer's and Other Complex Diseases. *Trends in Genetics*. Elsevier Ltd. https://doi.org/10.1016/j.tig.2019.02.005

Chen, R., Hu, Z., Wei, L., Qin, X., McCracken, C., & Copeland, J. R. (2008). Severity of depression and risk for subsequent dementia: cohort studies in China and the UK. *British Journal of Psychiatry*, *193*(05), 373–377. https://doi.org/10.1192/bjp.bp.107.044974

Cho, H., Choi, J. Y., Hwang, M. S., Kim, Y. J., Lee, H. M., Lee, H. S., … Lyoo, C. H. (2016). In vivo cortical spreading pattern of tau and amyloid in the Alzheimer disease spectrum. *Annals of Neurology*, *80*(2), 247–258. https://doi.org/10.1002/ana.24711

Chu, C., Hsu, A. L., Chou, K. H., Bandettini, P., & Lin, C. P. (2012). Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. *NeuroImage*, *60*(1), 59–70. https://doi.org/10.1016/j.neuroimage.2011.11.066

Corder, E. H., Saunders, A. M., Strittmatter, W. J., Schmechel, D. E., Gaskell, P. C., Small, G. W., … Pericak-Vance, M. A. (1993). Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science*, *261*(5123), 921–923. https://doi.org/10.1126/science.8346443

Crane, P. K., Carle, A., Gibbons, L. E., Insel, P., Mackin, R. S., Gross, A., … Initiative, for the A. D. N. (2012). Development and assessment of a composite score for memory in the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Brain Imaging and Behavior*, *6*(4), 502–516. https://doi.org/10.1007/s11682-012-9186-z

Crary, J. F., Trojanowski, J. Q., Schneider, J. A., Abisambra, J. F., Abner, E. L., Alafuzoff, I., … Nelson, P. T. (2014). Primary age-related tauopathy (PART): a common pathology associated with human aging. *Acta Neuropathologica*, *128*(6), 755–766. https://doi.org/10.1007/s00401-014-1349-0

Cui, Y., Liu, B., Luo, S., Zhen, X., Fan, M., Liu, T., … Initiative, the A. D. N. (2011). Identification of Conversion from Mild Cognitive Impairment to Alzheimer's Disease Using Multivariate Predictors. *PLoS ONE*, *6*(7), e21896. https://doi.org/10.1371/journal.pone.0021896

Cummings, J., Reiber, C., & Kumar, P. (2018). The price of progress: Funding and financing Alzheimer's disease drug development. *Alzheimer's and Dementia: Translational Research and Clinical Interventions*, *4*, 330–343. https://doi.org/10.1016/j.trci.2018.04.008

Davatzikos, C., Xu, F., An, Y., Fan, Y., & Resnick, S. M. (2009). Longitudinal progression of Alzheimer's-like patterns of atrophy in normal older adults: the SPARE-AD index. *Brain*, *132*(8), 2026–2035. https://doi.org/10.1093/brain/awp091

Davatzikos, Christos. (2019, August 15). Machine learning in neuroimaging: Progress and challenges. *NeuroImage*. Academic Press Inc. https://doi.org/10.1016/j.neuroimage.2018.10.003

Davatzikos, Christos, Sotiras, A., Fan, Y., Habes, M., Erus, G., Rathore, S., … Kontos, D. (2019). Precision diagnostics based on machine learning-derived imaging signatures. *Magnetic Resonance Imaging*, *64*, 49–61. https://doi.org/10.1016/j.mri.2019.04.012

De Leon, M. J., Convit, A., Wolf, O. T., Tarshish, C. Y., DeSanti, S., Rusinek, H., … Fowler, J. (2001). Prediction of cognitive decline in normal elderly subjects with 2-[18F]fluoro-2-deoxy-D-glucose/positron-emission tomography (FDG/PET). *Proceedings of the National Academy of Sciences of the United States of America*, *98*(19), 10966–10971. https://doi.org/10.1073/pnas.191044198

Defrancesco, M., Marksteiner, J., Kemmler, G., Fleischhacker, W. W., Blasko, I., & Deisenhammer, E. A. (2017). Severity of Depression Impacts Imminent Conversion from Mild Cognitive Impairment to Alzheimer's Disease. *Journal of Alzheimer's Disease*, *59*(4), 1439–1448. https://doi.org/10.3233/JAD-161135

Desikan, R. S., Fischl, B., Cabral, H. J., Dillon, W. P., Salat, D. H., Settecase, F., … Glastonbury, C. M. (2010). Automated MRI measures predict progression to Alzheimer's disease. *Neurobiology of Aging*, *31*(8), 1364–1374. https://doi.org/10.1016/j.neurobiolaging.2010.04.023

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., … Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, *31*(3), 968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021

Devanand, D. P., Pradhaban, G., Liu, X., Khandji, A., De Santi, S., Segal, S., … De Leon, M. J. (2007). Hippocampal and entorhinal atrophy in mild cognitive impairment: Prediction of Alzheimer disease. *Neurology*, *68*(11), 828–836. https://doi.org/10.1212/01.wnl.0000256697.20968.d7

Dickerson, B. C., Stoub, T. R., Shah, R. C., Sperling, R. A., Killiany, R. J., Albert, M. S., … Detoledo-Morrell, L. (2011). Alzheimer-signature MRI biomarker predicts AD dementia in cognitively normal adults. *Neurology*, *76*(16), 1395–1402. https://doi.org/10.1212/WNL.0b013e3182166e96

Dickerson, Bradford C., Bakkour, A., Salat, D. H., Feczko, E., Pacheco, J., Greve, D. N., … Buckner, R. L. (2009). The cortical signature of Alzheimer's disease: Regionally specific cortical thinning relates to symptom severity in very mild to mild AD dementia and is detectable in asymptomatic amyloid-positive individuals. *Cerebral Cortex*, *19*(3), 497–510. https://doi.org/10.1093/cercor/bhn113

Dong, A., Toledo, J. B., Honnorat, N., Doshi, J., Varol, E., Sotiras, A., … Davatzikos, C. (2016). Heterogeneity of neuroanatomical patterns in prodromal Alzheimer's disease: links to cognition, progression and biomarkers. *Brain*, *140*(3), aww319. https://doi.org/10.1093/brain/aww319

Dong, A., Toledo, J. B., Honnorat, N., Doshi, J., Varol, E., Sotiras, A., … Davatzikos, C. (2017). Heterogeneity of neuroanatomical patterns in prodromal Alzheimer's disease: links to cognition, progression and biomarkers. *Brain*, *140*(3), 735–747. https://doi.org/10.1093/brain/aww319

Donohue, M. C., Sperling, R. A., Salmon, D. P., Rentz, D. M., Raman, R., Thomas, R. G., …

Aisen, P. S. (2014). The Preclinical Alzheimer Cognitive Composite. *JAMA Neurology*, *71*(8), 961. https://doi.org/10.1001/jamaneurol.2014.803

Donovan, N. J., Amariglio, R. E., Zoller, A. S., Rudel, R. K., Gomez-Isla, T., Blacker, D., … Rentz, D. M. (2014). Subjective cognitive concerns and neuropsychiatric predictors of progression to the early clinical stages of Alzheimer disease. *American Journal of Geriatric Psychiatry*, *22*(12), 1642–1651. https://doi.org/10.1016/j.jagp.2014.02.007

Dubois, B., Hampel, H., Feldman, H. H., Scheltens, P., Aisen, P., Andrieu, S., … Jack, C. R. (2016). Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria. *Alzheimer's & Dementia*, *12*(3), 292–323. https://doi.org/10.1016/j.jalz.2016.02.002

Dukart, J., Sambataro, F., & Bertolino, A. (2015). Accurate prediction of conversion to Alzheimer's disease using imaging, genetic, and neuropsychological biomarkers. *Journal of Alzheimer's Disease*, *49*(4), 1143–1159. https://doi.org/10.3233/JAD-150570

Dumurgier, J., Hanseeuw, B. J., Hatling, F. B., Judge, K. A., Schultz, A. P., Chhatwal, J. P., … Gómez-Isla, T. (2017). Alzheimer's Disease Biomarkers and Future Decline in Cognitive Normal Older Adults. *Journal of Alzheimer's Disease*, *60*(4), 1451–1459. https://doi.org/10.3233/JAD-170511

Ebrahimighahnavieh, M. A., Luo, S., & Chiong, R. (2020). Deep learning to detect Alzheimer's disease from neuroimaging: A systematic literature review. *Computer Methods and Programs in Biomedicine*, *187*, 105242. https://doi.org/10.1016/j.cmpb.2019.105242

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., … Dean, J. (2019, January 1). A guide to deep learning in healthcare. *Nature Medicine*. Nature Publishing Group. https://doi.org/10.1038/s41591-018-0316-z

Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., … Jiang, T. (2016). The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture. *Cerebral Cortex*, *26*(8), 3508–3526. https://doi.org/10.1093/cercor/bhw157

Fan, Y., Batmanghelich, N., Clark, C. M., & Davatzikos, C. (2008). Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. *NeuroImage*, *39*(4), 1731–1743. https://doi.org/10.1016/j.neuroimage.2007.10.031

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Franzmeier, N., Neitzel, J., Rubinski, A., Smith, R., Strandberg, O., Ossenkoppele, R., … Raj, B. A. (2020). Functional brain architecture is associated with the rate of tau accumulation in Alzheimer's disease. *Nature Communications*, *11*(1), 1–17. https://doi.org/10.1038/s41467-019-14159-1

Frölich, L., Peters, O., Lewczuk, P., Gruber, O., Teipel, S. J., Gertz, H. J., … Kornhuber, J. (2017). Incremental value of biomarker combinations to predict progression of mild cognitive impairment to Alzheimer's dementia. *Alzheimer's Research and Therapy*, *9*(1), 84. https://doi.org/10.1186/s13195-017-0301-7

Furst, A. J., Rabinovici, G. D., Rostomian, A. H., Steed, T., Alkalay, A., Racine, C., … Jagust, W. J. (2012). Cognition, glucose metabolism and amyloid burden in Alzheimer's disease. *Neurobiology of Aging*, *33*(2), 215–225. https://doi.org/10.1016/j.neurobiolaging.2010.03.011

Gallucci, M., Di Battista, M. E., Battistella, G., Falcone, C., Bisiacchi, P. S., & Di Giorgi, E. (2018). Neuropsychological tools to predict conversion from amnestic mild cognitive impairment to dementia. The TREDEM Registry. *Aging, Neuropsychology, and Cognition*, *25*(4), 550–560. https://doi.org/10.1080/13825585.2017.1349869

Gibbons, L. E., Carle, A. C., Mackin, R. S., Harvey, D., Mukherjee, S., Insel, P., … Alzheimer's Disease Neuroimaging Initiative. (2012). A composite score for executive functioning, validated in Alzheimer's Disease Neuroimaging Initiative (ADNI) participants with baseline mild cognitive impairment. *Brain Imaging and Behavior*, *6*(4), 517–527. https://doi.org/10.1007/s11682-012-9176-1

Giorgio, J., Karlaftis, V. M., Wang, R., Shen, Y., Tino, P., Welchman, A., & Kourtzi, Z. (2018). Functional brain networks for learning predictive statistics. *Cortex*, *107*. https://doi.org/10.1016/j.cortex.2017.08.014

Giorgio, Joseph, Jagust, W. J., Baker, S., Landau, S. M., Tino, P., & Kourtzi, Z. (2020). Title: Predicting future regional tau accumulation in asymptomatic and early Alzheimer's disease. *BioRxiv*, 2020.08.15.252601. https://doi.org/10.1101/2020.08.15.252601

Giorgio, Joseph, Landau, S., Jagust, W., Tino, P., & Kourtzi, Z. (2020). Modelling prognostic trajectories of cognitive decline due to Alzheimer's disease. *NeuroImage: Clinical*, 102199. https://doi.org/10.1016/j.nicl.2020.102199

Glenner, G. G., & Wong, C. W. (1984). Alzheimer's disease: Initial report of the purification and characterization of a novel cerebrovascular amyloid protein. *Biochemical and Biophysical Research Communications*, *120*(3), 885–890. https://doi.org/10.1016/S0006-291X(84)80190-4

Goate, A., Chartier-Harlin, M. C., Mullan, M., Brown, J., Crawford, F., Fidani, L., … Hardy, J. (1991). Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature*, *349*(6311), 704–706. https://doi.org/10.1038/349704a0

Goldgaber, D., Lerman, M. I., McBride, O. W., Saffiotti, U., & Gajdusek, D. C. (1987). Characterization and chromosomal localization of a cDNA encoding brain amyloid of Alzheimer's disease. *Science*, *235*(4791), 877–880. https://doi.org/10.1126/science.3810169

Gordon, B. A., Blazey, T. M., Su, Y., Hari-Raj, A., Dincer, A., Flores, S., … Benzinger, T. L. S. (2018). Spatial patterns of neuroimaging biomarker change in individuals from families with autosomal dominant Alzheimer's disease: a longitudinal study. *The Lancet Neurology*, *17*(3), 241–250. https://doi.org/10.1016/S1474-4422(18)30028-0

Greenfield, E. A., & Moorman, S. M. (2019). Childhood Socioeconomic Status and Later Life Cognition: Evidence From the Wisconsin Longitudinal Study. *Journal of Aging and Health*, *31*(9), 1589–1615. https://doi.org/10.1177/0898264318783489

Grober, E., Hall, C. B., Lipton, R. B., Zonderman, A. B., Resnick, S. M., & Kawas, C. (2008). Memory impairment, executive dysfunction, and intellectual decline in preclinical Alzheimer's disease. *Journal of the International Neuropsychological Society*, *14*(2), 266–278.

Grundke-Iqbal, I., Iqbal, K., Tung, Y. C., Quinlan, M., Wisniewski, H. M., & Binder, L. I. (1986). Abnormal phosphorylation of the microtubule-associated protein tau (tau) in Alzheimer cytoskeletal pathology. *Proceedings of the National Academy of Sciences of the United States of America*, *83*(13), 4913–4917. https://doi.org/10.1073/pnas.83.13.4913

Hall, B., Mak, E., Cervenka, S., Aigbirhio, F. I., Rowe, J. B., & O'Brien, J. T. (2017, July 1). In vivo tau PET imaging in dementia: Pathophysiology, radiotracer quantification, and a systematic review of clinical findings. *Ageing Research Reviews*. Elsevier Ireland Ltd. https://doi.org/10.1016/j.arr.2017.03.002

Hall, S., Öhrfelt, A., Constantinescu, R., Andreasson, U., Surova, Y., Bostrom, F., … Hansson, O. (2012). Accuracy of a panel of 5 cerebrospinal fluid biomarkers in the differential diagnosis of patients with dementia and/or Parkinsonian disorders. *Archives of Neurology*, *69*(11), 1445–1452. https://doi.org/10.1001/archneurol.2012.1654

Hanseeuw, B. J., Betensky, R. A., Jacobs, H. I. L., Schultz, A. P., Sepulcre, J., Becker, J. A., … Johnson, K. (2019). Association of Amyloid and Tau with Cognition in Preclinical

Alzheimer Disease: A Longitudinal Study. *JAMA Neurology*, *76*(8), 915–924. https://doi.org/10.1001/jamaneurol.2019.1424

Hardy, J. A., & Higgins, G. A. (1992). Alzheimer's disease: The amyloid cascade hypothesis. *Science*. https://doi.org/10.1126/science.1566067

Harrison, T. M., La Joie, R., Maass, A., Baker, S. L., Swinnerton, K., Fenton, L., … Jagust, W. J. (2019). Longitudinal tau accumulation and atrophy in aging and alzheimer disease. *Annals of Neurology*, *85*(2), 229–240. https://doi.org/10.1002/ana.25406

Hedden, T., & Gabrieli, J. D. E. (2004). Insights into the ageing mind: a view from cognitive neuroscience. *Nature Reviews Neuroscience*, *5*(2), 87–96. https://doi.org/10.1038/nrn1323

Hinrichs, C., Singh, V., Xu, G., & Johnson, S. C. (2011). Predictive markers for AD in a multi-modality framework: An analysis of MCI progression in the ADNI population. *NeuroImage*, *55*(2), 574–589. https://doi.org/10.1016/J.NEUROIMAGE.2010.10.081

Hochstetler, H., Trzepacz, P. T., Wang, S., Yu, P., Case, M., Henley, D. B., … Lyketsos, C. G. (2015). Empirically Defining Trajectories of Late-Life Cognitive and Functional Decline. *Journal of Alzheimer's Disease*, *50*(1), 271–282. https://doi.org/10.3233/JAD-150563

Holmberg, B., Johnels, B., Blennow, K., & Rosengren, L. (2003, February 1). Cerebrospinal fluid Aβ42 is reduced in multiple system atrophy but normal in Parkinson's disease and progressive supranuclear palsy. *Movement Disorders*. https://doi.org/10.1002/mds.10321

Hong, Y. T., Veenith, T., Dewar, D., Outtrim, J. G., Mani, V., Williams, C., … Menon, D. K. (2014). Amyloid imaging with carbon 11 - Labeled pittsburgh compound B for traumatic brain injury. *JAMA Neurology*, *71*(1), 23–31. https://doi.org/10.1001/jamaneurol.2013.4847

Hutton, M., Lendon, C. L., Rizzu, P., Baker, M., Froelich, S., Houlden, H. H., … Heutink, P. (1998). Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17. *Nature*, *393*(6686), 702–704. https://doi.org/10.1038/31508

Ikonomovic, M. D., Klunk, W. E., Abrahamson, E. E., Mathis, C. A., Price, J. C., Tsopelas, N. D., … DeKosky, S. T. (2008). Post-mortem correlates of in vivo PiB-PET amyloid imaging in a typical case of Alzheimer's disease. *Brain : A Journal of Neurology*, *131*(Pt 6), 1630–1645. https://doi.org/10.1093/brain/awn016

Insel, P. S., Mattsson, N., Mackin, R. S., Kornak, J., Nosheny, R., Tosun-Turgut, D., … Alzheimer's Disease Neuroimaging Initiative, A. D. N. (2015). Biomarkers and

cognitive endpoints to optimize trials in Alzheimer's disease. *Annals of Clinical and Translational Neurology*, *2*(5), 534–547. https://doi.org/10.1002/acn3.192

Jack, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., … Contributors, R. (2018). NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia : The Journal of the Alzheimer's Association*, *14*(4), 535–562. https://doi.org/10.1016/j.jalz.2018.02.018

Jack, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Feldman, H. H., Frisoni, G. B., … Dubois, B. (2016, August 2). A/T/N: An unbiased descriptive classification scheme for Alzheimer disease biomarkers. *Neurology*. Lippincott Williams and Wilkins. https://doi.org/10.1212/WNL.0000000000002923

Jack, C. R., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., … Weiner, M. W. (2008, April). The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging*. NIH Public Access. https://doi.org/10.1002/jmri.21049

Jack, C. R., & Holtzman, D. M. (2013, December 18). Biomarker modeling of alzheimer's disease. *Neuron*. Cell Press. https://doi.org/10.1016/j.neuron.2013.12.003

Jack, C. R., Jr, Wiste, H. J., Vemuri, P., Weigand, S. D., Senjem, M. L., … Initiative, the A. D. N. (2010). Brain beta-amyloid measures and magnetic resonance imaging atrophy both predict time-to-progression from mild cognitive impairment to Alzheimer's disease. *Brain*, *133*(11), 3336. https://doi.org/10.1093/BRAIN/AWQ277

Jack, C. R., Knopman, D. S., Chételat, G., Dickson, D., Fagan, A. M., Frisoni, G. B., … Vos, S. J. B. (2016, February 1). Suspected non-Alzheimer disease pathophysiology-concept and controversy. *Nature Reviews Neurology*. Nature Publishing Group. https://doi.org/10.1038/nrneurol.2015.251

Jack, C. R., Knopman, D. S., Jagust, W. J., Petersen, R. C., Weiner, M. W., Aisen, P. S., … Trojanowski, J. Q. (2013). Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *The Lancet. Neurology*, *12*(2), 207–216. https://doi.org/10.1016/S1474-4422(12)70291-0

Jack, C. R., Knopman, D. S., Jagust, W. J., Shaw, L. M., Aisen, P. S., Weiner, M. W., … Trojanowski, J. Q. (2010). Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *The Lancet. Neurology*, *9*(1), 119–128. https://doi.org/10.1016/S1474-4422(09)70299-6

Jack, C. R., Knopman, D. S., Weigand, S. D., Wiste, H. J., Vemuri, P., Lowe, V., … Petersen, R. C. (2012). An operational approach to National Institute on Aging-

Alzheimer's Association criteria for preclinical Alzheimer disease. *Annals of Neurology*, *71*(6), 765–775. https://doi.org/10.1002/ana.22628

Jack, C. R., Wiste, H. J., Schwarz, C. G., Lowe, V. J., Senjem, M. L., Vemuri, P., … Petersen, R. C. (2018). Longitudinal tau PET in ageing and Alzheimer's disease. *Brain*, *141*(5), 1517–1528. https://doi.org/10.1093/brain/awy059

Jack, C. R., Wiste, H. J., Weigand, S. D., Therneau, T. M., Lowe, V. J., Knopman, D. S., … Petersen, R. C. (2017). Defining imaging biomarker cut points for brain aging and Alzheimer's disease. *Alzheimer's & Dementia : The Journal of the Alzheimer's Association*, *13*(3), 205–216. https://doi.org/10.1016/j.jalz.2016.08.005

Jagust, W. (2018). Imaging the evolution and pathophysiology of Alzheimer disease. *Nature Reviews Neuroscience*, 1. https://doi.org/10.1038/s41583-018-0067-3

Jagust, W., Gitcho, A., Sun, F., Kuczynski, B., Mungas, D., & Haan, M. (2006). Brain imaging evidence of preclinical Alzheimer's disease in normal aging. *Annals of Neurology*, *59*(4), 673–681. https://doi.org/10.1002/ana.20799

Jagust, W. J., Landau, S. M., Shaw, L. M., Trojanowski, J. Q., Koeppe, R. A., Reiman, E. M., … Mathis, C. A. (2009). Relationships between biomarkers in aging and dementia. *Neurology*, *73*(15), 1193–1199. https://doi.org/10.1212/WNL.0b013e3181bc010c

Jansen, W. J., Ossenkoppele, R., Knol, D. L., Tijms, B. M., Scheltens, P., Verhey, F. R. J., … Zetterberg, H. (2015). Prevalence of cerebral amyloid pathology in persons without dementia: A meta-analysis. *JAMA - Journal of the American Medical Association*, *313*(19), 1924–1938. https://doi.org/10.1001/jama.2015.4668

Jansen, W. J., Ossenkoppele, R., Tijms, B. M., Fagan, A. M., Hansson, O., Klunk, W. E., … Zetterberg, H. (2018). Association of Cerebral Amyloid-β Aggregation With Cognitive Functioning in Persons Without Dementia. *JAMA Psychiatry*, *75*(1), 84. https://doi.org/10.1001/jamapsychiatry.2017.3391

Janssen, R. J., Mourão-Miranda, J., & Schnack, H. G. (2018). Making Individual Prognoses in Psychiatry Using Neuroimaging and Machine Learning. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*(9), 798–808. https://doi.org/10.1016/J.BPSC.2018.04.004

Jo, T., Nho, K., & Saykin, A. J. (2019). Deep Learning in Alzheimer's Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data. *Frontiers in Aging Neuroscience*, *11*, 220. https://doi.org/10.3389/fnagi.2019.00220

Johnson, K. A., Schultz, A., Betensky, R. A., Becker, J. A., Sepulcre, J., Rentz, D., … Sperling, R. (2016). Tau positron emission tomographic imaging in aging and early

Alzheimer disease. *Annals of Neurology*, *79*(1), 110–119. https://doi.org/10.1002/ana.24546

Jollans, L., Boyle, R., Artiges, E., Banaschewski, T., Desrivières, S., Grigis, A., … Whelan, R. (2019). Quantifying performance of machine learning methods for neuroimaging data. *NeuroImage*, *199*, 351–365. https://doi.org/10.1016/j.neuroimage.2019.05.082

Jonsson, T., Atwal, J. K., Steinberg, S., Snaedal, J., Jonsson, P. V., Bjornsson, S., … Stefansson, K. (2012). A mutation in APP protects against Alzheimer‘s disease and age-related cognitive decline. *Nature*, *488*(7409), 96. https://doi.org/10.1038/nature11283

Joshi, S., Shenoy, D., Vibhudendra Simha, G. G., Rrashmi, P. L., Venugopal, K. R., & Patnaik, L. M. (2010). Classification of Alzheimer's disease and Parkinson's disease by using machine learning and neural network methods. In *ICMLC 2010 - The 2nd International Conference on Machine Learning and Computing* (pp. 218–222). https://doi.org/10.1109/ICMLC.2010.45

Jutten, R. J., Harrison, J., de Jong, F. J., Aleman, A., Ritchie, C. W., Scheltens, P., & Sikkes, S. A. M. (2017). A composite measure of cognitive and functional progression in Alzheimer's disease: Design of the Capturing Changes in Cognition study. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, *3*(1), 130–138. https://doi.org/10.1016/J.TRCI.2017.01.004

Jutten, R. J., Harrison, J., Lee Meeuw Kjoe, P. R., Opmeer, E. M., Schoonenboom, N. S. M., de Jong, F. J., … Sikkes, S. A. M. (2018). A novel cognitive-functional composite measure to detect changes in early Alzheimer's disease: Test–retest reliability and feasibility. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *10*, 153–160. https://doi.org/10.1016/J.DADM.2017.12.002

Kang, J., Lemaire, H. G., Unterbeck, A., Salbaum, J. M., Masters, C. L., Grzeschik, K. H., … Müller-Hill, B. (1987). The precursor of Alzheimer's disease amyloid A4 protein resembles a cell-surface receptor. *Nature*, *325*(6106), 733–736. https://doi.org/10.1038/325733a0

Karlaftis, V. M., Giorgio, J., Vértes, P. E., Wang, R., Shen, Y., Tino, P., … Kourtzi, Z. (2019). Multimodal imaging of brain connectivity reveals predictors of individual decision strategy in statistical learning. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-018-0503-4

Kero, M., Paetau, A., Polvikoski, T., Tanskanen, M., Sulkava, R., Jansson, L., … Tienari, P. J. (2013). Amyloid precursor protein (APP) A673T mutation in the elderly Finnish population. *Neurobiology of Aging*, *34*(5), 1518.e1-1518.e3.

https://doi.org/10.1016/j.neurobiolaging.2012.09.017

Kim, W. S., Li, H., Ruberu, K., Chan, S., Elliott, D. A., Low, J. K., … Garner, B. (2013). Deletion of Abca7 increases cerebral amyloid-β accumulation in the J20 mouse model of Alzheimer's disease. *Journal of Neuroscience*, *33*(10), 4387–4394. https://doi.org/10.1523/JNEUROSCI.4165-12.2013

Klunk, W. E., Engler, H., Nordberg, A., Wang, Y., Blomqvist, G., Holt, D. P., … Långström, B. (2004). Imaging Brain Amyloid in Alzheimer's Disease with Pittsburgh Compound-B. *Annals of Neurology*, *55*(3), 306–319. https://doi.org/10.1002/ana.20009

Knopman, D. S., Jack, C. R., Wiste, H. J., Weigand, S. D., Vemuri, P., Lowe, V. J., … Petersen, R. C. (2013). Brain injury biomarkers are not dependent on β-amyloid in normal elderly. *Annals of Neurology*, *73*(4), 472–480. https://doi.org/10.1002/ana.23816

Knopman, D. S., Lundt, E. S., Therneau, T. M., Vemuri, P., Lowe, V. J., Kantarci, K., … Jack, C. R. (2019). Entorhinal cortex tau, amyloid-β, cortical thickness and memory performance in non-demented subjects. *Brain*, *142*(4), 1148–1160. https://doi.org/10.1093/brain/awz025

Krishnan, A., Williams, L. J., McIntosh, A. R., & Abdi, H. (2011). Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review. *NeuroImage*, *56*(2), 455–475. https://doi.org/10.1016/j.neuroimage.2010.07.034

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). *ImageNet Classification with Deep Convolutional Neural Networks*. Retrieved from http://code.google.com/p/cuda-convnet/

Krut, J. J., Zetterberg, H., Blennow, K., Cinque, P., Hagberg, L., Price, R. W., … Gisslén, M. (2013). Cerebrospinal fluid Alzheimer's biomarker profiles in CNS infections. *Journal of Neurology*, *260*(2), 620–626. https://doi.org/10.1007/s00415-012-6688-y

Landau, S. M., Harvey, D., Madison, C. M., Reiman, E. M., Foster, N. L., Aisen, P. S., … Jagust, W. J. (2010). Comparing predictors of conversion and decline in mild cognitive impairment. *Neurology*, *75*(3), 230–238. https://doi.org/10.1212/WNL.0b013e3181e8e8b8

Landau, Susan M, Fero, A., Baker, S. L., Koeppe, R., Mintun, M., Chen, K., … Jagust, W. J. (2015). Measurement of longitudinal β-amyloid change with 18F-florbetapir PET and standardized uptake value ratios. *Journal of Nuclear Medicine : Official Publication, Society of Nuclear Medicine*, *56*(4), 567–574. https://doi.org/10.2967/jnumed.114.148981

Langbaum, J. B., Hendrix, S. B., Ayutyanont, N., Chen, K., Fleisher, A. S., Shah, R. C., … Reiman, E. M. (2014). An empirically derived composite cognitive test score with

improved power to track and evaluate treatments for preclinical Alzheimer's disease. *Alzheimer's & Dementia*, *10*(6), 666–674. https://doi.org/10.1016/J.JALZ.2014.02.002

LaPoint, M. R., Chhatwal, J. P., Sepulcre, J., Johnson, K. A., Sperling, R. A., & Schultz, A. P. (2017). The association between tau PET and retrospective cortical thinning in clinically normal elderly. *NeuroImage*, *157*, 612–622. https://doi.org/10.1016/j.neuroimage.2017.05.049

Leal, S. L., Lockhart, S. N., Maass, A., Bell, R. K., & Jagust, W. J. (2018). Subthreshold amyloid predicts tau deposition in aging. *Journal of Neuroscience*, *38*(19), 4482–4489. https://doi.org/10.1523/JNEUROSCI.0485-18.2018

Leandrou, S., Petroudi, S., Kyriacou, P. A., Reyes-Aldasoro, C. C., & Pattichis, C. S. (2018). Quantitative MRI Brain Studies in Mild Cognitive Impairment and Alzheimer's Disease: A Methodological Review. *IEEE Reviews in Biomedical Engineering*, *11*, 97–111. https://doi.org/10.1109/RBME.2018.2796598

Lecun, Y., Bengio, Y., & Hinton, G. (2015, May 27). Deep learning. *Nature*. Nature Publishing Group. https://doi.org/10.1038/nature14539

Lemere, C. A., Blusztajn, J. K., Yamaguchi, H., Wisniewski, T., Saido, T. C., & Selkoe, D. J. (1996). Sequence of deposition of heterogeneous amyloid β-peptides and APO E in down syndrome: Implications for initial events in amyloid plaque formation. *Neurobiology of Disease*, *3*(1), 16–32. https://doi.org/10.1006/nbdi.1996.0003

Lemere, Cynthia A., Lopera, F., Kosik, K. S., Lendon, C. L., Ossa, J., Saido, T. C., … Arango V., J. C. (1996). The E280A presenilin 1 Alzheimer mutation produces increased Aβ42 deposition and severe cerebellar pathology. *Nature Medicine*, *2*(10), 1146–1150. https://doi.org/10.1038/nm1096-1146

Li, D., Iddi, S., Aisen, P. S., Thompson, W. K., & Donohue, M. C. (2019). The relative efficiency of time-to-progression and continuous measures of cognition in presymptomatic Alzheimer's disease. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, *5*, 308–318. https://doi.org/10.1016/J.TRCI.2019.04.004

Lindsey, R., Daluiski, A., Chopra, S., Lachapelle, A., Mozer, M., Sicular, S., … Potter, H. (2018). Deep neural network improves fracture detection by clinicians. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(45), 11591–11596. https://doi.org/10.1073/pnas.1806905115

Liu, K., Chen, K., Yao, L., & Guo, X. (2017). Prediction of Mild Cognitive Impairment Conversion Using a Combination of Independent Component Analysis and the Cox Model. *Frontiers in Human Neuroscience*, *11*, 33.

https://doi.org/10.3389/fnhum.2017.00033

Liu, X., Faes, L., Kale, A. U., Wagner, S. K., Fu, D. J., Bruynseels, A., … Denniston, A. K. (2019). A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *The Lancet Digital Health*, *1*(6), e271–e297. https://doi.org/10.1016/S2589-7500(19)30123-2

Lowe, V. J., Curran, G., Fang, P., Liesinger, A. M., Josephs, K. A., Parisi, J. E., … Murray, M. E. (2016). An autoradiographic evaluation of AV-1451 Tau PET in dementia. *Acta Neuropathologica Communications*, *4*(1). https://doi.org/10.1186/s40478-016-0315-6

Maass, A., Landau, S., Baker, S. L., Horng, A., Lockhart, S. N., La Joie, R., … Jagust, W. J. (2017). Comparison of multiple tau-PET measures as biomarkers in aging and Alzheimer's disease. *NeuroImage*, *157*, 448–463. https://doi.org/10.1016/J.NEUROIMAGE.2017.05.058

Maass, A., Lockhart, S. N., Harrison, T. M., Bell, R. K., Mellinger, T., Swinnerton, K., … Jagust, W. J. (2018). Entorhinal tau pathology, episodic memory decline, and neurodegeneration in aging. *Journal of Neuroscience*, *38*(3), 530–543. https://doi.org/10.1523/JNEUROSCI.2028-17.2017

Mak, E., Gabel, S., Mirette, H., Su, L., Williams, G. B., Waldman, A., … O'Brien, J. (2017). Structural neuroimaging in preclinical dementia: From microstructural deficits and grey matter atrophy to macroscale connectomic changes. *Ageing Research Reviews*, *35*, 250–264. https://doi.org/10.1016/j.arr.2016.10.001

Marinescu, R. V., Oxtoby, N. P., Young, A. L., Bron, E. E., Toga, A. W., Weiner, M. W., … Consortium, the E. (2018). TADPOLE Challenge: Prediction of Longitudinal Evolution in Alzheimer's Disease. Retrieved from http://arxiv.org/abs/1805.03909

Maroco, J., Silva, D., Rodrigues, A., Guerreiro, M., Santana, I., & de Mendonça, A. (2011). Data mining methods in the prediction of Dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests. *BMC Research Notes*, *4*(1), 299. https://doi.org/10.1186/1756-0500-4-299

Marquié, M., Normandin, M. D., Vanderburg, C. R., Costantino, I. M., Bien, E. A., Rycyna, L. G., … Gómez-Isla, T. (2015). Validating novel tau positron emission tomography tracer [F-18]-AV-1451 (T807) on postmortem brain tissue. *Annals of Neurology*, *78*(5), 787–800. https://doi.org/10.1002/ana.24517

Masters, C. L., Simms, G., Weinman, N. A., Multhaup, G., McDonald, B. L., & Beyreuther,

K. (1985). Amyloid plaque core protein in Alzheimer disease and Down syndrome. *Proceedings of the National Academy of Sciences of the United States of America*, *82*(12), 4245–4249. https://doi.org/10.1073/pnas.82.12.4245

Mateos-Pérez, J. M., Dadar, M., Lacalle-Aurioles, M., Iturria-Medina, Y., Zeighami, Y., & Evans, A. C. (2018). Structural neuroimaging as clinical predictor: A review of machine learning applications. *NeuroImage: Clinical*, *20*, 506–522. https://doi.org/10.1016/j.nicl.2018.08.019

Matsuda, H. (2016). MRI morphometry in Alzheimer's disease. *Ageing Research Reviews*, *30*, 17–24. https://doi.org/10.1016/j.arr.2016.01.003

Mazzeo, S., Santangelo, R., Bernasconi, M. P., Cecchetti, G., Fiorino, A., Pinto, P., … Magnani, G. (2016). Combining Cerebrospinal Fluid Biomarkers and Neuropsychological Assessment: A Simple and Cost-Effective Algorithm to Predict the Progression from Mild Cognitive Impairment to Alzheimer's Disease Dementia. *Journal of Alzheimer's Disease*, *54*(4), 1495–1508. https://doi.org/10.3233/JAD-160360

McIntosh, A. R., & Lobaugh, N. J. (2004). Partial least squares analysis of neuroimaging data: Applications and advances. *NeuroImage*, *23*(SUPPL. 1), 250–263. https://doi.org/10.1016/j.neuroimage.2004.07.020

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Kawas, C. H., … Phelps, C. H. (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, *7*(3), 263–269. https://doi.org/10.1016/j.jalz.2011.03.005

Michaud, T. L., Su, D., Siahpush, M., & Murman, D. L. (2017). The risk of incident mild cognitive impairment and progression to dementia considering mild cognitive impairment subtypes. *Dementia and Geriatric Cognitive Disorders Extra*, *7*(1), 15–29. https://doi.org/10.1159/000452486

Milan, L., & Whittaker, J. (1995). Application of the Parametric Bootstrap to Models that Incorporate a Singular Value Decomposition. *Applied Statistics*, *44*(1), 31. https://doi.org/10.2307/2986193

Monteiro, J. M., Rao, A., Shawe-Taylor, J., & Mourão-Miranda, J. (2016). A multiple hold-out framework for Sparse Partial Least Squares. *Journal of Neuroscience Methods*, *271*, 182–194. https://doi.org/10.1016/j.jneumeth.2016.06.011

Mormino, E. C., Betensky, R. A., Hedden, T., Schultz, A. P., Amariglio, R. E., Rentz, D. M., … Sperling, R. A. (2014). Synergistic Effect of β-Amyloid and Neurodegeneration on

Cognitive Decline in Clinically Normal Individuals. *JAMA Neurology*, *71*(11), 1379. https://doi.org/10.1001/jamaneurol.2014.2031

Mullan, M., Crawford, F., Axelman, K., Houlden, H., Lilius, L., Winblad, B., & Lannfelt, L. (1992). A pathogenic mutation for probable Alzheimer's disease in the APP gene at the N–terminus of β–amyloid. *Nature Genetics*, *1*(5), 345–347. https://doi.org/10.1038/ng0892-345

Murray, M. E., Graff-Radford, N. R., Ross, O. A., Petersen, R. C., Duara, R., & Dickson, D. W. (2011). Neuropathologically defined subtypes of Alzheimer's disease with distinct clinical characteristics: A retrospective study. *The Lancet Neurology*, *10*(9), 785–796. https://doi.org/10.1016/S1474-4422(11)70156-9

Nelson, P. T., Head, E., Schmitt, F. A., Davis, P. R., Neltner, J. H., Jicha, G. A., … Scheff, S. W. (2011, May). Alzheimer's disease is not "brain aging": Neuropathological, genetic, and epidemiological human studies. *Acta Neuropathologica*. https://doi.org/10.1007/s00401-011-0826-y

Nho, K., Risacher, S. L., Crane, P. K., DeCarli, C., Glymour, M. M., Habeck, C., … Saykin, A. J. (2012). Voxel and surface-based topography of memory and executive deficits in mild cognitive impairment and Alzheimer's disease. *Brain Imaging and Behavior*, *6*(4), 551–567. https://doi.org/10.1007/s11682-012-9203-2

Ossenkoppele, R., Jansen, W. J., Rabinovici, G. D., Knol, D. L., van der Flier, W. M., van Berckel, B. N. M., … Brooks, D. J. (2015). Prevalence of amyloid PET positivity in dementia syndromes: a meta-analysis. *JAMA*, *313*(19), 1939–1949. https://doi.org/10.1001/jama.2015.4669

Öst, M., Nylén, K., Csajbok, L., Öhrfelt, A. O., Tullberg, M., Wikkelsö, C., … Nellgård, B. (2006). Initial CSF total tau correlates with 1-year outcome in patients with traumatic brain injury. *Neurology*, *67*(9), 1600–1604. https://doi.org/10.1212/01.wnl.0000242732.06714.0f

Oulhaj, A., Wilcock, G. K., Smith, A. D., & De Jager, C. A. (2009). Predicting the time of conversion to MCI in the elderly: Role of verbal expression and learning. *Neurology*, *73*(18), 1436–1442. https://doi.org/10.1212/WNL.0b013e3181c0665f

Palmqvist, S., Zetterberg, H., Blennow, K., Vestberg, S., Andreasson, U., Brooks, D. J., … Hansson, O. (2014). Accuracy of brain amyloid detection in clinical practice using cerebrospinal fluid β-Amyloid 42: A cross-validation study against amyloid positron emission tomography. *JAMA Neurology*, *71*(10), 1282–1289. https://doi.org/10.1001/jamaneurol.2014.1358

Palmqvist, S., Zetterberg, H., Mattsson, N., Johansson, P., Minthon, L., Blennow, K., … Hansson, O. (2015). Detailed comparison of amyloid PET and CSF biomarkers for identifying early Alzheimer disease. *Neurology*, *85*(14), 1240–1249. https://doi.org/10.1212/WNL.0000000000001991

Patterson, C. (2018). *World Alzheimer Report 2018 - The state of the art of dementia research: New frontiers*.

Pellegrini, E., Ballerini, L., Hernandez, M. del C. V., Chappell, F. M., González-Castro, V., Anblagan, D., … Wardlaw, J. M. (2018). Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: A systematic review. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *10*, 519–535. https://doi.org/10.1016/J.DADM.2018.07.004

Pereira, T., Ferreira, F. L., Cardoso, S., Silva, D., de Mendonça, A., Guerreiro, M., & Madeira, S. C. (2018). Neuropsychological predictors of conversion from mild cognitive impairment to Alzheimer's disease: a feature selection ensemble combining stability and predictability. *BMC Medical Informatics and Decision Making*, *18*(1), 137. https://doi.org/10.1186/s12911-018-0710-y

Pereira, T., Lemos, L., Cardoso, S., Silva, D., Rodrigues, A., Santana, I., … Madeira, S. C. (2017). Predicting progression of mild cognitive impairment to dementia using neuropsychological data: a supervised learning approach using time windows. *BMC Medical Informatics and Decision Making*, *17*(1), 110. https://doi.org/10.1186/s12911-017-0497-2

Pernet, C. R., Wilcox, R., & Rousselet, G. A. (2013). Robust Correlation Analyses: False Positive and Power Validation Using a New Open Source Matlab Toolbox. *Frontiers in Psychology*, *3*, 606. https://doi.org/10.3389/fpsyg.2012.00606

Petersen, R. C., Aisen, P. S., Beckett, L. A., Donohue, M. C., Gamst, A. C., Harvey, D. J., … Weiner, M. W. (2010). Alzheimer's Disease Neuroimaging Initiative (ADNI): Clinical characterization. *Neurology*, *74*(3), 201–209. https://doi.org/10.1212/WNL.0b013e3181cb3e25

Petersen, Ronald C., Doody, R., Kurz, A., Mohs, R. C., Morris, J. C., Rabins, P. V., … Winblad, B. (2001). Current Concepts in Mild Cognitive Impairment. *Archives of Neurology*, *58*(12), 1985. https://doi.org/10.1001/archneur.58.12.1985

Petersen, Ronald C. (2009). Early diagnosis of Alzheimer's disease: is MCI too late? *Current Alzheimer Research*, *6*(4), 324–330. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19689230

Pontecorvo, M. J., Devous, M. D., Kennedy, I., Navitsky, M., Lu, M., Galante, N., … Mintun, M. A. (2019). A multicentre longitudinal study of flortaucipir (18F) in normal ageing, mild cognitive impairment and Alzheimer's disease dementia. *Brain*, *142*(6), 1723–1735. https://doi.org/10.1093/brain/awz090

Poorkaj, P., Bird, T. D., Wijsman, E., Nemens, E., Garruto, R. M., Anderson, L., … Schellenberg, G. D. (1998). Tau is a candidate gene for chromosome 17 frontotemporal dementia. *Annals of Neurology*, *43*(6), 815–825. https://doi.org/10.1002/ana.410430617

Prasher, V. P., Farrer, M. J., Kessling, A. M., Fisher, E. M. C., West, R. J., Barber, P. C., & Butler, A. C. (1998). Molecular mapping of Alzheimer-type dementia in Down's syndrome. *Annals of Neurology*, *43*(3), 380–383. https://doi.org/10.1002/ana.410430316

Prince, M., Wimo, A., Guerchet, M., Gemma-Claire Ali, M., Wu, Y.-T., Prina, M., … Xia, Z. (2015). *World Alzheimer Report 2015 The Global Impact of Dementia: An analysIs of prevalence, Incidence, cost and trends*. Retrieved from www.alz.co.uk/worldreport2015corrections

Radua, J., Canales-Rodríguez, E. J., Pomarol-Clotet, E., & Salvador, R. (2014). Validity of modulation and optimal settings for advanced voxel-based morphometry. *NeuroImage*, *86*, 81–90. https://doi.org/10.1016/j.neuroimage.2013.07.084

Rafii, M. S., & Aisen, P. S. (2019). Alzheimer's Disease Clinical Trials: Moving Toward Successful Prevention. *CNS Drugs*, *33*(2), 99–106. https://doi.org/10.1007/s40263-018-0598-1

Rathore, S., Habes, M., Iftikhar, M. A., Shacklett, A., & Davatzikos, C. (2017, July 15). A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages. *NeuroImage*. Academic Press Inc. https://doi.org/10.1016/j.neuroimage.2017.03.057

Rebeck, W. G., Reiter, J. S., Strickland, D. K., & Hyman, B. T. (1993). Apolipoprotein E in sporadic Alzheimer's disease: Allelic variation and receptor interactions. *Neuron*, *11*(4), 575–580. https://doi.org/10.1016/0896-6273(93)90070-8

Resnick, S. M., & Sojkova, J. (2011). Amyloid imaging and memory change for prediction of cognitive impairment. *Alzheimer's Research & Therapy*, *3*(1), 3. https://doi.org/10.1186/alzrt62

Rovelet-Lecrux, A., Hannequin, D., Raux, G., Le Meur, N., Laquerrière, A., Vital, A., … Campion, D. (2006). APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nature Genetics*, *38*(1), 24–26. https://doi.org/10.1038/ng1718

Rowe, C. C., Ng, S., Ackermann, U., Gong, S. J., Pike, K., Savage, G., … Villemagne, V. L. (2007). Imaging β-amyloid burden in aging and dementia. *Neurology*, *68*(20), 1718–1725. https://doi.org/10.1212/01.wnl.0000261919.22630.ea

Samper-González, J., Burgos, N., Bottani, S., Fontanella, S., Lu, P., Marcoux, A., … Colliot, O. (2018). Reproducible evaluation of classification methods in Alzheimer's disease: Framework and application to MRI and PET data. *NeuroImage*, *183*, 504–521. https://doi.org/10.1016/j.neuroimage.2018.08.042

Sander, K., Lashley, T., Gami, P., Gendron, T., Lythgoe, M. F., Rohrer, J. D., … Årstad, E. (2016). Characterization of tau positron emission tomography tracer [18 F]AV-1451 binding to postmortem tissue in Alzheimer's disease, primary tauopathies, and other dementias. *Alzheimer's & Dementia*, *12*(11), 1116–1124. https://doi.org/10.1016/j.jalz.2016.01.003

Schneider, P., Biehl, M., & Hammer, B. (2009). Adaptive Relevance Matrices in Learning Vector Quantization. *Neural Computation*, *21*(12), 3532–3561. https://doi.org/10.1162/neco.2009.11-08-908

Schöll, M., Lockhart, S. N., Schonhaut, D. R., O'Neil, J. P., Janabi, M., Ossenkoppele, R., … Jagust, W. J. (2016). PET Imaging of Tau Deposition in the Aging Human Brain. *Neuron*, *89*(5), 971–982. https://doi.org/10.1016/j.neuron.2016.01.028

Schöll, M., Maass, A., Mattsson, N., Ashton, N. J., Blennow, K., Zetterberg, H., & Jagust, W. (2019, June 1). Biomarkers for tau pathology. *Molecular and Cellular Neuroscience*. Academic Press Inc. https://doi.org/10.1016/j.mcn.2018.12.001

Schultz, S. A., Gordon, B. A., Mishra, S., Su, Y., Perrin, R. J., Cairns, N. J., … Benzinger, T. L. S. (2018). Widespread distribution of tauopathy in preclinical Alzheimer's disease. *Neurobiology of Aging*, *72*, 177–185. https://doi.org/10.1016/j.neurobiolaging.2018.08.022

Schwarz, A. J., Yu, P., Miller, B. B., Shcherbinin, S., Dickson, J., Navitsky, M., … Mintun, M. S. (2016). Regional profiles of the candidate tau PET ligand 18F-AV-1451 recapitulate key features of Braak histopathological stages. *Brain : A Journal of Neurology*, *139*(Pt 5), 1539–1550. https://doi.org/10.1093/brain/aww023

Selkoe, D. J. (1991, April 1). The molecular pathology of Alzheimer's disease. *Neuron*. Cell Press. https://doi.org/10.1016/0896-6273(91)90052-2

Serrano-Pozo, A., Qian, J., Monsell, S. E., Blacker, D., Gómez-Isla, T., Betensky, R. A., … Hyman, B. T. (2014). Mild to moderate Alzheimer dementia with insufficient neuropathological changes. *Annals of Neurology*, *75*(4), 597–601.

https://doi.org/10.1002/ana.24125

Shi, Y., & Holtzman, D. M. (2018, December 1). Interplay between innate immunity and Alzheimer disease: APOE and TREM2 in the spotlight. *Nature Reviews Immunology*. Nature Publishing Group. https://doi.org/10.1038/s41577-018-0051-1

Shi, Y., Yamada, K., Liddelow, S. A., Smith, S. T., Zhao, L., Luo, W., … Holtzman, D. M. (2017). ApoE4 markedly exacerbates tau-mediated neurodegeneration in a mouse model of tauopathy. *Nature*, *549*(7673), 523–527. https://doi.org/10.1038/nature24016

Silva, D., Guerreiro, M., Santana, I., Rodrigues, A., Cardoso, S., Maroco, J., & De Mendonça, A. (2013). Prediction of long-term (5 years) conversion to dementia using neuropsychological tests in a memory clinic setting. *Journal of Alzheimer's Disease*, *34*(3), 681–689. https://doi.org/10.3233/JAD-122098

Singh, R., Kalra, M. K., Nitiwarangkul, C., Patti, J. A., Homayounieh, F., Padole, A., … Digumarthy, S. R. (2018). Deep learning in chest radiography: Detection of findings and presence of change. *PLOS ONE*, *13*(10), e0204155. https://doi.org/10.1371/journal.pone.0204155

Singh, V., Chertkow, H., Lerch, J. P., Evans, A. C., Dorr, A. E., & Kabani, N. J. (2006). Spatial patterns of cortical thinning in mild cognitive impairment and Alzheimer's disease. *Brain : A Journal of Neurology*, *129*(Pt 11), 2885–2893. https://doi.org/10.1093/brain/awl256

Sintini, I., Martin, P. R., Graff-Radford, J., Senjem, M. L., Schwarz, C. G., Machulda, M. M., … Whitwell, J. L. (2019). Longitudinal tau-PET uptake and atrophy in atypical Alzheimer's disease. *NeuroImage: Clinical*, *23*. https://doi.org/10.1016/j.nicl.2019.101823

Smith, C. D., Chebrolu, H., Wekstein, D. R., Schmitt, F. A., Jicha, G. A., Cooper, G., & Markesbery, W. R. (2007). Brain structural alterations before mild cognitive impairment. *Neurology*, *68*(16), 1268–1273. https://doi.org/10.1212/01.wnl.0000259542.54830.34

Spasov, S., Passamonti, L., Duggento, A., Liò, P., & Toschi, N. (2019). A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease. *NeuroImage*, *189*, 276–287. https://doi.org/10.1016/j.neuroimage.2019.01.031

Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., … Phelps, C. H. (2011). Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association

workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, *7*(3), 280–292. https://doi.org/10.1016/J.JALZ.2011.03.003

Spillantini, M. G., Van Swieten, J. C., & Goedert, M. (2000). Tau gene mutations in frontotemporal dementia anti parkinsonism linked to chromosome 17 (FTDP-17). *Neurogenetics*. https://doi.org/10.1007/s100489900084

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*(2), 245–251. https://doi.org/10.1037/0033-2909.87.2.245

Strozyk, D., Blennow, K., White, L. R., & Launer, L. J. (2003). CSF Aß 42 levels correlate with amyloid-neuropathology in a population-based autopsy study. *Neurology*, *60*(4), 652–656. https://doi.org/10.1212/01.WNL.0000046581.81650.D0

T., S., C., R., F., A., N., M., K., B., H., Z., … Zetterberg, H. (2014). Diagnostic performance of cerebrospinal fluid total tau and phosphorylated tau in Creutzfeldt-Jakob disease: Results from the Swedish Mortality Registry. *JAMA Neurology*, *71*(4), 476–483. https://doi.org/10.1001/jamaneurol.2013.6455 LK - http://ze6dt7rj9y.search.serialssolutions.com?sid=EMBASE&issn=21686149&id=doi:10.1001%2Fjamaneurol.2013.6455&atitle=Diagnostic+performance+of+cerebrospinal+fluid+total+tau+and+phosphorylated+tau+in+Creutzfeldt-Jakob+disease%3A+Results+from+the+Swedish+Mortality+Registry&stitle=JAMA+Neurol.&title=JAMA+Neurology&volume=71&issue=4&spage=476&epage=483&aulast=Skillb%C3%A4ck&aufirst=Tobias&auinit=T.&aufull=Skillb%C3%A4ck+T.&coden=&isbn=&pages=476-483&date=2014&auinit

Tabert, M. H., Manly, J. J., Liu, X., Pelton, G. H., Rosenblum, S., Jacobs, M., … Devanand, D. P. (2006). Neuropsychological prediction of conversion to alzheimer disease in patients with mild cognitive impairment. *Archives of General Psychiatry*, *63*(8), 916–924. https://doi.org/10.1001/archpsyc.63.8.916

Tang, E. Y. H., Harrison, S. L., Errington, L., Gordon, M. F., Visser, P. J., Novak, G., … Stephan, B. C. M. (2015). Current Developments in Dementia Risk Prediction Modelling: An Updated Systematic Review. *PLOS ONE*, *10*(9), e0136181. https://doi.org/10.1371/journal.pone.0136181

Tanveer, M., Richhariya, B., Khan, R. U., Rashid, A. H., Khanna, P., Prasad, M., & Lin, C. T. (2020). Machine Learning Techniques for the Diagnosis of Alzheimer's Disease. *ACM Transactions on Multimedia Computing, Communications, and Applications*, *16*(1s), 1–35. https://doi.org/10.1145/3344998

Tanzi, R. E., Gusella, J. F., Watkins, P. C., Bruns, G. A. P., St. George-Hyslop, P., Van

Keuren, M. L., … Neve, R. L. (1987). Amyloid β protein gene: CDNA, mRNA distribution, and genetic linkage near the Alzheimer locus. *Science*, *235*(4791), 880–884. https://doi.org/10.1126/science.2949367

Tapiola, T., Alafuzoff, I., Herukka, S. K., Parkkinen, L., Hartikainen, P., Soininen, H., & Pirttilä, T. (2009). Cerebrospinal fluid β-amyloid 42 and tau proteins as biomarkers of Alzheimer-type pathologic changes in the brain. *Archives of Neurology*, *66*(3), 382–389. https://doi.org/10.1001/archneurol.2008.596

Thung, K.-H., Yap, P.-T., Adeli, E., Lee, S.-W., & Shen, D. (2018). Conversion and time-to-conversion predictions of mild cognitive impairment using low-rank affinity pursuit denoising and matrix completion. *Medical Image Analysis*, *45*, 68–82. https://doi.org/10.1016/J.MEDIA.2018.01.002

Thung, K. H., Yap, P. T., Adeli, E., Lee, S. W., & Shen, D. (2018). Conversion and time-to-conversion predictions of mild cognitive impairment using low-rank affinity pursuit denoising and matrix completion. *Medical Image Analysis*, *45*, 68–82. https://doi.org/10.1016/j.media.2018.01.002

Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, *25*(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

Tosun, D., Landau, S., Aisen, P. S., Petersen, R. C., Mintun, M., Jagust, W., … Alzheimer's Disease Neuroimaging Initiative. (2017). Association between tau deposition and antecedent amyloid-β accumulation rates in normal and early symptomatic individuals. *Brain : A Journal of Neurology*, *140*(5), 1499–1512. https://doi.org/10.1093/brain/awx046

van der Kant, R., Goldstein, L. S. B., & Ossenkoppele, R. (2020, January 1). Amyloid-β-independent regulators of tau pathology in Alzheimer disease. *Nature Reviews Neuroscience*. Nature Research. https://doi.org/10.1038/s41583-019-0240-3

van Maurik, I. S., Vos, S. J., Bos, I., Bouwman, F. H., Teunissen, C. E., Scheltens, P., … van der Flier, W. M. (2019). Biomarker-based prognosis for people with mild cognitive impairment (ABIDE): a modelling study. *The Lancet Neurology*, *18*(11), 1034–1044. https://doi.org/10.1016/S1474-4422(19)30283-2

Vogel, J. W., Vachon-Presseau, E., Pichet Binette, A., Tam, A., Orban, P., La Joie, R., … Villeneuve, S. (2018). Brain properties predict proximity to symptom onset in sporadic Alzheimer's disease. *Brain*, *141*(6), 1871–1883. https://doi.org/10.1093/brain/awy093

Wang, L., Benzinger, T. L., Su, Y., Christensen, J., Friedrichsen, K., Aldea, P., … Ances, B. M. (2016). Evaluation of Tau imaging in staging Alzheimer disease and revealing

interactions between β-Amyloid and tauopathy. *JAMA Neurology*, *73*(9), 1070–1077. https://doi.org/10.1001/jamaneurol.2016.2078

Wang, Y., Haaksma, M. L., Ramakers, I. H. G. B., Verhey, F. R. J., Flier, W. M., Scheltens, P., … Melis, R. J. F. (2019). Cognitive and functional progression of dementia in two longitudinal studies. *International Journal of Geriatric Psychiatry*, gps.5175. https://doi.org/10.1002/gps.5175

Wei, W., Poirion, E., Bodini, B., Tonietto, M., Durrleman, S., Colliot, O., … Ayache, N. (2020). Predicting PET-derived Myelin Content from Multisequence MRI for Individual Longitudinal Analysis in Multiple Sclerosis. *NeuroImage*, 117308. https://doi.org/10.1016/j.neuroimage.2020.117308

Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Cedarbaum, J., … Trojanowski, J. Q. (2015). 2014 Update of the Alzheimer's Disease Neuroimaging Initiative: A review of papers published since its inception. *Alzheimer's and Dementia*, *11*(6), e1–e120. https://doi.org/10.1016/j.jalz.2014.11.001

Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., … Trojanowski, J. Q. (2017). Recent publications from the Alzheimer's Disease Neuroimaging Initiative: Reviewing progress toward improved AD clinical trials. *Alzheimer's and Dementia*, *13*(4), e1–e85. https://doi.org/10.1016/j.jalz.2016.11.007

Whitwell, J. L., Przybelski, S. A., Weigand, S. D., Knopman, D. S., Boeve, B. F., Petersen, R. C., & Jack, C. R. (2007). 3D maps from multiple MRI illustrate changing atrophy patterns as subjects progress from mild cognitive impairment to Alzheimer's disease. *Brain*, *130*(7), 1777–1786. https://doi.org/10.1093/brain/awm112

Wilkosz, P. A., Seltman, H. J., Devlin, B., Weamer, E. A., Lopez, O. L., DeKosky, S. T., & Sweet, R. A. (2010). Trajectories of cognitive decline in Alzheimer's disease. *International Psychogeriatrics*, *22*(2), 281–290. https://doi.org/10.1017/S1041610209991001

Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, *1*(1), 67–82. https://doi.org/10.1109/4235.585893

Woo, C.-W., Chang, L. J., Lindquist, M. A., & Wager, T. D. (2017). Building better biomarkers: brain models in translational neuroimaging. *Nature Neuroscience*, *20*(3), 365–377. https://doi.org/10.1038/nn.4478

Yesavage, J. A. (1988). Geriatric depression scale. *Psychopharmacol Bull*, *24*(4), 709–711.

Young, A. L., Marinescu, R. V., Oxtoby, N. P., Bocchetta, M., Yong, K., Firth, N. C., …

Furst, A. J. (2018). Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with Subtype and Stage Inference. *Nature Communications*, *9*(1), 4273. https://doi.org/10.1038/s41467-018-05892-0

Young, A. L., Oxtoby, N. P., Daga, P., Cash, D. M., Fox, N. C., Ourselin, S., … Alzheimer's Disease Neuroimaging Initiative. (2014). A data-driven model of biomarker changes in sporadic Alzheimer's disease. *Brain*, *137*(9), 2564. https://doi.org/10.1093/BRAIN/AWU176

Zhang, D., Shen, D., & Initiative, A. D. N. (2012). Predicting Future Clinical Changes of MCI Patients Using Longitudinal and Multimodal Biomarkers. *PLoS ONE*, *7*(3), e33182. https://doi.org/10.1371/journal.pone.0033182

Zhou, T., Thung, K.-H., Liu, M., Shi, F., Zhang, C., & Shen, D. (2018). Multi-modal Neuroimaging Data Fusion via Latent Space Learning for Alzheimer's Disease Diagnosis (pp. 76–84). Springer, Cham. https://doi.org/10.1007/978-3-030-00320-3_10

Zhu, X., Suk, H. Il, Lee, S. W., & Shen, D. (2016). Canonical feature selection for joint regression and multi-class identification in Alzheimer's disease diagnosis. *Brain Imaging and Behavior*, *10*(3), 818–828. https://doi.org/10.1007/s11682-015-9430-4