COMBINING NETWORK NEUROSCIENCE AND MACHINE LEARNING TO DISCOVER NEUROCOGNITIVE SUBGROUPS IN AGING INDIVIDUALS AT RISK OF OR DIAGNOSED WITH ALZHEIMER'S DISEASE

Marc Douglas Rudolph

A dissertation submitted to faculty at the University of North Carolina at Chapel Hill in fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology & Neuroscience in the College of Arts & Sciences.

Chapel Hill
2022

Approved by:

Jessica R. Cohen

Eran Dayan

Kelly S. Giovanello

Joseph B. Hopfinger

Keely A. Muscatell

ABSTRACT

Marc Douglas Rudolph: Combining network neuroscience and machine learning to discover neurocognitive subgroups in aging individuals at risk of or diagnosed with Alzheimer's disease (Under the direction of Jessica R. Cohen)

Dementia is a complicated medical condition that negatively impacts an individual's mental and physical well-being increasing the risk of early mortality. The risk of developing dementia increases with age, although rates vary according to several factors, such as sex, race, and genetics. While Alzheimer's disease (AD), the most common form of dementia, is often associated with episodic memory impairment, individuals with AD can present with diverse cognitive profiles. Similarly, heterogeneous subgroups of typically aging individuals may show impairments in both general and specific forms of attention, memory, and/or executive functioning. Individual variation in cognitive impairments experienced by aging individuals, and resistance or resilience to cognitive decline, can be linked to functional brain network organization and communication capacity. This study aimed to disentangle shared and unique aspects of cognitive impairment and functional network topology seen in healthy aging, early-stage or preclinical dementia, and AD. Specifically, Aim 1 sought to establish if weighted metrics that index redundancy in unthresholded functional brain networks, as a proxy of brain and cognitive reserve, support general and/or specific forms of cognition. Aim 2 sought to establish whether a combination of core demographic risk factors (age, sex, and education), cognitive measures, and weighted functional network metrics could accurately distinguish otherwise cognitively normal individuals (CN), from CN who will convert to AD (AD-C), and individuals diagnosed with AD, using supervised machine learning. Aim 3 specifically sought

quantify the presence of data-driven neurocognitive subgroups utilizing a combination of unsupervised and supervised machine learning. In the present study, redundancy-based metrics (global communicability and global clustering coefficient) were not predictive of overall cognitive functioning, nor were they the most informative predictors when attempting to distinguish between CN, AD-C, and AD participants using machine learning. When classifying older individuals with and without AD, neuropsychological measures were more informative than metrics assessing global network topology, including redundancy-based measures. Finally, while this study failed to identify cognitive subgroups previously reported in MCI and AD participants, AD converters correctly classified or misclassified as AD showed diverging neurocognitive profiles and may represent a subset of individuals with primarily executive, as opposed to memory-related impairments, respectively.

To Candice, this work is dedicated to you.

Thank you for your unwavering support and constant encouragement.

I could not have done this without you.

ACKNOWLEDGEMENTS

First and foremost, I want to thank my advisor and mentor, Dr. Jessica R. Cohen. Thank you for your support, guidance, and commitment over the last five years. To my committee members: Drs. Eran Dayan, Kelly Giovanello, Joseph Hopfinger, and Keely Muscatell – thank you for your support and feedback on critical aspects of this project.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AAL	Automated Anatomical Labeling
AD	Alzheimer's disease
AD-C	AD Converters
AD-C-AD	AD Converters Predicted as AD
AD-C-CN	AD Converters Predicted as Cognitively Normal
Aβ	Amyloid-Beta
ACh	Acetylcholine
aMCI	Amnestic MCI
ApoE	Apolipoprotein E
ANOVA	Analysis of Variance
ART	Artifact Detection Toolbox
AUROC	Area Under the Receiver Operating Curve
bvAD	Behavioral Variant of AD
BOLD	Blood Oxygen Level-Dependent
BOSTON	Boston Naming Test
CBS	Corticobasal Syndrome
CC	Clustering Coefficient
CDR	Clinical Dementias Rating
CDR-SOB	CDR Sum of Box Scores
CI	Confidence Interval
CN	Cognitively Normal
CNS	Central Nervous System

COG-GAP	Number Of Days in Between Cognitive Assessment and First fMRI Scan
COMM	Communicability
CSF	Cerebrospinal Fluid
DKEFS	Delis-Kaplan Executive Function System
DAN	Dorsal Attention Network
DAYS-CONV	Number Of Days in Between First fMRI Scan and Final AD Diagnosis
DIGIB	Digit Span Backward
DIGIF	Digit Span Forward
dMCI	Dysexecutive MCI
DMN	Default Mode Network
EC	Eigenvector Centrality
ED	Education
EF	Executive Function
exMCI	Executive dysfunction-MCI
FD	Framewise Displacement
FDR	False-Discovery Rate
FPN	Frontoparietal Cognitive Control Network
fMRI	Functional Magnetic Resonance Imaging
GE	Global Efficiency
L temporal	Lateral Temporal
LOGIMEM	Logical Memory
lvPPA	Logopenic Variant Primary Progressive Aphasia
MCI	Mild Cognitive Impairment

MDS	Multidimensional Scaling
MMSE	Mini-Mental State Exam
MOD	Modularity
MRI	Magnetic Resonance Imaging
MTL	Medial Temporal Lobe
mxMCI	Mixed or Multi-Domain MCI
naMCI	Non-amnestic MCI
NIR	No Information Rate
non-exMCI	No Executive Dysfunction MCI
OASIS	Open Access Of Imaging Series
PC	Participation Coefficient
PCA	Posterior Cortical Atrophy
PET	Positron Emission Tomography
PIB	Pittsburgh Compound B
p-NIR	No Information Rate P-Value
PT	Path Transitivity
ROC	Receiver Operating Curve
RMSE	Root Mean Square Error
SD	Standard Deviation
SENS	Sensitivity
SPEC	Specificity
SMOTE	Synthetic Minority Oversampling Technique
STAT	Test Statistic

STD	Standard
SUVR	Standard Uptake Ratios
SWP	Small-World Propensity
TRAILA	Trail Making Test Part A
TRAILB	Trail Making Test Part B
VEG	Vegetable Naming (Category Fluency)
WAIS	Wechsler Memory Scale
WAIS-R	Wechsler Memory Scale-Revised
WMD	Within-Module Degree Z-Score
XNAT	Extensible Neuroimaging Archive Toolkit

CHAPTER 1: INTRODUCTION

Dementia is a complicated medical condition that negatively impacts an individual's mental and physical well-being and increases the risk of early mortality. The risk of developing dementia increases with age (Jack et al., 2019; Jaul & Barron, 2017; Knopman et al., 2019) and is most common in individuals 65 years of age and older, although rates can vary according to several factors, such as sex, race, and genetic profile. The number of individuals living over the age of 65 is expected to double by 2050, reaching almost 1.5 billion people. The 2-fold increase in individuals living past the age of 65 poses substantial challenges for healthcare systems and social and economic infrastructures that provide sustainable living solutions to aging individuals, healthy or otherwise. Within the United States alone, nearly 6 million individuals are currently diagnosed with incurable dementia (https://population.un.org/wpp/Publications). Globally, approximately 50 million individuals live with some form of dementia and this number is expected to increase 3-fold by 2050, reaching an estimated 152 million affected persons. Alzheimer's disease (AD) is the most common cause of dementia, accounting for approximately 60-80% of all dementia cases (Bronzuoli et al., 2016). According to recent estimates, two-thirds of individuals living with AD in the United States are female ("2021 Alzheimer's Disease Facts and Figures," 2021). The higher proportion of females living with AD is most often attributed longer average lifespans. However, the extent to which females are at a greater risk for *developing* AD compared to males remains uncertain (Beam et al., 2018). Some studies suggest genetics (as discussed further below) may play an integral role (Altmann et al., 2014; Ungar et al., 2014). Two pathological hallmarks of AD, amyloid-beta (Aβ) and tau, are present in the

central nervous system (CNS) of otherwise healthy aging individuals, and in individuals with Mild Cognitive Impairment (MCI) (Okello et al., 2009), and found to correlate with both the likelihood and rate of conversion to AD (LaMontagne et al., 2019; Marcus et al., 2010). Further, the level of Aβ detected in non-demented aging individuals has been associated with cognitive impairment and decline in both general (e.g. fluid intelligence), and more specific aspects of cognitive functioning including measures of processing speed, episodic memory, and both executive and visuospatial functioning (Baker et al., 2017).

As the definition of MCI has evolved, several subtypes have emerged with respect to specific components of cognition (del Carmen Díaz-Mardomingo et al., 2017), specifically aMCI (amnestic MCI), naMCI (non-amnestic MCI), dMCI (dysexecutive MCI). mxMCI (mixed or multi-domain MCI), exMCI (executive dysfunction-MCI), and non-exMCI (no executive dysfunction-MCI). Critically, these same cognitive variants have been extended to individuals with AD (Jellinger, 2021; Mez et al., 2016; Stuss & Alexander, 2007; Vogel & Hansson, 2022), with the dysexecutive subtype gaining the most recent attention (Corriveau-Lecavalier et al., 2022; Mez et al., 2016; Townley et al., 2020). In a study by Edmonds and colleagues (2021), a combination of hierarchical clustering and latent-discriminant analyses were employed to identify and assess the validity of data-driven subgroups compared to groups defined according to clinical criteria (Edmonds et al., 2021). In this study, a sample of 738 individuals (CN: N = 334; MCI: N = 404) completed a battery of neuropsychological assessments. Genetic testing confirmed if participants were carriers of the Apolipoprotein E (ApoE) protein e4 allele, which conveys an added risk of developing AD (Altmann et al., 2014; Ungar et al., 2014). Five primary subgroups emerged. Otherwise cognitively normal individuals (N = 334), were split into two groups based on average (17.6%; N = 204, Mean age = 71.77 (7.30); % e4 positive = 31%) or

above average (27.6% N = 130, Mean age = 71.53 (6.95); % e4 positive = 38%) cognitive performance. Individuals with MCI (n = 404) were clustered into amnestic (14.1%; N = 216, Mean age = 71.53 (7.97); % e4 positive = 41%), non-amnestic (29.3%; N = 104, Mean age = 71.25 (7.54); % e4 positive = 33%), and mixed (global impairment; 11.4%; N = 84, Mean age = 74.08 (8.64); % e4 positive = 43%) types. Using the data-driven groups, binary classification accuracy was 81.4%. Post-classification, the presence of AD-related pathology was summarized and compared between data-driven subgroups. AD-related pathology was more abundant in the mixed and amnestic MCI subgroups as compared to the non-amnestic MCI and CN groups. Likewise, education was lowest in the mixed MCI group, however, age was not significantly different between the detected clusters.

As in MCI, several recent studies have assessed the presence and composition of data-driven cognitive subtypes in AD. In a study by Scheltens and colleagues (2017), using nonnegative matrix factorization, a non-parametric clustering approach often used in image segmentation problems, two distinct clusters of "memory-impaired" and "memory-spared" individuals with probable AD were detected across four large independent cohorts (Scheltens et al., 2017). Here, the terms memory-impaired and memory-spared were used to distinguish data-driven subgroups of individuals from those classified as amnestic or non-amnestic based on diagnostic classification. Compared to the memory-impaired groups (N = 1195, Mean age = 71.5 (8.7); % e4 positive = 68%), the memory-spared groups (N = 787, Mean age = 70.4 (8.4); % e4 positive = 58%) were on average comprised of more younger individuals with lower scores of the Mini-Mental State Exam (MMSE) (Folstein et al., 1975). Of note, across cohorts, the number of individuals in either group with a positive AD-biomarker status (92% positive-rate in both groups) did not differ. Building on these findings, Scheltens and colleagues (2018) further

reported that individuals who were clustered into the memory-spared group had a *faster* progression of disease and higher rates of mortality irrespective of age, sex, ApoE genotype, and baseline MMSE scores (Scheltens et al., 2018). The memory-spared groups from these studies are phenotypically similar to the established behavioral/dysexecutive subtype of AD (Mez et al., 2016; Stuss & Alexander, 2007). Thus, multiple cognitive subtypes have been detected in both non-demented aging individuals and those with AD. Though speculative, memory-impaired subtypes of AD may have neuropathology confined to, or more prominent in the hippocampus, whereas memory-spared subtypes may exhibit widespread changes in large-scale networks involving multiple regions (Ferreira et al., 2020; Vogel & Hansson, 2022).

Critically, however, individuals with Alzheimer's pathology can be asymptomatic, showing no clinical signs of severe cognitive dysfunction, and subsequently never developing dementia (Driscoll & Troncoso, 2011; Gomez-Isla & Frosch, 2019; Jack et al., 2019). Further, not all aging individuals experience cognitive impairment (e.g. superagers) (Burke et al., 2019).

As noted above, In addition to age, Aβ, and tau, genes that encode the ApoE protein are associated with different levels of risk for developing AD (Safieh et al., 2019). ApoE is a lipid transporter naturally occurring within astrocytes involved in clearing toxic levels of Aβ within the CNS. The presence of specific forms of ApoE (ε1, ε2, ε3, ε4) can lead to divergent trajectories of neuropathological spread in AD contributing to both white and gray matter pathology and cognitive dysfunction (Porrata-Doria et al., 2010; Ricciarelli & Fedele, 2017; Scheller et al., 2018). Existing research suggests that carriers of the ApoE-ε4 allele are at a greater risk of experiencing age-related cognitive dysfunction and developing dementia. Likewise, neuroimaging studies consistently document structural and functional differences in the brains of ε4 carriers versus non-carriers. Compared to non-carriers, ε4 carriers may exhibit

accelerated rates of cognitive decline and develop dementia at an earlier age (Reas et al., 2019; Scheller et al., 2018), in part, because ApoE begins to bind with Aβ, failing to clear it from post-synaptic terminals and thus facilitating aggregation (Zilberter & Zilberter, 2017). This risk is two-fold for those who are homozygous for the ε4 allele (e.g. having two copies; 44). While the neuroprotective role of the ε3 allele remains uncertain, the ε2 allele has been considered protective given its demonstrated role in clearing Aβ in both human and animal models (Bu, 2009; Conejero-Goldberg et al., 2014; Elias-Sonnenschein et al., 2018). Moreover, some studies suggest that older female ε4 carriers may be at a greater risk of developing AD (Altmann et al., 2014; Ungar et al., 2014). Mechanistically, women ε4 carriers may exhibit reduced levels of estrogen that contribute to more significant memory-related cognitive impairments. Importantly, however, the degree to which age and ApoE status may confer a greater risk for developing AD in females can depend on additional factors, including levels of education attainment (Hasselgren et al., 2020) and exposure to psychosocial stress and circulating levels of pro-inflammatory cytokines (Au et al., 2016).

Subtyping & Heterogeneity

As discussed, several atypical variants of AD have been identified in both living and recently-deceased individuals (Ferreira et al., 2019, 2020; Geifman et al., 2018; Murray et al., 2011; Vogel & Hansson, 2022). In a recent meta-analysis, Ferreira and colleagues (2020) provide a synopsis of both hypothesis- and data-driven approaches aimed at identifying robust biologically-based subtypes reported by clinicians (Ferreira et al., 2020). Across the studies reviewed, four primary subtypes of AD were commonly observed: (1) typical AD (55%), (2) limbic- or medial temporal lobe (MTL) predominant AD, (3) hippocampal- or MTL sparing AD (17%), and (4) minimal atrophy AD (15%). Whereas prototypical AD

neuropathology focuses on early MTL atrophy, each of these subtypes has patterns and trajectories of neuropathology that are distinct. Specifically, observed pathology may fall on a four-dimensional continuum of typicality and disease severity (see Figure 1) (Ferreira et al., 2020).

Vogel and colleagues (2022) expanded on the review by (Ferreira et al., 2020) to further categorize typical and atypical variants of neurobiologically-based AD subtypes based on patterns of tau pathology (Vogel & Hansson, 2022). First, even in typical late-onset AD, there is a good deal of heterogeneity such that not all individuals present with predominant memory impairments or exhibit tau deposition that is confined to the MTL. As noted by Ferreira and colleagues (2020), individuals with non-amnestic AD are on average younger, and more likely to be non-carriers (or heterozygotes) for the ApoE e4 allele (Ferreira et al., 2020). Concerning tau pathology, although there is some degree of overlap in cortical patterns of tau deposition observed across subtypes, the spatial distribution of tau appears to map onto regions that are associated with specific cognitive impairments (Vogel & Hansson, 2022).

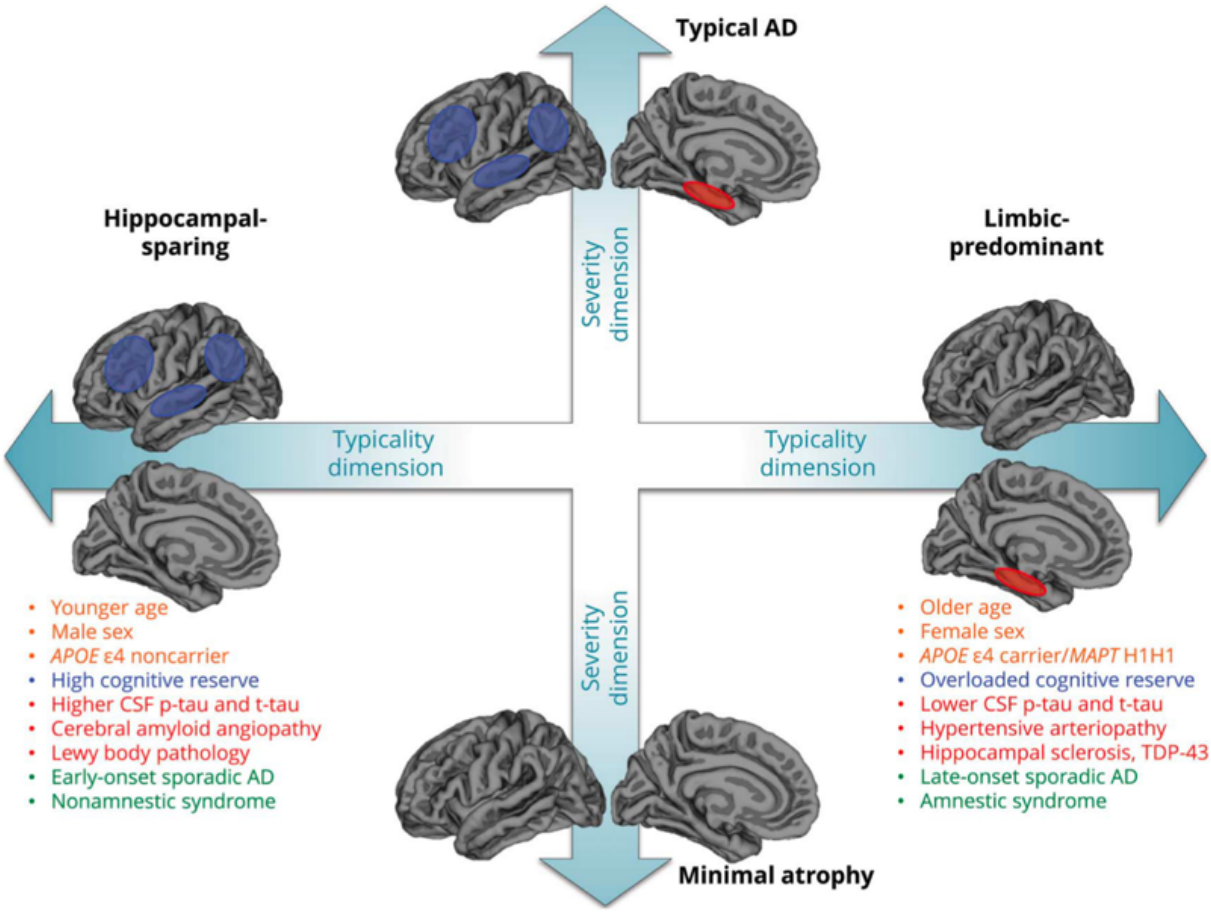


Figure 1. Variants of AD. A four-dimensional graph depicts variants of AD on a proposed scale of typicality versus severity. Patterns of brain atrophy, risk factors (age, sex, education, ApoE status, etc.), and cognition are associated with four distinct subtypes including typical AD, limbic-predominant AD, hippocampal-sparing AD, and minimal atrophy AD. Figure from Ferreira et al., 2020.

Specifically, consistent with prior work (Bondi et al., 2014; Edmonds et al., 2014, 2021; Jacobson et al., 2009; Jak et al., 2009; Scheltens et al., 2016, 2017, 2018), Vogel and colleagues (2022) highlight that in atypical or non-amnestic forms of AD, predominant cognitive impairment has been observed in the domains of visuospatial, language, and executive functioning, as well as a more rare behavioral version of AD (Vogel & Hansson, 2022). Importantly, consistent with work concerning pathological spread discussed earlier in the introduction (see the section *Communication-Based Models of Network Spread*), altered connectivity patterns between diverse sets of brain regions may contribute to and/or reflect

variation observed within and across subtypes (Figure 2). Consistent with this recent work, in the present study, the terms amnestic and non-amnestic are used to delineate subsets of individuals with and without primary memory-related cognitive impairments who may comprise one of several atypical variants of AD.

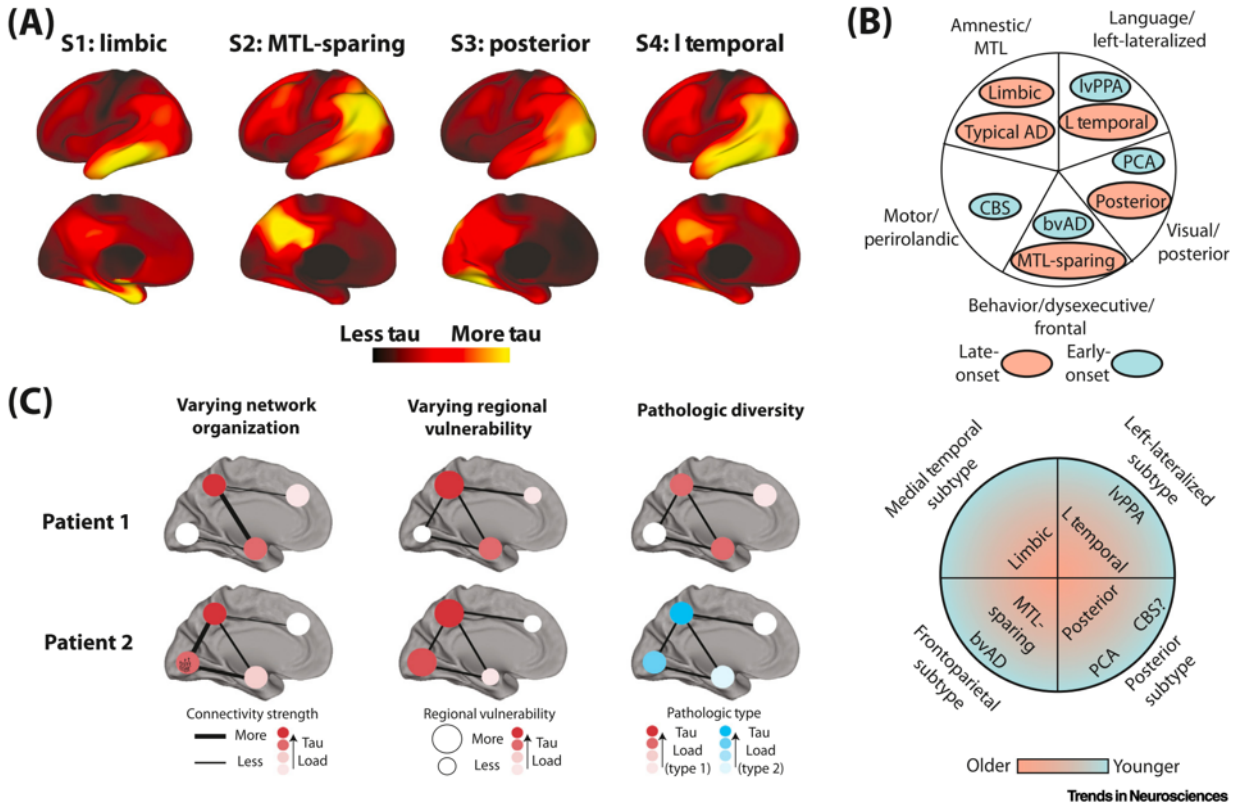


Figure 2. AD Variants and Tau Pathology. Commonly identified subtypes of AD based on neurobiology and cognitive functioning are illustrated. Panels a-c show (A) regional and global tau deposition, (B) cognitive phenotypes, and (C) hypothesized patterns of regional and network vulnerability. In addition to the prototypical amnestic subtype associated with late-onset AD, language, visuospatial, and dysexecutive subtypes are highlighted (panel B). Abbreviations: AD, Alzheimer's disease; bvAD, behavioral variant of AD; CBS, corticobasal syndrome; L temporal, lateral temporal; lvPPA, logopenic variant primary progressive aphasia; MTL, medial temporal lobe; PCA, posterior cortical atrophy. Figure taken from Vogel et al., 2022.

Reserve, Resistance, & Resilience in Cognitive Aging

Aging individuals experience normative cognitive decline in various cognitive domains including attention, memory, and executive function. Normative cognitive impairment and decline experienced by healthy aging individuals can negatively impact everyday functioning,

lowering the quality of living, and increasing the risk of developing additional mental and physical health ailments, including dementia. Theories and hypotheses of cognitive aging imply that, with age, there is (1) a general slowing of cognitive and motor processing speed, (2) degraded perceptual acuity and perceptual planning, (3) a decline in fluid, as opposed to crystallized, intelligence, and (4) impaired decision making and memory (Anderson & Craik, 2017; Bisiacchi et al., 2008; Cabeza et al., 2009; Campbell et al., 2010; Craik & Bialystok, 2006; Lustig et al., 2007; Monge & Madden, 2016). Additional deficits in inhibitory control and resistance to distraction have been observed and may further contribute to declines in cognitive performance in attention, memory, and executive functioning (Campbell et al., 2010; Ziegler et al., 2018). Individual variation in cognitive impairments experienced by aging individuals, and resistance or resilience to cognitive decline, can broadly be summarized by two interrelated theories of cognitive aging: brain reserve and cognitive reserve. Brain reserve is conceptualized in terms of underlying structural neuroanatomy ("hardware"), while cognitive reserve is conceptualized in terms of (1) cognitive processing and (2) physiological responses encoded at the functional level ("software") (Medaglia et al., 2017a; Stern, 2017; Stern et al., 2019). More specifically, brain reserve is the degree to which brain structure, at a microscopic level, is innately resilient to insult that would otherwise cause cognitive impairment. Examples of brain reserve include increased synaptic and neuronal density. Whereas brain reserve is a passive process, cognitive reserve is an active process (Medaglia et al., 2017b). Cognitive reserve theorizes that over time, with experience and education, individuals develop both generalized and specific strategies to solve complex problems (Stern, 2002). These strategies are supported by flexible and adaptive functional networks of interconnected brain regions (Stern, 2002; Stern et al., 2019). The capability of brain functional networks to adapt and reconfigure in response to

changing environmental and contextual demands may confer the degree to which an individual is more or less resilient to experiencing age- and disease-related cognitive deficits (Montine et al., 2019).

Educational and professional attainment are often used as indirect indicators of cognitive reserve. The level of education an individual receives in their lifetime is considered an indirect indicator of resistance to normative age-related cognitive impairment occurring later in life. A higher education level may also serve as a protective factor (Lövdén et al., 2020) that decreases the probability of developing dementia (Stern, 2017) by increasing resilience to AD pathology (Arenaza-Urquijo & Vemuri, 2018; Medaglia et al., 2017a; Montine et al., 2019; Stern et al., 2019). However, this is not true for all individuals. Specifically, while higher levels of education may be a protective factor, research has shown that for some individuals educational attainment can be associated with the rate at which an individual develops more severe forms of cognitive impairment and dementia (Contador et al., 2017). This and other work suggest that there is a *tipping point* at which time protective factors can become disadvantageous. Thus, it is unclear what factors interact with education to promote resistance or resilience to AD. The development of more direct measures that quantify brain and/or cognitive reserve may help parse some of this individual heterogeneity, and provide a means to identify individuals at greater risk for developing AD.

Network Neuroscience

Cognitive dysfunction is increasingly being thought of in terms of disrupted patterns in both structural and functional brain connectivity (Baronchelli et al., 2013; Bressler & Tognoli, 2006; Mill et al., 2017; Reuter-Lorenz & Park, 2014; Sporns, 2014). As a complex system, the brain is organized into discrete networks that communicate with each other to support flexible

behavior and cognition (Power et al., 2011; Singer, 2009; Sporns, 2013a). Changes in brain structure and function due to aging and dementia are widely distributed. With age, pathological markers of AD become toxic, contributing to a substantial loss of gray and white matter brain volume that can result in both mild and severe forms of cognitive and emotional dysfunction (Hardy & Higgins, 1992; Hedden et al., 2016; Ricciarelli & Fedele, 2017). Thus, mathematical tools based on graph theory allow researchers to move beyond examining changes in individual brain regions and to instead investigate patterns of change across whole-brain networks (Bullmore & Sporns, 2009; Hallquist & Hillary, 2018; Muldoon & Bassett, 2014; Sporns, 2013b; Wig, 2017) and are the optimal set of methodological tools to assess these distributed changes. Graph theoretical metrics can be applied to quantify the degree of network reconfiguration due to aging or the presence of pathology.

Extant research shows that functional brain networks supporting higher-order cognition (e.g., dorsal attention network (DAN), default mode network (DMN), and frontoparietal cognitive control network (FPN) change with age (Cabeza et al., 2009; Chan et al., 2017; Murray et al., 2010; Park & Reuter-Lorenz, 2009; Sala-Llonch, BartrÃs-Faz, et al., 2015), and are altered in dementia (Fornito & Bullmore, 2015; Franzmeier et al., 2020; Jones et al., 2016). With age, there are reported decreases in functional brain signal variability, increases in task-specific compensatory functional activation (Betzel et al., 2014; Cabeza et al., 2009; Lee et al., 2015, 2016; Monge et al., 2017; Petrican et al., 2017; Sala-Llonch, Bartrés-Faz, et al., 2015), and loss of specialization at both the regional and network-level (Chan et al., 2017; Koen & Rugg, 2019). This age-related dedifferentiation, or loss of specialized processing, is linked with two key properties of brain networks that are disrupted in aging and dementia, namely integration and segregation (Cao et al., 2014; Fornito et al., 2015; Koen & Rugg, 2019; Park et al., 2004; Zuo et

al., 2017). In general, networks reconfigure by becoming more integrated or segregated to ensure efficient use of neural resources supporting complex cognitive functions. Network integration relies on distributed neural processing between distant brain regions. Network segregation reflects the degree to which regions tend to cluster together to form distinct modules comprised of highly interconnected brain regions. To maintain an optimal balance of network integration and segregation, and thus strike a balance of cost and efficiency, structural and functional brain networks possess small-world properties (Achard & Bullmore, 2007). Small-world networks are comprised of distinct modules where there is both a high degree of clustering within modules and a greater propensity for short paths connecting regions between modules (Muldoon et al., 2016a). Further, the ability of functional networks to reconfigure declines with age (Avelar-Pereira et al., 2017; Chan et al., 2017; Grady, 2017) and is altered in dementia (Ferreira et al., 2019).

From a mechanistic perspective, the disruption of neuromodulatory systems may be a common thread linking cognitive impairment and decline in aging and dementia. The cholinergic hypothesis of aging and AD (Bartus et al., 1982) holds that a loss of cholinergic neurons in the basal forebrain and reduced synthesis of acetylcholine (ACh) alters ACh signaling leading to cognitive dysfunction. Specifically, loss of cholinergic functioning contributes to specific deficits in perceptual processing, selective attention, and memory encoding and recognition (Bentley et al., 2011). From a network neuroscience perspective, noradrenergic and cholinergic systems play a crucial role in maintaining a balance of network integration and segregation in the presence of neural noise through the coordination of excitatory and inhibitory neurons that modulate resting membrane potentials (Shine et al., 2016, 2018; Shine & Poldrack, 2018). Specifically, Shine (2018) elegantly describes how cholinergic signals emanating from the basal forebrain facilitate

neural gain (e.g., increasing signal-to-noise) to promote overall network segregation by selectively activating regions critical for a specific cognitive process (e.g., attentional selection, cue/target selection, visuospatial perception, etc.) (Shine et al., 2018). In contrast, the noradrenergic system tips the balance of functional network organization toward a more integrated state by coordinating activity in otherwise segregated regions (Shine et al., 2018; Shine & Poldrack, 2018; H. M. Snyder et al., 2017).

In AD, and perhaps as a part of normative aging, the degree of functional system segregation or integration observed across studies may also depend on the complex interplay of Aβ and tau pathology. The amyloid cascade hypothesis posits that AD-related neuropathology emerges primarily due to an initial accumulation of Aβ. To date, however, there is conflicting evidence of whether or not the spread of tau pathology is dependent on the presence of toxic Aβ (Aβ-40 & Aβ-42) (Weigand et al., 2020). Post-mortem histological studies and murine models suggest pathological tau is initially confined to specific subcortical brain structures (entorhinal cortex, locus coeruleus) and only later becomes distributed across heteromodal cortices in individuals with Aβ pathology (Aβ+) (Braak & Braak, 1991; Franzmeier et al., 2020). With age-related tauopathy, the aggregation of tau in healthy older individuals has been found to occur at non-pathological levels of Aβ (Aβ-) (Johnson et al., 2016; Sepulcre et al., 2016), challenging the amyloid cascade hypothesis. The network degeneration hypothesis of aging and AD posits that the spread of tau occurs through large-scale brain networks (Drzezga, 2018; Palop et al., 2006) and is supported by recent work by Sepulcre and colleagues (2016) suggesting tau deposition is positively correlated with the global deposition of Aβ (Sepulcre et al., 2016). Franzmeier and colleagues (2020) further suggest the spread of pathological tau occurs via network propagation (e.g., *transneuronal or transsynaptic)* and is thus an active, rather than a passive process of

diffusion (Franzmeier et al., 2020). Active network propagation implies complex spreading dynamics, such that pathological spreading occurs along both pathways in the brain that are subject to metabolic constraints or neural compensation in response to insult. That is, the presence of AD pathology may systematically alter the topology of brain networks to facilitate, rather than prevent or protect against the spread of disease (Drzezga, 2018; Lella et al., 2019; Lella & Estrada, 2020).

Communication-Based Models of Network Spread

As discussed above, network-based models can quantify shifts in structural and functional network topology observed over time with age and/or due to the presence of pathology, and elucidate how the change in topology facilitates complex patterns of network communication. Additional graph-theoretical metrics, based on information theory, can be used to describe how information "flows," or is transmitted, amongst elements embedded within a complex system via direct and indirect paths. These metrics may be critically important to pathological aging as they can potentially be used to indicate, formally model, or predict the likelihood of disease spread involving multiple redundant paths, and open up new avenues for quantifying reserve and resilience on a neurobiological level. The concepts of redundancy and reserve are thought to be critical to understanding how structural and functional brain networks maintain a balance of cost and efficiency in health and disease. In general, greater redundancy is hypothesized to quantify the degree to which individuals may be more or less resistant to injury or insult in normative and pathological aging (Di Lanzo et al., 2012; Langella, Sadiq, et al., 2021; Leistritz et al., 2013; Sadiq et al., 2021; Tononi et al., 1999). Thus, measures quantifying redundancy may be useful as markers of brain reserve.

Two commonly used communication-based metrics are redundancy (Di Lanzo et al., 2012) and communicability (Crofts & Higham, 2009). In this proposal, the redundancy metric proposed by Di Lanzo and colleagues (2012) is here forth referred to as network redundancy, to distinguish between the general concept of redundancy and specific graph-theoretical metrics used to quantify redundant elements in a system. Network redundancy quantifies the total number of possible direct and indirect paths connecting a pair of regions of a given length. With network redundancy, self-connections are excluded such that a particular connection or node cannot be traversed or revisited more than once. Di Lanzo and colleagues suggest that "self-loops" may not be biologically meaningful (Di Lanzo et al., 2012). Network redundancy is designed to work on sparse, binary graphs (e.g. unweighted connections) (Di Lanzo et al., 2012). Communicability, like network redundancy, also assesses the total number of paths connecting two regions. However, unlike network redundancy, communicability can be used on either binary or weighted graphs. In addition, nodes can be revisited more than once, and longer paths are down-weighted to control for redundant connections involving potentially spurious paths of infinite length (Crofts & Higham, 2009).

Several recent studies have sought to formally quantify brain network redundancy in aging individuals with and without dementia using functional magnetic resonance imaging (MRI) (Langella, Mucha, et al., 2021; Langella, Sadiq, et al., 2021; Sadiq et al., 2021). Langella et. al. (2021) restricted analyses to the hippocampus, a region impacted early in the progression of AD, and reported that posterior, but not anterior, hippocampal network redundancy was reduced in individuals in both the early and late stages of MCI) as compared to healthy controls (Langella, Sadiq, et al., 2021). In a follow-study, Langella and colleagues (2021) found that hippocampal network redundancy mediated the association between hippocampal volume and a

composite memory score comprising both encoding and recall stages of memory (Langella, Mucha, et al., 2021). Sadiq and colleagues (2021) found that global network redundancy tended to increase with age up until approximately 60 years and mediated age-related declines in executive function measured using the color-word inhibition task as part of the Delis-Kaplan Executive Functioning System (Baron, 2004; Sadiq et al., 2021).

Although longitudinal analysis of data that includes individuals younger than 60 years of age is warranted to provide further support for these findings, from both a theoretical and applied perspective, it is unclear what neurobiological mechanisms might both promote and support the emergence and maintenance of multiple redundant pathways associated with increased metabolic cost. The routing of information via a particular pathway, or set of pathways, depends on the relative balance of several fundamental network properties. In structural networks, where regions are connected by physical white matter pathways, previous work has found that greater communicability is associated with emerging brain pathology (Lella & Estrada, 2020; Mišić et al., 2015; Stam et al., 2016). In functional networks, the presence of additional pathways may permit or even facilitate the spread of diffuse neural activity that is not restricted to direct physical connections, nor bound by geodesic distance, however, this remains to be tested.

Further, it is unclear if various metrics that capture redundancy are associated with one another and/or provide unique information for distinguishing typically aging individuals, who have varying degrees of cognitive difficulty and AD-related pathology, from individuals who are at risk of converting to AD or who are clinically diagnosed with AD.

One promising method to distinguish different groups of individuals (e.g., typically aging, at risk of conversion to AD, AD) based on network topology, and cognition, is to use machine learning. Machine learning methods are designed to uncover meaningful patterns of

information to discover both linear and non-linear high-dimensional relationships between biological and non-biological factors at the individual level.

Machine Learning in AD

Several recent reviews highlight the utility and efficacy of machine learning algorithms to detect features that best distinguish cognitively unimpaired individuals (non-demented) from individuals with MCI or AD (Dwyer et al., 2018; Grueso & Viejo-Sobera, 2021). To accurately predict group status (e.g. CN, MCI, or AD), or conversion to AD from MCI, machine learning approaches have incorporated a range of biological and non-biological factors (Fisher et al., 2019). Specifically, in addition to core demographic factors such as age, sex, and education, other risk factors of AD assessed include markers of genetic expression (Grueso & Viejo-Sobera, 2021) and patterns of gray and/or white matter brain atrophy (Costafreda et al., 2011; Guan et al., 2017; Kwak et al., 2021). Other studies have assessed the utility of neuropsychiatric symptoms (Mallo et al., 2020), socio-demographic factors (Grassi et al., 2019), and clinical health records to assess diagnostic subtypes of individuals with AD (Alexander et al., 2021). As with the selection of predictive features, the choice of machine learning algorithm(s) used for prediction is also variable. Machine learning algorithms extend from simple decision trees (Ang et al., 2019) to the use of deep learning applications based on multilayer neural networks (Gao & Lima, 2022; Kwak et al., 2021; Liu et al., 2020; Suk & Shen, 2015) and ensemble-based approaches that combine several high-performing algorithms (Grassi et al., 2019; Lebedev et al., 2014; Naimi & Balzer, 2018).

Several studies have used machine learning techniques to study brain morphology associated with the progression of AD-related pathology, such as atrophy confined to the hippocampus (Costafreda et al., 2011) or in specific hippocampal subfields (Kwak et al., 2022).

In the study by Costafreda and colleagues (2011), an accuracy of 80% was achieved when predicting conversion from MCI to AD (one-year from diagnosis) using patterns of hippocampal morphology (Costafreda et al., 2011). Using cortical surface area and subcortical volumes, Guan and colleagues (2017) reported accuracies of 77%, 81%, and 70% when classifying amnestic MCI (Mean age = 78.36) from non-amnestic MCI (Mean age = 77.76), amnestic MCI from healthy controls (Mean age = 77.12), and non-amnestic MCI from healthy controls respectively (Guan et al., 2017). Here, reductions in subcortical volumes of the hippocampus and amygdala, and cortical surface area of the frontal pole contributed most to classification of individuals with amnestic MCI. Kwak and colleagues (2021) used deep learning to classify individuals (N=489) with MCI (Mean age = 72.15) as either cognitively normal (Mean age = 70.98) or AD (Mean age = 72.91) based solely on brain atrophy patterns in gray matter (Kwak et al., 2021). Gray matter density in the temporal lobe (medial and lateral) most impacted model performance (Accuracy = 93.75%). Specifically, individuals with MCI who were predicted as AD (MCI-AD), exhibited patterns of gray matter atrophy similar to the AD group.

In their meta-analysis, Grueso and Viejo-Sobera (2021) found that studies that incorporate multimodal neuroimaging data, specifically MRI and positron emission topography (PET; Mean accuracy: MRI = 74.5%; PET = 76.9%; Combined = 77.5%), tended to outperform single modality studies on average when classifying MCI from AD (Grueso & Viejo-Sobera, 2021). For studies that reported specific brain regions, although variable, selected features tended to correspond with known atrophy patterns when classifying older samples of MCI participants versus those with AD. For example, medial temporal regions (hippocampus, amygdala, entorhinal cortex), posterior cingulate gyrus, and the precuneus. Moreover, multidimensional

studies that combined neuroimaging data with demographic, clinical, cognitive, and other measures contributed to higher overall prediction rates.

Thus, machine learning studies combining demographic, clinical, diagnostic, and/or neuroimaging data have been able to successfully distinguish healthy individuals from those diagnosed with either MCI or AD, with accuracies reported in the range of 70 – 99%. Importantly, as highlighted in the meta-analysis by (Grueso & Viejo-Sobera, 2021), a direct comparison of model accuracy is not always possible between studies. Sample size, the number and type of features, the age of individuals, and other factors influence classification performance. A model with 99.99% accuracy in one dataset may not generalize to a different sample of participants. Moreover, brain atrophy patterns, or structural changes, that distinguish non-demented individuals from those with MCI or AD may not correspond or align with observed cognitive impairment. For example, in the study by Kwak and colleagues (2021), heterogeneous patterns of cognitive impairment were observed in both MCI-CN and MCI-AD groups. Specifically, deficits in language and memory were most prominent overall (MCI-CN: 32.9% and 38.6%; MCI-AD: 46.4% and 32.8% respectively), however, the MCI-CN group was comprised of 24.9% of otherwise unimpaired individuals. Moreover, 13.6% of individuals in the MCI-AD group had more prominent executive deficits. When performing a post-hoc cluster analysis using the neuropsychological assessments alone, four primary subgroups emerged, including a dysnomic (language; 37.43%), amnestic (36.63%), dysexecutive (6.95%), and otherwise cognitively normal (18.98%) group. Thus, there was a lack of agreement between groups defined by brain atrophy versus cognitive performance. As noted by the authors, further work is needed to better understand the link between emerging cognitive deficits and AD-related pathology that accounts for heterogeneity in both brain and cognition.

Machine Learning & Network Neuroscience

The studies discussed above focused on core demographic risk factors, indices of brain atrophy (e.g., cortical thickness, whole-brain, or hippocampal volume), and biomarkers of AD pathology (e.g., cerebrospinal fluid (CSF) and PET-based markers of Aß load and tau deposition**)** to distinguish between individuals with MCI and AD, or between healthy controls and MCI and/or AD. Given the theoretical relevance of network-level dysfunction, studies have also utilized a combination of network metrics to distinguish between groups or predict conversion to AD from MCI. Specifically, differences in network topology between individuals with and without MCI and/or AD have been observed at global and nodal levels and used as predictive features in supervised classification models. These studies report classification accuracies of 37% - 100%. Higher accuracy rates were observed when classification models compared healthy controls to individuals with either MCI or AD (Dyrba et al., 2015; Hojjati et al., 2017, 2019; Jie et al., 2014; Jitsuishi & Yamaguchi, 2022; Khazaee et al., 2015, 2016, 2017; Sun et al., 2019; Wang et al., 2013; Wee et al., 2012; Wei et al., 2016; Xu et al., 2020; L. Zhang et al., 2020; T. Zhang et al., 2021), in contrast to models that compared individuals with MCI or AD (Hojjati et al., 2017; Jitsuishi & Yamaguchi, 2022; Khazaee et al., 2015, 2016, 2017; Sun et al., 2019; Wei et al., 2016). Lower classification performance was observed for studies attempting to distinguish early- from late-MCI (Accuracy: 43% - 70%) (Jitsuishi & Yamaguchi, 2022) or stable versus progressive MCI (e.g. converters; Accuracy: 24% - 62%) (Hojjati et al., 2019). Concerning predictive features, global and nodal estimates of efficiency, clustering coefficient, and multiple measures of network centrality (e.g., betweenness centrality, eigenvector centrality, and page-rank centrality), were most consistently found to distinguish between healthy controls and individuals with MCI or AD (Dyrba et al., 2015; Hojjati et al., 2017, 2019; Jie et al., 2014;

Jitsuishi & Yamaguchi, 2022; Khazaee et al., 2015, 2016, 2017; Sun et al., 2019; Wang et al., 2013; Wee et al., 2012; Wei et al., 2016; Xu et al., 2020; L. Zhang et al., 2020; T. Zhang et al., 2021).

Concerning methodology, additional work has assessed the predictive performance of various machine learning algorithms along with different cross-validation strategies, feature selection methods, and graph-theoretical metrics computed using either structural or functional imaging data (Jitsuishi & Yamaguchi, 2022; Wang et al., 2013; Xu et al., 2020; L. Zhang et al., 2020). For example, Jie and colleagues (2014) were able to accurately classify healthy controls from individuals with MCI (Accuracy = 91.9%; AUC = .94) using a single weighted graph metric (clustering coefficient) calculated at the nodal level using resting-state fMRI (Jie et al., 2014). Likewise, using diffusion MRI and resting-state fMRI, Wee and colleagues (2012) noted comparable performance using the nodal clustering coefficient (Wee et al., 2012). Wang and colleagues (2013) used graph metrics that were calculated on frequency-specific functional connectivity matrices to distinguish healthy controls (N = 47) from individuals with amnestic MCI (N = 37; Accuracy = 85.7%) (Wang et al., 2013).

These studies demonstrate the predictive utility of combining machine learning and network neuroscience to distinguish individuals with and without MCI and/or AD. However, several non-trivial decision points may limit the generalizability of the reported findings and should be considered when drawing inferences concerning classification performance and forming hypotheses regarding specific features. No studies combining machine learning and graph theory reviewed reported metrics that quantify in-scanner micromovements (e.g., framewise displacement) (Ciric et al., 2017; Power et al., 2012, 2015; Satterthwaite et al., 2013). Thus, it is unknown if the reviewed studies had enough data (i.e., scan length) to estimate

functional connectivity robustly to derive graph-theoretical metrics. Half of the studies reviewed reported using unweighted metrics calculated on thresholded graphs to define network topology. As discussed earlier in the introduction, both the method and degree of thresholding applied to structural and functional graphs can distort differences in connectivity patterns between patient and non-patient groups (Garrison et al., 2015; Langer et al., 2013; van den Heuvel et al., 2017; Váša et al., 2018). One study directly assessed how graph thresholding impacted classification performance when distinguishing MCI participants from healthy controls (Zhang et al., 2020). Here, features included diagnostic measures (Mini-Mental State Exam, Clinical Diagnostic Rating scale) and global and nodal graph-theoretical metrics. The study by (Zhang et al., 2020) reported that classification performance fluctuated across thresholds (Accuracy = 86.3%-92.4%). Moreover, considering baseline classification accuracy achieved using diagnostic measures (Accuracy = 86.3%), the percentage increase in model performance after adding graph-theoretical metrics fluctuated in the range of 1%-6.4%. Three studies chose to limit analyses and only reported model performance for a single threshold (Khazaee et al., 2015, 2017; Wei et al., 2016).

The number of connections retained via thresholding depends on the number of brain regions included in a particular brain atlas. Eight of the fifteen studies generated brain graphs using either 90 or 115 regions from the structurally-defined AAL atlas. Studies also show that structural parcellations do not adequately capture the complexity of functional connectivity, as observed by (Wee et al., 2012). Khazaee and colleagues (2016) directly compared atlas-based model performance (Khazaee et al., 2016). Model performance suffered when graph metrics were estimated using a structurally-defined AAL atlas (90 regions) instead of a functionally-defined atlas consisting of 264 cortical and subcortical regions. In the studies reviewed, the

number of features ranged anywhere from 90 to 3773 features depending on the number of ROIs and graph metrics assessed. Thirteen of fifteen studies used feature selection to reduce a large number of potential variables (approximately one to twenty percent retained) for classification.

Feature selection (or reduction) is standard practice in high-dimensional machine learning problems. It typically helps improve model performance, reduce overfitting, and improve model generalizability when combined with cross-validation (Hastie et al., 2013). However, many studies were small-sample studies having as few as 10-40 subjects per group, although a large range of sample sizes was observed (Average sample size across studies reviewed: Healthy Controls: = 40 (Min =12; Max 93); MCI (MCI-C, MCI-NC; aMCI) = 41 (Min = 12; Max = 89; AD). Further, 6 of 14 of the studies used leave-one-out cross-validation. It is well-established that leave-one-out cross-validation produces biased (overly optimistic) predictions and models that are more likely to overfit and poorly generalize to unseen data (Kohavi, 1995; Poldrack, 2012). When feature selection methods are paired with leave-one-out cross-validation in small samples, the set of features retained for prediction might be less reliable (Kuncheva et al., 2020; Way et al., 2010). Lastly, studies assessing more than two groups (e.g., CN, MCI, AD, and either amnestic or non-amnestic subgroups) (Hojjati et al., 2017; Khazaee et al., 2015, 2016; Sun et al., 2019; Wei et al., 2016) relied on multiple independent binary classification models. These models do not capture the overlap or similarity between groups.

In addition to the limitations described above, none of the studies combined network metrics *and* cognitive performance measures from neuropsychological assessments as predictive features in their machine-learning models. This limitation is critical as cognition and network topology are intricately linked but may be differentially impacted by aging and AD. Further, no studies combining machine learning and network neuroscience have addressed the presence of

neurocognitive subgroups in individuals with and without AD. Thus, it is uncertain if cognition and network topology, along with core demographic risk factors of AD (age, education, sex), can identify cognitively normal individuals lacking a formal diagnosis of MCI who are at a greater risk of converting to AD. Moreover, it is unknown if cognitive subgroups, previously identified in individuals with MCI and AD, emerge when cognitive assessments are combined with measures assessing network topology. To address these gaps, this study built upon prior foundational work by utilizing a large open-source dataset with exceptional imaging quality and combined (1) weighted and unthresholded graph-theoretical metrics constructed using a well-validated functional atlas to capture redundancy and other critical aspects of network topology, (2) demographic measures and cognitive assessments evaluating attention, memory, and executive function, (3) and both supervised and unsupervised machine learning paired with repeated k-fold cross-validation to distinguish individuals with and without AD and to evaluate the presence of neurocognitive subgroups.

CHAPTER 2: PROJECT GOALS

While AD is often associated with episodic memory impairment, individuals with AD or who are diagnosed with one of several behavioral variants of dementia (e.g. multidomain, dysexecutive, amnestic, etc.) show diverse cognitive profiles. Similarly, heterogeneous subgroups of typically aging individuals may show impairments in both general and specific forms of attention, memory, and/or executive functioning. As with cognition, there is overlap between age and AD-related changes observed at the level of functional brain networks. The extent to which functional brain networks are altered may be predictive of cognitive impairment experienced by an individual, however, this has not been tested. Thus, several questions were raised. First, do metrics that quantify redundancy in functional brain networks, as a proxy of brain and cognitive reserve, support cognition? Second, is it possible to disentangle the overlap of brain changes and cognitive impairment seen in aging and early-stage or preclinical dementia? Third, are metrics that quantify redundancy in functional brain networks more sensitive than other traditional network measures? Fourth, can this information be used help to identify individuals who are either more resistant to normative and/or disease-related cognitive decline? Finally, what about individuals who are at risk of converting to AD? Do these markers depend on other risk factors of AD such as age, sex, and education level?

To address these questions, this study utilized an existing dataset collected as part of the Open Access Series of Imaging Studies (Marcus et al., 2010). The dataset was comprised of individuals with and without dementia who completed a series of structural and functional neuroimaging scans, a battery of neuropsychological assessments, and genetic testing (see

Methods for greater detail). As noted by Bondi and colleagues (2008), expansion beyond diagnostic entities enables researchers to delineate clinically-relevant subtypes based on the combination of multiple neuropsychological measures and rigorous statistical methodology (Bondi et al., 2008, 2014; del Carmen Díaz-Mardomingo et al., 2017). Data-driven statistical learning algorithms are designed to uncover complex high-dimensional patterns within and across datasets. Solutions from these predictive algorithms provide researchers with both sample- and cohort-specific probabilities of risk that can be used to disentangle multifaceted profiles of cognitive dysfunction in aging individuals with and without AD pathology. Thus, this study leveraged a large-open source dataset with functional imaging data of exceptional quality and sufficient duration to robustly estimate network properties, and multiple neuropsychological assessments to address the heterogeneity in individuals with and without AD using machine learning.

First, in Aim 1, I sought to establish whether metrics that index redundancy in functional brain networks were associated with cognition. Next, in Aim 2, I conducted predictive modeling to establish if a combination of core demographic factors (age, sex, and education), network metrics, and cognitive measures could accurately distinguish CN, AD converters (AD-C), and AD individuals using supervised machine learning. Finally, in Aim 3, I quantified the presence of data-driven neurocognitive subgroups using the same combination of demographic, graph-theoretical metrics, and neuropsychological assessments using unsupervised and supervised machine learning. Ultimately, this study aimed to identify a set of factors that could distinguish otherwise healthy aging individuals from those with or at risk of developing AD, as well as to identify individuals who may be at a greater risk of converting to AD using baseline assessments of cognition and brain network organization and communication capacity.

Aim 1 sought to establish whether metrics that quantify aspects of redundancy in weighted functional brain networks were associated with episodic memory and aspects of executive function, including working memory and mental flexibility. Specifically, this aim tested for the presence of linear relationships between global communicability and global clustering coefficient with performance on the (1) Logical Memory, (2) Digit Span Backward, and (3) Trail Making Part B tests. These assessments capture primary cognitive difficulties impacted by aging and AD that have previously been shown to differentially contribute to the detection of cognitive subtypes (Scheltens et al., 2016, 2017).

Hypothesis: Redundancy is a measure of reserve in brain networks. Prior work has shown that (1) whole-brain redundancy increases with age and is associated with better executive functioning, and (2) higher hippocampal redundancy is associated with better memory performance and is reduced in individuals with MCI (Langella, Mucha, et al., 2021; Langella, Sadiq, et al., 2021; Sadiq et al., 2021). Building off this seminal work, I predicted that higher levels of whole-brain redundancy, captured by global communicability and global clustering coefficient, would be associated with better overall performance on neuropsychological assessments. However, as older individuals with AD are more likely to have impaired memory, I predicted that the magnitude of association between whole-brain redundancy and memory performance may be greater in the AD group as compared to the AD-C and CN groups (AD > AD-C > CN).

Alternative Hypothesis: The magnitude of association between whole-brain redundancy and memory ability may not differ between groups. Specifically, this study assessed whole-brain redundancy, as opposed to hippocampal redundancy. As noted above, Langella and colleagues (2021) reported that hippocampal redundancy was positively associated with memory

performance (Langella, Mucha, et al., 2021; Langella, Sadiq, et al., 2021), however, Sadiq and colleagues (2021) found that whole-brain redundancy was positively associated with executive functioning (Sadiq et al., 2021). Moreover, whole-brain redundancy was found to mediate the negative relationship between age and executive function. Thus, whole-brain redundancy may equally serve as a protective mechanism against cognitive impairment in individuals with and without AD regardless of cognitive domain.

Aim 2 sought to establish whether a combination of core demographic risk factors (age, sex, and education), cognitive measures, and weighted functional network metrics could accurately distinguish CN, AD-C, and AD individuals using supervised machine learning. As discussed in the introduction, the magnitude of cognitive impairment and decline observed in otherwise cognitively normal individuals, and the subsequent risk of developing dementia is person-specific, not all aging individuals experience cognitive impairment. Likewise, graph-theoretical metrics describe, or summarize, particular interdependent aspects of a complex system. As such, it is not likely that a single brain metric can, or should, be used to predict dementia (Foo et al., 2021; C. Li et al., 2011; Telesford et al., 2010, 2013). Thus, in addition to global communicability and global clustering coefficient, network metrics that capture critical aspects of network topology, including measures of network integration and segregation, were estimated and included in the predictive models. Independently, these metrics are altered in a portion of individuals with AD as compared to non-demented adults (Ferreira et al., 2019; Sanz-Arigita et al., 2010; Sun et al., 2017) and individuals with non-AD dementia, such as dementia with Lewy bodies (Peraza et al., 2015)

As highlighted above, various studies have combined machine learning and network neuroscience to distinguish aging individuals with and without AD or MCI (see Machine

Learning in AD). In these studies, global efficiency, clustering coefficient, and centrality measures (e.g. betweenness centrality, Pagerank centrality) have most consistently been shown to contribute to classification performance. However, to my knowledge, no studies have combined network metrics, demographic factors, and cognitive performance measures to distinguish individuals who will convert to AD without a formal diagnosis of MCI (e.g. AD-C) from healthy controls and individuals already diagnosed with AD. As highlighted in the introduction, age, education, and sex are core demographic risk factors for AD. Older individuals who are less educated, and thus have lower cognitive reserve, are at a greater risk for developing dementia. As highlighted in the section on *Reserve, Resistance, & Resilience in Cognitive Aging,* with age and education, individuals develop both generalized and specific strategies to solve complex problems that are supported by adaptive networks of interconnected brain regions (Stern et al., 2019) captured by graph-theoretical measures. Individuals may experience variable forms or degrees of cognitive impairment, even when there are no detectable differences in network topology between groups. Moreover, as noted by Kwak and colleagues (2021), brain and cognitive profiles may not align when assessed independently (Kwak et al., 2021).

Thus, this study combined summary measures of network topology with cognitive performance measures that broadly tap cognitive domains differentially impacted by aging and AD. Specifically, using baseline neuropsychological assessments and functional neuroimaging scans, this study aimed to address gaps in the literature to understand how both brain and cognition interact when attempting to understand why some individuals are more at risk of converting to AD. Further, this aim sought to compare the degree to which cognitive and/or network metrics contribute to classification.

Hypothesis: I predicted that the combination of core demographic risk factors, cognitive measures, and network metrics would be able to distinguish the CN, AD-C, and AD groups with moderate success. I predicted that age would be the most informative demographic feature as it is a primary risk factor of AD ("2021 Alzheimer's Disease Facts and Figures," 2021; Guerreiro & Bras, 2015; Jack et al., 2018; LaMontagne et al., 2019). Neuropsychological assessments are designed to detect broad cognitive impairments due neurodegenerative disease (Horowitz et al., 2018; Silverstein, 2008), and are thus likely to significantly contribute to classification performance. However, given the older age range of the sample and greater prevalence of typical AD cases with predominant memory impairments, I predicted that memory-based cognitive metrics would be weighted more heavily than measures predominately assessing attention or executive function. Concerning network topology, studies combining machine learning and graph theory consistently find that global and nodal estimates of efficiency and clustering coefficient contribute to better model performance when classifying individuals with AD or MCI from healthy controls (Dyrba et al., 2015; Jie et al., 2014; Khazaee et al., 2015; Y. P. Li et al., 2013; Wang et al., 2013; Wee et al., 2012). Thus, in addition to global communicability, I hypothesized that global efficiency and global clustering coefficient would contribute more to classification performance. Finally, I expected that the AD-C group was likely comprised of individuals who are at vastly different stages in the AD continuum (Ferreira et al., 2020). Thus, I further hypothesized that individuals in the AD-C group were more likely to misclassified as either CN or AD (Hojjati et al., 2017; Kwak et al., 2021; Pereira et al., 2017).

Alternative Hypothesis: Whether based on cognitive performance or pathology, multiple subgroups have been identified in otherwise cognitively normal individuals and individuals diagnosed with either MCI or AD (Alexander et al., 2021; del Carmen Díaz-Mardomingo et al.,

2017; Edmonds et al., 2021; Ferreira et al., 2020; Guan et al., 2017; Jak et al., 2009; Mez et al., 2016; Murray et al., 2011; Scheltens et al., 2018, 2016, 2017; Vogel & Hansson, 2022). Thus, I hypothesized that individuals within each of the *a priori* groups (CN, AD-C, AD) may exhibit diverse neurocognitive profiles that would lead to poor overall classification performance.

Aim 3 sought to establish whether the combination of core demographic, network metrics, and cognitive measures would identify data-driven neurocognitive subgroups. Specifically, Aim 3 specifically sought to establish if previously identified cognitive subgroups would emerge (Bickerton et al., 2011; del Carmen Díaz-Mardomingo et al., 2017; Edmonds et al., 2021; Geifman et al., 2018; Kwak et al., 2021; Scheltens et al., 2016, 2017, 2018; Vogel & Hansson, 2022) using unsupervised learning.

Hypothesis: I predicted that multiple neurocognitive subgroups would emerge driven by greater heterogeneity in CN and AD-C participants. That is, I expected many of the AD participants to show more pronounced memory-related deficits, and/or exhibit impairment across the battery of neuropsychological assessments given the older age range of the sample. Based on prior work (Scheltens et al., 2017), I hypothesized that two primary groups, memory-impaired and memory-spared, would emerge. Prior work suggests that redundancy, as a mechanism of reserve, begins to decline around the age of 60 (Sadiq et al., 2021), and is reduced in individuals with MCI who have poorer episodic memory (Langella, Mucha, et al., 2021; Langella, Sadiq, et al., 2021). Additionally, typical AD is the most common biological subtype estimated to represent approximately 55% of all cases (Ferreira et al., 2019).

Participants with typical AD have impaired memory that is accompanied by hippocampal/medial-temporal lobe atrophy (Ferreira et al., 2019; Vogel & Hansson, 2022). Thus, I hypothesized that the memory-impaired group would be comprised primarily of AD

participants, and characterized by reduced whole-brain redundancy (global communicability and global clustering coefficient). Within the memory-spared group, individuals may be further stratified by dysexecutive and mixed/global impairment subtypes (Ferreira et al., 2019; Sanz-Arigita et al., 2010). Thus, I further hypothesized that the memory-spared group would be comprised of younger AD-C participants with poorer executive functioning corresponding with a reduction in global efficiency and global participation coefficient estimates (Avelar-Pereira et al., 2017; Chan et al., 2017; Grady, 2017).

Alternative Hypothesis: When implementing data-driven clustering of CN, MCI, and/or AD individuals, anywhere between two to nine subgroups have been identified. Thus, I hypothesized that multiple smaller subgroups may emerge where each subgroup is comprised of individuals with complex neurocognitive profiles that (1) do not perfectly correspond to previously observed subtypes, or (2) reflect diagnostic status. However, when assessing older individuals at a single time point, it may be difficult to capture robust and clinically meaningful neurocognitive subgroups over and above global cognitive impairment (Ferreira et al., 2020). In this instance, individuals may not be separated into distinct clusters, and/or clusters may simply reflect the original groups as defined by the OASIS consortium.

CHAPTER 3: METHODS AND MATERIALS

Dataset

Participants. This project utilized a large open-source dataset from the Open Access Of Imaging Series (OASIS-3) (LaMontagne et al., 2019; Marcus et al., 2010). The OASIS-3 dataset is comprised of community-dwelling individuals who were recruited over 15 years via the Charles F. and Joanne Knight Alzheimer Disease Research Center at Washington University in St. Louis, Missouri. Approximately 84% of individuals identified as White and 15% identified as Black. Greater detail about the original study protocol and subject recruitment is provided elsewhere (LaMontagne et al., 2019). Open-source data used for the current project was downloaded from https://www.oasis-brains.org using the *Extensible Neuroimaging Archive Toolkit* (XNAT) (Marcus et al., 2007). Prior to analyses, participants were excluded for having a history of neurological disorders, seizures, stroke, drug or substance use disorders, or other serious medical conditions using custom-built R code (R Core Team, 2017). From the total sample of 1098, 37 participants were removed for having non-AD dementia (n = 7 frontotemporal dementia, n=30 Parkinson's disease), and 552 participants were removed for meeting exclusionary criteria as described above. 509 individuals were available for further analysis.

Image acquisition. The OASIS-3 consortium collected several anatomical and functional MRI sequences. Participants were scanned while lying in the supine position and wore earplugs to reduce scanner noise. Foam pads were used to minimize head motion. All scans were acquired using a 16-channel head coil. A high-resolution T1-weighted anatomical scan (TR = 2.4 s,

TE = 3.08 ms, FOV = 256 × 256 mm, FA = 8°, voxel size 1 × 1 × 1 mm^3), two six-minute functional resting-state blood oxygen level-dependent (BOLD) scans (EPI; (164 volumes per run), TR = 2.2 s, TE = 27 ms, FOV = 240 × 240 mm, FA = 90°, voxel size 4 × 4 × 4 mm, 36 slices), and a PET scan were acquired. For each resting-state scan, participants were instructed to keep their eyes open and to lay as still and quiet as they can. PET data were acquired on a Siemens/CTI EXACT HR+ scanner or a Biograph 40 PET/CT scanner. Pittsburgh Compound B (PIB) and/or Florbetapir [18F] radiotracers were used to quantify amyloid (Aβ) levels in the brain. To capture the uptake window for PIB, a 60-minute dynamic PET scan (24 x 5-sec frames; 9 x 20-second frames; 10 x 1-minute frames; 9 x 5-minute frames) was acquired starting at the time of intravenous administration. Florbetapir PET scans were acquired using one of two possible methods depending on whether or not intravenous injection was (1) administered in the scanner and a 70-minute dynamic PET scan was used to capture uptake (4 x 15-second frames, 8 x 30sec frames, 9 x 1 minute, 2 x 3-minute frames, 10 x 5-minute frames) or (2) before image acquisition where a dynamic 20-minute scan was acquired 50-minutes after injection. All PET data was processed by the OASIS consortium using the PET unified pipeline. Standard uptake ratios (SUVR) for PIB and Florbetapir were calculated using the cerebellum as a reference region and normalized to a range of 0-100. For greater detail regarding PET acquisition and use of radiotracers see LaMontagne et al. (2019) and Marcus et al. (2010).

Image processing. Baseline neuroimaging data were preprocessed using standard processing procedures with the CONN connectivity toolbox (Whitfield-Gabrieli & Nieto-Castanon, 2012) in Matlab. Briefly, preprocessing of neuroimaging data included segmentation of high-resolution structural (T1-weighted) images into gray matter, white matter, and cerebrospinal fluid masks. Temporal and spatial artifacts were removed from the functional

BOLD data. Specifically, functional images underwent realignment and unwarping, slice-timing correction, co-registration to structural images, spatial normalization, and motion outlier identification. Spurious, non-neuronal noise from white-matter and CSF masks, as well as 6 motion realignment parameters and their first-order derivatives, were used as nuisance regressors. Temporal band-pass filtering was employed to remove BOLD signal frequencies below 0.008 Hz or above 0.09 Hz. Motion artifacts were identified and removed from the Artifact Detection Toolbox (ART) as part of the standard CONN preprocessing pipeline. A combination of global BOLD signal change and framewise displacement (FD) was used to detect the presence of sub-millimeter movement known to negatively impact data quality and alter functional connectivity estimates (Dosenbach et al., 2017; Power et al., 2012). Specifically, data for the current project was processed using conservative motion correction criteria in which outlier volumes were identified if (1) global BOLD signal change (Z) was greater than 3 standard deviations from the global average (standard deviation units), and (2) if relative frame-to-frame motion (e.g., framewise displacement [FD]), in six directions (3 translations: x, y, z; 3 rotations: pitch, roll, yaw) was greater than .5 mm. By default in CONN, FD is calculated in reference to six control points placed in three-dimensional space within a bounding box surrounding the brain (Whitfield-Gabrieli & Nieto-Castanon, 2012). Control points are centered on each face of the three-dimensional bounding box. This method differs slightly from FD as originally defined by Power et al. (2012), in which movement is relative to displacement calculated on the surface of a 50 mm radius sphere. Participants with less than 6 minutes of data (N=75), excluding outlier volumes, were removed from all analyses to ensure any findings were robust to the deleterious effects motion has on functional connectivity) (Birn et al., 2013; Ciric et al., 2017; Power et al.,

2012, 2015, 2017; Satterthwaite et al., 2012). Finally, six participants were removed due to errors during preprocessing due to poor image quality and failed registration.

Final Participant Sample. The final clean sample of low-motion participants who completed cognitive assessments within one year of their first fMRI scan (N =296) was comprised of cognitively normal individuals (CN; N=159), individuals who were cognitively normal at the time of assessment who will convert to AD at some point during the study (AD-C; N=51), and individuals diagnosed with AD at study onset (AD; N = 86; Table 1). All diagnoses were based on the Clinical Dementias Rating (CDR) scale. Global CDR scores were based on a scale of 0–3: no dementia (CDR = 0), questionable dementia (CDR = 0.5), mild dementia (CDR = 1), moderate dementia (CDR = 2), and dementia with severe cognitive impairment (CDR = 3). The CDR Sum of Box scores (CDR-SOB) represents a continuous measure of dementia-related cognitive impairment based on scores summed across specific categories that describe cognitive and functional impairments (Crane et al., 2012; McDougall et al., 2021; O'Bryant et al., 2008). The CDR-SOB was scored on a scale of 0-18 (relationship between the scales: CDR 0 = CDR-SOB 0; CDR 0.5 = CDR-SOB 0.5-4; CDR 1 = CDR-SOB 4.5-9; CDR 2 = CDR-SOB 9.5-15.5; CDR 3 = CDR-SOB 16-18). Of note, the CN group, as defined by OASIS, is comprised of individuals who do not have dementia and have not been diagnosed with AD based on the CDR (Table 2). However, these individuals may still display age-related cognitive deficits on neuropsychological assessments. In addition to the CDR, education level (years completed) and performance on the MMSE (Folstein et al., 1975) were collected. The MMSE is a 16-item questionnaire used to assess cognitive functioning in older individuals across general domains of orientation, attention, memory, language, and visual-spatial skills.

Neuropsychological assessments were completed at different sessions than the neuroimaging acquisition. All participants completed neuropsychological assessments within one-year of their baseline functional resting-state scan. The gap between administration of neuropsychological assessments and baseline neuroimaging scans ranged from 237-356 days (CN: Mean = -42.6 days; AD-C: Mean = -43.0 days; AD = -63 days), and was not statistically different between groups (Table 1).

To permit comparisons with prior work, a breakdown of ApoE polymorphisms for this sample is provided in Table 3. Two participants were missing genetic testing data. From the sample of participants with ApoE information (N = 433), ApoE-ε3 (81%) was the most common polymorphism followed by ApoE-ε4 (38%). Approximately 38% of the sample was homozygous for the ApoE 33 polymorphism. The proportion of ApoE-ε4 positive individuals was greatest in AD participants (65%), followed by AD-C (41%), and CN (29%).

Table 1. Participants characteristics and data availability are presented for the final dataset for each *a priori* group.

	CN	**AD-C**	**AD**	**f/x²***	***p*****
Age (Years)	65.91 (10.17)	78.78 (8.05)	75.72 (8.31)	79.37	**< .001**
Education (Years) [a]	16.23 (2.44)	15.62 (2.33)	14.45 (3.19)	76.19	**< .001**
Sex (Male/Female)	126/164	27/31	34/52	.742	.690
MMSE[b]	29.22 (3.31)	27.07 (3.64)	23.54 (4.20)	178.20	**< .001**
CDR-SOB[c]	0.01 (0.07)	1.28 (1.87)	4.32 (2.64)	336.45	**< .001**
Average Motion					
Minutes Remaining	10.1 (1.37)	11.2 (.74)	10.3 (1.32)	14.31	**< .001**
Mean FD (mm) [d]	0.211 (.058)	0.122 (.042)	0.209 (.059)	64.574	**< .001**
COG-GAP [e]	-42.6 (102)	-43 (111)	-63 (73.2)	1.28	.278
DAYS-CONV [f]	---	1785.11 (1197.89)	---	---	---

[a] *Education level in years [7-22]*
[b] *MMSE: Mini Mental-State Examination [0-30]* (Folstein et al., 1975)
[c] *CDR-SOB: Clinical dementia rating scale sum of box scores [0-30]*
[d] Mean FD: *Mean remaining framewise displacement* (Power et al., 2012)
[e] *COG-GAP: Number of days in between cognitive assessment and first fMRI scan [0-365]*
[f] *DAYS-CONV: Number of days in between first fMRI scan and final AD diagnosis*
** Chi-square (x^2) tests were used to test for group differences on categorical variables (e.g. sex and CDR scores); Omnibus f-tests (f) were used from a one-way analysis of variance to assess group differences on continuous measures.*

*** Bolded values are significant at alpha = .05 (uncorrected)*

Table 2. Global CDR Scores. The total number of individuals for a given CDR category is presented.

	Global CDR				
	0	**0.5**	**1**	**2**	**3**
CN	290	--	--	--	--
AD-C	27	28	2	1	0
AD	0	45	36	4	1

[a] *CN: cognitively normal*
[b] *AD-C: AD converters*
[c] *AD: Alzheimer's disease*

Table 3. ApoE polymorphism tally. The total number of individuals with a given ApoE polymorphism is displayed for each a priori group. The percentage of individuals, relative to the total sample with ApoE data (N=434), is provided in the parentheses.

	22	**23**	**24**	**33**	**34**	**44**
CN[a]	2 (.46)	30 (6.91)	9 (2.07)	164 (37.79)	73 (16.82)	12 (2.76)
AD-C[b]	1 (.23)	7 (1.61)	1 (.23)	25 (5.76)	20 (4.61)	4 (.92)
AD-C	0	3 (.69)	1 (.23)	26 (5.99)	45 (10.37)	11 (2.53)

[a] *CN: cognitively normal*
[b] *AD-C: AD converters*
[c] *AD: Alzheimer's disease*

Neuropsychological assessments

Participants completed a battery of cognitive assessments testing general aspects of cognition in the domains of memory, executive function, and attention. Assessments included the Logical Memory Subtest (Story A) from the Wechsler Memory Scale-Revised (Wechsler, 1981), the Digit Symbol and Digit Span tasks from the Wechsler Adult Intelligence Scale-Revised (Wechsler, 1981), the Category Fluency Test (Lezak et al., 2004), the Boston Naming Test (Kaplan, 1983), and Trail Making Test parts A & B from the Delis-Kaplan Executive Function System (D-KEFS; Baron, 2004). Brief descriptions for each task are provided below.

Logical Memory II (Story A). A measure of episodic memory that requires participants to recall as many details as they can from a short story read aloud to them by a qualified neuropsychologist. Participants must recall information after a 30-minute delay. Performance is

captured by the number of elements recalled from the story with scores ranging from 0 (no recall) to 25 (complete recall).

WAIS-R Digit Symbol Test. A general measure of several cognitive functions including attention, visuoperceptual ability, and motor speed. Participants are provided a key and are required to match as many symbols and numbers as they can in 90 seconds. Performance is captured by the total number of digit symbol pairs completed in 90 seconds.

Digit Span Test. A measure of attention and working memory that requires participants to repeat a series of digits either forwards or backward. In the forward-span condition, participants are asked to repeat a series of digits in sequential order. In the backward-span condition, participants are asked to repeat a series of digits in reverse sequential order. Digit span performance was captured by the number of trials repeated correctly for both the digit-span forward and digit-span backward conditions.

***Trail Making Test*:** The D-KEFS Trail Making Test assesses motor functioning and general aspects of executive functioning including inhibition and interference control. Participants are asked to connect a series of dots, by drawing a line, connecting numbers (1-26) in part A, or a series of alternating numbers and letters (1-A-2-B) in part B. Performance is captured by the total time taken to connect the series of dots for Part A and Part B, as well as the number of omission and commission (e.g. sequencing) errors for each part.

Category Fluency. A measure of semantic memory. Participants were asked to name as many animals or vegetables as they could in 60 seconds. Performance was captured by the total number of words correctly identified for each category.

Boston Naming Test. A measure of verbal fluency. Participants were asked to name each of 60 black and white line drawings representing everyday common objects. Performance was captured by the total number of drawings correctly named.

Graph Theory Metrics

Communicability. Communicability is a decentralized measure of information transfer within a given system that quantifies the total number of paths, direct or indirect, that can be traversed by a random walker between any pair of regions (Estrada & Hatano, 2008). Communicability can be applied to either binary or weighted graphs. For weighted graphs, the communicability of a matrix is equal to the weighted connectivity matrix (*W)* raised to a power (*k)* defined as the degree of a given target region. Communicability is algebraically equivalent to taking the matrix exponential (e^W) of an undirected graph. Communicability, as proposed by Estrada and Hatano (2008), down-weights the contribution of longer walks by dividing the matrix exponential (A^k) by the total number of paths *k!* present in the network (Estrada and Hatano, 2008). Global communicability is derived by taking the sum of the communicability matrix as given by Equation 1:

$$Global\ Communicability = \sum_{k=0}^{\infty} (\frac{W^k}{k!})_{ij} = (e^W)_{ij} . \quad (1)$$

Clustering Coefficient. The weighted clustering coefficient measures the degree of connectedness between neighboring nodes in a weighted graph (Onnela et al., 2005). It is the probability that two neighbors of a node are themselves connected, forming a closed triangle. The global clustering coefficient is the average of the clustering coefficient for all nodes in a network and is given by Equation 2:

$$Global\ Clustering\ Coefficient = \frac{\Sigma_{i,j,k} (w_{ij} w_{jk} w_{ki})^{1/3}}{\Sigma_i k_i (k_i - 1)} , \quad (2)$$

where *w* is the weighted edges connecting node *i* to neighboring nodes *j* and *k*. Values are averaged across all nodes to obtain a global summary estimate of the clustering coefficient.

Global Efficiency. Global efficiency measures the relative speed of information transfer across a network via shortest path routing. Shortest paths are the most direct paths connecting a pair of regions. Global efficiency is given by Equation 3:

$$Global\ Efficiency = \frac{1}{n}\sum_{i\in N}\frac{\sum j\in N, j\neq i\ (d_{ij}^{W})}{n-1}\ , \tag{3}$$

where *N* is the number of nodes in the network, *d* is a geodesic distance matrix, and d_{ij} is the shortest weighted *w* path between node *i* and node *j.*

Path Transitivity. Path transitivity measures the proportion of detours or alternate routes along the shortest path connecting a pair of regions (Goni et al., 2014). Global path transitivity is the average path transitivity for all nodes *n* in a graph as given by Equation 4:

$$Global\ Path\ Transitivity = \frac{2\sum_{i\in\Omega}\ \sum_{j\in\Omega}\ m_{ij}}{|\Omega|\ (\ |\Omega|-1)}\ . \tag{4}$$

where Ω is the sequence of nodes along the shortest path connecting nodes *i* and *j*, and *m* is the matching index that describes the set of shared input and output links (e.g. detours) along the shortest path connecting nodes *i* and *j.*

Eigenvector centrality. Eigenvector centrality is a centrality metric that captures the overall importance of highly connected and influential nodes, or vertices *v,* in a graph (Newman, 2010). Nodes with high centrality are connected to other nodes with high centrality and are weighted more heavily. Global eigenvector centrality is the average eigenvector centrality for all nodes n in a graph as defined by Equation 5:

$$Global\ Eigenvector\ Centraliity = \frac{1}{\lambda}\Sigma_{t\in M(v)}x_t\ , \tag{5}$$

where *n* is the total number of nodes and *N* is the set of nodes in a network. λ = is a constant

term, *t* is a node and is connected to a set of neighboring nodes *M* for node *v*. *x* is a vector of eigenvalues that represent each nodes weighted centrality.

Modularity. Modularity represents the degree to which a system or network can be partitioned into distinct communities, or modules, that contain subsets of nodes that are highly interconnected to one another. The Louvain algorithm (Blondel et al., 2008) from the Brain Connectivity Toolbox (www.brain-connectivity-toolbox.net) was applied to estimate the modularity quotient *Q* over a pre-defined set of 14 well-stablished functional networks (Seitzman et al., 2018). Modularity is defined by Equation 6:

$$Q = \frac{1}{l^w}\sum_{i,j}\left[w_{ij} - \frac{k_i^w k_j^w}{l^w}\right]\delta(m_i, m_j)\,, \tag{6}$$

where W_{ij} is an entry of the weighted connectivity matrix containing the weight of an edge between nodes *i* and *j, k_i and k_j is the* sum of the weights *w* of the edges attached to node *i* and *j*, *m* is the community or module to which node *i* or *j* is assigned, and *l* is the total number of links. δ is the Kronecker delta function.

Participation Coefficient. The participation coefficient is a measure of network integration and represents the degree to which a brain region communicates with other regions outside of its own module. The global participation coefficient is the average participation coefficient for all nodes n in a graph and is given by Equation 7:

$$Global\ Participation\ Coefficient = \frac{1}{n}\sum_{i=1}^{N}\left[1 - \sum_{s=1}^{M}\left(\frac{k_{is}}{k_i}\right)^2\right]\,, \tag{7}$$

where k_i is the sum of connection weights belonging to node *i*, k_{is} is the sum of region *i's* connection weights within its own module *s*, and *M* is the total number of modules for a given network partition. The participation coefficient is equal to 0 when a region communicates exclusively within a given module and 1 if all connections are between modules.

Small-World Propensity. The small-world propensity (Muldoon et al., 2016b) is a global measure of small-worldness (Watts & Strogatz, 1998), that captures the degree to which networks exhibit a balance between high clustering and short path lengths while controlling for network density. In general, a higher small-world propensity represents a more efficient system compared to random chance, and is given by Equation 8:

$$Small\text{-}World\ Propensity = 1 - \sqrt{\frac{\Delta_C^2 + \Delta_L^2}{2}}, \quad (8)$$

where C is the average clustering coefficient, and L is the average path length of a network. Δ_C^2 and Δ_L^2 are ratios and represent the degree to which C and L deviate from lattice and random network models, constructed with the same number of nodes and degree distribution as the comparison network, respectively.

Within-module degree. Within-module degree represents how connected a node is to all other nodes within its own module. The global within-module degree is the average within-module degree for all nodes n in a graph and is given by Equation 9:

$$Within\text{-}module\ degree = \frac{1}{n}\sum_{i=1}^{N}\left[\frac{k_{is} - \overline{k}_{si}}{\sigma_{k_{si}}}\right], \quad (9)$$

where k_{is} is the degree of node *i* to all other nodes within its own module (*s*), k_{si} is the average degree of all nodes in *s.* Within-module degree is a standardized measure, where $\sigma_{k_{si}}$ is the standard deviation of the degree of all nodes in *s*.

Machine Learning

Random forest classification. A random forest is a collection, or ensemble, of decision trees. Decision trees are appropriately named for their tree-like appearance and are used to group similar observations into classes based on a set of descriptive features (see Figure 3). Observations can be people, music, movies, food, and so on. In the case of people, descriptive

features could include age, gender, weight, heart rate, personality traits, likes, and dislikes. Class labels might represent categories such as old or young, sick or healthy. As a machine learning tool, a decision tree is tasked with learning a set of rules that best divides a dataset into a set of known class labels, or into data-driven subgroups comprised of similar observations (e.g. individuals). In unsupervised learning problems, no class labels (e.g. CN or AD) are provided *a priori*. That is, a model is tasked with learning the degree to which individual observations cluster together from the data (Breiman, 2001). For both supervised and unsupervised classification problems, this set of rules can later be used to make predictions as to which class a new, unfamiliar observation object might be classified.

Within a given decision tree, a dataset is split into branches (connections) and nodes (features) in a hierarchical manner. Starting at the root node (top of a tree), the data is split on a feature that best differentiates a set of observations. From the top of the tree, branches extend downward to form connections with nodes that contain other features that next best differentiate a set of observations. This process occurs until no more splits are possible and a decision is reached at a final terminal node (see Figure 3). To grasp how a decision tree works in practice, consider how a person may decide to go outside and exercise or stay home and read. The decision to go outside and exercise or not can be based on features that describe the weather, the person's current mood state, health status, love of reading, and so on. In this example, the most obvious choice to exercise or not depends on whether or not the person is feeling ill. So, starting at the root node with health status the decision process begins by asking if the person sick? If not, is it raining? If not, is it too windy? If it is not too windy, does the person love reading too

much? If not, go exercise!

To assess which factors best distinguish a group of cognitively normal individuals from those diagnosed with AD, decisions can be based on demographic, diagnostic, genetic, neuroimaging, or other data types. In Figure 3, the decision process begins with age, such that an individual may be younger (blue) or older (red) than 65 years of age. For the younger individual (blue), a higher education level and absence of the ApoE-e4 allele results in a final decision of cognitively normal. For the older individual (red), a greater abundance of *Aß* and observed memory impairment results in a final decision of AD. In this simple example, the final decision can only be one of two known choices (e.g. CN or AD) and is thus considered a form of supervised machine learning. In unsupervised machine learning, there are no predefined class labels. The goal of unsupervised learning is to identify individuals who share a common set of attributes. As discussed in the introduction, different forms of unsupervised learning have been used to detect subsets of individuals with and without AD, who were labeled as being memory-impaired or memory-spared (Scheltens et al.,

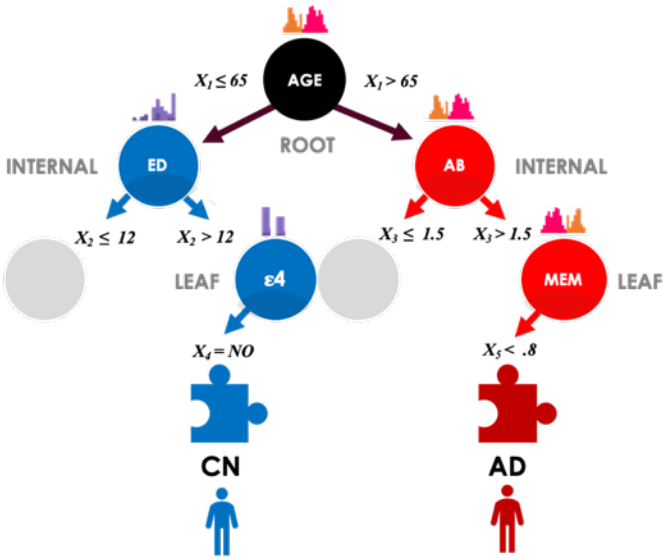


Figure 3. Simple Decision Tree. This figure provides a graphical representation of a decision tree and its constituent parts. Within a given decision tree, a dataset is split into branches (connections) and nodes (features) in a hierarchical manner. Starting at the root node (top of a tree), the data is split on a feature (Age) that best differentiates a set of observations. From the top of the tree, branches extend downward to form connections with nodes that contain other features that next best differentiate a set of observations. This process occurs until no more splits are possible and a decision is reached at a final terminal node. Abbreviations: ED = Education; AB = Amyloid-beta; ε4 = ApoE-ε4; MEM = Memory; CN = Cognitively Normal; AD = Alzheimer's disease

2016, 2017). For an unsupervised decision tree, the final decision represents the probability that an observation is similar to other observations previously assessed.

Decision trees can learn from data by revealing values for each feature that best split observations into their respective class. The learned values at each split in the tree are the set of rules that can guide future predictions. Most decision processes are not as simple as in the example above. Typically, decisions in classification problems involve a large number of features that can vary greatly across observations or individuals. Further, features may interact with one another in unexpected ways. This variation cannot be accounted for using a single decision tree with a single set of rules, or even many different trees using the same set of rules. To circumvent this issue, the random forest uses many decision trees. Each decision tree is constructed using a different, randomly selected subset of observations and features, a process known as bootstrapping. This process also helps to ensure that a set of decision trees are not correlated with one another or overly biased towards a specific item or feature within a dataset. The unbiased trees are then combined or aggregated (bootstrap aggregating). The most predictive features are those which contribute the greatest to classification across the set of decision trees. Random forest ensemble methods have been shown to outperform various classification algorithms on datasets varying in sample size and dimensionality (Briand et al., 2009). Further, the random forest has been found to perform well in small sample studies with at least 40 observations per class (Feczko et al., 2018; Luan et al., 2020).

Random forest application. The random forest is designed to handle complex high-dimensional interactions and is well-suited for combining imaging and non-imaging metrics measuring function brain network organization and cognition respectively (Altman & Krzywinski, 2017; Breiman, 2001; Bzdok et al., 2018; Denisko & Hoffman, 2018). In the present

study, a combination of supervised and unsupervised classification modeling was used: (1) to assess if *a priori* defined groups (i.e., CN, AD-C, AD) could be accurately distinguished using a combination of nine weighted functional network metrics and nine cognitive performance measures (see the *Neuropsychological Assessment* section above for specific features), and (2) to identify data-driven subgroups with no *a priori* defined class labels based on the same combination of predictive features. In exploratory data analysis, Pearson's correlation was used to assess multicollinearity between features. Features that were highly correlated ($r > .85$) were to be excluded from the model to eliminate redundant predictors (C. Li et al., 2011). For both supervised and unsupervised models, the set of decision trees is aggregated to generate forests that minimize a cost function, the Gini impurity index. The Gini impurity index measures how "pure" a set of classifications are for a decision point at each root node within a given decision tree and represents the probability of misclassifying new data. Minimizing the cost function ensures that groups are maximally separated (i.e. better classification accuracy).

Supervised random forest. For supervised models, to assess classification performance repeated k-fold cross-validation was performed. K-fold cross-validation randomly partitions data into *k* independent subsamples which are then used to train and test model performance. When combined with the built-in bootstrapping technique (see Modeling & Prediction), cross-validation helps to further reduce overfitting and sampling bias. This study implemented 5-fold cross-validation with ten repeats. With 5-fold cross-validation, a dataset is randomly partitioned into 5 subsamples. Within each fold, 4 of the subsamples are used to train the random forest, and the remaining subsample is set aside and used to test model performance. This process was repeated 10 times to ensure all observations were subsampled and to reduce test error (Borra & Di Ciaccio, 2010). The optimal set of hyperparameters, including the number of splitting

variables and the number of decision trees, was determined via grid search whereby cross-validated models were fit using a combination of parameters (Combrisson & Jerbi, 2015; Kohavi, 1995). Parameters from the model with the lowest average out-of-sample prediction error were then selected to construct a final supervised model.

Unsupervised random forest. For unsupervised random forest classification problems, no class labels are provided *a priori*. Thus, a randomly generated dataset (called a synthetic dataset) is created on each bootstrap iteration. Specifically, on each iteration for a given decision tree, data points are randomly sampled for each variable selected to split the set of observations in a training dataset. Observations from the true dataset are labeled 1. Observations from the synthetic dataset are labeled 2. The random forest then proceeds to classify observations as true or synthetic (Breiman, 2001; Cutler et al., 2007). From the final aggregate unsupervised random forest model, a proximity, or similarity matrix is generated. The proximity matrix is an N x N symmetrical matrix in which each cell represents the number of times a pair of true observations (e.g. individuals) occupies the same terminal node (e.g. prediction) across the set of decision trees. Random forest models, 5-fold cross-validation, and permutation tests were carried out in R (R Core Team, 2020) using the randomForest (Liaw & Wiener, 2002), caret (Kuhn, 2008), and the rfUtilities (Murphy et al., 2010) packages, respectively. The proximity matrix was visualized using the *MDSplot* function within the randomForest package (Liaw & Wiener, 2002).

Cluster analysis. To determine the presence of data-driven neurocognitive subgroups, cluster analysis was performed on the proximity matrix using k-means clustering. The K-means algorithm attempts to minimize the total distance, and thus dissimilarity, between a set of data points in geographical space. For every cluster k, a random starting point (e.g. centroid) is assigned. The point of origin within a given cluster is reassigned until the mean distance between

all points in that cluster is minimized. The *NbClust* (Charrad et al., 2015) and *factoextra* (Kassambara & Mundt, 2020) packages were used to objectively select and visualize the optimal number of clusters. To both estimate and validate an unbiased number of optimal clusters, the *NbClust* package calculates and compares up to thirty indices most commonly applied in cluster optimization problems. For example, the elbow and gap statistic methods are two of the most widely used indices. With the elbowing method, the average distance of each data point (x-axis) to its nearest *k* neighbors (y-axis) is plotted in ascending order. The sharpest inflection point in the graph, where the reduction in error from adding additional clusters is minimized, is used to designate the optimal number of clusters. The gap statistic formally assesses variation in the number of clusters. Specifically, for a set of given data points, the total variation within a set of clusters is computed across a range of *k* clusters and compared to a random null distribution. The larger the gap statistic, the greater the difference (or dissimilarity) there is between empirically and randomly derived cluster solutions. Each cluster-based index provides a solution, or vote, with respect to the optimal number of clusters. The final number of clusters is the solution with the most votes across all indices.

Cluster-based classification. Following cluster analysis, a supervised random forest was used to validate the data-driven clusters. Specifically, using the same approach as described above, a supervised random forest classification model assessed the degree to which individuals could be distinguished from one another using the data-driven class labels obtained via cluster analysis (cluster-then-predict method) (Tibshirani & Walther, 2005). Classification performance was assessed as described above.

Model performance. For supervised random forest models, permutation tests were used to assess model significance compared to random chance. Here, the distribution of model

misclassification rates averaged across all independent (e.g., held out) test sets was compared to a random null model. To generate empirical null models derived from the data, each step in the random forest process was repeated and class labels (e.g. CN, AD-C, AD) were randomly shuffled. Variable importance scores (e.g. prediction weights), average classification accuracy, sensitivity and specificity, and precision and recall were reported. Additional metrics based on predicted class probabilities including Informedness, the No Information Rate, and Cohen's Kappa were also quantified to compare performance of supervised models (Kuhn, 2008). Informedness represents how informed a set of predictions are relative to chance, defined as the difference between true and false-positive rates. Informedness values range from -1 to 1. Values less than chance (0) indicate a model which is strongly biased resulting in a greater proportion of false positive and false negative predictions. The No Information Rate (NIR) is a statistic based on a one-sided binomial test of proportions and was used to compare observed model accuracy to accuracy expected by chance given the total number of possible true cases. That is, a *p*-value (p-NIR) under the specified alpha (.05) was used to indicate if cross-validated model accuracy was statistically significant compared to the total number of sampled cases in the majority class (e.g. CN). NIR was used to assess sampling bias, whereby the proportion of classes may differ across cross-validation folds. In addition to the NIR, Kappa was used to quantify the degree to which class imbalance may impact model fit. Specifically, Kappa calculates the accuracy of a model's predictions while controlling for accuracy obtained via random chance. Values can be less than or equal to 1. Higher values suggest a better model fit.

Class imbalance. To ensure model validity, and account for selection bias, it is important to ensure adequate sampling of individuals within and across groups. Machine learning applications make extensive use of cross-validation to build and test models on unique

subsamples of data. In imbalanced classification problems, one or more groups may have a disproportionality larger number of individuals. In such cases, model predictions are skewed to the larger dominant class, and simple measures of accuracy are thus biased.

Several procedures exist that deal with class imbalance issues by either over-sampling from a minority class (or classes), under-sampling from the majority class, or some combination of over- and under-sampling during the model training phase (Altini, 2015). Synthetic minority oversampling technique (SMOTE) was used in the current study. SMOTE attempts to balance the selection of observations from minority classes (AD-C, AD), as compared to a majority class (CN) when constructing a training sample. Within each cross-validation fold, SMOTE is applied so that the randomly selected training data comprises a greater number of minority-like cases, using synthetically created data points. It is important to note that techniques that have been developed to mitigate the influence of class imbalance in machine learning models do not always result in improved model performance (Harrell, 2020). In addition to the magnitude of class imbalance, the true prevalence or proportion of instances in the population must be considered. Thus, to assess the degree to which class imbalance biases classification performance, supervised models will be generated with and without the use of subsampling. In addition, balanced accuracy and the F1 score will be reported for all models. Balanced accuracy and the F1 score are averages of sensitivity and specificity and precision and recall respectively.

Model Comparisons. Supervised random forest models with and without SMOTE were assessed for their respective ability to classify individuals into either *a priori* (CN, AD-C, or AD) or data-driven neurocognitive subgroups. Models were (1) qualitatively compared using the performance metrics as described above, and (2) quantitively compared using the area under the receiver operating curve (AUROC). The AUROC is a single numerical summary describing the

proportion of correctly identified cases (sensitivity, or true positive rate) to those correctly classified as healthy controls (specificity, or precision) for a given classification model. The AUROC was used to compared models using the non-parametric Delong test (DeLong et al., 1988). Pairwise comparisons were corrected for multiple comparisons via the FDR (Benjamini & Hochberg, 1995). ROC metrics and plots utilized the caret (Kuhn, 2008) and plotROC (Sachs, 2017) packages in R.

Missing Data Imputation

Missing data was observed for the WAIS (n = 7) and Trail Making Part B scores (n = 21). The majority of missing values were from AD participants (WAIS: CN = 0 of 159; AD-C = 2 of 56; AD = 5 of 81; Trail Making Part B: CN = 1 of 159; AD-C = 2 of 56; AD = 18 of 81). Random forest imputation was used to impute missing data. Distributions for Trail Making Part B scores were assessed before and after random forest imputation with a standard two-sample t-test. No significant differences were observed between the distributions before and after missing data imputation in the whole sample, or when comparing scores within each *a priori* group for the WAIS (whole sample: $t(582.84) = 0.251$, $p = .802$; CN = $t(316.00) = 0.00$, $p = .999$; AD-C = $t(107.96) = 0.043$, $p = .965$; AD = $t(154.16) = 0.055$, $p = .956$) and Trail Making Part B (whole sample: $t(568.76) = -1.233$, $p = .218$; CN = $t(314.97) = -0.002$, $p = .998$; AD-C = $t(128.00) = -0.063$, $p = .950$; AD = $t(128.00) = -0.870$, $p = .386$) assessments.

CHAPTER 4: RESEARCH PLAN

Aim 1: Identifying relationships between functional network redundancy and cognitive performance. Aim 1 used a subset of individuals from the primary dataset with high-quality neuroimaging data (N=435) who completed the neuropsychological battery within one year of their functional MRI scan (N=296: CN=159; AD-C=56; AD=81). Three cognitive metrics indexing episodic memory (Logical Memory), working memory (Digit Span Test-Backward), and mental flexibility (Trail Making Test B), and two weighted functional redundancy metrics (global communicability and global clustering coefficient) were used for Aim 1 analyses.

Prior to analyses, a series of quality control steps were conducted. Many neuropsychological measures suffer from ceiling effects and may exhibit low variance. Thus, the magnitude of association between network metrics and cognitive scores can be dependent on the relative distribution of scores across tests. Where appropriate, variables with a high degree of skew (e.g. log or exponentially distributed) were transformed to better approximate a normal distribution for linear analyses. The optimal transformation (e.g. log, boxcox, Yeo-Johnson) for a given variable was determined via the *bestNormalize* package (Peterson, 2021). For a given transformation algorithm, the *bestNormalize* function used 10-fold cross-validation with 5 repeats to generate average estimates for each data point. The final transformation was then subjected to a normality significance test. Pearson's correlation (*r*) was used to compare a given transformed variable to a standard normal distribution. The algorithm with the lowest estimated normality statistic (p-value corrected for the sample degrees of freedom) was selected.

Linear regression analyses tested for general associations between indices of cognitive performance and weighted functional redundancy metrics using the base R linear regression function (R: lm). An interaction term was specified within each independent regression model to test for a difference in slopes between groups. Model fit statistics, including r-squared, adjusted r-squared, root mean square error (RMSE), and FDR-corrected p-values, were reported. Model fit and coefficient estimates were considered significant if p was less than the pre-specified alpha level of .05. For models with a significant interaction term, diagnostic residual plots were assessed using base R plotting functions (R: plot) to ensure model estimates and regression coefficients were unbiased. Censored regression was used to assess the robustness of a significant interaction in the presence of either floor or ceiling effects using the *censReg* package in R (Henningsen, 2022).

Aim 2: Assessing if AD converters can be distinguished from neurotypically aging individuals and individuals diagnosed with AD. Supervised classification was used to assess if (1) *a priori*-defined diagnostic groups could be accurately distinguished using a combination of 22 predictive features, and (2) to identify the set of features that best contributed to accurate classification (Figure 4). Specifically, a multi-class supervised random forest was fit to the data to classify participants into one of three *a priori* groups (CN, AD-C, AD) as defined by the OASIS-3 consortium. Predictive features included age, education, sex, nine cognitive performance measures, and nine weighted functional network metrics. Specifically, cognitive performance measures and redundancy-based metrics described in Aim 1 (global communicability and global clustering coefficient) were combined with seven additional network metrics that assess core properties of overall functional network topology: global efficiency, global path transitivity, global eigenvector centrality, modularity, global participation coefficient,

small-world propensity, and within-module degree. These metrics were calculated to capture the overall importance of weight- and path-based functional connections, as well as the degree of integration and segregation between a set of data-driven modules. All network metrics were computed on an individual's unthresholded weighted correlation matrix (positive connections only).

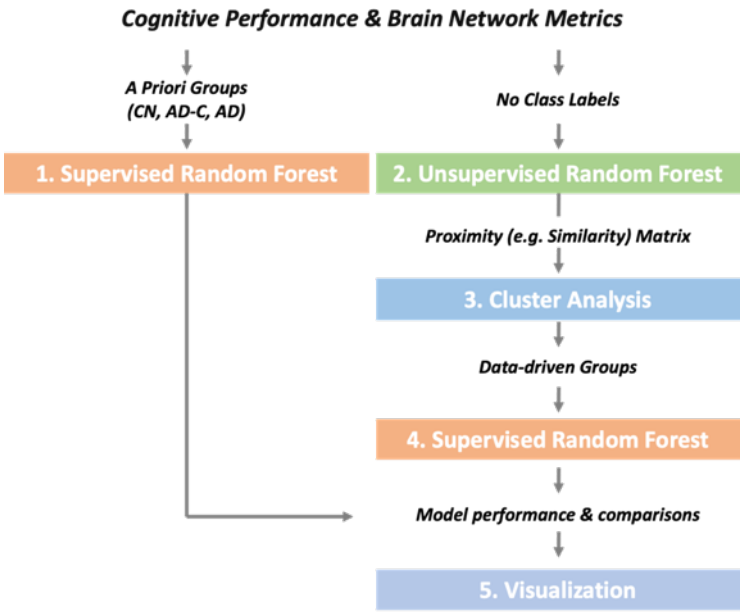


Figure 4. Analysis pipeline. (1) A supervised classification model is first fit to the data to identify features that best distinguish individuals within a priori-defined diagnostic groups. (2) An unsupervised classification model is then fit to the data to generate a proximity matrix. Cluster analysis is then used to identify data-driven subgroups of individuals with similar neurocognitive profiles. Groups (e.g. class labels) identified via cluster analysis are then used to build a final supervised model to obtain classification performance metrics. Models predicting the data-driven subgroups will then be compared to models generated using the diagnostic groups as defined by the OASIS-3 consortium. Abbreviations: CN = Cognitively Normal; AD-C = AD Converters; AD = Alzheimer's disease

Prior to classification modeling, multicollinearity was assessed between all continuous features to ensure model performance estimates and feature importance rankings were not biased by including redundant features. No pair of features exhibited a Pearson's correlation coefficient above .85; thus, no features were excluded from analyses (see the *Discussion* section for potential limitations with this approach). For supervised models, performance metrics were reported in table form. Demographics, neuropsychological performance, and graph-theoretical metrics were visualized using the R: ggplot2 package and reported in table form. In addition to neuropsychological performance and graph-theoretical metrics, variables not used for

classification that inform clinical diagnosis (MMSE scores, CDR ratings) were reported and summarized by group. Supervised classification models were generated with and without controlling for class imbalance using SMOTE (see Methods).

Aim 3. Characterizing data-driven neurocognitive profiles. Unsupervised machine learning was used to identify the presence and composition of data-driven neurocognitive subgroups using the same features described in Aim 2. Specifically, to retain both linear and non-linear interactions between features, an unsupervised random forest model was first applied to the dataset to generate a proximity matrix. Next, cluster analysis was performed on the resulting proximity matrix to identify data-driven neurocognitive subgroups. Finally, to formally assess the validity of the data-driven groups derived via unsupervised learning, a supervised random forest model was used generated using the data-driven clusters. Feature importance and model performance metrics were reported as described in Aim 2.

CHAPTER 5: AIM 1 RESULTS

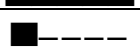
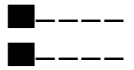
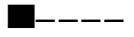
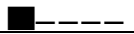
Aim 1 sought to establish if metrics that index redundancy in weighted functional brain networks predicted cognitive performance. Specifically, a series of linear regression analyses tested if, independently, global communicability and global clustering coefficient significantly predicted performance on cognitive measures indexing episodic memory (Logical Memory), working memory (Digit Span Test-Backward), and mental flexibility (Trail Making Test B). Additionally, linear models tested the hypothesis that redundancy-based network metrics would be more positively correlated with cognitive performance in the AD-C and AD groups, with the strongest association observed between the redundancy metrics and Logical Memory performance in the AD group.

Quality Control

Assessment of missing data and variable distributions by a priori group. Scores on neuropsychological assessments (Logical Memory, Digit Span Backward, Trail Making Part B) were non-normally distributed within the CN and AD groups (all Shapiro-Wilk's $p < .05$; see Table 4). In the AD-C group, only Trail-Making Part B scores were non-normally distributed (Shapiro-Wilk's $p = .014$). As hypothesized, Trail-Making Part B scores exhibited a high degree of skew before transformation. The high degree of skew suggested the presence of ceiling effects in the CN and AD-C groups. In the AD group, Trail-Making Part B scores exhibited a bimodal distribution indicating the presence of both floor, and ceiling effects. Assessing redundancy-based graph-theoretical metrics, the global clustering coefficient was normally distributed across

groups (Shapiro-Wilk's $p > .05$; see Table 4). Consistent with previous work, global communicability exhibited a power-law degree distribution (Betzel et al., 2016; Estrada & Hatano, 2008; Langella, Sadiq, et al., 2021; Mišić et al., 2015) and was thus non-normally distributed across groups (all Shapiro-Wilk's $p < .05$) Variable distributions are shown before transformation in Table 4.

Table 4. Summary statistics of cognitive performance and network topology by a priori group.

		M	SD	MED	MIN	MAX	SKEW	KURT	HIST	p [a]
Logical Memory	**CN**	13.69	3.63	14.00	3.00	21.00	-0.35	-0.50		**.004**
	AD-C	11.38	4.36	12.00	0.00	20.00	-0.58	-0.19		.062
	AD	5.36	4.17	5.00	0.00	16.00	0.63	-0.43		**< .001**
Digit-Span Backward	**CN**	6.60	1.94	6.00	3.00	12.00	0.57	-0.03		**< .001**
	AD-C	6.04	1.87	6.00	2.00	11.00	0.28	-0.03		.095
	AD	4.67	2.20	4.00	0.00	11.00	0.50	0.09		**.008**
Trail Making Part B	**CN**	82.16	33.92	75	36.00	300.00	2.31	10.08		.606
	AD-C	133.26	80.58	100.00	52	300.00	1.14	-0.15		**.014**
	AD	184.56	92.75	157.00	47.00	300.00	-0.22	-1.51		**< .001**
Global Communicability[c]	**CN**	2.26+18	2.84+19	1.68+11	2.56+8	3.58+20	12.37	152.03		**< .001**
	AD-C	1.48+13	5.39+13	1.05+11	2.10+8	3.60+14	5.02	27.73		**< .001**
	AD	7.13+13	2.94+14	1.44+11	1.90+8	1.95+15	4.70	22.79		**< .001**
Global Clustering Coefficient	**CN**	.068	.011	.070	.041	.093	-0.02	-0.35		.803
	AD-C	.065	.008	.060	.046	.085	0.31	-0.24		.440
	AD	.065	.009	.070	.043	.097	0.37	1.22		.167

Note: Statistical moments are displayed along with visual representations of raw variable distributions prior to transformation. Abbreviations. AD = Alzheimer's Disease; AD-C = AD Converters; CN = Cognitively Normal; HIST = Histogram; KURT = Kurtosis; M = Mean; MED = Median; MIN = Minimum; SD = Standard Deviation. [a] Significance of Shapiro-Wilk's normality test; [b] Trail Making Part B after random forest missing data imputation; [c] Global Communicability is logarithmically distributed as expected (Mišić et al., 2015).

Data transformation to approximate a normal distribution for linear analyses. The Yeo-Johnson algorithm was selected as the optimal transformation for each of the cognitive outcome variables using the *bestNormalize* R package (Estimated Normality Statistics: Logical Memory = 1.57; Digit Span Backward = 2.57; Trail Making Part B = 1.52). To be consistent with prior work and not overly distort the underlying distribution, global communicability was

log-transformed for linear analyses (Mišić et al., 2015). Trail making part B scores were reverse coded with and without imputation before transformation, to be consistent with the Logical Memory and Digit Span Backward assessments.

Descriptive Statistics & Group Comparisons. *A priori* groups were compared across the neuropsychological assessments and graph theoretical metrics of interest (Figure 5). Groups were compared across measures using a one-way analysis of variance test correcting for multiple comparisons with FDR (Benjamini & Hochberg, 1995). Significant group differences were found on the Logical Memory (CN: M = 13.69, SD = 3.63; AD-C: M = 11.38, SD = 4.36; AD: M = 5.36, SD = 4.17; $F(2, 293) = 121.04$, $p < .001$), Digit Span Backward (CN: M = 6.60, SD = 1.94; AD-C: M = 6.04, SD = 1.94; AD: M = 4.67, SD = 2.20; $F(2, 293) = 25.05$, $p < .001$), and Trail Making Part B (CN: M = 218.83, SD = 33.81; AD-C: M = 166.77, SD = 81.11; AD: M = 103.31, SD = 85.94; $F(2, 293) = 93.57$, $p < .001$) assessments. Exploratory post-hoc comparisons were conducted using an independent t-test (uncorrected). Post-hoc comparisons indicated that CN, AD-C, and AD participants scored significantly different from one another on the Logical Memory and Trail-Making Part B Assessments (CN > AD-C > AD; all $p < .05$). On the Digit Span Backward assessment, AD participants scored lower compared to AD-C ($p < .05$) and CN participants ($p < .05$). No significant differences were bound between CN and AD-C participants on the Digit Span Backward assessment ($p > .05$).

With respect to the graph theoretical metrics assessed in Aim 1, global communicability, log-transformed, was not found to be significantly different between groups (CN: M = 27.25, SD = 4.43; AD-C: M = 25.61, SD = 3.43; AD: M = 25.93, SD = 3.69; $F(2, 293) = 1.60$, $p = 0.203$; 95% CI [0.00, 1.00]). However, global clustering coefficient was significantly different between the CN, AD-C, and AD groups (CN: M = .068, SD = .011; AD-C: M = .065, SD = .008; AD: M

= .065, SD = .009; F(2, 293) = 3.71, p = 0.026; 95% CI [1.64e-03, 1.00]). Post-hoc comparisons indicated that CN participants had significantly higher global clustering coefficient estimates compared to both the CN (p < .05) and AD (p < .05) groups. CN and AD participants did not differ from one another (p > .05). Global communicability (log-transformed) and global clustering coefficient were positively correlated (Whole sample: r = .56, p < .001; CN: r = .58, p < .001; AD-C: r = .46, p < .001 AD: r = .52, p < .001).

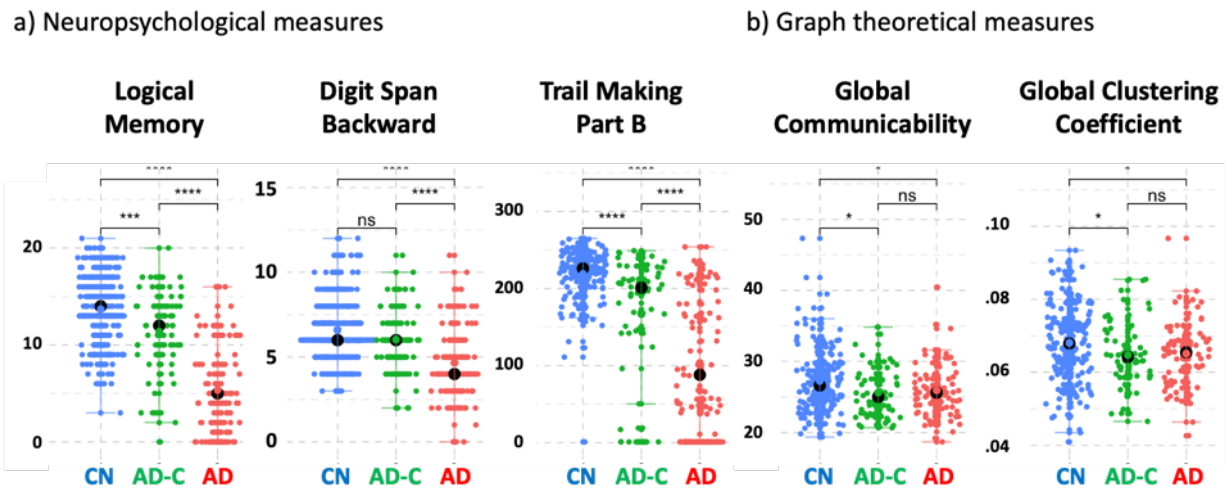


Figure 5. ***Variable distributions & group comparisons.*** a) Performance on the Logical Memory, Digit Span Backward, and Trail Making Part B assessments are displayed by *a priori* group. Logical memory, Digit Span Backward, and Trail Making Part B scores were significantly different across the *a priori* groups (all p-fdr < .05). Post-hoc comparisons (uncorrected) indicated that CN, AD-C, and AD participants scored significantly different from one another on the Logical Memory and Trail-Making Part B Assessments (CN > AD-C > AD; all p < .05). On the Digit Span Backward assessment, AD participants scored lower compared to AD-C (p < .05) and CN participants (p < .05). No significant differences were bound between CN and AD-C participants on the Digit Span Backward assessment (p > .05). b) Estimates of global communicability (log-transformed) and global clustering coefficient are displayed by *a priori* group. Global clustering coefficient (p < .05), but not log-transformed global communicability (p = 0.203) estimates were significantly different across the *a priori* groups. Post-hoc comparisons indicated that CN participants had significantly higher global clustering coefficient estimates compared to both the CN (p < .05) and AD (p < .05) groups. CN and AD participants did not differ from one another (p > .05). *A priori* groups are color-coded (CN = blue; AD-C = green; AD = red). Black spheres indicate group means for each of the neuropsychological assessments and graph theoretical metrics. Abbreviations: AD = Alzheimer's Disease; AD-C = AD Converters; CN = Cognitively Normal. **** = p < .0001; *** = p < .001; ** = p < .01; * = p < .05 (uncorrected); ns = non-significant.

Linear Regression Analyses

Primary analysis. Six independent models were estimated using ordinary least-squares regression, with FDR (Benjamini & Hochberg, 1995), to assess if global communicability (Figure 6, panel a) and/or global clustering coefficient (Figure 6, panel b) were associated with performance on the Logical Memory, Digit Span Backward, and Trail Making Part B assessments. An interaction term was specified within each model to test for a difference in slopes between the CN, AD-C, and AD groups.

Global Communicability

Logical Memory. The model predicting performance on the Logical Memory assessment from global communicability explained a statistically significant proportion of variance (R2 = 0.45, F(5, 290) = 47.95, p-fdr = < .0001, adj. R2 = 0.44). The model's intercept, corresponding to log-transformed global communicability = 0 and group = CN, was equal to 0.05 (95% CI [-0.68, 0.78], t(290) = 0.13, p = 0.894). No significant interaction effects were observed. A statistically significant and negative main effect of group was found for AD participants only (beta = -1.84, 95% CI [-3.21, -0.47], t(290) = -2.64, p = 0.009; Std. beta = -1.53, 95% CI [-1.74, -1.33]). The main effect of log-transformed global communicability was not significant (p = .200). All model effects are displayed in Table 5.

Digit Span Backward. The model predicting performance on the Digit Span Backward assessment from global communicability explained a statistically significant proportion of variance (R2 = 0.16, F(5, 290) = 10.78, p-fdr < .0001, adj. R2 = 0.14). The model's intercept, corresponding to global communicability (log) = 0 and group = CN, was equal to 0.08 (95% CI [-0.83, 0.98], t(290) = 0.17, p = 0.866). No significant interaction or main effects were observed (p > .05). All model effects are displayed in Table 5.

Trail Making Part B. The model predicting performance on the Trail Making Part B assessment from global communicability (log-transformed) explained a statistically significant proportion of variance (R2 = 0.43, F(5, 290) = 43.35, p-fdr < .001, adj. R2 = 0.42).The model's intercept, corresponding to log-transformed global communicability = 0 and group = CN, was equal to 0.62 (95% CI [-0.13, 1.36], t(290) = 1.63, p = 0.105). A significant interaction effect was observed (p = .001), such that the magnitude of association between log-transformed global communicability and Trail Making Part B performance was greater in the AD-C group as compared to both the CN and AD groups (see Table 5). That is, there was a significant positive interaction effect such that log-transformed global communicability in AD-C group predicted better performance on the Trail Making Part B assessment (beta = 0.11, 95% CI [0.05, 0.18], t(290) = 3.36, p = .0001; Std. beta = 0.46, 95% CI [0.19, 0.72]). Main effects of group were statistically significant and negative (AD-C: beta = -3.52, 95% CI [-5.22, -1.82], t(290) = -4.07, p < .001; Std. beta = -0.57, 95% CI [-0.81, -0.32]); (AD: beta = -2.45, 95% CI [-3.86, -1.05], t(290) = -3.44, p < .001; Std. beta = -1.45, 95% CI [-1.66, -1.24]).

Table 5. Model coefficients from regression models predicting cognition from global communicability and global clustering coefficient

a. Global Communicability

	Logical Memory			Digit Span Backward			Trail Making Part B		
	Beta	**95% CI[1]**	**p-value**	**Beta**	**95% CI[1]**	**p-value**	**Beta**	**95% CI[1]**	**p-value**
CN	—	—		—	—		—	—	
AD-C	-0.36	-2.0, 1.3	.700	-0.81	-2.9, 1.2	.400	-3.52	-4.9, -1.3	**<.001**
AD	-1.8	-3.2, -0.47	**.009**	-0.64	-2.3, 1.1	.500	-2.45	-3.3, -0.34	**.016**
COMM	0.02	-0.01, 0.04	.200	0.01	-0.02, 0.04	.600	0	-0.03, 0.02	.800
AD-C * COMM	0	-0.07, 0.06	.900	0.02	-0.06, 0.10	.600	0.11	0.05, 0.18	**<.001**
AD * COMM	0.01	-0.04, 0.06	.700	-0.01	-0.07, 0.05	.800	0.04	-0.02, 0.09	.200

b. Global Clustering Coefficient

	Logical Memory			Digit Span Backward			Trail Making Part B		
	Beta	**95% CI[1]**	**p-value**	**Beta**	**95% CI[1]**	**p-value**	**Beta**	**95% CI[1]**	**p-value**

CN	—	—		—	—		—	—	
AD-C	-0.22	-1.9, 1.4	.800	-0.4	-2.4, 1.6	.700	-1.7	-3.4, -0.05	**.043**
AD	-2.5	-3.9, -1.0	**.001**	-1.1	-2.9, 0.72	.200	-2.9	-4.4, -1.4	**<.001**
CC	1.2	-9.8, 12	.800	-1.4	-15, 12	.800	.55	-11, 12	.900
AD-C * CC	-3.4	-28, 21	.800	2.2	-29, 33	.900	17	-8.9, 42	.200
AD * CC	14	-8.2, 36	.200	2.7	-25, 30	.800	22	-0.93, 44	.060

Note: a) Interaction and main effects are presented for regression models assessing linear relationships between performance on the Logical Memory, Digit Span Backward, and Trail Making Part B assessments with global communicability. b) Interaction and main effects are presented for regression models assessing linear relationships between performance on the Logical Memory, Digit Span Backward, and Trail Making Part B assessments with global clustering coefficient. Abbreviations: AD = Alzheimer's Disease; AD-C = AD Converters; CC = Global Clustering Coefficient; CI = Confidence Interval; CN = Cognitively Normal; COMM = Global Communicability. Bold values indicate a statistically significant effect at $p < .05$.

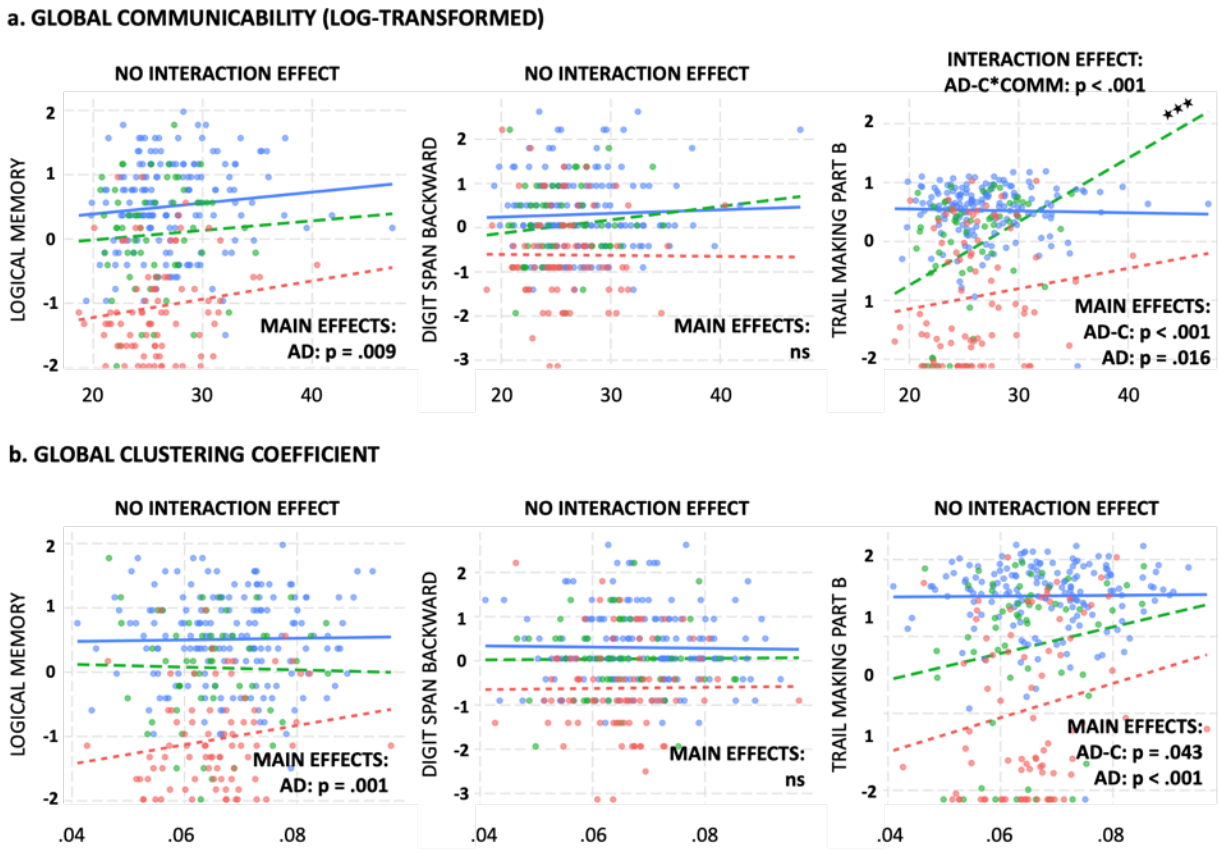


Figure 6. Interaction plots assessing the influence of *a priori* group membership on the relationship between redundancy-based metrics and cognition. Interaction plots are presented for regression models assessing linear relationships between performance on the Logical Memory, Digit Span Backward, and Trail Making Part B assessments with log-transformed global communicability (top row: panels a, b, and c) and global clustering coefficient (bottom row: panels d, e, and f). A significant interaction effect ($p < .001$) was

observed when predicting Trail Making Part B performance from log-transformed global communicability by group. Specifically, higher levels of global communicability predicted better Trail Making Part B performance for AD-C participants. Main effects of group indicated that AD-C ($p < .001$) and AD ($p = .016$) participant had performed lower on average compared to CN participants. No other interaction effects were observed, nor were there main effects of either global communicability or global clustering coefficient for any of the cognitive assessments (all $p < .05$). Additional main effects were observed on the Logical Memory and Trail Making Part B assessments. Specifically, AD participants scored significantly lower on both the Logical Memory and Trail Making Part B assessments. Abbreviations: AD = Alzheimer's Disease; AD-C = AD Converters; CC = Global Clustering Coefficient; CN = Cognitively Normal; COMM = Global Communicability. *** = $p < .001$; ns = non-significant.

Global Clustering Coefficient

Logical Memory. The model predicting performance on the Logical Memory assessment from global clustering coefficient explained a statistically significant proportion of variance ($R2 = 0.45$, $F(5, 290) = 47.61$, p-fdr = < .0001, adj. $R2 = 0.44$). The model's intercept, corresponding to global clustering coefficient = 0 and group = CN, was equal to 0.43 (95% CI [-0.34, 1.19], $t(290) = 1.10$, $p = 0.272$). No significant interaction effects were observed ($p > .05$). A statistically significant and negative main effect of group was found for AD participants only (beta = -2.46, 95% CI [-3.91, -1.00], $t(290) = -3.31$, $p = .001$; Std. beta = -1.54, 95% CI [-1.74, -1.34]). All model effects are displayed in Table 5.

Digit Span Backward. The model predicting Digit Span Backward performance from global clustering coefficient explained a statistically significant proportion of variance ($R2 = 0.15$, $F(5, 290) = 10.57$, p-fdr = < .0001, adj. $R2 = 0.14$). The model's intercept, corresponding to group = CN and global clustering coefficient = 0, was equal to 0.39 (95% CI [-0.55, 1.34], $t(290) = 0.82$, $p = 0.415$). No significant interaction or main effects were observed ($p > .05$). All model effects are displayed in Table 5.

Trail Making Part B. The model predicting performance on the Trail Making Part B assessment from global clustering coefficient explained a statistically significant proportion of

variance (R2 = 0.41, F(5, 290) = 40.70, p-fdr = < .0001, adj. R2 = 0.40). The model's intercept, corresponding to group = CN and global clustering coefficient = 0, was equal to 0.49 (95% CI [-0.30, 1.28], t(290) = 1.22, p = 0.222). No significant interaction effects were observed (p > .05). A statistically significant and negative main effect of group was found for both AD-C (beta = -1.75, 95% CI [-3.44, -0.05], t(290) = -2.03, p = 0.043; Std. beta = -0.63, 95% CI [-0.87, -0.39]) and AD participants (beta = -2.89, 95% CI [-4.40, -1.38], t(290) = -3.76, p < .001; Std. beta = -1.44, 95% CI [-1.65, -1.23]). All model effects are displayed in Table 5.

Post-Hoc Exploratory Analyses

Trail Making Part B. A significant group by global communicability (log-transformed) interaction effect was observed for the model predicting Trail Making Part B performance. To ensure the significant interaction effect was not due to the use of imputation or transformation, additional regression models were run using only observations with Trail Making Part B data, before and after transformation. Here, the interaction between group status and global communicability in the model predicting Trail Making Part B performance remained significant when constructed on data without missing data imputation, and both before and after transformation (before transformation: R2 = 0.35, F(5, 269) = 29.31, p < .001, adj. R2 = 0.34; after transformation: R2 = 0.36, F(5, 269) = 30.00, p < .001, adj. R2 = 0.35). Model intercepts, corresponding to global communicability = 0 and group = CN, were 0.52 (95% CI [-0.26, 1.30], t(290) = 1.31, p = 0.190) before transformation, and at 0.54 (95% CI [-0.25, 1.33], t(269) = 1.34, p = 0.183) after transformation. All model effects are shown in Table 6.

Table 6. Coefficients from the post-hoc regression models predicting Trail Making Part B performance from global communicability and global clustering coefficient

	No NA			No NA + Transformed		
	Beta	**95% CI**	**p-value**	**Beta**	**95% CI**	**p-value**

CN	—	—	—	—	—	—
AD-C	-279	-416, -141	**<.001**	-3.6	-5.4, -1.8	**<.001**
AD	-194	-326, -62	**.004**	-2.4	-4.1, -0.65	**.007**
COMM	-0.30	-2.5, 1.9	.800	0	-0.03, 0.03	.800
AD-C * COMM	8.90	3.6, 14	**.001**	0.11	0.04, 0.18	**.001**
AD * COMM	3.50	-1.5, 8.5	.200	0.04	-0.03, 0.11	.200

Note: Interaction and main effects are presented for regression models assessing linear relationships between performance on the Trail Making Part B assessments with global communicability after removing observations with missing data. Coefficients are shown for the models conducted without missing data imputation, before and after transformation. Abbreviations: AD = Alzheimer's Disease; AD-C = AD Converters; CI = Confidence Interval; CC = Global Clustering Coefficient; CN = Cognitively Normal; COMM = Global Communicability. Bold values indicate a statistically significant effect at $p < .05$.

Additionally, as stated in the *Research Plan* section, diagnostics were also inspected for the model with a significant interaction term. Although transformation overall resulted in improved residual diagnostics, the transformation may still not be accounting for floor effects observed in the Trail Making Part B assessment, as indicated by the vertical grouping of data points in the residuals vs fitted plot (Figure 7). The curvature in the residuals may also indicate the presence of non-linear relationships between Trail Making Part B performance and global communicability. Thus, a censored regression model was next fit to the data to assess if the significant positive relationship between global communicability in the AD-C group was robust to floor effects present in the data.

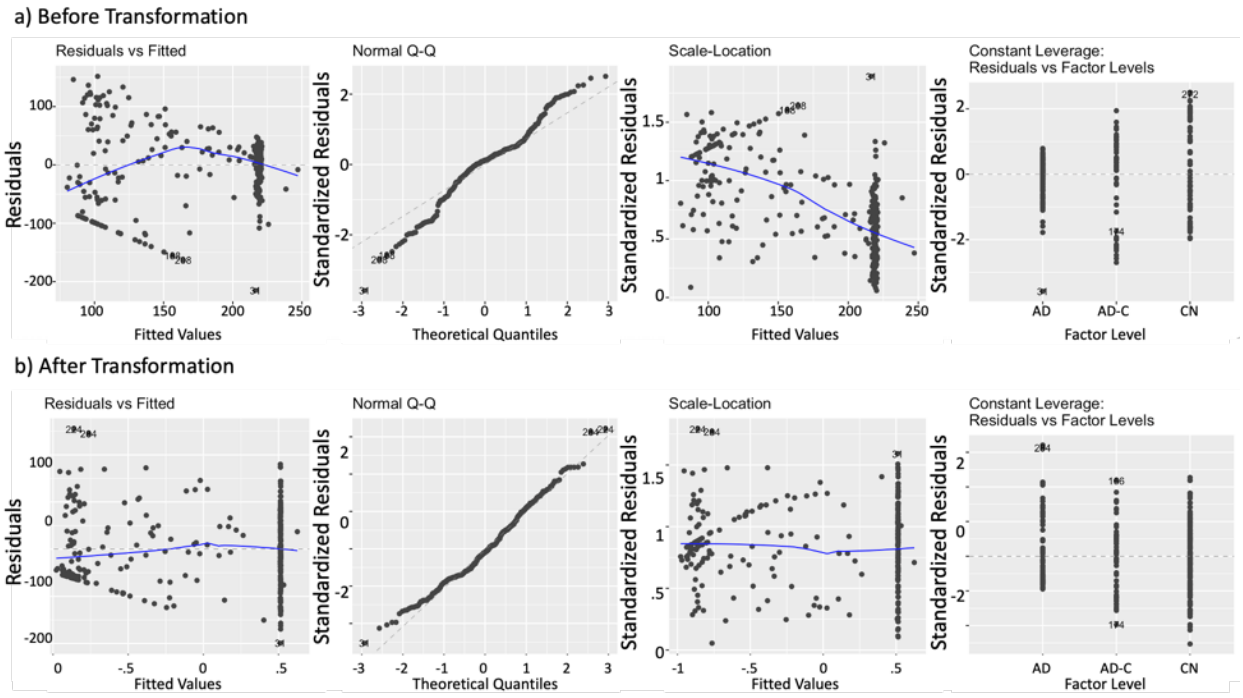


Figure 7. Diagnostic plots for the regression model predicting Trail Making Part B performance from global communicability before and after transformation. Model diagnostics were inspected to assess the robustness of the interaction effect. a) Prior to transformation, diagnostic plots indicate the presence of non-normally distributed residuals. b), after transformation, diagnostic plots indicate a relatively acceptable degree of residual normality. However, the model may be biased by the presence of floor effects in Trail Making Part B scores.

Trail Making Part B Censored Regression Model. The censored regression model predicting performance on the Trail Making Part B assessment from global communicability (log-transformed) explained a statistically significant proportion of variance (Pseudo R2 = 0.24, Wald-statistic (5, 290) = 117.5, p-fdr = < .001).The model's intercept, corresponding to log-transformed global communicability = 0 and group = CN, was equal to 0.56 (95% CI [0.14, 0.98], t(290) = 2.64, p = 0.008). A significant interaction effect was observed (p = .01), such that the magnitude of association between log-transformed global communicability and Trail Making Part B performance was greater in the AD-C group as compared to both the CN and AD groups (see Table 7). That is, there was a significant positive interaction effect with log-transformed global communicability for the AD-C group only (beta = 0.05, 95% CI [0.11, 0.09], z(290) =

2.55, p = .01). A statistically significant and negative main effect was observed for the AD-C group only (AD-C: beta = -1.62, 95% CI [-2.63, -0.62], z(290) = -3.16, p = .001).

Table 7. Coefficients from the post-hoc censored regression model predicting Trail Making Part B performance from global communicability.

	Trail Making Part B		
	Beta	**95% CI**	**p-value**
CN	—	—	
AD-C	-1.62	-2.63, -0.62	**.002**
AD	-0.77	-1,72, 0.18	.113
COMM	0	-0.02, 0.15	.940
AD-C * COMM	0.05	0.01, 0.09	**.011**
AD * COMM	0.001	-0.34, 0.38	.924

Note: a) Interaction and main effects for the censored regression model. After accounting for floor effects, the interaction between Trail Making Part B performance and global communicability remained statistically significant for the AD-C group (p = .011). A main effect of group for AD-C participants was also observed such that AD-C participants exhibited lower levels of global communicability compared to both CN and AD participants (p = .002). No main effect was observed for AD participants after accounting for floor effects (p = .113). Abbreviations: AD = Alzheimer's Disease; AD-C = AD Converters; CI = Confidence Interval; CN = Cognitively Normal; COMM = Global Communicability. Bold values indicate a statistically significant effect at p < .05.

CHAPTER 6: AIM 2 RESULTS

Aim 2 sought to establish whether a combination of core demographic risk factors (age, sex, and education), cognitive measures, and weighted functional network metrics could accurately distinguish CN, AD-C, and AD individuals using supervised machine learning. A primary goal of Aim 2 was to understand which participants may be at a greater risk of converting to AD using baseline assessments.

Quality Control

Before analyses, multicollinearity was assessed between the set of 3 demographic factors, 9 metrics assessing cognitive performance, and 9 graph-theoretical metrics. Multicollinearity was assessed across the whole sample, as well as for each *a priori* group separately (Figure 8). No pair of features exhibited a positive correlation of .85 or greater (see Methods) when assessed across the whole sample, or within *a priori* groups. Thus no features were excluded prior to analysis. Descriptive statistics for all features used in the supervised and unsupervised learning models, as well as MMSE scores, and CDR-SB ratings are provided in Table 8 and Figure 9.

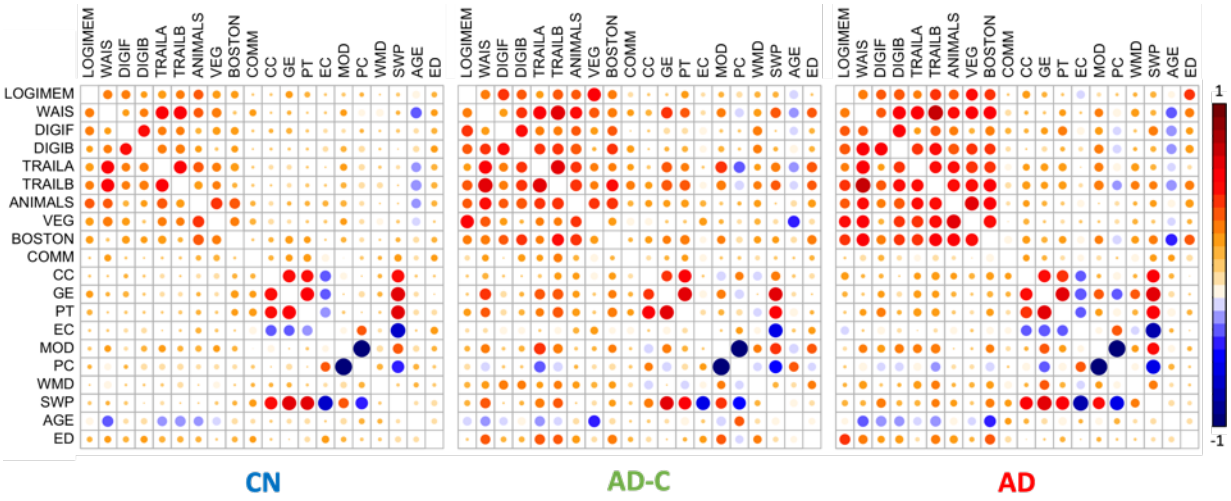


Figure 8. Feature correlation plots by *a priori* group. For each of the *a priori* defined groups (CN, AD-C, AD), pairwise correlations are shown for the set of 3 demographic factors, 9 metrics assessing cognitive performance, and 9 graph-theoretical metrics used for supervised and unsupervised classification. Warmer colors signify features that are more positively correlated with one another. Cooler colors signify features that are more negatively correlated with one another. The size of a circle indicates the strength of correlation between features. Abbreviations: AD = Alzheimer's disease; AD-C = AD Converters; CN = Cognitively Normal; DIGIF = Digit Span Forward; DIGIB = Digit Span Backward; CC = Clustering Coefficient; COMM = Communicability; EC = Eigenvector Centrality; ED = Education; GE = Global Efficiency; LOGIMEM = Logical Memory; MOD = Modularity; PC = Participation Coefficient; PT = Path Transitivity; SWP = Small-World Propensity; TRAILA = Trail Making Test Part A; TRAILB = Trail Making Test Part B; VEG = Vegetable Naming (Category Fluency); WAIS = Weschler Adult Intelligence Scale; WMD = Within-Module Degree Z-Score.

Descriptive Statistics

Features used for supervised and unsupervised learning, and measures important for diagnostic classification (MMSE and CDR-SB) not used for prediction, were compared across the *a priori* groups (CN, AD-C, AD) using a one-way analysis of variance test corrected for multiple comparisons with FDR (Benjamini & Hochberg, 1995). Groups were significantly different on age (CN: M = 71, SD = 6; AD-C: M = 79, SD = 8; AD: M = 76, SD = 8; $F(2, 293) = 31.07$, p-fdr < .001; Eta2 = 0.17, 95% CI [0.11, 1.00]) and education level (CN: M = 9.95, SD = 2.58; AD-C: M = 9.64, SD = 2.36; AD: M = 8.47, SD = 3.24; $F(2, 293) = 7.96$, p-fdr < .001; Eta2 = 0.05, 95% CI [0.02, 1.00]). Sex was not significantly different between groups (CN:

Female = 53%, Male = 47%; AD-C: Female = 52%, Male = 48%; AD: Female = 59%, Male = 41%; $x^2(2) = 7.96$, p-fdr = .600). The CN, AD-C, and AD groups performed significantly different on the battery of neuropsychological assessments (all p-fdr < .001).

Concerning graph-theoretical metrics, CN, AD-C, and AD participants significantly differed on global clustering coefficient (CN: M = .068, SD = .011; AD-C: M = .065, SD = .008; AD: M = .065, SD = .009; F(2, 293) = 3.71, p = 0.026; 95% CI [1.64e-03, 1.00]), global efficiency (CN: M = .182, SD = .014; AD-C: M = .175, SD = .012; AD: M = .176, SD = .012; F(2, 293) = 6.55, p-fdr = .002; Eta2 = 0.04, 95% CI [0.01, 1.00]), global path transitivity (CN: M = .697, SD = .020; AD-C: M = .689, SD = .019; AD: M = .690, SD = .019; F(2, 293) = 4.95, p-fdr = 0.010; Eta2 = 0.03, 95% CI [5.10e-03, 1.00]), global modularity (CN: M = .090, SD = .021; AD-C: M = .096, SD = .025; AD: M = .085, SD = .019; F(2, 293) = 4.98, p-fdr = 0.01; Eta2 = 0.03, 95% CI [5.20e-03, 1.00]), and global participation coefficient estimates (CN: M = .838, SD = .014; AD-C: M = .835, SD = .015; AD: M = .842, SD = .012; F(2, 293) = 5.65, p-fdr = .006; Eta2 = 0.04, 95% CI [7.34e-03, 1.00]). *A priori* groups did not significantly differ on global communicability (CN: M = 27.25, SD = 4.43; AD-C: M = 25.61, SD = 3.43; AD: M = 25.93, SD = 3.69; F(2, 293) = 1.60, p-fdr = 0.203; 95% CI [0.00, 1.00]), global eigenvector centrality (CN: M = 0.050, SD = 0.003; AD-C: M = 0.0523, SD = 0.003; AD: M = 0.053, SD = 0.003; F(2, 293) = 1.03, p-fdr = 0.400, Eta2 = 6.95e-03, 95% CI [0.00, 1.00]), within-module degree z-score (CN: M = -0.05, SD = 0.99; AD-C: M = 0.14, SD = 0.87; AD: M = 0.00, SD = 1.09; F(2, 293) = 0.79, p-fdr = 0.500, Eta2 = 5.33e-03, 95% CI [0.00, 1.00]), or small-world propensity (CN: M = 0.60, SD = 0.05; AD-C: M = 0.61, SD = 0.05; AD: M = 0.59, SD = 00.05; F(2, 293) = 2.86, p-fdr = 0.071, Eta2 = 0.02, 95% CI [0.00, 1.00]).

Exploratory post-hoc comparisons were conducted using independent t-tests (uncorrected). Several significant findings were observed. AD-C participants were older on average (Mean age = 79) compared to both the CN (Mean age = 71; p-fdr < .001) and AD (Mean age = 76; p-fdr = .009) groups. The AD group had the lowest education levels compared to the CN (p-fdr < .001) and AD-C (p-fdr = .009) groups. In addition to demographic factors, CN and AD participants were found to differ on all neuropsychological assessments (all p-fdr < .001) and graph-theoretical metrics (p < .05) except for global communicability, global eigenvector centrality, and within-module degree z-score (p-fdr > .05). Overall, AD-C scored lower than CN, but higher than AD on the battery of neuropsychological assessments (CN > AD-C > AD-C; Table 8; Figure 8). AD-C differed significantly from CN and AD on 7 of 9 neuropsychological assessments (Logical Memory, WAIS, Trail Making Part A, Trail Making Part B, Animal Naming, Vegetable Naming, Boston Naming Test; all p-fdr < .001). AD-C did not differ from either the CN or AD groups on the Digit Span Forward (DIGIF; p-fdr = .400) or Digit Span Backward (DIGIB; p-fdr = .072) assessments.

Several patterns were observed when comparing network topology across the CN, AD-C, and AD *a priori* groups. AD-C participants had lower global clustering coefficient (p = .03), global efficiency (p = .003), and path transitivity (p = .008) compared to CN, and lower estimates of global participation coefficient as compared to AD participants (p = 001). Conversely, compared to AD participants only, AD-C participants had higher estimates of modularity (p-fdr = 002) and small-world propensity (p-fdr = 038). No significant differences were observed when comparing *a priori* groups on global communicability, global eigenvector centrality, or within-module degree z-score (all p-fdr > .05). MMSE scores were found to be

significantly lower (CN > AD-C > AD, and CDR-SB ratings significantly higher (CN < AD-C < AD), in the patient groups (all p-fdr = < .001).

Table 8. General statistical comparison of features across *a priori* groups.

Predictive Features	CN N = 159[1]	AD-C N = 56[1]	AD N = 81[1]	STAT[2]	p-fdr[3]	CN vs. AD-C	CN vs. AD	AD-C vs. AD
Age	71 (6)	79 (8)	76 (8)	31	**<.001**	**<.001**	**<.001**	**.009**
Education	9.95 (2.58)	9.64 (2.36)	8.47 (3.24)	8	**<.001**	.500	**<.001**	**.014**
Sex	--	--	--	1	.600	.800	.400	.400
Female	85 (53%)	29 (52%)	48 (59%)	--	--	--	--	--
Male	74 (47%)	27 (48%)	33 (41%)	--	--	--	--	--
Logical Memory	13.7 (3.6)	11.4 (4.4)	5.4 (4.2)	121	**<.001**	**<.001**	**<.001**	**<.001**
WAIS	58 (11)	45 (13)	34 (15)	97	**<.001**	**<.001**	**<.001**	**<.001**
Digit Span Forward	8.52 (2.06)	8.29 (1.69)	7.27 (1.97)	11	**<.001**	.400	**<.001**	**.002**
Digit Span Backward	6.60 (1.94)	6.04 (1.87)	4.67 (2.20)	25	**<.001**	.072	**<.001**	**<.001**
Trail Making Part A	149 (11)	137 (25)	120 (41)	36	**<.001**	**.002**	**<.001**	**<.001**
Trail Making Part B	219 (34)	167 (81)	103 (86)	94	**<.001**	**<.001**	**<.001**	**<.001**
Animal Naming	21 (5)	17 (5)	12 (5)	86	**<.001**	**<.001**	**<.001**	**<.001**
Vegetable Naming	14.8 (4.1)	11.7 (4.1)	8.3 (3.9)	69	**<.001**	**<.001**	**<.001**	**<.001**
Boston Naming Test	27.7 (2.4)	26.2 (3.8)	21.3 (5.8)	76	**<.001**	**.013**	**<.001**	**<.001**
Global Communicability	26.9 (4.6)	25.8 (3.5)	26.5 (3.5)	1.6	.200	.076	.500	.300
Global Clustering Coefficient	0.068 (0.011)	0.065 (0.009)	0.065 (0.009)	3.7	**.033**	**.030**	**.028**	.800
Global Efficiency	0.182 (0.014)	0.175 (0.012)	0.176 (0.012)	6.5	**.002**	**.002**	**.006**	.600
Global Path Transitivity	0.697 (0.020)	0.689 (0.019)	0.690 (0.019)	4.9	**.010**	**.008**	**.018**	.600
Global Eigenvector Centrality	0.05 (0.003)	0.0523 (0.003)	0.053 (0.003)	1	.400	.500	.300	.200
Global Modularity	0.090 (0.021)	0.096 (0.025)	0.085 (0.019)	5	**.010**	.067	.058	**.002**
Global Participation Coefficient	0.838 (0.014)	0.835 (0.015)	0.842 (0.012)	5.7	**.006**	.081	**.027**	**.001**
Within-Module Degree Z-Score	-0.05 (0.99)	0.14 (0.87)	0.00 (1.09)	0.79	.500	.200	.700	.400
Small-World Propensity	0.60 (0.05)	0.61 (0.05)	0.59 (0.05)	2.9	.071	.600	**.037**	**.038**
MMSE†	29.1 (1.2)	27.0 (3.7)	23.7 (4.1)	98	**<.001**	**<.001**	**<.001**	**<.001**
CDR-SB†	0.02 (0.09)	1.32 (1.88)	4.27 (2.64)	189	**<.001**	**<.001**	**<.001**	**<.001**

Note: Average performance on neuropsychological assessments and estimates of graph-theoretical metrics are compared across the *a priori* groups. CN and AD participants differed on all neuropsychological assessments and graph-theoretical metrics except for global communicability (COMM), global eigenvector centrality (EC), and within-module degree z-score (WMD) (all p < .05). No differences were observed across groups for COMM, EC, or WMD. AD-C differed from CN and AD on all neuropsychological assessments except Digit Span Forward (DIGIF) and Digit Span Backward (DIGIB). AD-C participants had lower

estimates of global clustering coefficient (CC), global efficiency (GE), and path transitivity (PT) compared to CN. AD-C participants had higher estimates of modularity (MOD) and small-world propensity (SWP), and lower estimates of global participation coefficient (PC) compared to AD participants. MMSE scores were significantly lower, and CDR-SB ratings significantly higher in the patient groups. All significance levels are set to an uncorrected alpha of .05 for visual comparisons. Abbreviations: AD = Alzheimer's disease; AD-C = AD Converters; CDR-SB = Clinical Dementia Rating Scale Sum of Boxes ; CN = Cognitively Normal; MMSE = Mini-mental state examination; STAT = Test Statistic. *[1]Mean (SD); n (%); [2] One-way ANOVA; Pearson's Chi-squared test; [3]False discovery rate correction for multiple testing; † Not used in any prediction models. Bold values indicate a statistically significant at p < .05*

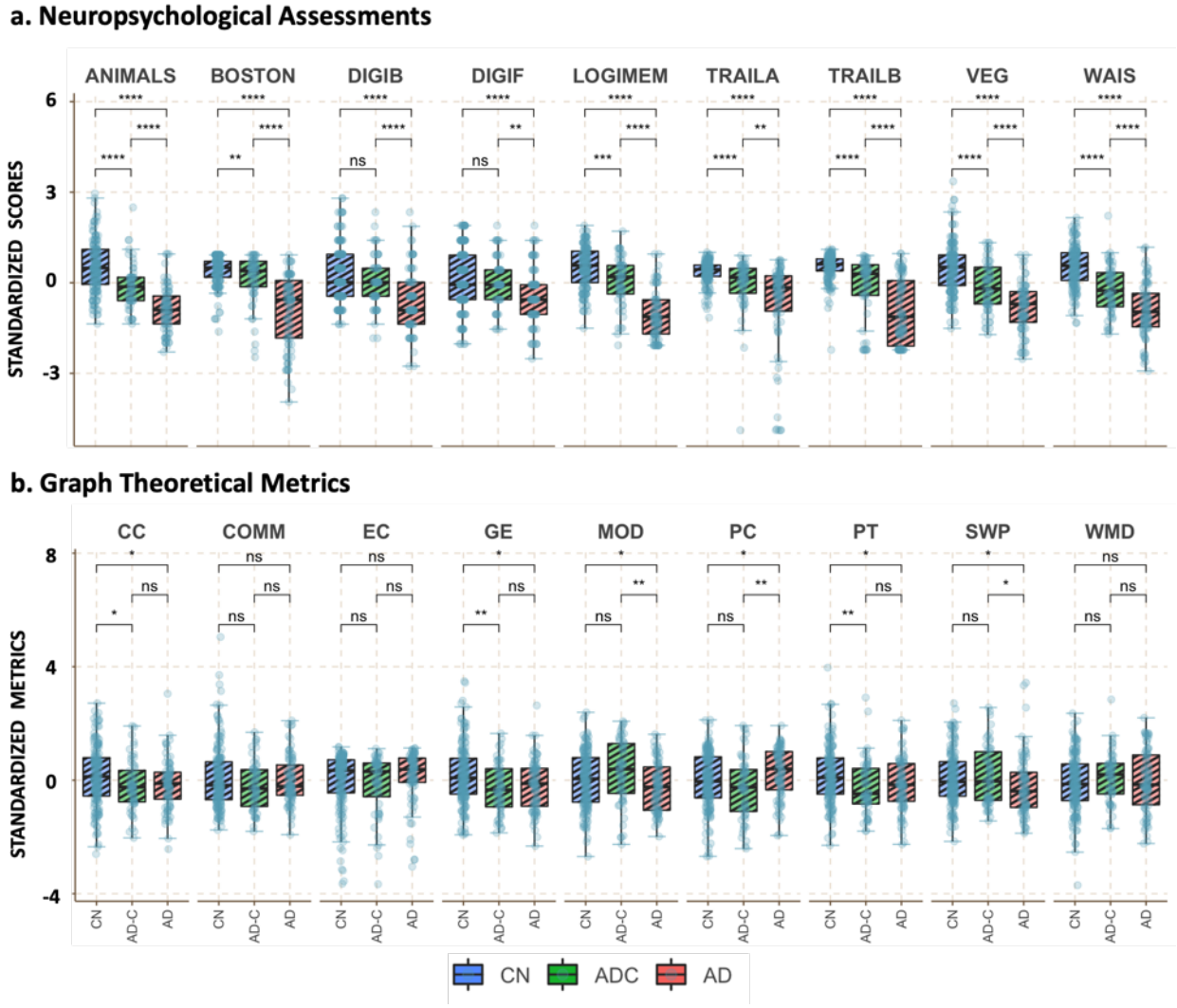


Figure 9. General statistical comparison of features across *a priori* groups. Average performance on neuropsychological assessments and estimates of graph theoretical metrics are compared across the *a priori* groups. CN and AD participants differed on all neuropsychological assessments ($p < .05$) and graph theoretical metrics except for global communicability (COMM), global eigenvector centrality (EC), and within-module degree z-score (WMD). No differences were observed across groups for COMM, EC, or WMD. AD-C differed from CN and AD on all

neuropsychological assessments except Digit Span Forward (DIGIF) and Digit Span Backward (DIGIB). AD-C participants had lower global clustering coefficient (CC), global efficiency (GE), and path transitivity (PT) compared to CN. AD-C participants had higher estimates of modularity (MOD) and small-world propensity (SWP), and lower estimates of global participation coefficient (PC) compared to AD participants. All significance levels are set to an uncorrected alpha of .05 for visual comparisons. For descriptive statistics and corrected p-values, see Table 8. All variables have been centered and scaled for visualization purposes. Colors represent *a priori* groups an CN (blue), AD-C (green), AD (red). Abbreviations: AD = Alzheimer's disease; AD-C = AD Converters; CN = Cognitively Normal; DIGIF = Digit Span Forward; DIGIB = Digit Span Backward; CC = Clustering Coefficient; COMM = Communicability; EC = Eigenvector Centrality; ED = Education; GE = Global Efficiency; LOGIMEM = Logical Memory; MOD = Modularity; PC = Participation Coefficient; PT = Path Transitivity; SWP = Small-World Propensity; TRAILA = Trail Making Test Part A; TRAILB = Trail Making Test Part B; VEG = Vegetable Naming (Category Fluency); WAIS = Weschler Adult Intelligence Scale; WMD = Within-Module Degree Z-Score. **** = $p < .0001$; *** = $p < .001$; ** = $p < .01$; * = $p < .05$ (uncorrected); ns = non-significant.

Supervised Learning To Classify Participants Into A Priori Groups

Model 1: Prediction of a priori group status without subsampling. A supervised random forest model containing all participants comprised of 3 demographic factors (age, education, and sex), 9 metrics assessing cognitive performance (Logical Memory: total elements correctly recalled; WAIS-R Digit Symbol Test: total number of symbols correctly matched; Digit Span Forward: number of digits recalled in sequential order; Digit Span Backward: number of digits recalled in reverse order; Trail Making Test Parts A & B: total number of errors; Category Fluency: total animals and vegetables named; Boston Naming Test: total objects correctly identified), and 9 graph-theoretical metrics (global communicability, global clustering coefficient, global efficiency, global path transitivity, global eigenvector centrality, global modularity, global participation coefficient, small-world propensity, within-module degree z-score) was used to classify participants into *a priori* groups (CN: N = 159; AD-C: N = 56, and AD: N = 81).

Average cross-validated model accuracy was 74.4%, with an AUC of .87, sensitivity of .63, specificity of .84, kappa of .54, F1 score of .63, and a balanced accuracy of 73.4%.

Permutation testing with 1000 iterations indicated that the final model was significant compared to random chance (out-of-bag error: 0.253; Random model out-of-bag error: 0.507; permutation $p < .0001$). Model 1 classified 220 of 296 participants correctly: 147 of 159 CN (92.45%), 10 of 56 AD-C (17.86%), and 63 of 81 AD (77.78%). CN participants were equally likely to be misclassified as either AD-C (N=5) or AD (N=7). Of the 18 AD participants misclassified, 16 were misclassified as CN, and two were misclassified as AD-C. Here, AD-C participants were more likely to be misclassified as CN (N = 32) or AD (N = 14).

The 5 highest-ranked predictors included performance on the Logical Memory (21.66%), WAIS (19.10%), Trail Making Part B (15.34%), and animal (18.12%) and vegetable (14.50%) naming (Category Fluency) assessments (Figure 10, panel a). As a multiclass classification problem, in addition to overall predictive performance, features are ranked by the degree to which they can distinguish a specific class. Assessing class-specific feature importance, features were best able to discriminate CN and AD participants (Average feature importance: CN = 6.36%; AD = 4.93%). However, consistent with the low classification accuracy for AD-C participants (Accuracy = 17.86%), most feature importance scores were near or below 0. Thus, most features were not useful when attempting to distinguish AD-C from CN and/ AD participants (Average feature importance = 0.74%). The top-ranked predictors, specific to the AD-C group, were age (9.63 %), modularity (6.60%), vegetable naming (VEG; 2.58%), and global path transitivity (PT; 2.53%) most contributed to accurate classification of AD-C (N=10) participants.

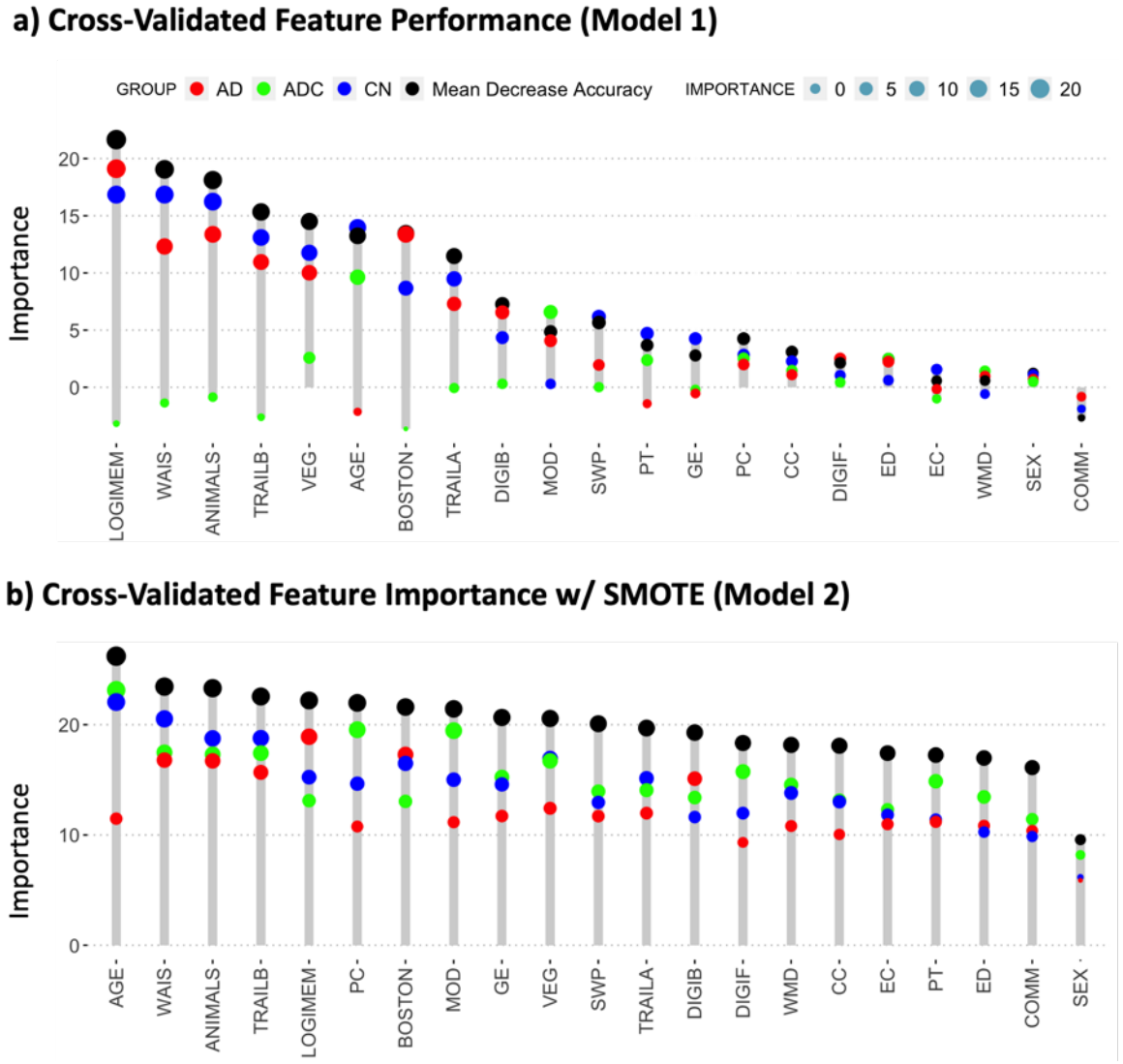


Figure 10. Feature importance with and without subsampling. Feature importance weights are displayed for the supervised random forest models classifying *a priori* groups, before and after SMOTE subsampling. Feature importance is displayed here as the expected drop in accuracy should that feature be removed from the respective model. The mean decrease in accuracy (black) represents the decrease in classification accuracy averaged across classes. Scores are scaled by their standard deviation and thus can be negative. Class-specific feature importance weights are color-coded by *a priori* classes (CN = blue; AD-C = green, AD = red). a) Features are shown for Model 1 without correcting for the imbalance of observations between *a priori* groups. All features, except global communicability (COMM), had average importance scores above 0, and thus contributed to the classification of *a priori* groups. b) Feature importance is shown for Model 2. In Model 2, SMOTE subsampling was applied to each training sample during cross-validation. Adjusting for class imbalance with SMOTE, features were able to distinguish between all three classes with feature importance scores greater than 0.Age was the top-performing feature, and both global participation coefficient (PC) and modularity (MOD) were weighted more heavily. Abbreviations: AD = Alzheimer's disease; AD-C = AD Converters; CN = Cognitively Normal; DIGIF = Digit Span Forward; DIGIB = Digit Span

Backward; CC = Clustering Coefficient; COMM = Communicability; EC = Eigenvector Centrality; ED = Education; GE = Global Efficiency; LOGIMEM = Logical Memory; MOD = Modularity; PC = Participation Coefficient; PT = Path Transitivity; SWP = Small-World Propensity; TRAILA = Trail Making Test Part A; TRAILB = Trail Making Test Part B; VEG = Vegetable Naming (Category Fluency); WAIS = Weschler Adult Intelligence Scale; WMD = Within-Module Degree Z-Score.

Assessing characteristics of AD-C participants who were classified correctly to those misclassified as CN or AD. After supervised random forest modeling, AD-C participants who were correctly classified (AD-C) were compared to those who were misclassified as CN (AD-C–CN) or AD (AD-C–AD) on demographic, neuropsychological, and graph-theoretical metrics used for classification, as well as the MMSE and CDR-SB. Omnibus tests were conducted using a one-way analysis of variance corrected for multiple comparisons using the false-discovery rate (Benjamini & Hochberg, 1995). AD-C participants correctly classified or misclassified as either CN or AD differed significantly on age and across the battery of neuropsychological assessments (all p-fdr < .05; Table 9). AD-C participants correctly classified or misclassified as either CN (AD-C–CN) or AD (AD-C–AD) did not significantly differ on education (AD-C–CN: M = 10.09; SD = 2.22; AD-C: M = 9.00; SD = 1.63; AD-C–AD: M = 9.07; SD = 2.97; F(2, 53) = 1.38, p-fdr = 0.400; Eta2 = 0.05, 95% CI [0.00, 1.00]) or sex (AD-C–CN: Female = 56%; Male = 44%; AD-C: Female = 50%; Male = 50%; AD-C–AD: Female = 43%; Male = 57%; $x^2(2)$ = 0.72, p-fdr = .900).

Exploratory follow-up comparisons were conducted using pairwise t-tests (uncorrected). Compared to AD-C participants who were misclassified as CN (AD-C–CN), participants correctly classified (AD-C) were on average older (AD-C: M = 83; SD = 6; AD-C–CN: M = 77; SD = 8; t(19.81) = 2.75, p = 0.019), and scored lower on three out of nine neuropsychological assessments (Table 9; Figure 11, panel a). Specifically, AD-C participants who were correctly classified had poorer attention, executive functioning, and semantic fluency as assessed by the

WAIS (AD-C: M = 36; SD = 8; AD-C–CN: M = 53; SD = 11; t(19.88) = 5.02, p < .001), Trail Making Part B (AD-C: M = 144; SD = 72; AD-C–CN: M = 210; SD = 45; t(19.88) = 5.02, p = .005), and animal naming (Category Fluency) (AD-C: M = 15.1; SD = 3.0; AD-C–CN: M = 19.5; SD = 4.6; t(23.48) = 3.54, p = .004) assessments.

Compared to those misclassified as CN (AD-C–CN), AD-C participants misclassified as AD (AD-C–AD) were older (AD-C–AD: M = 82; SD = 7; AD-C–CN: M = 77; SD = 8; t(28.28) = 2.23, p = .033) and performed significantly worse on all neuropsychological assessments (all p < .05; Table 9). Additionally, compared to correctly classified AD-C participants, AD-C participants misclassified as AD (AD-C–AD) performed worse on the Logical Memory (AD-C: M = 11.8; SD = 4.0; AD-C–AD: M = 6.9; SD = 4.4; t(20.76) = -2.82; p < .002), Trail Making Part B (AD-C: M = 144; SD = 72; AD-C–AD: M = 85; SD = 86; t(21.35) = -1.80; p < .028), vegetable naming (Category Fluency) (AD-C: M = 12.2; SD = 3.1 AD-C–AD: M = 8.7; SD = 4.1; t(21.89) = -2.38; p < .032), and Boston Naming (AD-C: M = 27.1; SD = 3.3; AD-C–AD: M = 22.4; SD = 5.1; t(21.87) = -2.76; p < .001) assessments.

No omnibus group differences were observed for any of the graph-theoretical metrics assessed (all p-fdr > .05; (Table 9; Figure 11, panel b). In exploratory post-hoc comparisons, compared to AD-C misclassified as CN (AD-C–CN), correctly classified AD-C participants exhibited significant reductions in global clustering coefficient (AD-C: M = .060; SD = .007; AD-C–CN: M = .067; SD = .009; t(19.30) = -2.50, p = .028) and global efficiency (AD-C: M = .168; SD = .014; AD-C–CN: M = .178; SD = .011; t(12.49) = -2.15, p = .017). Additionally, AD-C participants misclassified as AD (AD-C–AD), exhibited reductions in global path transitivity compared to AD-C misclassified as CN (AD-C–CN) (AD-C–CN: M = .694; SD = .021; AD-C–AD: M = .681; SD = .014; t(38.97) = -2.39, p = .047). No significant pairwise differences were

observed for any of the graph metrics assessed between correctly classified AD-C participants and those misclassified as AD (all $p > .05$).

Concerning diagnostic variables not used for prediction, no significant group differences were observed on either the MMSE (AD-C–CN: M = 27.19; SD = 2.46; AD-C: M = 25.60; SD = 7.14; AD-C–AD: M = 27.71; SD = 2.37; F(2, 52) = 1.02, p-fdr = 0.500; Eta2 = 0.04, 95% CI [0.00, 1.00]) or CDR-SB (AD-C–CN: M = 1.22; SD = 1.52; AD-C: M = 1.60; SD = 3.18; AD-C–AD: M = 1.36; SD = 1.56; F(2, 53) = 0.15, p-fdr = 0.900; Eta2 = 5.79e-03, 95% CI [0.00, 1.00]). All statistical comparisons are reported in Table 9.

Table 9. Comparing AD-C participant characteristics correctly classified as AD-C or misclassified as CN or AD.

Predictive Features	AD-C–CN N = 32[1]	AD-C N = 10[1]	AD-C–AD, N = 14[1]	STAT[2]	p-fdr[3]	AD-C–CN vs. AD-C	AD-C–CN vs. AD-C–AD	AD-C vs. AD-C–AD
Age	77 (8)	83 (6)	82 (7)	4.2	**.049**	**.019**	**.033**	.700
Education	10.09 (2.22)	9.00 (1.63)	9.07 (2.97)	1.4	.400	.200	.200	>.900
Sex					.900	.700	.400	.700
Female	18 (56%)	5 (50%)	6 (43%)					
Male	14 (44%)	5 (50%)	8 (57%)					
Logical Memory	13.2 (2.9)	11.8 (4.0)	6.9 (4.4)	15	**<.001**	.300	**<.001**	**.002**
WAIS	53 (11)	36 (8)	35 (9)	20	**<.001**	**<.001**	**<.001**	.700
Digit Span Forward	8.8 (1.6)	8.0 (1.5)	7.2 (1.5)	5.45	**.018**	.140	**.002**	.200
Digit Span Backward	6.53 (1.85)	6.10 (1.45)	4.86 (1.75)	4.4	**.047**	.500	**.005**	.094
Trail Making Part A	146 (11)	132 (22)	121 (39)	6.1	**.015**	.092	**.001**	.300
Trail Making Part B	210 (45)	144 (72)	85 (86)	20	**<.001**	**.005**	**<.001**	**.028**
Animal Naming	19.5 (4.6)	15.1 (3.0)	13.1 (3.0)	14	**<.001**	**.004**	**<.001**	.200
Vegetable Naming	12.8 (3.9)	12.2 (3.1)	8.7 (4.1)	5.8	**.016**	.600	**.001**	**.032**
Boston Naming Test	27.6 (1.8)	27.1 (3.3)	22.4 (5.1)	14	**<.001**	.700	**<.001**	**<.001**
Global Communicability[4]	25.8 (3.7)	26.6 (2.7)	25.2 (3.8)	0.46	0.7	.500	.600	.300
Global Clustering Coefficient	0.067 (0.009)	0.060 (0.007)	0.062 (0.008)	3.2	.092	**.028**	.088	.500
Global Efficiency	0.178 (0.011)	0.168 (0.014)	0.173 (0.012)	3.3	.092	**.017**	.200	.300
Global Path Transitivity	0.694 (0.021)	0.681 (0.016)	0.682 (0.013)	3.2	.092	**.054**	**.047**	.900
Global Eigenvector Centrality	0.052 (0.004)	0.052 (0.003)	0.053 (0.002)	0.53	.700	>.900	.300	.500
Global Modularity	0.099 (0.025)	0.100 (0.021)	0.087 (0.025)	1.3	.400	>.900	.130	.200

Global Participation Coefficient	0.833 (0.015)	0.832 (0.015)	0.840 (0.015)	1	.500	.900	.200	.200
Within-Module Degree Z-Score	0.01 (1.10)	0.19 (0.66)	-0.15 (1.00)	0.33	.800	.600	.600	.400
Small-World Propensity	0.62 (0.06)	0.59 (0.05)	0.59 (0.04)	2.7	.130	.100	.054	>.900
MMSE†	27.19 (2.46)	25.60 (7.14)	27.71 (2.37)	1	.500	.200	.700	.200
CDR-SB†	1.22 (1.52)	1.60 (3.18)	1.36 (1.56)	0.15	.900	.600	.800	.800

Note: A supervised random forest model was used to classify participants into *a priori* groups based on the combination of demographic, neuropsychological, and graph-theoretical metrics. Post-hoc analyses assessed if AD-C participants who were correctly classified (AD-C), differed from those misclassified as CN (AD-C–CN) or AD (AD-C–CN) across the set of features used for classification. Groups were also compared on the MMSE and CDR-SB. Overall, the predicted groups significantly differed on the battery neuropsychological assessments. In contrast, groups were found to have similar network topology on average. However, post-hoc comparisons suggested that the predicted groups may exhibit different levels of global clustering coefficient and global efficiency. The predicted AD-C groups were not found to differ on either the MMSE or CDR-SB assessments. Abbreviations: AD = Alzheimer's disease; AD-C = AD Converters; CDR-SB = Clinical Dementia Rating Scale Sum of Boxes ; CN = Cognitively Normal; MMSE = Mini-mental state examination; STAT = Test Statistic. *[1]Mean (SD); n (%); [2]One-way ANOVA; Pearson's Chi-squared test; [3]False discovery rate correction for multiple testing; [3] Global Communicability was log-transformed before calculation of descriptive statistics; † Not used in any prediction models. Bold values indicate a statistically significant at* $p < .05$.

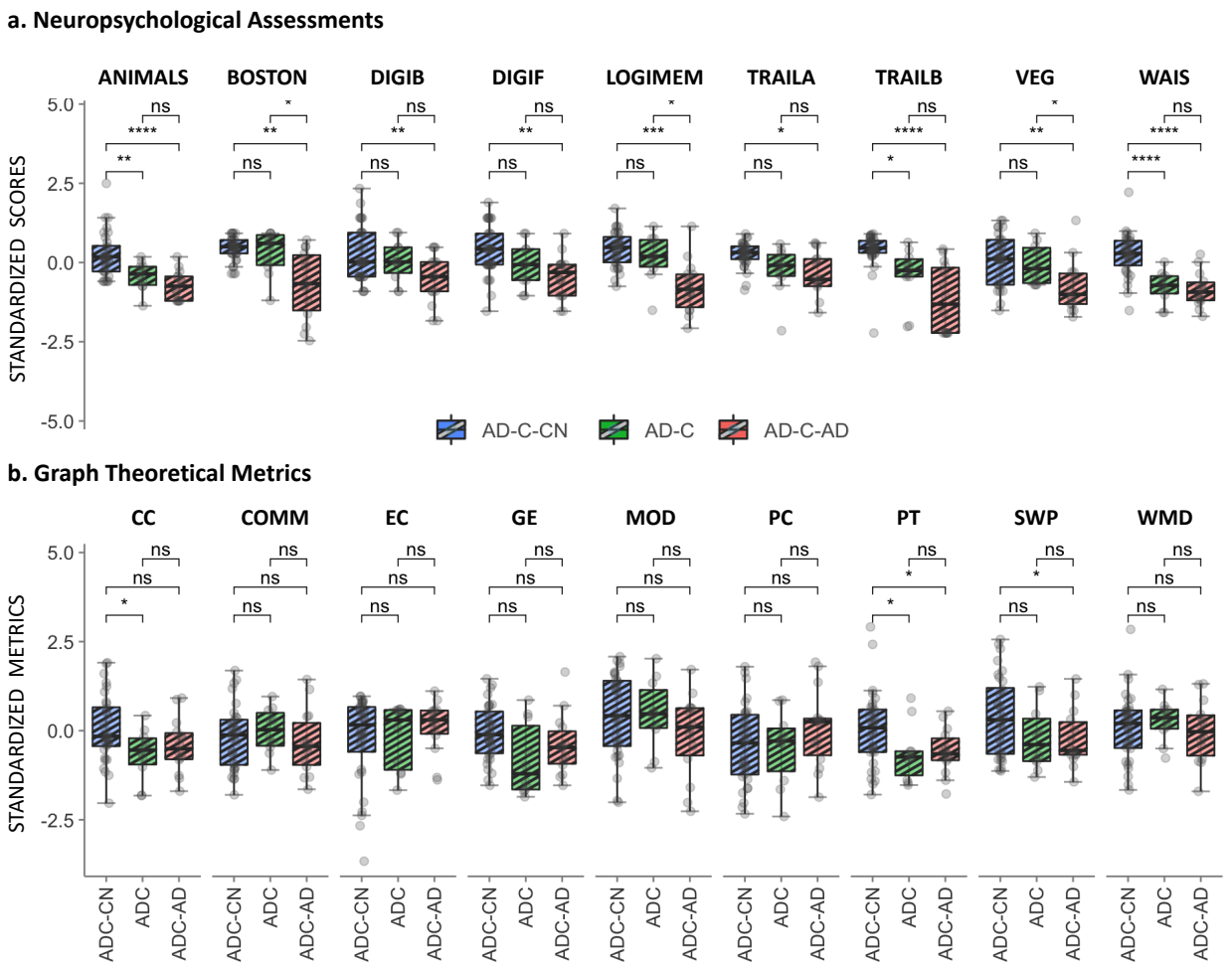


Figure 11. Supervised random forest classification of AD-C participants. a) AD-C participants correctly classified (AD-C; N = 10; 17.86%), are compared to AD-C classified as CN (AD-C–CN; N = 30; 53.57%), or AD (AD-C–AD; N = 16; 28.57%) on neuropsychological assessments. b) AD-C participants correctly classified as CN, are compared to AD-C classified as CN (AD-C–CN) or AD (AD-C–AD) on graph-theoretical metrics. No omnibus group differences were observed for any of the graph-theoretical metrics assessed. However, post-hoc comparisons suggest that AD-C participants exhibit reductions in global clustering coefficient and global efficiency compared to AD-C classified as CN (AD-C–CN), and that AD-C–CN participants have increased small-world propensity and global path transitivity compared to AD-C participants who were classified correctly of as AD (AD-C–AD). All variables have been centered and scaled for visualization purposes. Colors represent the *a priori* group an AD-C participant was classified as. In the legend, AD-C classified as CN (AD-C–CN) are colored blue, AD-C correctly classified in green, and AD-C classified as AD (AD-C–AD) in red. Abbreviations: AD = Alzheimer's disease; AD-C = AD Converters; AD-C–CN = AD-C Classified as Cognitively Normal; AD-C–AD = AD-C Classified as AD; CN = Cognitively Normal; DIGIF = Digit Span Forward; DIGIB = Digit Span Backward; CC = Clustering Coefficient; COMM = Communicability; EC = Eigenvector Centrality; ED = Education; GE = Global Efficiency; LOGIMEM = Logical Memory; MOD = Modularity; PC = Participation Coefficient; PT = Path Transitivity; SWP = Small-World Propensity; TRAILA = Trail Making Test Part A; TRAILB = Trail Making Test Part B; VEG = Vegetable Naming (Category Fluency); WAIS = Weschler Adult Intelligence Scale; WMD = Within-Module Degree Z-Score.

Model 2. Prediction of a priori group status with subsampling to account for class imbalance. As in Model 1, a supervised random forest model, comprised of 3 demographic factors, 9 metrics assessing cognitive performance, and 9 graph-theoretical metrics was used to classify participants into *a priori* groups (CN: N = 159; AD-C: N = 56, and AD: N = 8181). Here, to address the imbalance between classes, SMOTE subsampling was performed on each training sample during cross-validation. With SMOTE resampling, the minority class (AD-C) was over-sampled using synthetic data. Model 2 classified 89.5% of participants correctly (152 of 159 CN (92.12%); 154 of 168 AD-C (91.67%); 45 of 59 (76.27%) AD). Average cross-validated model accuracy was 69.6% with an AUC of .86, sensitivity of .65, specificity of .84, kappa of .50, F1 score of .65, and a balanced accuracy of 75%.

From the cross-validated model, consisting of both real and synthetic observations, a final model was tested on the full sample of observed participants in the CN, AD-C, and AD groups. Final model performance achieved an accuracy of 89.5% (95% CI: 86.10% - .92.40%), with an average sensitivity 0f .87, average specificity of .94, kappa of .83, and balanced accuracy of 90%. Class-specific model performance metrics are shown in Table 10 for both Model 1 and Model 2 before and after SMOTE subsampling. The 5 highest-ranked predictors included age (26.21%), performance on the WAIS (23.45%), animal naming test (Category Fluency; 23.31%), Trail Making Test Part B (22.56%), and Logical Memory (22.20%; Figure 11, panel b). Permutation testing with 1000 iterations indicated that the final model was significant compared to random chance (out-of-bag error: 0.247; Random model out-of-bag error: 0.503; permutation $p < .0001$).

Table 10. Class-specific model performance metrics before and after accounting for class imbalance when classifying participants into a priori groups.

	Model 1	Model 2 (SMOTE)

Performance Metric	CN	AD-C	AD	CN	AD-C	AD
Sensitivity	0.9245	0.17857	0.7778	0.9212	0.9167	0.7627
Specificity	0.6496	0.97083	0.9023	0.9339	0.9018	0.988
Precision	0.7538	0.58824	0.75	0.9102	0.875	0.9184
Recall	0.9245	0.17857	0.7778	0.9212	0.9167	0.7627
F1 Score	0.8305	0.27397	0.7636	0.9157	0.8953	0.8333
Positive Predictive Value	0.7538	0.58824	0.75	0.9102	0.875	0.9184
Negative Predictive Value	0.8812	0.83513	0.9151	0.9422	0.9352	0.9592
Detection Rate	0.4966	0.03378	0.2128	0.3878	0.3929	0.1148
Balanced Accuracy	0.7871	0.5747	0.8401	0.9276	0.9092	0.8753

Note: Predictions for individual classes are shown for Models 1 and 2, before and after SMOTE subsampling. Relative to Model 1, SMOTE subsampling (Model 2) balanced predictions and improved control of true positive and false negative rates across classes. In Model 2, with SMOTE subsampling, detection rates (within-class predictions) were close to the true proportion of cases used to train the models during cross-validation. Abbreviations: AD = Alzheimer's disease; AD-C = AD Converters; CN = Cognitively Normal; SMOTE = Synthetic Minority Oversampling Technique.

Model Comparisons. To further compare model performance with and without SMOTE subsampling, AUROC curves were compared across models for each respective class using Delong's test (see Methods). Significant differences were found when comparing the AUROC of CN vs. AD, and AD-C vs. AD models with and without SMOTE subsampling. Specifically, classification performance was significantly better for the model with SMOTE (CN vs. AD: $D = 10.457$, $p < .0001$; AD-C vs AD: $D = 4.910$, $p < .0001$). No significant difference was found when comparing the AUROC of CN vs. AD-C with and without SMOTE subsampling ($D = 0.359$, $df = 466.59$, $p = 0.720$).

CHAPTER 7: AIM 3 RESULTS

Aim 3 sought to establish whether the combination of core demographic, cognitive measures, and network metrics could identify data-driven neurocognitive subgroups. Aim 3 specifically sought to establish if cognitive subgroups previously identified in the literature (e.g. memory-impaired, memory-spared, dysexecutive) emerge using unsupervised learning. First, an unsupervised random forest model was used to generate a proximity matrix of scores that indicate how similar participants are to one another, as compared to random chance. Next, cluster analysis was used to determine the presence and number of data-driven groups in the data. Finally, cluster-based prediction was performed. Specifically, a supervised random forest was used to classify participants into the data-driven clusters to determine the overall fit of the clusters to the data.

Unsupervised Random Forest To Assess The Presence Of Data-Driven Subgroups

Unsupervised Random Forest. An unsupervised random forest model comprised of the same observations and set of features used for supervised modeling (Aim 2), was used to generate a proximity matrix for cluster analysis. To qualitatively assess the presence of data-driven groups before formal cluster analysis, the proximity matrix was visualized in two-dimensional space using multidimensional scaling (MDS) and principal components analysis (PCA; Figure 12, panels a & b). Low-dimensional representations of the data indicated a high degree of overlap between participants across *a priori* groups. Overall, a greater degree of separation was observed between the CN and AD groups. AD-C participants were again

interspersed among CN and/or AD participants with respect to *a priori* groups. Qualitatively, these observations suggested the presence of two large clusters, where AD-C participants were more likely to be assigned to clusters comprised of either CN or AD participants similar to the supervised classification model in Aim 2. From the unsupervised random forest model, the relative contribution of features used to separate true versus synthetic data points is visualized in Figure 12, panels a & b, are displayed in Figure 12, panel c. Overall, neuropsychological assessments were relatively matched on average feature importance (Neuropsychological assessments: Mean = 16.06; SD = 2.83); Graph-theoretical metrics: Mean = 14.75; SD = 3.93). The top 5 features consisted of performance on the Trail Making Test Part B (19.69%), small-world propensity (19.45%), performance on the WAIS (19.43%), global participation coefficient (19.03%), and modularity (18.28%).

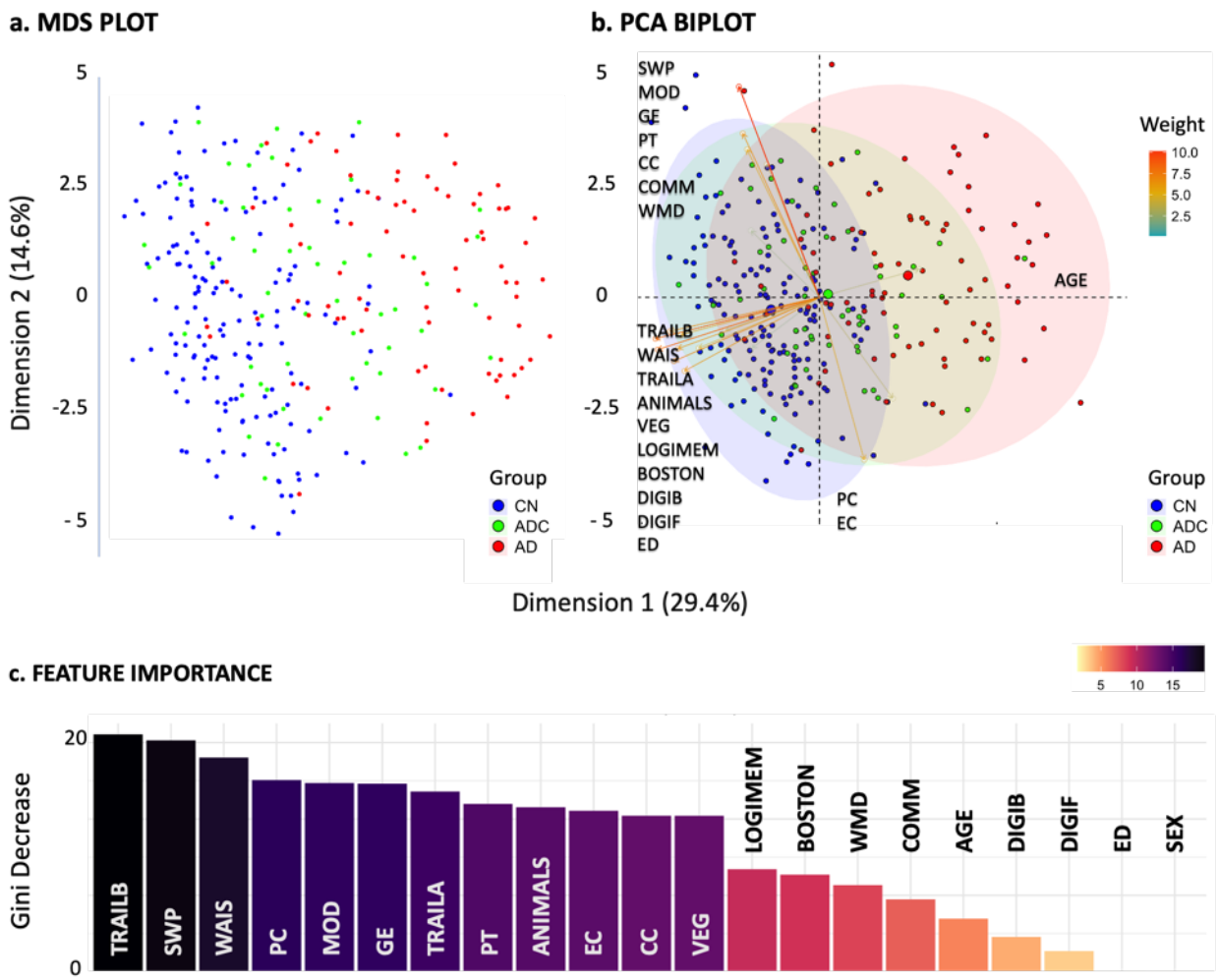


Figure 12. Unsupervised random forest to assess the presence of data-driven subgroups. An unsupervised random forest model was used to generate a proximity matrix that described the degree of similarity between participants based on high-dimensional interactions between the set of demographic, neuropsychological, and graph-theoretical features. a) A multi-dimensional scaling plot (MDS) visualizes the distance between data points in two-dimensional space computed directly from the proximity matrix. Here, data points are color-coded by their a priori classes (CN, AD-C, AD). b) Principal components analysis (PCA) was used to reduce the dimension of the data and assess the amount of variance explained by the set of features in two dimensions. Data points are outlined by their a priori classes (CN, AD-C, AD). The width and length of a line represent the amount of variance explained by a given predictor for a given dimension. c) Unsupervised random forest feature importance scores are ranked from left to right. Here, variable importance represents the mean decrease in Gini importance (i.e. the average decrease in node purity for splitting data) when classifying real versus synthetic data points. Abbreviations: DIGIF = Digit Span Forward; DIGIB = Digit Span Backward; CC = Clustering Coefficient; COMM = Communicability; EC = Eigenvector Centrality; ED = Education; GE = Global Efficiency; LOGIMEM = Logical Memory; MDS = Multidimensional Scaling; MOD = Modularity; PC = Participation Coefficient; PCA = Principle Components Analysis; PT = Path Transitivity; SWP = Small-World Propensity; TRAILA = Trail Making Test

Part A; TRAILB = Trail Making Test Part B; VEG = Vegetable Naming (Category Fluency); WAIS = Weschler Adult Intelligence Scale; WMD = Within-Module Degree Z-Score.

Cluster analysis. K-means clustering, implemented via the *NbClust* R package (Charrad et al., 2015), was used to determine the optimal number of clusters that could be derived from the proximity matrix. The majority voting procedure selected a two cluster solution (Charrad et al., 2015). Ten out of twenty-three cluster optimization indices voted for a two-cluster solution (43.5%). Of note, eight indices (34.8%) voted for a three-cluster solution. Thus, a two-cluster solution was selected. However, the three-cluster solution was selected for follow-up exploratory analyses.

Two-cluster solution. Concerning *a priori* groups, Cluster 1 was comprised of 13% CN (N = 12), 23% AD-C (N = 21), and 64% AD (N = 59). Cluster 2 was 72% CN (N = 147), 17% AD-C (N = 35), and 11% AD (N = 22). Independent two-sided t-tests, corrected for multiple comparisons, found that Clusters 1 and 2 were significantly different (all p-fdr < .05) on all measures assessed except sex, global communicability, and within-module degree z-score (p-fdr > .05; Table 11). Participants assigned to Cluster 1 were on average older (Cluster 1: M = 78.68; SD = 7.0; Cluster 2: M = 71.71; SD = 7.1; t(177.81) = 7.95, p-fdr < .001) and less educated (Cluster 1: M = 8.47; SD = 3.1; Cluster 2: M = 9.95; SD = 2.5; t(149.91) = -4.04, p-fdr = < .001).

Participants assigned to Cluster 1 scored lower on all neuropsychological assessments (all p-fdr < .001; Figure 13, panel a). With respect to network topology (Figure 13, panel b), Cluster 1 exhibited reduced global clustering coefficient (t(224.27) = -3.38, p-fdr = .003), global efficiency (t(187.29) = -4.69, p-fdr = .003), global path transitivity (t(181.80) = -4.42, p-fdr < .001), modularity (t(195.16) = -4.42, p-fdr < .001), and small-world propensity (t(205.19) = -5.73, p-fdr < .001). Cluster 1 had higher estimates of global eigenvector centrality (t(213.77) = 5.63, p-fdr = .002) and global participation coefficient (t(213.77) = 5.63, p-fdr < .001). No

between-cluster differences were observed for global communicability or within-module-degree-zscore (p-fdr > .05). Additionally, Cluster 1 had significantly lower MMSE scores (p-fdr < .001) and higher CDR-SB ratings (p-fdr < .001). All statistical comparisons, along with means and standard deviations are shown in Table 11. Clusters are visualized in Figure 14.

Table 11. Participant characteristics for the two-cluster solution.

Characteristic	Cluster 1 N = 92[1]	Cluster 2 N = 204[1]	Statistic[2]	p-fdr[3]
Age	78.68 (7.0)	71.71 (7.1)	7.9	**<.001**
Education	8.47 (3.1)	9.95 (2.5)	-4.0	**<.001**
Sex	--	--	0.01	>.900
Female	50 (54%)	112 (55%)	--	--
Male	42 (46%)	92 (45%)	--	--
Logical Memory	6.21 (4.9)	13.12 (3.9)	-12	**<.001**
WAIS	31.49 (12.5)	56.55 (11.1)	-17	**<.001**
Digit Span Forward	6.88 (1.7)	8.70 (1.9)	-8.2	**<.001**
Digit Span Backward	4.35 (1.8)	6.69 (1.9)	-10	**<.001**
Trail Making Test Part A	115.86 (40.0)	148.81 (9.7)	-7.8	**<.001**
Trail Making Test Part B	86.52 (79.1)	218.34 (30.2)	-15	**<.001**
Animal Naming	11.83 (4.3)	20.55 (5.4)	-15	**<.001**
Vegetable Naming	8.15 (3.9)	14.40 (4.1)	-13	**<.001**
Boston Naming Test	21.04 (5.5)	27.72 (2.2)	-11	**<.001**
Global Communicability[4]	26.26 (3.67)	26.72 (4.30)	-0.95	.400
Global Clustering Coefficient	0.064 (0.008)	0.068 (0.010)	-3.4	**.002**
Global Efficiency	0.174 (0.013)	0.181 (0.014)	-4.7	**<.001**
Path Transitivity	0.686 (0.019)	0.697 (0.020)	-4.4	**<.001**
Eigenvector Centrality	0.054 (0.002)	0.052 (0.003)	3.7	**.002**
Modularity	0.081 (0.019)	0.094 (0.021)	-5.1	**<.001**
Participation Coefficient	0.845 (0.011)	0.836 (0.014)	5.6	**<.001**
Within-Module Degree Z-Score	0.000 (0.000)	0.000 (0.000)	-0.61	.500
Small-World Propensity	0.577 (0.043)	0.610 (0.051)	-5.7	<.001
MMSE†	23.92 (4.4)	28.67 (1.9)	-9.8	**<.001**
CDR-SB†	3.38 (3.0)	0.55 (1.4)	8.6	**<.001**

Note: On average, clusters were significantly different on all demographic, neuropsychological, and network-based metrics assessed except for sex, communicability, and within-module-degree-zscore. Cluster 1 was comprised of participants who, on average, (1) have higher scores across all neuropsychological assessments, and (2) exhibit lower scores on all graph-theoretical metrics assessed, except global eigenvector centrality and global participation coefficient. Abbreviations:

CDR-SB = Clinical Dementia Rating Scale Sum of Boxes; MMSE = Mini-mental state examination. [1] *n (%); Mean (SD);* [2] *Two Sample t-test; Pearson's Chi-squared test*; [3] *False discovery rate correction for multiple testing;* [4] *Global Communicability was log-transformed before calculation of descriptive statistics; † Not used in prediction analyses.*

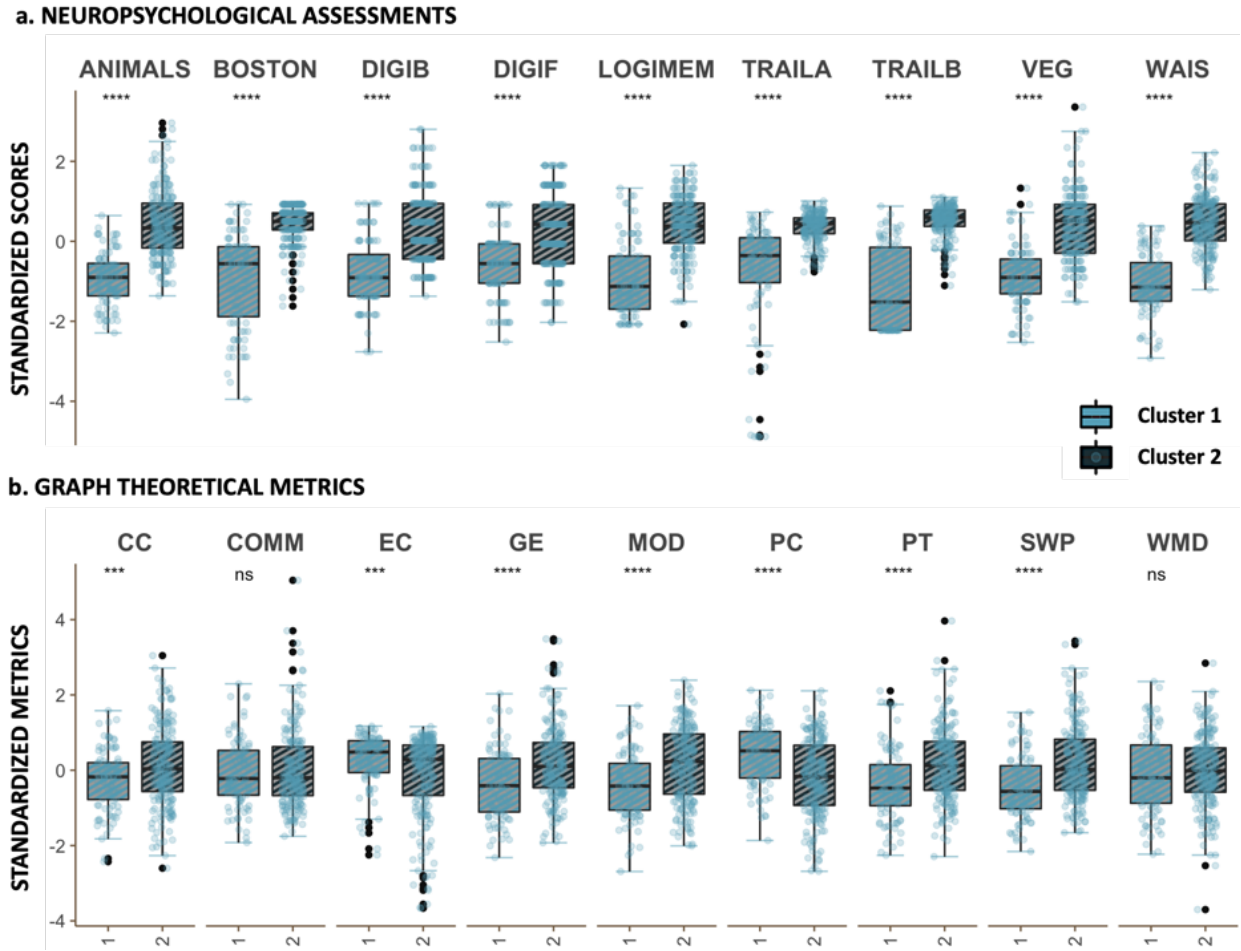


Figure 13. Predictive Features Summarized by Cluster. In panel a, distributions for neuropsychological assessments are displayed by cluster. In panel b, distributions for graph-theoretical metrics are displayed cluster. Cluster 1 is comprised of participants who, on average, (1) have higher scores across all neuropsychological assessments (all p-fdr < .05), and (2) exhibit lower estimates of global clustering coefficient, global efficiency, modularity, global path transitivity, and small-world propensity (all p-fdr < .05), and higher estimates of global eigenvector centrality and global participation coefficient (p-fdr > .05). Clusters did not significantly differ on global communicability (COMM; p-fdr > .05) or within-module-degree-Zscore (WMD; p-fdr > .05). Abbreviations: CC = Clustering Coefficient; COMM = Communicability; EC = Eigenvector Centrality; ED = Education; GE = Global Efficiency; LOGIMEM = Logical Memory; MOD = Modularity; PC = Participation Coefficient; PT = Path Transitivity; SWP = Small-World Propensity; TRAILA = Trail Making Test Part A; TRAILB = Trail Making Test Part B; VEG = Vegetable Naming (Category Fluency); WAIS = Weschler

Adult Intelligence Scale; WMD = Within-Module Degree Z-Score. **** = p < .0001; *** = p < .001; ** = p < .01; * = p < .05 (uncorrected); ns = non-significant.

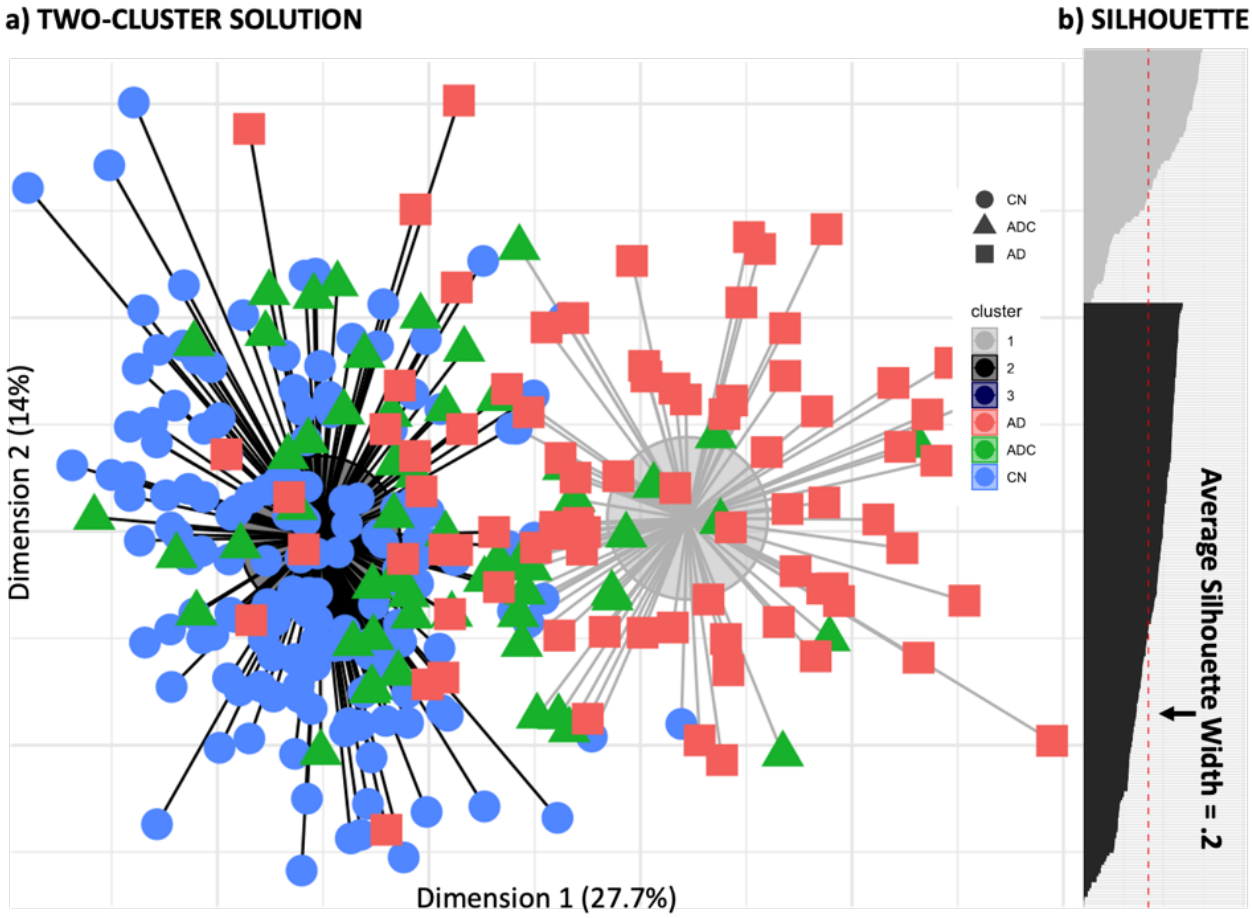


Figure 14. Visualizing the two-cluster solution. a) Clusters are depicted in two-dimensional space where dimensions represent principle components. Participants are shown in proximity to the center of their respective cluster assignment. Observations that are closer to the center of a given cluster are more representative of that cluster. Data points are represented by shapes and color-coded by a priori group (CN, AD-C, AD). Two components explain 41.7% of the data, and clusters are best distinguished by the first principle component. b) A silhouette plot is used to visualize and quantify the distance for each observation to the center of a cluster (average of all distances). Observations with higher values are more representative of that cluster. An average silhouette width <= .2 is considered a poor fit such that observations within a cluster are not homogenous and there is a high degree of variability within the clusters. Clusters are depicted in grayscale.

Three-cluster solution. One-way analysis of variance tests, corrected for multiple comparisons, found that Clusters significantly differed on all neuropsychological measures assessed (all p-fdr < .001; Figure 15, panel a), as well as all graph-theoretical metrics (all p-fdr <

.001; Figure 15, panel b) except within-module degree z-score (F(2, 293) = 0.57, p-fdr = 0.600; Eta2 = 3.90e-03, 95% CI [0.00, 1.00]) (Table 12). Concerning *a priori* groups, Cluster 1 was comprised primarily of AD participants, and Cluster 2 and 3 were comprised primarily of CN participants. Specifically, Cluster 1 (N = 70) was comprised of 0.6% CN (N = 4), 15.7% AD-C (N = 11), and 78.6% AD (N = 55). Cluster 2 (N = 132) was comprised of 71.2% CN (N = 94), 18.2% AD-C (N = 24), and 10.6% AD (N = 14). Cluster 3 (N = 94) was comprised of 64.9% CN (N = 61), 22.3% AD-C (N = 21), and 12.8% AD (N = 12). Participants assigned to Cluster 1 were on average older and less educated compared to Clusters 2 and 3 (all p-fdr < .001). The proportion of males and females was again not significantly different across the clusters (p-fdr = .400). Additionally, Cluster 1 had significantly lower MMSE scores (p-fdr < .001) and higher CDR-SB ratings (p-fdr < .001) compared to Cluster 2 and Cluster 3. All comparisons are shown in Table 12.

Post-hoc two-sided t-tests (uncorrected) found that Cluster 1 (78.6% AD) scored significantly lower on all neuropsychological assessments compared to Clusters 2 and 3 (all p < .05). Clusters 2 and 3 did not significantly differ on any of the neuropsychological assessments (all p > .05; Table 12). With respect to network topology, Cluster 1 exhibited reduced modularity compared to Cluster 2 (t(140.11) = -2.16, p = 0.035). Cluster 1 and Cluster 2 did not significantly differ on any of the other graph-theoretical metrics assessed (p > .05). Conversely, Clusters 2 and 3 significantly differed on all graph-theoretical metrics (all p < .001) except for within-module degree z-score (t(217.91) = -1.10, p = 0.300). Compared to Cluster 3, Cluster 2 had reduced estimates on six of the nine graph theoretical metrics assessed (all p < .001), including global communicability, global clustering coefficient, global efficiency, global path transitivity, modularity and small-world propensity. Global eigenvector centrality and global participation

coefficient were significantly higher in Cluster 2 (Global eigenvector centrality: t(213.77) = 5.63, p-fdr = .002; global participation coefficient: t(213.77) = 5.63, p-fdr < .001). Clusters are visualized in (Figure 16).

Table 12. Participant characteristics for the three-cluster solution.

Characteristic	Cluster 1 N = 70[1]	Cluster 2 N = 132[1]	Cluster 3 N = 94[1]	STAT[2]	p-fdr[3]	Cluster 1 vs Cluster 2	Cluster 1 vs. Cluster 3	Cluster 2 vs. Cluster 3
Age	78.89 (7.1)	72.03 (6.9)	72.73 (7.8)	22	**<.001**	**<.001**	**<.001**	.500
Education	8.07 (2.9)	10.26 (2.5)	9.46 (2.7)	15	**<.001**	**<.001**	**.001**	**.027**
Sex				1.9	.400			
Female	41 (59%)	75 (57%)	46 (49%)					
Male	29 (41%)	57 (43%)	48 (51%)					
Logical Memory	5.09 (4.5)	12.81 (4.1)	12.77 (4.0)	92	**<.001**	**<.001**	**<.001**	>.900
WAIS	28.61 (11.8)	54.47 (11.8)	55.77 (12.1)	132	**<.001**	**<.001**	**<.001**	.400
Digit Span Forward	6.83 (1.7)	8.67 (2.0)	8.35 (1.8)	22	**<.001**	**<.001**	**<.001**	.200
Digit Span Backward	4.00 (1.7)	6.55 (1.9)	6.61 (2.0)	51	**<.001**	**<.001**	**<.001**	.800
Trail Making Test Part A	108.76 (43.2)	147.81 (10.0)	147.78 (10.8)	78	**<.001**	**<.001**	**<.001**	>.900
Trail Making Test Part B	61.99 (69.5)	212.64 (37.4)	213.76 (35.9)	281	**<.001**	**<.001**	**<.001**	.900
Animal Naming	10.80 (4.2)	20.20 (5.3)	19.77 (5.6)	85	**<.001**	**<.001**	**<.001**	.500
Vegetable Naming	7.39 (3.8)	14.31 (3.9)	13.63 (4.4)	73	**<.001**	**<.001**	**<.001**	.200
Boston Naming Test	19.64 (5.4)	27.47 (2.3)	27.55 (2.5)	148	**<.001**	**<.001**	**<.001**	.900
Global Communicability[4]	25.224 (3.437)	24.572 (2.379)	30.395 (3.889)	100	**<.001**	0.2	**<.001**	**<.001**
Global Clustering Coefficient	0.064 (0.008)	0.063 (0.009)	0.073 (0.009)	37	**<.001**	0.6	**<.001**	**<.001**
Global Efficiency	0.175 (0.013)	0.174 (0.011)	0.189 (0.012)	49	**<.001**	0.7	**<.001**	**<.001**
Path Transitivity	0.687 (0.020)	0.687 (0.016)	0.708 (0.018)	45	**<.001**	0.8	**<.001**	**<.001**
Eigenvector Centrality	0.054 (0.002)	0.054 (0.002)	0.050 (0.004)	67	**<.001**	0.15	**<.001**	**<.001**
Modularity	0.081 (0.020)	0.088 (0.020)	0.100 (0.021)	19	**<.001**	**0.035**	**<.001**	**<.001**
Participation Coefficient	0.845 (0.012)	0.842 (0.011)	0.830 (0.014)	35	**<.001**	0.076	**<.001**	**<.001**
Within-Module Degree Z-Score	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.57	.600	>.900	.400	.300
Small-World Propensity	0.578 (0.043)	0.572 (0.026)	0.654 (0.038)	175	**<.001**	.200	**<.001**	**<.001**
MMSE†	23.40 (4.1)	28.22 (2.7)	28.56 (2.3)	74	**<.001**	**<.001**	**<.001**	.400
CDR-SB†	3.84 (3.0)	0.70 (1.8)	0.65 (1.4)	65	**<.001**	**<.001**	**<.001**	.900

Note: A series of independent one-way analysis of variance tests found that clusters significantly differed on age (p-fdr = < .001) and education level (p-fdr = < .001). The proportion of males and females was not significantly different across clusters (p-fdr = .400). Additionally, clusters significantly differed on all neuropsychological measures assessed (all p-fdr < .05). Post-hoc t-tests found that Cluster 1 (5.7% CN; 15.7% AD-C; 78.6% AD) scored significantly lower on all neuropsychological assessments (all p < .05) compared to Cluster 2 (71.2% CN; 18.2% AD-C; 10.6% AD) and Cluster 3 (64.9% CN; 22.3% AD-C; 12.8% AD). Conversely, Clusters 2 and 3 did not significantly differ on any of the neuropsychological assessments (all p > .05). In panel b, distributions for graph-theoretical metrics are displayed by cluster. Except for within-module degree z-score (WMD; p-fdr = .600), clusters significantly differed on the set of graph-theoretical metrics assessed (all p-fdr < .001). Post-hoc t-tests found that modularity was significantly lower in Cluster 1 compared to Cluster 2 (p = .035). No other significant differences were observed between Cluster 1 and 2 (all p > .05). Cluster 3 had significantly higher estimates of global communicability (COMM), global clustering coefficient (CC), global efficiency (GE), global path transitivity (PT), modularity, (MOD), and small-world propensity (SWP) as compared to both Cluster 1 and Cluster 2 (all p < .05). Conversely, global eigenvector centrality (EC) and global participation coefficient (PC) were significantly lower in Cluster 3 compared to Cluster 1 and 2 (p < .05). Abbreviations: CDR-SB = Clinical Dementia Rating Scale Sum of Boxes; MMSE = Mini-Mental State Exam; STAT = Test Statistic. *[1]Mean (SD); n (%); [2]One-way ANOVA; Pearson's Chi-squared test; [3]False discovery rate correction for multiple testing; [4] Global Communicability was log-transformed before calculation of descriptive statistics † Not used in any prediction models. Bold values indicate a statistically significant at p < .05.*

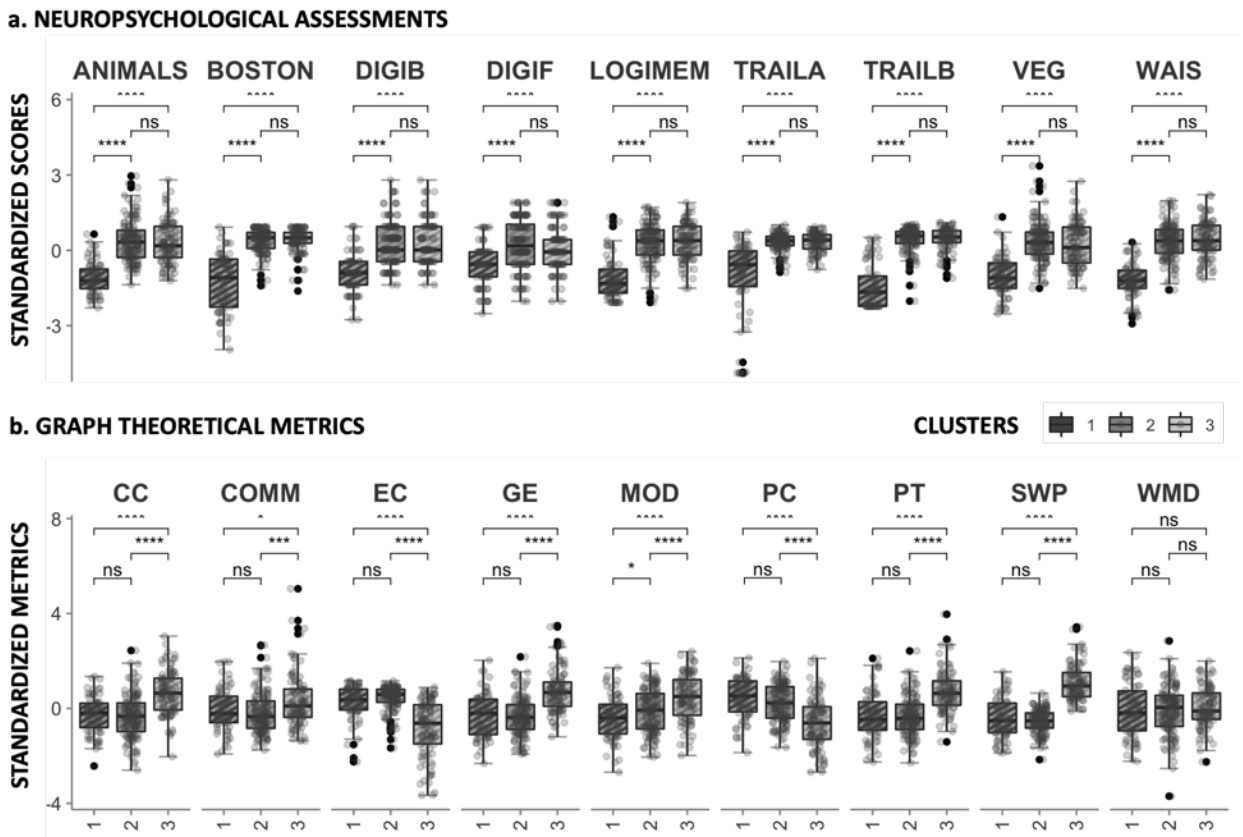


Figure 15 . Predictive features summarized by cluster. In panel a, distributions of neuropsychological assessments are displayed by cluster. Clusters significantly differed on all neuropsychological measures assessed via a series of independent one-way analysis of variance tests (all p-fdr < .001). Post-hoc t-tests found that Cluster 1 (5.7% CN; 15.7% AD-C; 78.6% AD) scored significantly lower on all neuropsychological assessments (all $p < .05$). Conversely, Cluster 2 (71.2% CN; 18.2% AD-C; 10.6% AD) and Cluster 3 (64.9% CN; 22.3% AD-C; 12.8% AD) did not significantly differ on any of the neuropsychological assessments (all $p > .05$). In panel b, distributions for graph-theoretical metrics are displayed by cluster. Except for within-module degree z-score (WMD; p-fdr = .600), clusters significantly differed on the set of graph-theoretical metrics assessed (all p-fdr < .001). Post-hoc t-tests found that Cluster 3 had significantly higher estimates of most graph-theoretical metrics assessed compared to both Clusters 1 and 2 (all $p < .001$). Conversely, eigenvector centrality (EC; $p < .001$) and participation coefficient (PC; $p < .001$) were significantly lower in Cluster 3 compared to Clusters 1 and 2. Only modularity was significantly different between Clusters 1 and 2 ($p = .035$). Abbreviations: CC = Clustering Coefficient; COMM = Communicability; EC = Eigenvector Centrality; GE = Global Efficiency; LOGIMEM = Logical Memory; MOD = Modularity; PC = Participation Coefficient; PT = Path Transitivity; SWP = Small-World Propensity; TRAILA = Trail Making Test Part A; TRAILB = Trail Making Test Part B; VEG = Vegetable Naming (Category Fluency); WAIS = Weschler Adult Intelligence Scale; WMD =

Within-Module Degree Z-Score. **** = p < .0001; *** = p < .001; ** = p < .01; * = p < .05 (uncorrected); ns = non-significant (p > .05).

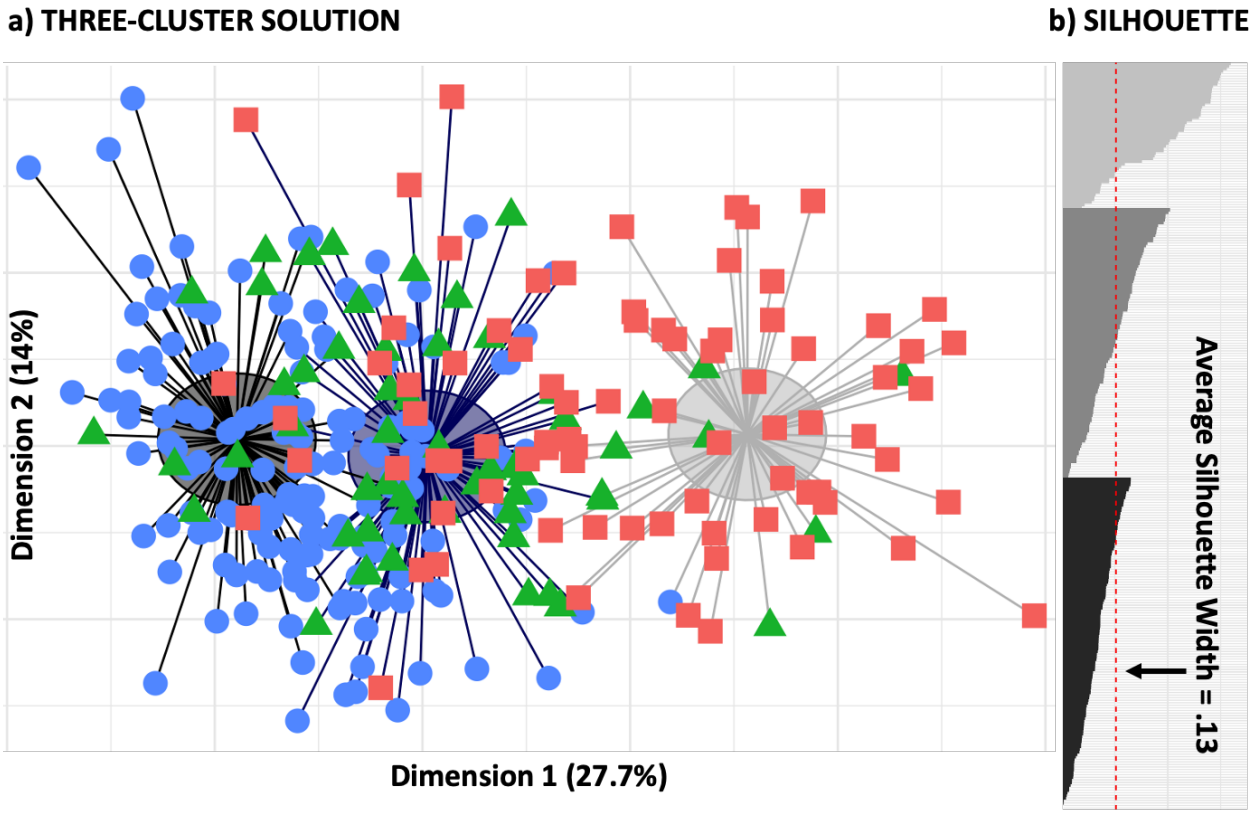


Figure 16. Visualizing the three-cluster solution. a) Clusters are depicted in two-dimensional space where dimensions represent principle components. Participants are shown in proximity to the center of their respective cluster assignment. Observations that are closer to the center of a given cluster are more representative of that cluster. Data points are represented by shapes and color-coded by a priori group (CN, AD-C, AD). Two components explain 41.7% of the data, and clusters are best distinguished by the first principle component. b) A silhouette plot is used to visualize and quantify the distance for each observation to the center of a cluster (average of all distances). Observations with higher values are more representative of that cluster. An average silhouette width <= .2 is considered a poor fit such that observations within a cluster are not homogenous and there is a high degree of variability within the clusters. Clusters are depicted in grayscale.

Cluster-Based Prediction

Two-cluster solution. Following unsupervised classification analysis, a supervised random forest classification model was used to generate estimates of model accuracy to quantify how well the data-driven clusters fit the data using cross-validation. Overall, the cluster-based prediction model produced highly accurate and balanced predictions. The supervised

model correctly classified 280 of 296 cases (Cluster 1: 77 of 92; Cluster 2: 201 of 204). Average cross-validated model accuracy was 94.1%, with an AUC of .99, sensitivity of .85, specificity of .98, kappa of .86, F1 score of .90, and a balanced accuracy of 91.1%. In addition, the detection rate of 26.7% is close to the true prevalence of cases in the positive class (Cluster 1: N = 92 of 296; 31.1%), further suggesting the model is capturing cases in the minority class (cluster 1). The final model was significant compared to random chance (out-of-bag error: 0.064; Random model out-of-bag error: 0.338; permutation $p < .0001$). Performance estimates are based on assigning cluster one as the positive class (case) and cluster 2 as the negative class (control). Overall, neuropsychological measures contributed more to model performance (Feature importance: Mean = 15.33; SD = 3.86) than graph-theoretical metrics (Feature importance: Mean = 10.41; SD = 3.81) as displayed in Figure 17 (panel a). Trail Making Test Part B (20.70%), WAIS (19.95%), Boston Naming Test (17.49%), and the animal naming condition (16.30) from the Category Fluency assessment were the top 5 ranked measures. Small-world propensity (8.59%) was the highest-ranked graph-theoretical metric, although it was ranked ninth overall and lower than eight of the nine neuropsychological assessments.

Three-cluster solution. As an exploratory analysis, a multiclass supervised random forest classification model was used to generate estimates of model accuracy to quantify how well the three-cluster solution fit the data using cross-validation. Overall, the three-cluster-based prediction model produced highly accurate and balanced predictions. The supervised model correctly classified 274 of 296 cases (Cluster 1: 60 of 70; Cluster 2: 129 of 132; Cluster 3: 85 of 94). Average cross-validated model accuracy was 92.6%, with an AUC of .99, sensitivity of .92, specificity of .96, kappa of .88, F1 score of .92, and balanced accuracy of 93.7%. The final model was significant compared to random chance (out-of-bag error: 0.078; Random model out-

of-bag error: 0.605; permutation p < .0001). Performance estimates are based on assigning Cluster 1 as the positive class (case) and Cluster 2 and 3 as the negative classes (control). Detection rates approached the true prevalence of cases in each class (Cluster 1: N = 92 of 296; 31.1%) suggesting the model is capturing cases in both the majority (Cluster 2) and minority classes (Cluster 1 and Cluster 3). Overall, neuropsychological measures contributed more to model performance (Feature importance: Mean = 13.74; SD = 1.55) than graph-theoretical metrics (Feature importance: Mean = 11.59; SD = 8.16) as displayed in Figure 17 (panel b). Small-world propensity (27.53), Trail Making Test Part B (21.11%), eigenvector centrality (18.51%), Boston Naming Test (17.12%), and the WAIS (16.61) were the top 5 ranked measures. Small-world propensity was again the highest-ranked graph-theoretical metric, although here it was the top-ranked feature overall.

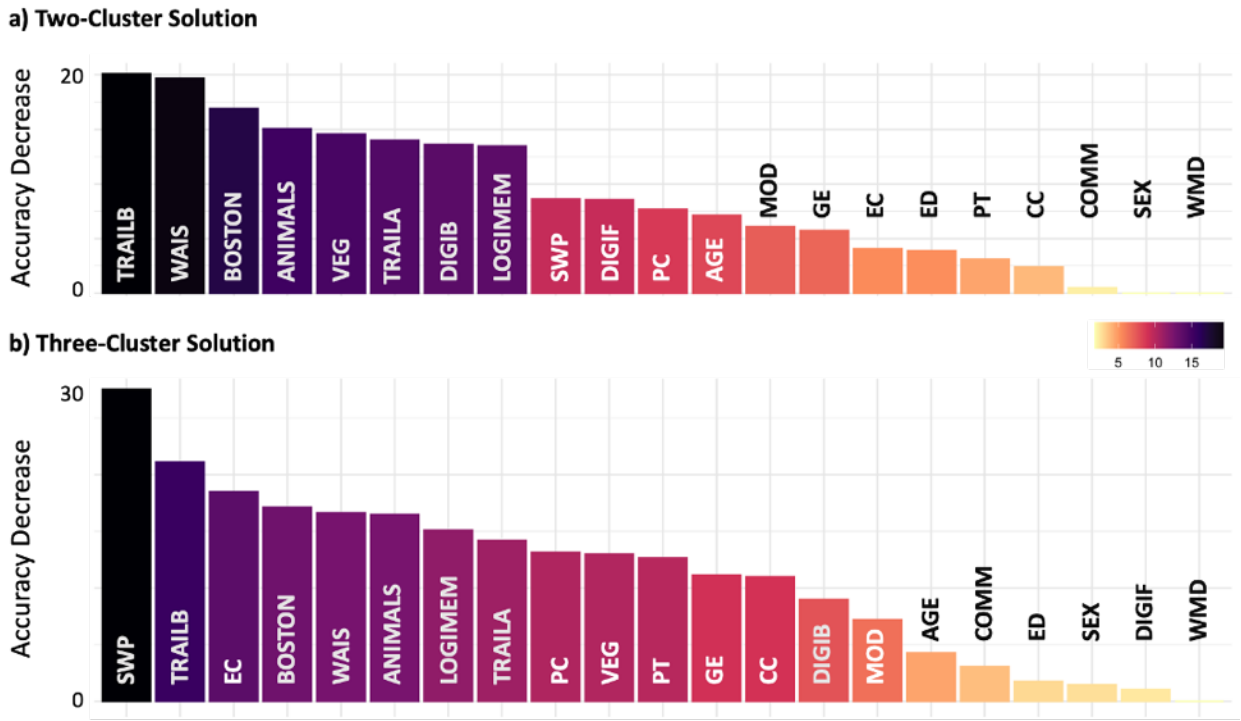


Figure 17. Cluster-based prediction feature importance. Feature importance represents the expected decrease in average classification accuracy should a given feature be removed from the model. a) Features are ranked by the degree to which they contribute to classification performance for the two-cluster solution. Numerically, neuropsychological measures contributed more to model performance (Feature importance: Mean = 15.33; SD = 3.86) than graph-

theoretical metrics (Feature importance: Mean = 10.41; SD = 3.81). b) Features are ranked by the degree to which they contribute to classification performance for the three-cluster solution. Numerically, neuropsychological measures contributed more overall to model performance (Feature importance: Mean = 15.33; SD = 3.86), than graph-theoretical metrics (Feature importance: Mean = 15.33; SD = 3.86). Of note, however, small-world propensity was the top-ranked feature. Abbreviations: AD = Alzheimer's disease; AD-C = AD Converters; CN = Cognitively Normal; DIGIF = Digit Span Forward; DIGIB = Digit Span Backward; CC = Clustering Coefficient; COMM = Communicability; EC = Eigenvector Centrality; ED = Education; GE = Global Efficiency; LOGIMEM = Logical Memory; MDS = Multidimensional Scaling; MOD = Modularity; PC = Participation Coefficient; PCA = Principle Components Analysis; PT = Path Transitivity; SWP = Small-World Propensity; TRAILA = Trail Making Test Part A; TRAILB = Trail Making Test Part B; VEG = Vegetable Naming (Category Fluency); WAIS = Weschler Adult Intelligence Scale; WMD = Within-Module Degree Z-Score.

CHAPTER 8: DISCUSSION

This study aimed to disentangle shared and unique aspects of cognitive impairment and functional network topology seen in healthy aging, early-stage or preclinical dementia, and AD. Specifically, Aim 1 sought to establish if weighted metrics that index redundancy in unthresholded functional brain networks, as a proxy of brain and cognitive reserve, support general and/or specific forms of cognition. Aim 2 sought to establish whether a combination of core demographic risk factors (age, sex, and education), cognitive measures, and weighted functional network metrics could accurately distinguish CN, AD-C, and AD individuals using supervised machine learning. Aim 3 specifically sought to quantify the presence of data-driven neurocognitive subgroups utilizing a combination of demographic, graph-theoretical metrics, and neuropsychological assessments using both unsupervised and supervised machine learning. Contrary to the proposed hypotheses, redundancy-based metrics (global communicability and global clustering coefficient) were not predictive of overall cognitive functioning, nor were they the most informative predictors when attempting to distinguish between CN, AD-C, and AD participants using machine learning.

In Aim 1, performance on the Trail Making Test Part B was positively associated with global communicability in the AD-C group only. No associations were observed between the redundancy-based metrics and either Logical Memory or Digit Span Backward performance. In Aim 2, using machine learning, a combination of core demographic risk factors of AD, metrics assessing cognitive performance, and graph-theoretical metrics assessing network topology were

able to accurately distinguish between healthy controls and individuals with AD. However, the AD-C group was more challenging to classify as predicted. AD-C participants were correctly classified as AD converters or misclassified as cognitively normal or AD. Neuropsychological assessments were overall more predictive than graph-theoretical metrics that summarize global network topology when distinguishing between CN and AD participants. The top predictive neuropsychological features were the Logical Memory (total elements correctly recalled), Trail Making Part B (total number of sequence and set loss errors), and WAIS Digit Symbol (total number of symbols correctly matched) assessments. Among the graph metrics assessed, modularity, global participation coefficient, and global efficiency were the most predictive. Global communicability and global clustering coefficient were not among the top-ranked predictive features as predicted. In Aim 3, unsupervised learning produced two large clusters as hypothesized; however, the detected clusters did not represent amnestic versus non-amnestic subtypes as predicted. Instead, participants were stratified into groups of high- and low-cognitive performers irrespective of the cognitive domain. The cluster comprised of low-performers (Cluster 1) was predominantly AD (64%) and AD-C (23%), with only 13% CN. Conversely, the cluster comprised of high-performers (Cluster 2) was predominantly CN (72%) and AD-C (17%), with only 11% AD. Network topology was also significantly different between Clusters 1 and 2. Consistent with the supervised classification model, AD-C participants were distributed between Clusters 1 and 2, primarily comprised of either AD or CN participants, respectively.

In the following paragraphs, I discuss the results for each aim in greater detail. I then provide a final synopsis and address the study's limitations before concluding with a general statement of significance.

Aim 1

Aim 1 sought to establish whether metrics that quantify redundancy in weighted functional brain networks are associated with attention, memory, and aspects of executive function. Specifically, three cognitive metrics indexing episodic memory (Logical Memory), working memory (Digit Span Test-Backward), and mental flexibility (Trail Making Test B), and two weighted functional redundancy metrics (global communicability and global clustering coefficient) were used for Aim 1 analyses. Global communicability and global clustering coefficient were found to be positively correlated. A significant positive association was observed regardless of whether metrics were computed for the entire sample or within the a priori groups. Thus, to a moderate degree, both metrics capture similar but not identical aspects of redundancy in weighted functional networks at the global level. Comparing levels of redundancy across *a priori* groups, global communicability was numerically but not significantly reduced in AD converters compared to CN and AD participants. Global clustering coefficient was significantly reduced in both the AD-C and AD groups compared to CN participants. The reduction in global clustering coefficient is consistent with prior work finding that redundancy is reduced in individuals with both early- and late-MCI compared to healthy controls (Langella, Mucha, et al., 2021; Langella, Sadiq, et al., 2021). The lack of difference observed between groups in levels of global communicability was unexpected.

Numerically, CN and AD participants had higher levels of global communicability compared to AD-C participants. Prior work estimating communicability on structural networks (e.g., white matter) has suggested that individuals with AD have increased levels of communicability (Lella et al., 2019; Lella & Estrada, 2020; Mišić et al., 2015). From this point of view, individuals in advanced stages of AD may have higher levels of whole-brain

redundancy estimated as estimated by global communicability that facilitates rather than prevents the further spread of pathology (Franzmeier et al., 2020; Hoenig et al., 2018; Vogel et al., 2019). Importantly, however, this has not been established in functionally-defined brain networks and may not be biologically plausible. However, persistent functional hyperconnectivity between brain regions is metabolically costly and may leave brain networks more vulnerable to pathology, as suggested above (Hillary & Grafman, 2017). Thus, while CN and AD had numerically similar levels of global communicability, the degree to which higher levels of communicability can be interpreted as a protective marker is uncertain. Future work will be required to test how communicability, estimated from weighted networks, changes over time in individuals at risk for developing AD. Further, as discussed below, redundancy-based metrics may be equated at the global level but likely diverge at the node or network levels.

Overall, the weighted redundancy-based metrics (global communicability and global clustering coefficient) were not predictive of overall cognitive ability as hypothesized. Moreover, neither global communicability nor global clustering coefficient was more strongly associated with memory performance in the AD group as predicted. However, global communicability (log-transformed) was significantly positively associated with performance on the Trail Making Part B assessment in the AD-C group. That is, higher levels of redundancy, as captured by global communicability, were associated with Trail Making Part B performance in the AD-C group.

As suggested by Sadiq and colleagues (2021), whole-brain redundancy may be a marker of age-related cognitive impairment in specific components of executive functioning (Sadiq et al., 2021). In (Sadiq et al., 2021), whole-brain redundancy was associated with better color-word inhibition performance but not verbal fluency. Inhibition, cognitive flexibility/task-switching/shifting, and working memory/updating are core components of executive function

(Miyake et al., 2000; H. R. Snyder et al., 2015). Verbal and category fluency measures may invoke executive operations but also involve language and memory components. Thus these are broad measures of general cognitive functioning (Gustavson et al., 2019; Kesler et al., 2017; Whiteside et al., 2015). However, compared to verbal fluency, the Trail Making Part B assessment is more closely aligned with executive processes. Thus, the finding that redundancy, here measured by global communicability, is specific to trail-making performance and therefore executive functioning, may be consistent with the results reported by (Sadiq et al., 2021). However, this effect may also be specific to a subset of converters as discussed further below.

Aim 2

Aim 2 sought to establish whether a combination of core demographic risk factors of AD (age, sex, and education), nine cognitive measures, and nine weighted functional network metrics could accurately distinguish CN, AD-C, and AD individuals using supervised machine learning. Average cross-validated model accuracy for the multiclass model was 74.4%, with an AUC of .87, indicating a moderate level of classification performance. The top predictive features included performance on the Logical Memory, WAIS, and Trail Making Part B assessments. Among the graph metrics assessed, modularity, global participation coefficient, and global efficiency were the most predictive. Although, as noted above, neuropsychological assessments were more predictive overall than graph metrics.

Prior studies combining machine learning and network neuroscience to distinguish healthy controls from MCI or AD participants have reported model accuracies in the range of 47-100% (Dyrba et al., 2015; Hojjati et al., 2017, 2019; Jie et al., 2014; Jitsuishi & Yamaguchi, 2022; Khazaee et al., 2015, 2016, 2017; Sun et al., 2019; Wang et al., 2013; Wee et al., 2012, 2013; Wei et al., 2016; Xu et al., 2020; L. Zhang et al., 2020). As discussed in the *Machine*

Learning and AD section, several factors contribute to classification performance, including sample size, use of two-class versus multiclass classification, method of cross-validation, and use of feature selection. The most successful classification models compared healthy controls to AD participants (Khazaee et al., 2015). In the study by Khazaee and colleagues (2015), 100% accuracy was reported when classifying a small sample of healthy controls (N = 20) against AD participants (N = 20) (Khazaee et al., 2015). Although CN and AD participants were not classified separately from converters in the present study, average cross-validated model accuracy was approximately 87.5% when considering CN and AD participants only.

Regardless, the use of multiclass classification also permits a comprehensive comparison of individuals who were misclassified. AD converters, who at the time of study were considered cognitively normal, to CN and AD participants. As predicted, participants in the AD-C group were more challenging to classify. Specifically, AD-C participants were just as likely to be classified as CN (N = 32) as either AD-C (N = 10) or AD (N = 14). Interestingly, compared to AD-C classified as CN (AD-C–CN), correctly classified converters were older and exhibited more impaired executive functioning. These individuals had comparable scores to the CN group on the Logical Memory, Boston Naming Test, and vegetable naming (Category Fluency) assessments. These tests primarily assess memory recall, phonetic and semantic fluency, and semantic memory, respectively, which are cognitive functions most impacted in typical and late-onset AD. Interestingly, AD-C individuals misclassified as AD (AD-C–AD) exhibited more memory-related impairments than those misclassified as CN, with lower performance on the Logical Memory, Boston Naming Test, and vegetable naming (Category Fluency) assessments.

Concerning graph-theoretical metrics, AD-C misclassified as CN (AD-C–CN) had higher estimates of global path transitivity and small-world propensity compared to those misclassified

as AD (AD-C–AD). AD-C who were correctly classified, exhibited lower estimates of global clustering coefficient and global efficiency when compared to individuals predicted as CN (AD-C–CN). However, AD-C participants correctly classified did not significantly differ from those misclassified as AD (AD-C–AD) on any of the graph-theoretical metrics assessed. These findings suggest that AD-C participants who were correctly classified had similar network topology to AD participants but presented with different cognitive profiles.

The subset of AD-C participants correctly classified may be more likely to be diagnosed with a dysexecutive subtype of AD (once they convert) (Corriveau-Lecavalier et al., 2022; Mez et al., 2016; Townley et al., 2020; Vogel & Hansson, 2022), whereas those misclassified as AD more closely resemble typical AD participants with primary memory impairments. While non-memory-related impairments are more common in early-onset AD, late-onset non-amnestic cases are not uncommon (Harrington et al., 2013; Mez et al., 2016; Vogel & Hansson, 2022). An alternative interpretation is that, regardless of age, the small subset of AD-C participants correctly classified are just at an earlier stage in their trajectory toward developing typical AD. Though rare, cases where executive function deficits proceed memory impairment have been reported in individuals eventually diagnosed with typical AD (Harrington et al., 2013).

Moreover, AD converters misclassified as CN (AD-C–CN) may represent a set of cases with AD pathology in the absence of severe cognitive impairment (Roe et al., 2007; Santacruz et al., 2011; Visser & Tijms, 2017). Concerning diagnostic measures (not used for prediction), neither MMSE scores nor CDR-SB ratings were significantly different between AD-C participants correctly or incorrectly classified. Thus it is uncertain if the MMSE and CDR-SB can aid in predicting when and why otherwise cognitively normal individuals will convert to AD.

Critically, future work assessing AB and tau-related pathology will be required to test these assumptions and confirm diagnoses.

As noted above, classification was highly accurate for CN and AD participants. Thus, given the limited number of CN and AD participants misclassified, the differences reported are qualitative. As such, CN participants were equally likely to be misclassified as either AD-C (N=5) or AD (N=7). The predicted CN groups did not differ numerically in MMSE scores or CDR-SB ratings. Numerically, CN misclassified as AD-C or AD were older, less educated, and exhibited lower estimates of modularity and small-world propensity and higher estimates of participation coefficient. Of the 18 AD participants misclassified, 16 were misclassified as CN, and two were misclassified as AD-C. AD-C misclassified as CN were younger and had better performance across the battery of assessments, had higher MMSE scores, lower CDR-SB ratings, and exhibited network topology more similar to CN participants. Additional work will be required to validate the diagnosis of these individuals based on currently established biological criteria (Jack et al., 2018).

Addressing class-imbalance. Cognitively normal individuals were over-represented in the current dataset. Thus the predictions above may be overly biased toward correctly classifying CN participants. SMOTE resampling was used to oversample observations from the minority classes to account for this bias. As expected, SMOTE resampling improved overall classification performance. The final model shows that applying parameters learned during cross-validation to the total sample produced balanced predictions across a priori classes. Features were ranked relatively consistently between classification models with and without SMOTE resampling. However, with SMOTE resampling, age was the most predictive feature, and Trail Making Part B was ranked higher in importance than Logical Memory. This result is not unexpected. AD-C

participants were the oldest participants on average, and as described above, a subset of participants appeared to have more pronounced executive impairments. Likewise, in Aim 1, the interaction between Trail Making Part B performance and redundancy was specific to the AD-C group.

In addition, when artificially balancing the dataset, metrics including modularity, global participation coefficient, and small-world propensity contributed more to classification performance. As highlighted in the introduction, both aging and AD have been associated with distributed changes in functional brain network organization that can be linked to neural dedifferentiation, including (1) reduced clustering, (2) disrupted network integration and segregation, and (3) loss of small-world properties, each of which contributes to a less efficient system. Although global clustering coefficient was significantly lower, on average, in the AD-C and AD groups, it did not contribute to classification of participants into their *a priori* groups. Moreover, global communicability and global efficiency were not overall critical to model performance as hypothesized.

Caution is warranted, however, when interpreting model performance using SMOTE resampling. Resampling techniques may lead to biased prediction at the participant level. If one is interested in assessing participant-level predictions (e.g., predicted probabilities), additional work will be necessary to ensure models are adequately calibrated (Steyerberg et al., 2010) and validated out-of-sample, particularly if models are to be used for risk assessment (Goldstein et al., 2020). Notably, the AD-C group does not represent a single diagnostic entity. Participants in the AD-C group are progressively at risk because individuals will knowingly convert to AD at different times. Here, multiclass modeling without subsampling (Model 1) permits a targeted and unbiased investigation of why specific individuals may be classified into distinct classes.

Additionally, results may further highlight that cognitive impairment observed in otherwise cognitively normal individuals, who will convert to AD, is not restricted to memory deficits common in typical AD.

Aim 3

Aim 3 sought to establish whether combining core demographic factors, network metrics, and cognitive measures could identify data-driven neurocognitive subgroups using unsupervised learning. In the unsupervised random forest, neuropsychological assessments and graph-theoretical metrics equally differentiated the sample of participants from synthetic data. Applying cluster analysis to the proximity matrix, a two-cluster solution was voted most optimal, receiving 10 of 23 votes. Contrary to hypotheses, the derived clusters did not correspond to amnestic and non-amnestic cognitive subtypes. Rather, the sample was simply stratified into two groups of low (Cluster 1) and high (Cluster 2) cognitive performers.

Cluster 1 also had significantly lower MMSE scores and higher CDR-SB ratings. Cluster 1 was 13% CN (N = 12), 23% AD-C (N = 21), and 64% AD (N = 59). Cluster 2 was 72% CN (N = 147), 17% AD-C (N = 35), and 11% AD (N = 22). Following cluster analysis, post-hoc comparisons found that Clusters 1 and 2 significantly differed on all neuropsychological assessments and most graph-theoretical metrics (no differences on either global communicability (log-transformed) or within-module degree Z-score were observed). This finding, although contrary to hypotheses informed by prior work (Ferreira et al., 2019; Kwak et al., 2021; Scheltens et al., 2016, 2017, 2018; Vogel & Hansson, 2022), is not unexpected. That is, with respect to the *a priori* groups, AD participants exhibited more global than domain-specific cognitive impairment and scored significantly lower on all cognitive assessments. Cognitive performance was more heterogeneous among AD-C participants who were evenly split across

the detected clusters. Overall, AD-C (and CN) assigned to Cluster 1 are participants whose cognitive ability and network topology more closely resemble AD participants on average.

While the two-cluster solution was voted most optimal (10 of 23 votes), a three-cluster solution was a close second (8 of 23 votes). Since there are three *a priori* groups, exploratory analyses were conducted using the three-cluster solution. As with the two-cluster solution, Cluster 1 was again comprised primarily of AD participants (79%; N = 55). Consistent with both supervised and unsupervised models, AD-C participants were distributed across clusters. However, here, CN participants were split between Clusters 2 and 3. Cluster 1 (N = 70) was comprised of 5.7% CN (N = 4), 15.7% AD-C (N = 11), and 78.6% AD (N = 55). Cluster 2 (N = 132) was comprised of 71.2% CN (N = 94), 18.2% AD-C (N = 24), and 10.6% AD (N = 14). Cluster 3 (N = 94) was comprised of 64.9% CN (N = 61), 22.3% AD-C (N = 21), and 12.8% AD (N = 12). Cluster 1 was defined by lower scores across the battery of neuropsychological assessments. Clusters 2 and 3 were not significantly different on any cognitive assessments. Conversely, Clusters 1 and 2 had similar estimates of network topology and differed only on modularity (Cluster 2 higher). Numerically, Cluster 3 had higher estimates for all graph-theoretical metrics, except for the global participation coefficient. Of note, higher estimates of redundancy (e.g., global clustering coefficient and communicability) were also observed in Cluster 3.

The two- and three-cluster solutions are thus capturing different subsets of CN and AD-C participants who (1) are most similar to AD across the set of demographic, cognitive, and network features, or (2) may have similar cognitive profiles but differing network topology. Some AD-C participants showed cognitive impairment and altered network topology using baseline assessments consistent with the supervised model. Conversely, different subsets of AD-

C and CN participants show altered network topology without severe cognitive impairment that may impact daily functioning. When matched on cognitive performance as in Clusters 2 and 3, differences in network topology helped distinguish a subset of AD-C and CN participants.

As noted above, Cluster 2 had lower estimates for most graph-theoretical measures, including redundancy-based metrics. However, it is unclear if Cluster 2 contains a subset of non-demented participants with altered network topology at a higher risk of experiencing future cognitive impairment. Longitudinal work will be required to assess if the observed patterns track and or predict conversion to AD in these individuals. Recent work conducted in a sizeable cross-cohort study of older cognitively unimpaired participants (N = 1325; 7 Cohorts) found that CN, positive for both *Aß* and tau (biomarker levels determined via PET imaging), are at a greater risk for converting to MCI and potentially AD over a 3-5-year period (Ossenkoppele et al., 2022). Thus, future work can assess if functional network topology, can help distinguish and/or predict if non-demented CN and AD-C participants who have a positive biomarker status will progress to MCI or AD. Modularity, global participation coefficient, and to an extent small-world propensity, may be particularly important.

As highlighted in the introduction, with age, functional networks tend to become more integrated (e.g. less modular). This increase in functional network integration maybe compensatory, enabling efficient use of available neural resources when confronted with challenging cognitive demands (Avelar-Pereira et al., 2017; Chan et al., 2017; Grady, 2017). The ability to maintain an optimal balance of network integration and segregations, further compromised in AD, is mechanistically tied to the disruption of neuromodulatory systems (e.g. cholinergic hypothesis of AD). As discussed further below, AD converters exhibited higher

levels of modularity and levels of global participation coefficient, which was the opposite pattern found in AD participants.

General Conclusion

Consistent with prior work combining machine learning and network neuroscience, classification models accurately distinguished cognitively normal aging individuals from those with AD. However, neuropsychological assessments were more predictive than graph-theoretical metrics assessing global aspects of network topology. As hypothesized, AD participants exhibited the most significant memory-related impairments (Logical Memory; memory recall). However, AD participants scored significantly lower on all neuropsychological assessments administered by the OASIS-3 consortium. Thus, in AD participants, a pattern of global rather than domain-specific (e.g., memory) cognitive impairment was observed. This result may not be all that unexpected, as neuropsychological assessments are designed to detect more broad and severe forms of cognitive impairment due to neurodegenerative disease (Horowitz et al., 2018; Silverstein, 2008). In typical AD and amnestic MCI, memory-related cognitive impairments are often detected earlier and tend to decline faster (Caselli et al., 2020; Mistridis et al., 2015; Wilson et al., 2011). The average age of participants included in this study was 74. Participants in the AD-C and AD groups were older than CN on average (CN: M = 71, SD = 6; AD-C: M = 79, SD = 8; AD = 76, SD = 8). With older age and in advanced stages of dementia due to AD, global cognitive impairment is more common (Amieva et al., 2014; Ferreira et al., 2020; Qiu et al., 2019; Wilson et al., 2011).

Critically, although both supervised and unsupervised classification models could distinguish CN from AD participants, non-demented individuals who will convert to AD were more challenging to classify. Approximately half of the AD-C participants were misclassified as

CN, while the remaining half were classified correctly as converters or misclassified as AD. Although the current study did not uncover any of the cognitive subtypes previously identified in either CN or AD participants, one subset of AD converters may resemble the non-amnestic/dysexecutive subtype of AD (Corriveau-Lecavalier et al., 2022; Mez et al., 2016; Townley et al., 2020), while another subset may be best defined by typical amnestic AD. Specifically, correctly classified AD-C participants were the oldest participants and showed more significant impairments in executive functioning and other related cognitive processes (attention, processing speed). This subset of AD-C participants showed comparable performance to AD-C classified as CN (AD-C–CN) on most memory- and fluency-based assessments. Conversely, AD converters misclassified as AD (AD-C–AD) exhibited more significant memory-related impairments and network topology more similar to AD participants. Critically, AD-C participants classified correctly versus incorrectly did not statistically differ in the number of days between their cognitive assessment and first fMRI scan. Thus, whether AD-C participants were correctly or incorrectly classified cannot be attributed to differences in time of assessment.

Concerning network topology, in classification analyses, modularity and global participation coefficient were consistently more predictive than other graph-theoretical metrics assessed. Moreover, when comparing general patterns of network topology across the CN, AD-C, and AD groups, post-hoc analyses identified several unique patterns. In general, compared to the CN group, AD-C and AD participants exhibited reduced small-world properties, including reduced levels of clustering coefficient and global efficiency consistent with prior work (Dyrba et al., 2015; Hallquist & Hillary, 2018; Khazaee et al., 2015; Y. P. Li et al., 2013; Sanz-Arigita et al., 2010; Wang et al., 2013; Zhao et al., 2012). In the current study, network topology was more segregated and less efficient in the AD-C group when compared to CN participants.

Interestingly, the overall pattern of increased network segregation in older AD-C participants is *inconsistent* with aging studies that find greater functional network integration in older non-demented adults (Bagarinao et al., 2019; Foo et al., 2021; Madden et al., 2020; Stumme et al., 2020; Voss et al., 2013) that support compensatory cognitive processes (Chan et al., 2017; Fornito et al., 2015; Meunier et al., 2014; Sala-Llonch, BartrÃs-Faz, et al., 2015). On the other hand, consistent with aging studies, AD participants exhibited greater network integration overall (lowest estimates of modularity, highest estimates of global participation coefficient) compared to the AD-C group and reduced clustering and global efficiency compared to CN participants. The contrasting patterns of network integration/segregation in the AD-C and AD groups may reflect the presence and or magnitude of pathological burden. Specifically, as highlighted in the introduction, the spread of pathological tau is activity-dependent (e.g., neural activation). Prior work suggests that early *Aß* deposition is associated with an initial stage of hyperconnectivity in specific association networks (e.g., default mode, salience, dorsal attention), followed by periods of hypoconnectivity in these same networks due to increasing levels of subcortical and neocortical tau (Jones et al., 2016; Keller & Christopher, 2017; Schultz et al., 2017). Thus, differences in network topology between AD-C and AD participants may represent a biologically plausible reorganization of functional brain networks attributed to the dynamic interplay of AD pathology. Knowledge of biomarker status (e.g., AB+/tau+; AB-/tau+, etc.) can help test these assumptions systematically. Importantly, however, differences in network topology observed between groups (*a priori* [Aim 2] or data-driven [Aim 3]) were small in magnitude. Moreover, graph-theoretical metrics did not significantly contribute to classification performance. Thus caution is warranted when interpreting differences in network topology between the *a priori* [Aim 2] or data-driven [Aim 3] groups.

Overall, findings may suggest that estimates of global network topology derived using functional MRI may not be informative when diagnosing AD or identifying cognitively normal individuals who will convert to AD in older cohorts. This finding may be relevant to clinicians and researchers alike as MRI scans are costly and time-consuming. Here, neuropsychological assessments may be sufficient if the goal is to diagnose AD in older participants. However, an additional consideration is that global estimates of network topology may obscure regional and network-level changes that track with pathology. Prior work combining machine learning and graph theory found that regional metrics may be more sensitive when classifying individuals with and without AD (Hojjati et al., 2019; Khazaee et al., 2015, 2017; Xu et al., 2020) or predicting conversion from MCI to AD (Hojjati et al., 2017; Sun et al., 2019; Wei et al., 2016; T. Zhang et al., 2021). Future work can expand these analyses to combine global and nodal graph-theoretical metrics as in previous work. However, it will be critical that such studies incorporate multiple measures of cognitive functioning capable of capturing divergent neurocognitive profiles and heterogeneous disease pathways that may or may not correspond with established biologically-based subtypes of AD. Thus more work is needed to understand how network metrics impact prediction and subgroup detection, especially at an earlier age and in earlier stages of the disease process before the cumulative impacts of aging and AD-related pathology have sufficiently altered network topology.

Interestingly, global communicability, global clustering coefficient, and education level did not significantly contribute to classification performance. As discussed in the introduction, some individuals are more resilient to AD pathology. Thus, markers of brain and cognitive reserve (e.g., redundancy, education level, etc.) may also be more discriminative at earlier points in time before the cumulative effects of age- and AD-related pathologies have reached a point of

criticality. As discussed above, although graph-theoretical metrics were not as informative as neuropsychological assessments in this sample, altered network topology observed in AD-C and AD participants can theoretically be linked to the opposing effects of *Aß* and tau. The utility of subtle but specific changes in network topology may be essential for classifying converters that do not necessarily conform to any of the previously identified subtypes or resemble the typical amnestic form of AD. Additionally, the present study focused on cognitive functioning and global network topology. Thus, more work is needed to tease apart how differences in functional network topology, observed at the regional, network, and global levels, are associated with pathological change and emerging cognitive impairment. Finally, it will also be critical to consider additional factors that can shape the trajectory and rate of cognitive decline in otherwise cognitively normal individuals at risk of developing AD, such as vascular and metabolic risk factors (diabetes, high blood pressure, obesity, etc.) (Hughes & Craft, 2016; MacIntosh et al., 2020; Perera et al., 2020).

Limitations and Future Directions

A fundamental limitation of this study is the lack of participants under the age of 65 with cognitive and neuroimaging data. Except for a small subset of AD-C participants, as discussed above, there was no clear indication that participants could be stratified into any of the previously defined cognitive subgroups. Participants with atypical and/or non-amnestic subtypes of AD are, on average younger individuals with early-onset dementia (<60 years of age) (Townley et al., 2020; Vogel & Hansson, 2022). Thus the ability to detect cognitive subgroups may have been impacted by the older age range of the sample. Moreover, the current study took exceptional care to exclude individuals without a primary diagnosis of AD and without additional comorbidities. The extent to which this impacted results is unclear.

Analytically, the number of subgroups detected may depend not only on the age or size of the sample but also on the number and scope of cognitive assessments administered, choice of statistical analysis (parametric, non-parametric), and whether or not the groups have been cross-validated (Ferreira et al., 2020). The lack of correspondence to previous work may be due to differing methodology. The unsupervised random forest attempts to best partition a dataset with respect to randomly generated synthetic data. The two-cluster solution simply maximized differences across the full set of features. Thus, with a limited number of features and clear set of [cognitive] patterns between groups, the random forest-based approach may not be optimal. In future work, cluster analysis can be performed directly on the raw data consistent with prior studies (Kwak et al., 2021; Scheltens et al., 2016, 2017, 2018). The resulting clusters can then be compared with the proximity-based approach taken here to assess the robustness and clinical utility of the derived solutions more formally.

Likewise, in the present study, 31 of 56 AD-C participants had global CDR scores of .5 or higher and would thus likely be diagnosed with MCI if using currently established thresholds. Although neither global CDR scores or CDR-SB ratings were different between the predicted AD-C groups, future work can assess the degree to which a formal diagnosis of MCI impacts classification of non-demented individuals who convert to AD. However, as discussed in the introduction, caution may be warranted with relying on traditional diagnostic criteria that are susceptible to false-positives to define MCI (Bondi et al., 2008, 2014; Edmonds et al., 2015, 2021; Jacobson et al., 2009). More comprehensive approaches that combine neuropsychological scores, assessment of daily-functioning activities, and biological data are more likely to yield improved early identification of impaired individuals, particularly those who do not conform to

conventional or typical subtypes of either MCI or AD (Edmonds et al., 2015, 2021; Ferreira et al., 2020; Jellinger, 2021).

Further, as previously discussed, neuropsychological assessments are designed to detect broad, non-specific cognitive impairments and simultaneously tap multiple executive and non-executive processes. Such assessments are also prone to floor and ceiling effects that can distort the true nature and magnitude of cognitive impairment detected between healthy participants and individuals with or at risk of developing AD (Duff, 2012; Duff et al., 2011). Gomez-Isla and colleagues (2019) noted that while the diagnosis of AD has become more reliant on biomarkers, several studies have failed to find consistent associations between markers of AD pathology and cognitive impairment or decline (Gomez-Isla & Frosch, 2019). Although several health-related factors may contribute to the lack of association, the non-specific nature and lack of sensitivity of standard neuropsychological assessments may also limit their predictive utility in younger, more diverse samples. Thus, future work should consider cognitive assessments that permit assessing specific subcomponents of cognition and allow for a deconstruction of performance based on accuracy and response time distributions. Moving the focus beyond single summary scores that describe gross characteristics of cognitive functioning may enable researchers to capture both general and specific cognitive impairments earlier in the time course of AD. In this study, 54% of AD converters were classified as CN; thus, more sensitive cognitive measures may be needed to capture the emergence of cognitive impairment and to help identify individuals at greater risk of developing dementia.

In Aim 1, linear models did not control for in-scanner motion or core demographic risk factors (age, sex, education). Future work can formally assess if linear relationships between weighted redundancy-based metrics and cognition are dependent on these factors. In the present

sample, the number of timepoints remaining after motion censoring (see *Image Processing*) was well beyond guidelines that recommend an average of 5-7 minutes (minimum) when estimating functional connectivity (Birn et al., 2013). Regardless*, a priori* groups differed on average in-scanner motion, driven by *lower* motion in the AD-C group. No motion-related differences were observed between CN and AD participants. Additional work may thus be needed to understand if motion effects before and after motion correction contribute to null findings in the CN and AD groups. Importantly, core demographic risk factors are respectively associated with the progression of AD pathology, aging and AD-related cognitive decline, and general prevalence of AD in the population. As noted above, education level and sex were not overall important for classifying individuals with or without AD. However, this may be attributed to several factors including, but not limited to the older age of the sample and the cross-sectional nature of the analyses as discussed above. Education level and sex may be particularly important for predicting risk of AD when both the onset and progression of additional health-related risk factors (e.g., cardiometabolic disease, etc.) can be accounted for (Altmann et al., 2014; Hasselgren et al., 2020; Letenneur et al., 1999). An additional consideration is that controlling for many influential variables associated with heterogeneous disease processes such as AD may lead to biased inference. To better account for omitted variable bias, measures that quantify pathology need to be explicitly modeled via additional interaction terms. In such cases, path-based (e.g., Structural-Equation Modeling) analyses, non-parametric models, or machine learning algorithms (e.g., random forest) that handle higher-order interaction terms may be most appropriate. Specifically, in addition to assessing the influence of individual covariates in a null hypothesis testing framework, higher-order interactions in machine learning models can be further investigated, as can marginal effects using partial dependence, automated local effects,

and individual conditional expectation analyses (Hastie et al., 2013; Henninger et al., 2022; Welchowski et al., 2022). In classification problems, these methods quantify the probability that an observation will be assigned to a given class at specific values for a given pair of features. In addition, model-based feature selection or reduction methods (see the *Machine Learning in AD* section) can be applied during cross-validation to eliminate features that do not consistently contribute to model performance, subsequently reducing the number of possible interactions to explore.

In addition to age, sex, and education, Aβ, tau, and genes encoding the ApoE protein interact to determine an individual's level of risk for developing AD (Safieh et al., 2019). Compared to non-carriers, ε4 carriers may exhibit accelerated rates of cognitive decline and develop dementia at an earlier age (Reas et al., 2019; Scheller et al., 2018). This risk is two-fold for individuals homozygous for the ε4 allele (e.g., having two copies; 44). More work is thus needed to understand how pathological AD markers, including ApoE status, interact with demographic risk factors, cognitive performance, and network topology to predict conversion in non-demented individuals. Moreover, the OASIS dataset is 84.3% white non-Hispanic. Black and other non-white individuals are at an increased risk of developing AD due to cardiovascular complications and comorbid pathologies (diabetes, obesity, etc.) (Lockhart et al., 2022), as well as other non-biological factors, including lack of health care access and social support structures (“2021 Alzheimer’s Disease Facts and Figures,” 2021; Barnes & Bennett, 2014; Chin et al., 2011). Future work should strive to ensure samples are representative of the population and that any proposed subtypes of AD, defined biologically and/or cognitively, capture variation due to race and ethnicity.

As noted above, future work can incorporate graph-theoretical metrics estimated at the nodal level to account for regional differences associated with the spread of pathology. Related, communicability and network redundancy are edgewise metrics, meaning there is an estimate of redundancy for every connection within a given graph. Though correlated at the global level, metrics that quantify redundancy in functional networks may not equally weight individual nodes or regions. Thus, global descriptions of redundancy, which take the sum or average overall connections, may distort the true variability present across regions and networks that track specific cognitive functions impacted by aging and AD. Furthermore, global summary scores of edgewise metrics may be highly correlated with average network strength (e.g., weighted degree) and/or mean estimates of functional connectivity. That is, while communicability in particular attempts to reduce bias driven by high-degree nodes (Andreotti et al., 2014; Estrada & Hatano, 2008), summing across connections, including the diagonal entries of the communicability matrix containing estimates of node centrality, may negate any degree-based normalization (Oldham et al., 2019).

Related, the present study strategically analyzed weighted graph-theoretical metrics to circumvent density-related issues with graph thresholding, as described in the introduction. However, an additional concern is that fully weighted unthresholded matrices contain noisy, potentially spurious connections (van den Heuvel et al., 2017). Future work can assess if weighted graph-theoretical metrics found to differ between patient and non-patient groups are robust to estimation on graphs with and without the use of thresholding. Here, single-thresholding methods, such as the minimal spanning tree (Dimitriadis et al., 2017), may be preferred over traditional thresholding approaches (e.g., absolute and proportional thresholding) that require multiple statistical comparisons.

As data-driven models, informed by theory, grow in complexity, it will be increasingly important to evaluate if models generalize out-of-sample. In the review by Grueso and colleagues (2021), only 3 of 234 studies using machine learning to predict conversion to AD (see the *Machine Learning & AD* section) were able to validate their models in a completely independent external dataset (Grueso & Viejo-Sobera, 2021). Although the current study utilized repeated 5-fold cross-validation to assess average classification performance and improve model generalizability, models were not externally validated. For prediction models to be clinically useful, they must be deployed and validated in independent and diverse datasets.

Concerning feature importance, the cutoff used to identify highly correlated variables was chosen somewhat arbitrarily given no single optimal threshold exists. The chosen threshold may thus be too lenient. In random forest classification, subsets of features are randomly selected and used for splitting a pair of observations at each node within a given decision tree (see *Random Forest Application*). Given the lack of correlation between decision trees in a random forest ensemble, classification performance is relatively robust to the presence of correlated features (Breiman, 2001). However, given the limited number of features included in the present analyses, feature importance rankings may be impacted (Genuer et al., 2010; Strobl et al., 2008). Thus, future work can ensure feature importance scores are robust to multicollinearity problems by testing a range of exclusionary thresholds (.75-.95) commonly used in the machine learning literature (Dormann et al., 2013).

Likewise, exploratory post-hoc analyses were not corrected for multiple comparisons to facilitate discovery of general trends emerging in secondary analyses (Murray et al., 2011; Scheltens et al., 2017). Thus it is uncertain if any inferences from the present results generalize to independent data, and caution is warranted with overinterpretation of mean differences

reported between predicted groups. Finally, the random forest is one of several versatile and robust machine learning algorithms. Future work can assess how different machine learning algorithms may boost prediction performance and improve early detection of individuals at a higher risk of converting to AD later in life.

In conclusion, this project aimed to identify and assess markers of reserve and resilience in functional brain networks that can help differentiate typical from atypical age-related cognitive impairment. Analyses sought to illuminate high-dimensional relationships between metrics that capture functional brain network topology and multiple cognitive domains using a large-high-quality open-access neuroimaging dataset. Here, weighted redundancy-based graph-theoretical metrics did not represent general markers of reserve and were not overall associated with cognitive functioning. The combination of demographic factors, cognitive performance measures, and graph-theoretical metrics summarizing global network topology could accurately distinguish CN from AD participants. However, overall classification performance was modest given that the AD-C group did not represent a single diagnostic entity. When classifying older individuals with and without AD, neuropsychological measures were more informative than metrics assessing global network topology, including redundancy-based measures. Among the graph-theoretical metrics assessed, modularity, global participation coefficient, and small-world propensity measures were most predictive and helped distinguish AD-C participants correctly classified or misclassified as either CN or AD. Importantly, in AD converters (AD-C), functional network redundancy was predictive of better executive functioning. Further, AD converters classified correctly exhibited reduced redundancy on average compared to converters misclassified as either CN or AD. Finally, while this study failed to identify cognitive subgroups previously reported in MCI and AD participants, AD converters correctly classified or

misclassified as AD show diverging neurocognitive profiles and may represent a subset of individuals with primarily executive, as opposed to memory-related impairments, respectively. Future longitudinal work incorporating pathological markers of AD will be required however to assess the robustness of these findings.

Ultimately, it is hoped that future extensions of this work will contribute to early-detection research and facilitate the discovery of neurocognitive profiles that better capture individuals at risk of experiencing mild to more severe forms of cognitive impairment who are at greater risk of developing dementia.

REFERENCES

2021 Alzheimer's disease facts and figures. (2021). *Alzheimer's and Dementia*, *17*(3), 327–406. https://doi.org/10.1002/alz.12328

Achard, S., & Bullmore, E. (2007). Efficiency and cost of economical brain functional networks. *PLoS Computational Biology*, *3*(2), 0174–0183. https://doi.org/10.1371/journal.pcbi.0030017

Alexander, N., Alexander, D. C., Barkhof, F., & Denaxas, S. (2021). Identifying and evaluating clinical subtypes of Alzheimer's disease in care electronic health records using unsupervised machine learning. *BMC Medical Informatics and Decision Making*, *21*(1). https://doi.org/10.1186/s12911-021-01693-6

Altini, M. (2015). *Dealing with imbalanced data: undersampling, oversampling and proper cross-validation*. https://www.marcoaltini.com/blog/dealing-with-imbalanced-data-undersampling-oversampling-and-proper-cross-validation

Altman, N., & Krzywinski, M. (2017). Points of Significance: Ensemble methods: bagging and random forests. *Nature Methods*, *14*(10), 933–934. https://doi.org/10.1038/nmeth.4438

Altmann, A., Tian, L., Henderson, V. W., & Greicius, M. D. (2014). Sex modifies the APOE-related risk of developing Alzheimer disease. *Annals of Neurology*, *75*(4), 563–573. https://doi.org/10.1002/ana.24135

Amieva, H., Mokri, H., Le Goff, M., Meillon, C., Jacqmin-Gadda, H., Foubert-Samier, A., Orgogozo, J. M., Stern, Y., & Dartigues, J. F. (2014). Compensatory mechanisms in higher-educated subjects with Alzheimer's disease: A study of 20 years of cognitive decline. *Brain*, *137*(4), 1167–1175. https://doi.org/10.1093/brain/awu035

Anderson, N. D., & Craik, F. I. M. (2017). 50 Years of Cognitive Aging Theory. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, *72*(1), 1–6. https://doi.org/10.1093/geronb/gbw108

Andreotti, J., Jann, K., Melie-Garcia, L., Giezendanner, S., Abela, E., Wiest, R., Dierks, T., & Federspiel, A. (2014). Validation of network communicability metrics for the analysis of brain structural networks. *PLoS ONE*, *9*(12). https://doi.org/10.1371/journal.pone.0115503

Ang, T. F. A., An, N., Ding, H., Devine, S., Auerbach, S. H., Massaro, J., Joshi, P., Liu, X., Liu, Y., Mahon, E., Au, R., & Lin, H. (2019). Using data science to diagnose and characterize heterogeneity of Alzheimer's disease. *Alzheimer's and Dementia: Translational Research and Clinical Interventions*, *5*, 264–271. https://doi.org/10.1016/J.TRCI.2019.05.002

Arenaza-Urquijo, E. M., & Vemuri, P. (2018). Resistance vs resilience to Alzheimer disease: Clarifying terminology for preclinical studies. *Neurology*, *90*(15), 695. https://doi.org/10.1212/WNL.0000000000005303

Au, A., Feher, A., McPhee, L., Jessa, A., Oh, S., & Einstein, G. (2016). Estrogens, inflammation

and cognition. *Frontiers in Neuroendocrinology*, *40*, 87–100. https://doi.org/10.1016/j.yfrne.2016.01.002

Avelar-Pereira, B., Bäckman, L., Wåhlin, A., Nyberg, L., & Salami, A. (2017). Age-Related Differences in Dynamic Interactions Among Default Mode, Frontoparietal Control, and Dorsal Attention Networks during Resting-State and Interference Resolution. *Frontiers in Aging Neuroscience*, *9*, 152. https://doi.org/10.3389/fnagi.2017.00152

Bagarinao, E., Watanabe, H., Maesawa, S., Mori, D., Hara, K., Kawabata, K., Yoneyama, N., Ohdake, R., Imai, K., Masuda, M., Yokoi, T., Ogura, A., Taoka, T., Koyama, S., Tanabe, H. C., Katsuno, M., Wakabayashi, T., Kuzuya, M., Ozaki, N., … Sobue, G. (2019). Reorganization of brain networks and its association with general cognitive performance over the adult lifespan. *Scientific Reports*, *9*(1). https://doi.org/10.1038/S41598-019-47922-X

Baker, J. E., Lim, Y. Y., Pietrzak, R. H., Hassenstab, J., Snyder, P. J., Masters, C. L., & Maruff, P. (2017). Cognitive impairment and decline in cognitively normal older adults with high amyloid-β: A meta-analysis. *Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring*, *6*, 108–121. https://doi.org/10.1016/j.dadm.2016.09.002

Barnes, L. L., & Bennett, D. A. (2014). Alzheimer's disease in African Americans: Risk factors and challenges for the future. *Health Affairs*, *33*(4), 580–586. https://doi.org/10.1377/hlthaff.2013.1353

Baron, I. S. (2004). Delis-Kaplan Executive Function System. In *Child Neuropsychology* (Vol. 10, Issue 2). Psychological Corp. https://doi.org/10.1080/09297040490911140

Baronchelli, A., Ferrer-i-Cancho, R., Pastor-Satorras, R., Chater, N., & Christiansen, M. H. (2013). *Networks in Cognitive Science*. *17*(7), 348–360. https://doi.org/10.1016/j.tics.2013.04.010

Bartus, R. T., Dean, R. L., Beer, B., & Lippa, A. S. (1982). The cholinergic hypothesis of geriatric memory dysfunction. In *Science* (Vol. 217, Issue 4558, pp. 408–417). Science. https://doi.org/10.1126/science.7046051

Beam, C. R., Kaneshiro, C., Jang, J. Y., Reynolds, C. A., Pedersen, N. L., & Gatz, M. (2018). Differences between Women and Men in Incidence Rates of Dementia and Alzheimer's Disease. *Journal of Alzheimer's Disease*, *64*(4), 1077–1083. https://doi.org/10.3233/JAD-180141

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, *57*(1), 289–300. https://doi.org/10.2307/2346101

Bentley, P., Driver, J., & Dolan, R. J. (2011). Cholinergic modulation of cognition: Insights from human pharmacological functional neuroimaging. In *Progress in Neurobiology* (Vol. 94, Issue 4, pp. 360–388). Elsevier. https://doi.org/10.1016/j.pneurobio.2011.06.002

Betzel, R. F., Byrge, L., He, Y., Goñi, J., Zuo, X. N., & Sporns, O. (2014). Changes in structural and functional connectivity among resting-state networks across the human lifespan. *NeuroImage*, *102*(P2), 345–357. https://doi.org/10.1016/j.neuroimage.2014.07.067

Betzel, R. F., Gu, S., Medaglia, J. D., Pasqualetti, F., & Bassett, D. S. (2016). Optimally controlling the human connectome: The role of network topology. *Scientific Reports*, *6*(1), 30770. https://doi.org/10.1038/srep30770

Bickerton, W. L., Samson, D., Williamson, J., & Humphreys, G. W. (2011). Separating Forms of Neglect Using the Apples Test Validation and Functional Prediction in Chronic and Acute Stroke. *Neuropsychology*, *25*(5), 567–580. https://doi.org/10.1037/a0023501

Birn, R. M., Molloy, E. K., Patriat, R., Parker, T., Meier, T. B., Kirk, G. R., Nair, V. A., Meyerand, M. E., & Prabhakaran, V. (2013). The effect of scan length on the reliability of resting-state fMRI connectivity estimates. *NeuroImage*, *83*, 550. https://doi.org/10.1016/J.NEUROIMAGE.2013.05.099

Bisiacchi, P. S., Tarantino, V., & Ciccola, A. (2008). Aging and prospective memory: The role of working memory and monitoring processes. *Aging Clinical and Experimental Research*, *20*(6), 569–577. https://doi.org/10.1007/BF03324886

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Bondi, M. W., Edmonds, E. C., Jak, A. J., Clark, L. R., Delano-Wood, L., McDonald, C. R., Nation, D. A., Libon, D. J., Au, R., Galasko, D., & Salmon, D. P. (2014). Neuropsychological Criteria for Mild Cognitive Impairment Improves Diagnostic Precision, Biomarker Associations, and Progression Rates. *Journal of Alzheimer's Disease*, *42*(1), 275–289. https://doi.org/10.3233/JAD-140276

Bondi, M. W., Jak, A. J., Delano-Wood, L., Jacobson, M. W., Delis, D. C., & Salmon, D. P. (2008). Neuropsychological contributions to the early identification of Alzheimer's disease. *Neuropsychology Review*, *18*(1), 73–90. https://doi.org/10.1007/s11065-008-9054-1

Borra, S., & Di Ciaccio, A. (2010). Measuring the prediction error. A comparison of cross-validation, bootstrap and covariance penalty methods. *Computational Statistics and Data Analysis*, *54*(12), 2976–2989. https://doi.org/10.1016/j.csda.2010.03.004

Braak, H., & Braak, E. (1991). Neuropathological stageing of Alzheimer-related changes. In *Acta Neuropathologica* (Vol. 82, Issue 4, pp. 239–259). Springer-Verlag. https://doi.org/10.1007/BF00308809

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Bressler, S. L., & Tognoli, E. (2006). Operational principles of neurocognitive networks. *International Journal of Psychophysiology*, *60*(2), 139–148.

https://doi.org/10.1016/j.ijpsycho.2005.12.008

Briand, B., Ducharme, G. R., Parache, V., & Mercat-Rommens, C. (2009). A similarity measure to assess the stability of classification trees. *Computational Statistics and Data Analysis*, *53*(4), 1208–1217. https://doi.org/10.1016/j.csda.2008.10.033

Bronzuoli, M. R., Iacomino, A., Steardo, L., & Scuderi, C. (2016). Targeting neuroinflammation in Alzheimer's disease. In *Journal of Inflammation Research* (Vol. 9, pp. 199–208). Dove Medical Press Ltd. https://doi.org/10.2147/JIR.S86958

Bu, G. (2009). Apolipoprotein E and its receptors in Alzheimer's disease: pathways, pathogenesis and therapy. *Nature Reviews. Neuroscience*, *10*(5), 333–344. https://doi.org/10.1038/nrn2620

Bullmore, E., & Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, *10*(3), 186–198. https://doi.org/10.1038/nrn2575

Burke, S. N., Mormino, E. C., Rogalski, E. J., Kawas, C. H., Willis, R. J., & Park, D. C. (2019). What are the later life contributions to reserve, resilience, and compensation? *Neurobiology of Aging*, *83*, 140–144. https://doi.org/10.1016/J.NEUROBIOLAGING.2019.03.023

Bzdok, D., Altman, N., & Krzywinski, M. (2018). Points of Significance: Statistics versus machine learning. *Nature Methods*, *15*(4), 233–234. https://doi.org/10.1038/nmeth.4642

Cabeza, R., Nyberg, L., & Park, D. C. (2009). Cognitive Neuroscience of Aging: Emergence of a New Discipline. In *Cognitive Neuroscience of Aging: Linking cognitive and cerebral aging* (pp. 3–15). https://doi.org/10.1093/acprof:oso/9780195156744.003.0001

Campbell, K. L., Hasher, L., & Thomas, R. C. (2010). Hyper-binding: A unique age effect. *Psychological Science*, *21*(3), 399–405. https://doi.org/10.1177/0956797609359910

Cao, M., Wang, J. H., Dai, Z. J., Cao, X. Y., Jiang, L. L., Fan, F. M., Song, X. W., Xia, M. R., Shu, N., Dong, Q., Milham, M. P., Castellanos, F. X., Zuo, X. N., & He, Y. (2014). Topological organization of the human brain functional connectome across the lifespan. *Developmental Cognitive Neuroscience*, *7*, 76–93. https://doi.org/10.1016/j.dcn.2013.11.004

Caselli, R. J., Langlais, B. T., Dueck, A. C., Chen, Y., Su, Y., Locke, D. E. C., Woodruff, B. K., & Reiman, E. M. (2020). Neuropsychological decline up to 20 years before incident mild cognitive impairment. *Alzheimer's and Dementia*, *16*(3), 512–523. https://doi.org/10.1016/j.jalz.2019.09.085

Chan, M. Y., Alhazmi, F. H., Park, D. C., Savalia, N. K., & Wig, G. S. (2017). Resting-State Network Topology Differentiates Task Signals across the Adult Life Span. *The Journal of Neuroscience*, *37*(10), 2734–2745. https://doi.org/10.1523/JNEUROSCI.2406-16.2017

Charrad, M., Ghazzali, N., Boiteau, V., & Maintainer, A. N. (2015). Determining the Best

Number of Clusters in a Data Set. In *Cran*. https://sites.google.com/site/malikacharrad/research/nbclust-package

Chin, A. L., Negash, S., & Hamilton, R. (2011). Diversity and disparity in dementia: The impact of ethnoracial differences in Alzheimer disease. *Alzheimer Disease and Associated Disorders*, *25*(3), 187–195. https://doi.org/10.1097/WAD.0b013e318211c6c9

Ciric, R., Wolf, D. H., Power, J. D., Roalf, D. R., Baum, G. L., Ruparel, K., Shinohara, R. T., Elliott, M. A., Eickhoff, S. B., Davatzikos, C., Gur, R. C., Gur, R. E., Bassett, D. S., & Satterthwaite, T. D. (2017). Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. *NeuroImage*, *154*(March), 174–187. https://doi.org/10.1016/j.neuroimage.2017.03.020

Combrisson, E., & Jerbi, K. (2015). Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *Journal of Neuroscience Methods*, *250*, 126–136. https://doi.org/10.1016/j.jneumeth.2015.01.010

Conejero-Goldberg, C., Gomar, J. J., Bobes-Bascaran, T., Hyde, T. M., Kleinman, J. E., Herman, M. M., Chen, S., Davies, P., & Goldberg, T. E. (2014). APOE2 enhances neuroprotection against alzheimer's disease through multiple molecular mechanisms. *Molecular Psychiatry*, *19*(11), 1243–1250. https://doi.org/10.1038/mp.2013.194

Contador, I., Bermejo-Pareja, F., Pablos, D. L., Villarejo, A., & Benito-León, J. (2017). High education accelerates cognitive decline in dementia: A brief report from the population-based NEDICES cohort. *Dementia e Neuropsychologia*, *11*(3), 297–300. https://doi.org/10.1590/1980-57642016dn11-030012

Corriveau-Lecavalier, N., Machulda, M. M., Botha, H., Graff-Radford, J., Knopman, D. S., Lowe, V. J., Fields, J. A., Stricker, N. H., Boeve, B. F., Jack, C. R., Petersen, R. C., & Jones, D. T. (2022). Phenotypic subtypes of progressive dysexecutive syndrome due to Alzheimer's disease: a series of clinical cases. *Journal of Neurology*, *1*, 1–19. https://doi.org/10.1007/S00415-022-11025-X/FIGURES/2

Costafreda, S. G., Dinov, I. D., Tu, Z., Shi, Y., Liu, C. Y., Kloszewska, I., Mecocci, P., Soininen, H., Tsolaki, M., Vellas, B., Wahlund, L. O., Spenger, C., Toga, A. W., Lovestone, S., & Simmons, A. (2011). Automated hippocampal shape analysis predicts the onset of dementia in mild cognitive impairment. *NeuroImage*, *56*(1), 212–219. https://doi.org/10.1016/j.neuroimage.2011.01.050

Craik, F. I. M., & Bialystok, E. (2006). Cognition through the lifespan: Mechanisms of change. *Trends in Cognitive Sciences*, *10*(3), 131–138. https://doi.org/10.1016/j.tics.2006.01.007

Crane, P. K., Carle, A., Gibbons, L. E., Insel, P., Mackin, R. S., Gross, A., Jones, R. N., Mukherjee, S., Curtis, S. M. K., Harvey, D., Weiner, M., & Mungas, D. (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

Crofts, J. J., & Higham, D. J. (2009). A weighted communicability measure applied to complex brain networks. *Journal of the Royal Society Interface*, *6*(33), 411–414. https://doi.org/10.1098/rsif.2008.0484

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, *88*(11), 2783–2792. https://doi.org/10.1890/07-0539.1

del Carmen Díaz-Mardomingo, M., García-Herranz, S., Rodríguez-Fernández, R., Venero, C., & Peraita, H. (2017). Problems in classifying mild cognitive impairment (MCI): One or multiple syndromes? *Brain Sciences*, *7*(9). https://doi.org/10.3390/brainsci7090111

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics*, *44*(3), 837. https://doi.org/10.2307/2531595

Denisko, D., & Hoffman, M. M. (2018). Classification and interaction in random forests. *Proceedings of the National Academy of Sciences*, *115*(8), 1690–1692. https://doi.org/10.1073/pnas.1800256115

Di Lanzo, C., Marzetti, L., Zappasodi, F., De Vico Fallani, F., & Pizzella, V. (2012). Redundancy as a graph-based index of frequency specific MEG functional connectivity. *Computational and Mathematical Methods in Medicine*, *2012*, 1–9. https://doi.org/10.1155/2012/207305

Dimitriadis, S. I., Salis, C., Tarnanas, I., & Linden, D. E. (2017). Topological filtering of dynamic functional brain networks unfolds informative chronnectomics: A novel data-driven thresholding scheme based on orthogonal minimal spanning trees (OMSTs). *Frontiers in Neuroinformatics*, *11*, 28. https://doi.org/10.3389/fninf.2017.00028

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., Mcclean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 027–046. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Dosenbach, N. U. F., Koller, J. M., Earl, E. A., Miranda-Dominguez, O., Klein, R. L., Van, A. N., Snyder, A. Z., Nagel, B. J., Nigg, J. T., Nguyen, A. L., Wesevich, V., Greene, D. J., & Fair, D. A. (2017). Real-time motion analytics during brain MRI improve data quality and reduce costs. *NeuroImage*, *161*, 80–93. https://doi.org/10.1016/j.neuroimage.2017.08.025

Driscoll, I., & Troncoso, J. (2011). Asymptomatic Alzheimer's Disease: A Prodrome or a State of Resilience? *Current Alzheimer Research*, *999*(999), 1–6. https://doi.org/10.2174/156721121222594205O

Drzezga, A. (2018). The network degeneration hypothesis: Spread of neurodegenerative patterns along neuronal brain networks. *Journal of Nuclear Medicine*, *59*(11), 1645–1648.

https://doi.org/10.2967/jnumed.117.206300

Duff, K. (2012). Current topics in science and practice evidence-based indicators of neuropsychological change in the individual patient: Relevant concepts and methods. *Archives of Clinical Neuropsychology*, *27*(3), 248–261. https://doi.org/10.1093/arclin/acr120

Duff, K., Lyketsos, C. G., Beglinger, L. J., Chelune, G., Moser, D. J., Arndt, S., Schultz, S. K., Paulsen, J. S., Petersen, R. C., & McCaffrey, R. J. (2011). Practice Effects Predict Cognitive Outcome in Amnestic Mild Cognitive Impairment. *The American Journal of Geriatric Psychiatry*, *19*(11), 932–939. https://doi.org/10.1097/JGP.0b013e318209dd3a

Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine Learning Approaches for Clinical Psychology and Psychiatry. *Ssrn*, *14*(1), 91–118. https://doi.org/10.1146/annurev-clinpsy-032816-045037

Dyrba, M., Grothe, M., Kirste, T., & Teipel, S. J. (2015). Multimodal analysis of functional and structural disconnection in Alzheimer's disease using multiple kernel SVM. *Human Brain Mapping*, *36*(6), 2118–2131. https://doi.org/10.1002/hbm.22759

Edmonds, E. C., Delano-Wood, L., Clark, L. R., Jak, A. J., Nation, D. A., Mcdonald, C. R., Libon, D. J., Au, R., Galasko, D., Salmon, D. P., & Bondi, M. W. (2015). Susceptibility of the conventional criteria for MCI to false positive diagnostic errors HHS Public Access. *Alzheimers Dement*, *11*(4), 415–424. https://doi.org/10.1016/j.jalz.2014.03.005.Susceptibility

Edmonds, E. C., Delano-Wood, L., Galasko, D. R., Salmon, D. P., & Bondi, M. W. (2014). Subjective cognitive complaints contribute to misdiagnosis of mild cognitive impairment. *Journal of the International Neuropsychological Society*, *20*(8), 836–847. https://doi.org/10.1017/S135561771400068X

Edmonds, E. C., Smirnov, D. S., Thomas, K. R., Graves, L. V., Bangen, K. J., Delano-Wood, L., Galasko, D. R., Salmon, D. P., & Bondi, M. W. (2021). Data-Driven vs Consensus Diagnosis of MCI: Enhanced Sensitivity for Detection of Clinical, Biomarker, and Neuropathologic Outcomes. *Neurology*, *97*(13), e1288–e1299. https://doi.org/10.1212/WNL.0000000000012600

Elias-Sonnenschein, L. S., Viechtbauer, W., Ramakers, I., Ukoumunne, O., Verhey, F. R. J., & Visser, P. J. (2018). APOE-4 allele for the diagnosis of Alzheimer's and other dementia disorders in people with mild cognitive impairment in a community setting. *Cochrane Database of Systematic Reviews*, *2018*(8). https://doi.org/10.1002/14651858.CD010948.pub2

Estrada, E., & Hatano, N. (2008). Communicability in complex networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, *77*(3), 036111. https://doi.org/10.1103/PhysRevE.77.036111

Feczko, E., Balba, N. M., Miranda-Dominguez, O., Cordova, M., Karalunas, S. L., Irwin, L.,

Demeter, D. V., Hill, A. P., Langhorst, B. H., Grieser Painter, J., Van Santen, J., Fombonne, E. J., Nigg, J. T., & Fair, D. A. (2018). Subtyping cognitive profiles in Autism Spectrum Disorder using a Functional Random Forest algorithm. *NeuroImage*, *172*, 674–688. https://doi.org/10.1016/j.neuroimage.2017.12.044

Ferreira, D., Nordberg, A., & Westman, E. (2020). Biological subtypes of Alzheimer disease: A systematic review and meta-analysis. *Neurology*, *94*(10), 436–448. https://doi.org/10.1212/WNL.0000000000009058

Ferreira, D., Pereira, J. B., Volpe, G., & Westman, E. (2019). Subtypes of Alzheimer's Disease Display Distinct Network Abnormalities Extending Beyond Their Pattern of Brain Atrophy. *Frontiers in Neurology*, *10*, 524. https://doi.org/10.3389/fneur.2019.00524

Fisher, C. K., Smith, A. M., Walsh, J. R., Simon, A. J., Edgar, C., Jack, C. R., Holtzman, D., Russell, D., Hill, D., Grosset, D., Wood, F., Vanderstichele, H., Morris, J., Blennow, K., Marek, K., Shaw, L. M., Albert, M., Weiner, M., Fox, N., … Kubick, W. (2019). Machine learning for comprehensive forecasting of Alzheimer's Disease progression. *Scientific Reports*, *9*(1). https://doi.org/10.1038/S41598-019-49656-2

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

Foo, H., Thalamuthu, A., Jiang, J., Koch, F., Mather, K. A., Wen, W., & Sachdev, P. S. (2021). Age- and Sex-Related Topological Organization of Human Brain Functional Networks and Their Relationship to Cognition. *Frontiers in Aging Neuroscience*, *13*, 897. https://doi.org/10.3389/fnagi.2021.758817

Fornito, A., & Bullmore, E. T. (2015). Connectomics: A new paradigm for understanding brain disease. *European Neuropsychopharmacology*, *25*(5), 733–748. https://doi.org/10.1016/j.euroneuro.2014.02.011

Fornito, A., Zalesky, A., & Breakspear, M. (2015). The connectomics of brain disorders. *Nature Reviews Neuroscience*, *16*(3), 159–172. https://doi.org/10.1038/nrn3901

Franzmeier, N., Neitzel, J., Rubinski, A., Smith, R., Strandberg, O., Ossenkoppele, R., Hansson, O., & Ewers, M. (2020). Functional brain architecture is associated with the rate of tau accumulation in Alzheimer's disease. *Nature Communications*, *11*(1), 347. https://doi.org/10.1038/s41467-019-14159-1

Gao, S., & Lima, D. (2022). A review of the application of deep learning in the detection of Alzheimer's disease. *International Journal of Cognitive Computing in Engineering*, *3*, 1–8. https://doi.org/10.1016/j.ijcce.2021.12.002

Garrison, K. A., Scheinost, D., Finn, E. S., Shen, X., & Constable, R. T. (2015). The (in)stability of functional brain network measures across thresholds. *NeuroImage*, *118*, 651–661. https://doi.org/10.1016/j.neuroimage.2015.05.046

Geifman, N., Kennedy, R. E., Schneider, L. S., Buchan, I., & Brinton, R. D. (2018). Data-driven identification of endophenotypes of Alzheimer's disease progression: Implications for clinical trials and therapeutic interventions. *Alzheimer's Research and Therapy*, *10*(1), 4. https://doi.org/10.1186/s13195-017-0332-0

Genuer, R., Poggi, J. M., & Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters*, *31*(14), 2225–2236. https://doi.org/10.1016/j.patrec.2010.03.014

Goldstein, B. A., Cerullo, M., Krishnamoorthy, V., Blitz, J., Mureebe, L., Webster, W., Dunston, F., Stirling, A., Gagnon, J., & Scales, C. D. (2020). Development and Performance of a Clinical Decision Support Tool to Inform Resource Utilization for Elective Operations. *JAMA Network Open*, *3*(11), e2023547. https://doi.org/10.1001/jamanetworkopen.2020.23547

Gomez-Isla, T., & Frosch, M. P. (2019). The Challenge of Defining Alzheimer Disease Based on Biomarkers in the Absence of Symptoms. *JAMA Neurology*, *76*(10), 1143–1144. https://doi.org/10.1001/jamaneurol.2019.1667

Goni, J., Van Den Heuvel, M. P., Avena-Koenigsberger, A., De Mendizabal, N. V., Betzel, R. F., Griffa, A., Hagmann, P., Corominas-Murtra, B., Thiran, J. P., & Sporns, O. (2014). Resting-brain functional connectivity predicted by analytic measures of network communication. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(2), 833–838. https://doi.org/10.1073/pnas.1315529111

Grady, C. L. (2017). *Age differences in functional connectivity during tasks and at rest*.

Grassi, M., Rouleaux, N., Caldirola, D., Loewenstein, D., Schruers, K., Perna, G., & Dumontier, M. (2019). A novel ensemble-based machine learning algorithm to predict the conversion from mild cognitive impairment to Alzheimer's disease using socio-demographic characteristics, clinical information, and neuropsychological measures. *Frontiers in Neurology*, *10*(JUL), 756. https://doi.org/10.3389/fneur.2019.00756

Grueso, S., & Viejo-Sobera, R. (2021). Machine learning methods for predicting progression from mild cognitive impairment to Alzheimer's disease dementia: a systematic review. *Alzheimer's Research & Therapy 2021 13:1*, *13*(1), 1–29. https://doi.org/10.1186/S13195-021-00900-W

Guan, H., Liu, T., Jiang, J., Tao, D., Zhang, J., Niu, H., Zhu, W., Wang, Y., Cheng, J., Kochan, N. A., Brodaty, H., Sachdev, P., & Wen, W. (2017). Classifying MCI subtypes in community-dwelling elderly using cross-sectional and longitudinal MRI-based biomarkers. *Frontiers in Aging Neuroscience*, *9*(SEP), 309. https://doi.org/10.3389/fnagi.2017.00309

Guerreiro, R., & Bras, J. (2015). The age factor in Alzheimer's disease. *Genome Medicine*, *7*(1), 1–3. https://doi.org/10.1186/s13073-015-0232-5

Gustavson, D. E., Panizzon, M. S., Franz, C. E., Reynolds, C. A., Corley, R. P., Hewitt, J. K., Lyons, M. J., Kremen, W. S., & Friedman, N. P. (2019). Integrating Verbal Fluency with

Executive Functions: Evidence from Twin Studies in Adolescence and Middle Age. *Journal of Experimental Psychology. General*, *148*(12), 2104. https://doi.org/10.1037/XGE0000589

Hallquist, M. N., & Hillary, F. G. (2018). Graph theory approaches to functional network organization in brain disorders: A critique for a brave new small-world. *Netowrk Neuroscience*, 243741. https://doi.org/10.1101/243741

Hardy, J. A., & Higgins, G. A. (1992). Alzheimer's disease: The amyloid cascade hypothesis. In *Science* (Vol. 256, Issue 5054, pp. 184–185). American Association for the Advancement of Science. https://doi.org/10.1126/science.1566067

Harrell, F. (2020). *Classification vs. Prediction | Statistical Thinking*. https://www.fharrell.com/post/classification/

Harrington, M. G., Chiang, J., Pogoda, J. M., Gomez, M., Thomas, K., Marion, S. D. B., Miller, K. J., Siddarth, P., Yi, X., Zhou, F., Lee, S., Arakaki, X., Cowan, R. P., Tran, T., Charleswell, C., Ross, B. D., & Fonteh, A. N. (2013). Executive function changes before memory in preclinical Alzheimer's pathology: A prospective, cross-sectional, case control study. *PLoS ONE*, *8*(11). https://doi.org/10.1371/journal.pone.0079378

Hasselgren, C., Ekbrand, H., Halleröd, B., Mellqvist Fässberg, M., Zettergren, A., Johansson, L., Skoog, I., & Dellve, L. (2020). Sex differences in dementia: On the potentially mediating effects of educational attainment and experiences of psychological distress. *BMC Psychiatry*, *20*(1), 1–13. https://doi.org/10.1186/s12888-020-02820-9

Hastie, T., Tibshirani, R., & Friedman, J. (2013). *The Elements of Statistical Learning*. https://doi.org/10.1109/SITIS.2013.106

Hedden, T., Schultz, A. P., Rieckmann, A., Mormino, E. C., Johnson, K. A., Sperling, R. A., & Buckner, R. L. (2016). Multiple Brain Markers are Linked to Age-Related Variation in Cognition. *Cerebral Cortex*, *26*(4), 1388–1400. https://doi.org/10.1093/cercor/bhu238

Henninger, M., Debelak, R., Rothacher, Y., & Strobl, C. (2022). Interpretable machine learning for psychological research: Opportunities and pitfalls. *Manuscript Submitted for Publication*. https://psyarxiv.com/xe83y

Henningsen, A. (2022). *censReg: Censored Regression (Tobit) Models*. http://www.sampleselection.org

Hojjati, S. H., Ebrahimzadeh, A., & Babajani-Feremi, A. (2019). Identification of the early stage of alzheimer's disease using structural mri and resting-state fmri. *Frontiers in Neurology*, *10*(AUG), 904. https://doi.org/10.3389/fneur.2019.00904

Hojjati, S. H., Ebrahimzadeh, A., Khazaee, A., & Babajani-Feremi, A. (2017). Predicting conversion from MCI to AD using resting-state fMRI, graph theoretical approach and SVM. *Journal of Neuroscience Methods*, *282*, 69–80. https://doi.org/10.1016/j.jneumeth.2017.03.006

Horowitz, T. S., Suls, J., & Treviño, M. (2018). A Call for a Neuroscience Approach to Cancer-Related Cognitive Impairment. *Trends in Neurosciences*, *41*(8), 493–496. https://doi.org/10.1016/j.tins.2018.05.001

Hughes, T. M., & Craft, S. (2016). The role of insulin in the vascular contributions to age-related dementia. *Biochimica et Biophysica Acta - Molecular Basis of Disease*, *1862*(5), 983–991. https://doi.org/10.1016/j.bbadis.2015.11.013

Jack, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., Holtzman, D. M., Jagust, W., Jessen, F., Karlawish, J., Liu, E., Molinuevo, J. L., Montine, T., Phelps, C., Rankin, K. P., Rowe, C. C., Scheltens, P., Siemers, E., Snyder, H. M., … Silverberg, N. (2018). NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. In *Alzheimer's and Dementia* (Vol. 14, Issue 4, pp. 535–562). Elsevier Inc. https://doi.org/10.1016/j.jalz.2018.02.018

Jack, C. R., Therneau, T. M., Weigand, S. D., Wiste, H. J., Knopman, D. S., Vemuri, P., Lowe, V. J., Mielke, M. M., Roberts, R. O., Machulda, M. M., Graff-Radford, J., Jones, D. T., Schwarz, C. G., Gunter, J. L., Senjem, M. L., Rocca, W. A., & Petersen, R. C. (2019). Prevalence of Biologically vs Clinically Defined Alzheimer Spectrum Entities Using the National Institute on Aging-Alzheimer's Association Research Framework. *JAMA Neurology*, *76*(10), 1174–1183. https://doi.org/10.1001/jamaneurol.2019.1971

Jacobson, M. W., Delis, D. C., Peavy, G. M., Wetter, S. R., Bigler, E. D., Abildskov, T. J., Bondi, M. W., & Salmon, D. P. (2009). The emergence of cognitive discrepancies in preclinical Alzheimer's disease: A six-year case study. *Neurocase*, *15*(4), 278–293. https://doi.org/10.1080/13554790902729465

Jak, A. J., Bondi, M. W., Delano-Wood, L., Wierenga, C., Corey-Bloom, J., Salmon, D. P., & Delis, D. C. (2009). Quantification of five neuropsychological approaches to defining mild cognitive impairment. *American Journal of Geriatric Psychiatry*, *17*(5), 368–375. https://doi.org/10.1097/JGP.0b013e31819431d5

Jaul, E., & Barron, J. (2017). Age-Related Diseases and Clinical and Public Health Implications for the 85 Years Old and Over Population. *Frontiers in Public Health*, *5*. https://doi.org/10.3389/fpubh.2017.00335

Jellinger, K. A. (2021). Pathobiological Subtypes of Alzheimer Disease. *Dementia and Geriatric Cognitive Disorders*, *49*(4), 321–333. https://doi.org/10.1159/000508625

Jie, B., Zhang, D., Wee, C. Y., & Shen, D. (2014). Topological graph kernel on multiple thresholded functional connectivity networks for mild cognitive impairment classification. *Human Brain Mapping*, *35*(7), 2876–2897. https://doi.org/10.1002/hbm.22353

Jitsuishi, T., & Yamaguchi, A. (2022). Searching for optimal machine learning model to classify mild cognitive impairment (MCI) subtypes using multimodal MRI data. *Scientific Reports*, *12*(1), 4284. https://doi.org/10.1038/s41598-022-08231-y

Johnson, K. A., Schultz, A., Betensky, R. A., Becker, J. A., Sepulcre, J., Rentz, D., Mormino, E.,

Chhatwal, J., Amariglio, R., Papp, K., Marshall, G., Albers, M., Mauro, S., Pepin, L., Alverio, J., Judge, K., Philiossaint, M., Shoup, T., Yokell, 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

Jones, D. T., Knopman, D. S., Gunter, J. L., Graff-Radford, J., Vemuri, P., Boeve, B. F., Petersen, R. C., Weiner, M. W., Jack, C. R., & Alzheimer's Disease Neuroimaging Initiative. (2016). Cascading network failure across the Alzheimer's disease spectrum. *Brain : A Journal of Neurology*, *139*(Pt 2), 547–562. https://doi.org/10.1093/brain/awv338

Kaplan, E. (1983). *Boston naming test*. Lea & Febiger.

Kassambara, A., & Mundt, F. (2020). factoextra: Extract and Visualize the Results of Multivariate Data Analyses. Retrieved from https://cran.r-project.org/web/packages/factoextra/index.html. In *CRAN- R Package*. https://cran.r-project.org/package=factoextra

Keller, A. S., & Christopher, L. (2017). Distinct Phases of Tau, Amyloid, and Functional Connectivity in Healthy Older Adults. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *37*(37), 8857–8859. https://doi.org/10.1523/JNEUROSCI.1687-17.2017

Kesler, S. R., Rao, A., Blayney, D. W., Oakley-Girvan, I. A., Karuturi, M., & Palesh, O. (2017). Predicting Long-Term Cognitive Outcome Following Breast Cancer with Pre-Treatment Resting State fMRI and Random Forest Machine Learning. *Frontiers in Human Neuroscience*, *11*, 555. https://doi.org/10.3389/fnhum.2017.00555

Khazaee, A., Ebrahimzadeh, A., & Babajani-Feremi, A. (2015). Identifying patients with Alzheimer's disease using resting-state fMRI and graph theory. *Clinical Neurophysiology*, *126*(11), 2132–2141. https://doi.org/10.1016/j.clinph.2015.02.060

Khazaee, A., Ebrahimzadeh, A., & Babajani-Feremi, A. (2016). Application of advanced machine learning methods on resting-state fMRI network for identification of mild cognitive impairment and Alzheimer's disease. *Brain Imaging and Behavior*, *10*(3), 799–817. https://doi.org/10.1007/s11682-015-9448-7

Khazaee, A., Ebrahimzadeh, A., & Babajani-Feremi, A. (2017). Classification of patients with MCI and AD from healthy controls using directed graph measures of resting-state fMRI. *Behavioural Brain Research*, *322*, 339–350. https://doi.org/10.1016/j.bbr.2016.06.043

Knopman, D. S., Petersen, R. C., & Jack, C. R. (2019). A brief history of "Alzheimer disease": Multiple meanings separated by a common name. In *Neurology* (Vol. 92, Issue 22, pp. 1053–1059). Lippincott Williams and Wilkins. https://doi.org/10.1212/WNL.0000000000007583

Koen, J. D., & Rugg, M. D. (2019). Neural Dedifferentiation in the Aging Brain. *Trends in Cognitive Sciences*, *23*(7), 547–559. https://doi.org/10.1016/J.TICS.2019.04.012

Kohavi, R. (1995). *A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection*. http://robotics.stanford.edu/~ronnyk

Kuhn, M. (2008). Building predictive models in R using the caret package. In *Journal of Statistical Software* (Vol. 28, Issue 5). https://doi.org/10.18637/jss.v028.i05

Kuncheva, L. I., Matthews, C. E., Arnaiz-González, Á., & Rodríguez, J. J. (2020). *Feature Selection from High-Dimensional Data with Very Low Sample Size: A Cautionary Tale*. https://doi.org/10.48550/arxiv.2008.12025

Kwak, K., Giovanello, K. S., Bozoki, A., Styner, M., & Dayan, E. (2021). Subtyping of mild cognitive impairment using a deep learning model based on brain atrophy patterns. *Cell Reports Medicine*, *2*(12), 100467. https://doi.org/10.1016/J.XCRM.2021.100467

Kwak, K., Niethammer, M., Giovanello, K. S., Styner, M., & Dayan, E. (2022). Differential Role for Hippocampal Subfields in Alzheimer's Disease Progression Revealed with Deep Learning. *Cerebral Cortex (New York, N.Y. : 1991)*, *32*(3), 467–478. https://doi.org/10.1093/cercor/bhab223

LaMontagne, P. J., Benzinger, T. L. S., Morris, J. C., Keefe, S., Hornbeck, R., Xiong, C., Grant, E., Hassenstab, J., Moulder, K., Vlassenko, A. G., Raichle, M. E., Cruchaga, C., & Marcus, D. (2019). OASIS-3: Longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease. *MedRxiv*, 2019.12.13.19014902. https://doi.org/10.1101/2019.12.13.19014902

Langella, S., Mucha, P. J., Giovanello, K. S., & Dayan, E. (2021). The association between hippocampal volume and memory in pathological aging is mediated by functional redundancy. *Neurobiology of Aging*, *108*, 179–188. https://doi.org/10.1016/J.NEUROBIOLAGING.2021.09.002

Langella, S., Sadiq, M. U., Mucha, P. J., Giovanello, K. S., & Dayan, E. (2021). Lower functional hippocampal redundancy in mild cognitive impairment. *Translational Psychiatry*, *11*(1), 1–12. https://doi.org/10.1038/s41398-020-01166-w

Langer, N., Pedroni, A., & Jäncke, L. (2013). The Problem of Thresholding in Small-World Network Analysis. *PLoS ONE*, *8*(1), e53199. https://doi.org/10.1371/journal.pone.0053199

Lebedev, A. V., Westman, E., Van Westen, G. J. P., Kramberger, M. G., Lundervold, A., Aarsland, D., Soininen, H., Kłoszewska, I., Mecocci, P., Tsolaki, M., Vellas, B., Lovestone, S., & Simmons, A. (2014). Random Forest ensembles for detection and prediction of Alzheimer's disease with a good between-cohort robustness. *NeuroImage: Clinical*, *6*, 115–125. https://doi.org/10.1016/j.nicl.2014.08.023

Lee, A., Ratnarajah, N., Tuan, T. A., Chen, S.-H. A., & Qiu, A. (2015). Adaptation of Brain Functional and Structural Networks in Aging. *PLOS ONE*, *10*(4), e0123462. https://doi.org/10.1371/journal.pone.0123462

Lee, A., Tan, M., & Qiu, A. (2016). Distinct Aging Effects on Functional Networks in Good and

Poor Cognitive Performers. *Frontiers in Aging Neuroscience*, *8*, 215. https://doi.org/10.3389/fnagi.2016.00215

Leistritz, L., Weiss, T., Bär, K. J., De VicoFallani, F., Babiloni, F., Witte, H., & Lehmann, T. (2013). Network Redundancy Analysis of Effective Brain Networks; a Comparison of Healthy Controls and Patients with Major Depression. *PLoS ONE*, *8*(4), e60956. https://doi.org/10.1371/journal.pone.0060956

Lella, E., Amoroso, N., Diacono, D., Lombardi, A., Maggipinto, T., Monaco, A., Bellotti, R., & Tangaro, S. (2019). Communicability Characterization of Structural DWI Subcortical Networks in Alzheimer's Disease. *Entropy*, *21*(5), 475. https://doi.org/10.3390/e21050475

Lella, E., & Estrada, E. (2020). Communicability distance reveals hidden patterns of Alzheimer's disease. *Network Neuroscience*, 1–23. https://doi.org/10.1162/netn_a_00143

Letenneur, L., Gilleron, V., Commenges, D., Helmer, C., Orgogozo, J. M., & Dartigues, J. F. (1999). Are sex and educational level independent predictors of dementia and Alzheimer's disease? Incidence data from the PAQUID project. *Journal of Neurology Neurosurgery and Psychiatry*, *66*(2), 177–183. https://doi.org/10.1136/jnnp.66.2.177

Lezak, M., Howieson, D., Loring, D., & Fischer, J. (2004). *Neuropsychological assessment*. https://books.google.com/books?hl=en&lr=&id=FroDVkVKA2EC&oi=fnd&pg=PA1&ots=q7Yh-RQk9L&sig=28i6RRA2EebwRWOf0dLhQd4C8FE

Li, C., Wang, H., De Haan, W., Stam, C. J., & Van Mieghem, P. (2011). The correlation of metrics in complex networks with applications in functional brain networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2011*(11), P11018. https://doi.org/10.1088/1742-5468/2011/11/P11018

Li, Y. P., Qin, Y., Chen, X., & Li, W. (2013). Exploring the Functional Brain Network of Alzheimer's Disease: Based on the Computational Experiment. *PLoS ONE*, *8*(9), e73186. https://doi.org/10.1371/journal.pone.0073186

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. In *R News* (Vol. 2, Issue 3). http://www.stat.berkeley.edu/

Liu, M., Lian, C., & Shen, D. (2020). Anatomical-Landmark-Based Deep Learning for Alzheimer's Disease Diagnosis with Structural Magnetic Resonance Imaging. *Intelligent Systems Reference Library*, *171*, 127–147. https://doi.org/10.1007/978-3-030-32606-7_8

Lockhart, S. N., Schaich, C. L., Craft, S., Sachs, B. C., Rapp, S. R., Jung, Y., Whitlow, C. T., Solingapuram Sai, K. K., Cleveland, M., Williams, B. J., Burke, G. L., Bertoni, A., Hayden, K. M., & Hughes, T. M. (2022). Associations among vascular risk factors, neuroimaging biomarkers, and cognition: Preliminary analyses from the Multi-Ethnic Study of Atherosclerosis (MESA). *Alzheimer's and Dementia*, *18*(4), 551–560. https://doi.org/10.1002/alz.12429

Lövdén, M., Fratiglioni, L., Glymour, M. M., Lindenberger, U., & Tucker-Drob, E. M. (2020).

Education and Cognitive Functioning Across the Life Span. *Psychological Science in the Public Interest*, *21*(1), 6–41. https://doi.org/10.1177/1529100620920576

Luan, J., Zhang, C., Xu, B., Xue, Y., & Ren, Y. (2020). The predictive performances of random forest models with limited sample size and different species traits. *Fisheries Research*, *227*, 105534. https://doi.org/10.1016/J.FISHRES.2020.105534

Lustig, C., Hasher, L., & Zacks, R. T. (2007). Inhibitory deficit theory: Recent developments in a "new view." In *Inhibition in cognition.* (Issue 571, pp. 145–162). https://doi.org/10.1037/11587-008

MacIntosh, B. J., Shirzadi, Z., Atwi, S., Detre, J. A., Dolui, S., Bryan, R. N., Launer, L. J., & Swardfager, W. (2020). Metabolic and vascular risk factors are associated with reduced cerebral blood flow and poorer midlife memory performance. *Human Brain Mapping*, *41*(4), 855–864. https://doi.org/10.1002/hbm.24844

Madden, D. J., Jain, S., Monge, Z. A., Cook, A. D., Lee, A., Huang, H., Howard, C. M., & Cohen, J. R. (2020). Influence of structural and functional brain connectivity on age-related differences in fluid cognition. *Neurobiology of Aging*, *96*, 205–222. https://doi.org/10.1016/j.neurobiolaging.2020.09.010

Mallo, S. C., Valladares-Rodriguez, S., Facal, D., Lojo-Seoane, C., Fernández-Iglesias, M. J., & Pereiro, A. X. (2020). Neuropsychiatric symptoms as predictors of conversion from MCI to dementia: A machine learning approach. *International Psychogeriatrics*, *32*(3), 381–392. https://doi.org/10.1017/S1041610219001030

Marcus, D. S., Fotenos, A. F., Csernansky, J. G., Morris, J. C., & Buckner, R. L. (2010). Open access series of imaging studies: Longitudinal MRI data in nondemented and demented older adults. *Journal of Cognitive Neuroscience*, *22*(12), 2677–2684. https://doi.org/10.1162/jocn.2009.21407

Marcus, D. S., Olsen, T. R., Ramaratnam, M., & Buckner, R. L. (2007). The extensible neuroimaging archive toolkit: An informatics platform for managing, exploring, and sharing neuroimaging data. *Neuroinformatics*, *5*(1), 11–33. https://doi.org/10.1385/NI:5:1:11

McDougall, F., Edgar, C., Mertes, M., Delmar, P., Fontoura, P., Abi-Saab, D., Lansdall, C. J., Boada, M., & Doody, R. (2021). Psychometric Properties of the Clinical Dementia Rating — Sum of Boxes and other Cognitive and Functional Outcomes in a Prodromal Alzheimer's Disease Population. *Journal of Prevention of Alzheimer's Disease*, *8*(2), 151–160. https://doi.org/10.14283/jpad.2020.73

Medaglia, J. D., Pasqualetti, F., Hamilton, R. H., Thompson-Schill, S. L., & Bassett, D. S. (2017a). Brain and cognitive reserve: Translation via network control theory. *Neuroscience and Biobehavioral Reviews*, *75*, 53–64. https://doi.org/10.1016/j.neubiorev.2017.01.016

Medaglia, J. D., Pasqualetti, F., Hamilton, R. H., Thompson-Schill, S. L., & Bassett, D. S. (2017b). Brain and cognitive reserve: Translation via network control theory. *Neuroscience and Biobehavioral Reviews*, *75*, 53–64. https://doi.org/10.1016/j.neubiorev.2017.01.016

Meunier, D., Stamatakis, E. A., & Tyler, L. K. (2014). Age-related functional reorganization, structural changes, and preserved cognition. *Neurobiology of Aging*, *35*(1), 42–54. https://doi.org/10.1016/j.neurobiolaging.2013.07.003

Mez, J., Mukherjee, S., Thornton, T., Fardo, D. W., Trittschuh, E., Sutti, S., Sherva, R., Kauwe, J. S., Naj, A. C., Beecham, G. W., Gross, A., Saykin, A. J., Green, R. C., & Crane, P. K. (2016). The executive prominent/memory prominent spectrum in Alzheimer's disease are highly heritable. *Neurobiology of Aging*, *41*, 115–121. https://doi.org/10.1016/j.neurobiolaging.2016.02.015

Mill, R. D., Ito, T., & Cole, M. W. (2017). From connectome to cognition: The search for mechanism in human functional brain networks. *NeuroImage*, *160*, 124–139. https://doi.org/10.1016/j.neuroimage.2017.01.060

Mišić, B., Betzel, R. F., Nematzadeh, A., Goñi, J., Griffa, A., Hagmann, P., Flammini, A., Ahn, Y.-Y., & Sporns, O. (2015). Cooperative and Competitive Spreading Dynamics on the Human Connectome. *Neuron*, *86*(6), 1518–1529. https://doi.org/10.1016/J.NEURON.2015.05.035

Mistridis, P., Krumm, S., Monsch, A. U., Berres, M., & Taylor, K. I. (2015). The 12 Years Preceding Mild Cognitive Impairment Due to Alzheimer's Disease: The Temporal Emergence of Cognitive Decline. *Journal of Alzheimer's Disease*, *48*(4), 1095. https://doi.org/10.3233/JAD-150137

Miyake, A., Emerson, M. J., & Friedman, N. P. (2000). Assessment of Executive Functions in Clinical Settings: Problems and Recommendations. *Seminars in Speech and Language*, *Volume 21*(Number 02), 0169–0183. https://doi.org/10.1055/s-2000-7563

Monge, Z. A., Geib, B. R., Siciliano, R. E., Packard, L. E., Tallman, C. W., & Madden, D. J. (2017). Functional modular architecture underlying attentional control in aging. *NeuroImage*, *155*(May), 257–270. https://doi.org/10.1016/j.neuroimage.2017.05.002

Monge, Z. A., & Madden, D. J. (2016). Linking cognitive and visual perceptual decline in healthy aging: The information degradation hypothesis. *Neuroscience and Biobehavioral Reviews*, *69*, 166–173. https://doi.org/10.1016/j.neubiorev.2016.07.031

Montine, T. J., Cholerton, B. A., Corrada, M. M., Edland, S. D., Flanagan, M. E., Hemmy, L. S., Kawas, C. H., & White, L. R. (2019). Concepts for brain aging: Resistance, resilience, reserve, and compensation. *Alzheimer's Research and Therapy*, *11*(1), 22. https://doi.org/10.1186/s13195-019-0479-y

Muldoon, S. F., & Bassett, D. S. (2014). Why network neuroscience? Compelling evidence and current frontiers. Comment on "Understanding brain networks and brain organization" by Luiz Pessoa. *Physics of Life Reviews*, *11*(3), 455–457. https://doi.org/10.1016/j.plrev.2014.06.006

Muldoon, S. F., Bridgeford, E. W., & Bassett, D. S. (2016a). Small-world propensity and weighted brain networks. *Scientific Reports*, *6*(1), 22057. https://doi.org/10.1038/srep22057

Muldoon, S. F., Bridgeford, E. W., & Bassett, D. S. (2016b). Small-world propensity and weighted brain networks. *Scientific Reports*, *6*(1), 22057. https://doi.org/10.1038/srep22057

Murphy, M. A., Evans, J. S., & Storfer, A. (2010). Quantifying Bufo boreas connectivity in Yellowstone National Park with landscape genetics. *Ecology*, *91*(1), 252–261. https://doi.org/10.1890/08-0879.1

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

Murray, M. E., Senjem, M. L., Petersen, R. C., Hollman, J. H., Preboske, G. M., Weigand, S. D., Knopman, D. S., Ferman, T. J., Dickson, D. W., & Jack, C. R. (2010). Functional impact of white matter hyperintensities in cognitively normal elderly subjects. *Archives of Neurology*, *67*(11), 1379–1385. https://doi.org/10.1001/archneurol.2010.280

Naimi, A. I., & Balzer, L. B. (2018). Stacked generalization: an introduction to super learning. *European Journal of Epidemiology*, *33*(5), 459–464. https://doi.org/10.1007/s10654-018-0390-z

Newman, M. E. J. (2010). Mathematics of networks. *Networks*, 109–167. https://doi.org/10.1093/acprof:oso/9780199206650.003.0006

O'Bryant, S. E., Waring, S. C., Cullum, C. M., Hall, J., Lacritz, L., Massman, P. J., Lupo, P. J., Reisch, J. S., & Doody, R. (2008). Staging dementia using clinical dementia rating scale sum of boxes scores: A Texas Alzheimer's research consortium study. *Archives of Neurology*, *65*(8), 1091–1095. https://doi.org/10.1001/archneur.65.8.1091

Okello, A., Koivunen, J., Edison, P., Archer, H. A., Turkheimer, F. E., Någren, K., Bullock, R., Walker, Z., Kennedy, A., Fox, N. C., Rossor, M. N., Rinne, J. O., & Brooks, D. J. (2009). Conversion of amyloid positive and negative mci to ad over 3 years: An c-pib pet study symbol. *Neurology*, *73*(10), 754–760. https://doi.org/10.1212/WNL.0b013e3181b23564

Oldham, S., Fulcher, B., Parkes, L., Arnatkeviciūtė, A., Suo, C., & Fornito, A. (2019). Consistency and differences between centrality measures across distinct classes of networks. *PLoS ONE*, *14*(7). https://doi.org/10.1371/journal.pone.0220061

Onnela, J. P., Saramäki, J., Kertész, J., & Kaski, K. (2005). Intensity and coherence of motifs in weighted complex networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, *71*(6), 065103. https://doi.org/10.1103/PhysRevE.71.065103

Ossenkoppele, R., Pichet Binette, A., Groot, C., Smith, R., Strandberg, O., Palmqvist, S., Stomrud, E., Tideman, P., Ohlsson, T., Jögi, J., Johnson, K., Sperling, R., Dore, V., Masters, C. L., Rowe, C., Visser, D., Van Berckel, B. N. M., Van Der Flier, W. M., Baker, S., … Australia, V. (2022). Amyloid and Tau PET positive cognitively unimpaired individuals: Destined to decline? *MedRxiv*, 2022.05.23.22275241. https://doi.org/10.1101/2022.05.23.22275241

Palop, J. J., Chin, J., & Mucke, L. (2006). A network dysfunction perspective on neurodegenerative diseases. *Nature*, *443*(7113), 768–773. https://doi.org/10.1038/nature05289

Park, D. C., Polk, T. A., Park, R., Minear, M., Savage, A., & Smith, M. R. (2004). Aging reduces neural specialization in ventral visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(35), 13091–13095. https://doi.org/10.1073/pnas.0405148101

Park, D. C., & Reuter-Lorenz, P. (2009). The Adaptive Brain: Aging and Neurocognitive Scaffolding. *Annual Review of Psychology*, *60*(1), 173–196. https://doi.org/10.1146/annurev.psych.59.103006.093656

Peraza, L. R., Taylor, J. P., & Kaiser, M. (2015). Divergent brain functional network alterations in dementia with Lewy bodies and Alzheimer's disease. *Neurobiology of Aging*, *36*(9), 2458–2467. https://doi.org/10.1016/j.neurobiolaging.2015.05.015

Pereira, T., Lemos, L., Cardoso, S., Silva, D., Rodrigues, A., Santana, I., De Mendonça, A., Guerreiro, M., & 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

Perera, G., Rijnbeek, P. R., Alexander, M., Ansell, D., Avillach, P., Duarte-Salles, T., Gordon, M. F., Lapi, F., Mayer, M. A., Pasqua, A., Pedersen, L., Van Der Lei, J., Visser, P. J., & Stewart, R. (2020). Vascular and metabolic risk factor differences prior to dementia diagnosis: A multidatabase case-control study using European electronic health records. *BMJ Open*, *10*(11), 1–9. https://doi.org/10.1136/bmjopen-2020-038753

Peterson, R. A. (2021). Finding Optimal Normalizing Transformations via bestNormalize. *R Journal*, *13*(1), 310–329. https://doi.org/10.32614/RJ-2021-041

Petrican, R., Taylor, M. J., & Grady, C. L. (2017). Trajectories of brain system maturation from childhood to older adulthood: Implications for lifespan cognitive functioning. *NeuroImage*, *163*(September), 125–149. https://doi.org/10.1016/j.neuroimage.2017.09.025

Poldrack, R. A. (2012). *The perils of leave-one-out crossvalidation for individual difference analyses*. Russpoldrack.Org. http://www.russpoldrack.org/2012/12/the-perils-of-leave-one-out.html

Porrata-Doria, T., Matta, J. L., & Acevedo, S. F. (2010). Apolipoprotein E Allelic Frequency Altered in Women with Early-onset Breast Cancer. *Breast Cancer : Basic and Clinical Research*, *4*, 43–48. http://www.ncbi.nlm.nih.gov/pubmed/20697532

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage*, *59*(3), 2142–2154. https://doi.org/10.1016/j.neuroimage.2011.10.018

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., Vogel, A. C., Laumann, T. O., Miezin, F. M., Schlaggar, B. L., & Petersen, S. E. (2011). Functional Network Organization of the Human Brain. *Neuron*, *72*(4), 665–678. https://doi.org/10.1016/j.neuron.2011.09.006

Power, J. D., Plitt, M., Laumann, T. O., & Martin, A. (2017). Sources and implications of whole-brain fMRI signals in humans. *NeuroImage*, *146*, 609–625. https://doi.org/10.1016/j.neuroimage.2016.09.038

Power, J. D., Schlaggar, B. L., & Petersen, S. E. (2015). Recent progress and outstanding issues in motion correction in resting state fMRI. *NeuroImage*, *105*, 536–551. https://doi.org/10.1016/j.neuroimage.2014.10.044

Qiu, Y., Jacobs, D. M., Messer, K., Salmon, D. P., & Feldman, H. H. (2019). Cognitive heterogeneity in probable Alzheimer disease: Clinical and neuropathologic features. *Neurology*, *93*(8), e778–e790. https://doi.org/10.1212/WNL.0000000000007967

Reas, E. T., Laughlin, G. A., Bergstrom, J., Kritz-Silverstein, D., Barrett-Connor, E., & McEvoy, L. K. (2019). Effects of APOE on Cognitive Aging in Community-Dwelling Older Adults. *Neuropsychology*, *33*(3), 406. https://doi.org/10.1037/NEU0000501

Reuter-Lorenz, P. A., & Park, D. C. (2014). How Does it STAC Up? Revisiting the Scaffolding Theory of Aging and Cognition. *Neuropsychology Review*, *24*(3), 355–370. https://doi.org/10.1007/s11065-014-9270-9

Ricciarelli, R., & Fedele, E. (2017). The Amyloid Cascade Hypothesis in Alzheimer's Disease: It's Time to Change Our Mind. *Current Neuropharmacology*, *15*(6), 926. https://doi.org/10.2174/1570159x15666170116143743

Roe, C. M., Xiong, C., Miller, J. P., & Morris, J. C. (2007). Education and Alzheimer disease without dementia: support for the cognitive reserve hypothesis. *Neurology*, *68*(3), 223–228. https://doi.org/10.1212/01.WNL.0000251303.50459.8A

Sachs, M. C. (2017). Plotroc: A tool for plotting ROC curves. *Journal of Statistical Software*, *79*. https://doi.org/10.18637/jss.v079.c02

Sadiq, M. U., Langella, S., Giovanello, K. S., Mucha, P. J., & Dayan, E. (2021). Accrual of functional redundancy along the lifespan and its effects on cognition. *NeuroImage*, *229*, 117737. https://doi.org/10.1016/j.neuroimage.2021.117737

Safieh, M., Korczyn, A. D., & Michaelson, D. M. (2019). ApoE4: an emerging therapeutic target for Alzheimer's disease. *BMC Medicine*, *17*(1), 64. https://doi.org/10.1186/s12916-019-1299-4

Sala-Llonch, R., BartrÃs-Faz, D., & JunquÃ, C. (2015). Reorganization of brain networks in aging: a review of functional connectivity studies. *Frontiers in Psychology*, *6*, 663. https://doi.org/10.3389/fpsyg.2015.00663

Sala-Llonch, R., Bartrés-Faz, D., & Junqué, C. (2015). Reorganization of brain networks in aging: a review of functional connectivity studies. *Frontiers in Psychology*, *6*, 663. https://doi.org/10.3389/fpsyg.2015.00663

Santacruz, K. S., Sonnen, J. A., Pezhouh, M. K., Desrosiers, M. F., Nelson, P. T., & Tyas, S. L. (2011). Alzheimer Disease Pathology in Subjects Without Dementia in Two Studies of Aging: The Nun Study and the Adult Changes in Thought Study. *Journal of Neuropathology and Experimental Neurology*, *70*(10), 832. https://doi.org/10.1097/NEN.0B013E31822E8AE9

Sanz-Arigita, E. J., Schoonheim, M. M., Damoiseaux, J. S., Rombouts, S. A. R. B., Maris, E., Barkhof, F., Scheltens, P., & Stam, C. J. (2010). Loss of "Small-World" Networks in Alzheimer's Disease: Graph Analysis of fMRI Resting-State Functional Connectivity. *PLoS ONE*, *5*(11). https://doi.org/10.1371/journal.pone.0013788

Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughead, J., Calkins, M. E., Eickhoff, S. B., Hakonarson, H., Gur, R. C., Gur, R. E., & Wolf, D. H. (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage*, *64*(1), 240–256. https://doi.org/10.1016/j.neuroimage.2012.08.052

Satterthwaite, T. D., Wolf, D. H., Loughead, J., Ruparel, K., Elliott, M. A., Hakonarson, H., Gur, R. C., & Gur, R. E. (2012). Impact of in-scanner head motion on multiple measures of functional connectivity: Relevance for studies of neurodevelopment in youth. *NeuroImage*, *60*(1), 623–632. https://doi.org/10.1016/j.neuroimage.2011.12.063

Scheller, E., Schumacher, L. V., Peter, J., Lahr, J., Wehrle, J., Kaller, C. P., Gaser, C., & Klöppel, S. (2018). Brain aging and APOE ε4 interact to reveal potential neuronal compensation in healthy older adults. *Frontiers in Aging Neuroscience*, *10*(MAR), 74. https://doi.org/10.3389/fnagi.2018.00074

Scheltens, N. M. E., Galindo-Garre, F., Pijnenburg, Y. A. L., Van Der Vlies, A. E., Smits, L. L., Koene, T., Teunissen, C. E., Barkhof, F., Wattjes, M. P., Scheltens, P., & Van Der Flier, W. M. (2016). The identification of cognitive subtypes in Alzheimer's disease dementia using latent class analysis. *Journal of Neurology, Neurosurgery and Psychiatry*, *87*(3), 235–243. https://doi.org/10.1136/jnnp-2014-309582

Scheltens, N. M. E., Tijms, B. M., Heymans, M. W., Rabinovici, G. D., Cohn-Sheehy, B. I., Miller, B. L., Kramer, J. H., Wolfsgruber, S., Wagner, M., Kornhuber, J., Peters, O., Scheltens, P., & Van Der Flier, W. M. (2018). Prominent non-memory deficits in Alzheimer's disease are associated with faster disease progression. *Journal of Alzheimer's Disease*, *65*(3), 1029–1039. https://doi.org/10.3233/JAD-171088

Scheltens, N. M. E., Tijms, B. M., Koene, T., Barkhof, F., Teunissen, C. E., Wolfsgruber, S., Wagner, M., Kornhuber, J., Peters, O., Cohn-Sheehy, B. I., Rabinovici, G. D., Miller, B. L., Kramer, J. H., Scheltens, P., & van der Flier, W. M. (2017). Cognitive subtypes of probable Alzheimer's disease robustly identified in four cohorts. *Alzheimer's and Dementia*, *13*(11),

1226–1236. https://doi.org/10.1016/j.jalz.2017.03.002

Schultz, A. P., Chhatwal, J. P., Hedden, T., Mormino, E. C., Hanseeuw, B. J., Sepulcre, J., Huijbers, W., LaPoint, M., Buckley, R. F., Johnson, K. A., & Sperling, R. A. (2017). Phases of Hyperconnectivity and Hypoconnectivity in the Default Mode and Salience Networks Track with Amyloid and Tau in Clinically Normal Individuals. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *37*(16), 4323–4331. https://doi.org/10.1523/JNEUROSCI.3263-16.2017

Seitzman, B. A., Gratton, C., Marek, S., Raut, R. V, Dosenbach, N. U., Schlaggar, B. L., Petersen, S. E., & Greene, D. J. (2018). A set of functionally-defined brain regions with improved representation of the subcortex and cerebellum. *BioRxiv*, 450452. https://doi.org/10.1101/450452

Sepulcre, J., Schultz, A. P., Sabuncu, M., Gomez-Isla, T., Chhatwal, J., Becker, A., Sperling, R., & Johnson, K. A. (2016). In vivo tau, amyloid, and gray matter profiles in the aging brain. *Journal of Neuroscience*, *36*(28), 7364–7374. https://doi.org/10.1523/JNEUROSCI.0639-16.2016

Shine, J. M., Aburn, M. J., Breakspear, M., & Poldrack, R. A. (2018). The modulation of neural gain facilitates a transition between functional segregation and integration in the brain. *ELife*, *7*. https://doi.org/10.7554/eLife.31130

Shine, J. M., Bissett, P. G., Bell, P. T., Koyejo, O., Balsters, J. H., Gorgolewski, K. J., Moodie, C. A., & Poldrack, R. A. (2016). The Dynamics of Functional Brain Networks: Integrated Network States during Cognitive Task Performance. *Neuron*, *92*(2), 544–554. https://doi.org/10.1016/j.neuron.2016.09.018

Shine, J. M., & Poldrack, R. A. (2018). Principles of dynamic network reconfiguration across diverse brain states. *NeuroImage*, *180*, 396–405. https://doi.org/10.1016/j.neuroimage.2017.08.010

Silverstein, S. M. (2008). Measuring specific, rather than generalized, cognitive deficits and maximizing between-group effect size in studies of cognition and cognitive change. *Schizophrenia Bulletin*, *34*(4), 645–655. https://doi.org/10.1093/schbul/sbn032

Singer, W. (2009). The brain, a complex self-organizing system. *European Review*, *17*(2), 321–329. https://doi.org/10.1017/S1062798709000751

Snyder, H. M., Ahles, T., Calderwood, S., Carrillo, M. C., Chen, H., Chang, C.-C., Craft, S., Jager, P. De, Driver, J. A., Fillet, H., Knopman, D., Lotze, M., Tierney, M. C., Petanceska, S., Saykin, A., Seshadri, S., Shineman, D., & Ganguli, M. (2017). Exploring the Nexus of Alzheimer's Disease and Related Dementias with Cancer and Cancer Therapies. *Alzheimer's & Dementia : The Journal of the Alzheimer's Association*, *13*(3), 267. https://doi.org/10.1016/J.JALZ.2016.11.002

Snyder, H. R., Miyake, A., & Hankin, B. L. (2015). Advancing understanding of executive function impairments and psychopathology: Bridging the gap between clinical and

cognitive approaches. *Frontiers in Psychology*, *6*(MAR), 328. https://doi.org/10.3389/fpsyg.2015.00328

Sporns, O. (2013a). Structure and function of complex brain networks. *Dialogues in Clinical Neuroscience*, *15*(3), 247–262. https://doi.org/10.1137/S003614450342480

Sporns, O. (2013b). The human connectome: Origins and challenges. *NeuroImage*, *80*, 53–61. https://doi.org/10.1016/j.neuroimage.2013.03.023

Sporns, O. (2014). Contributions and challenges for network models in cognitive neuroscience. *Nature Neuroscience*, *17*(5), 652–660. https://doi.org/10.1038/nn.3690

Stam, C. J., van Straaten, E. C. W., Van Dellen, E., Tewarie, P., Gong, G., Hillebrand, A., Meier, J., & Van Mieghem, P. (2016). The relation between structural and functional connectivity patterns in complex brain networks. *International Journal of Psychophysiology*, *103*, 149–160. https://doi.org/10.1016/j.ijpsycho.2015.02.011

Stern, Y. (2002). What is cognitive reserve? Theory and research application of the reserve concept. *Journal of the International Neuropsychological Society*, *8*(3), 448–460. https://doi.org/10.1017/S1355617702813248

Stern, Y. (2017). An approach to studying the neural correlates of reserve. *Brain Imaging and Behavior*, *11*(2), 410–416. https://doi.org/10.1007/s11682-016-9566-x

Stern, Y., Barnes, C. A., Grady, C., Jones, R. N., & Raz, N. (2019). Brain reserve, cognitive reserve, compensation, and maintenance: operationalization, validity, and mechanisms of cognitive resilience. *Neurobiology of Aging*, *83*, 124–129. https://doi.org/10.1016/j.neurobiolaging.2019.03.022

Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M. J., & Kattan, M. W. (2010). Assessing the performance of prediction models: A framework for traditional and novel measures. In *Epidemiology* (Vol. 21, Issue 1, pp. 128–138). NIH Public Access. https://doi.org/10.1097/EDE.0b013e3181c30fb2

Strobl, C., Boulesteix, A. L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, *9*, 1–11. https://doi.org/10.1186/1471-2105-9-307

Stumme, J., Jockwitz, C., Hoffstaedter, F., Amunts, K., & Caspers, S. (2020). Functional network reorganization in older adults: Graph-theoretical analyses of age, cognition and sex. *NeuroImage*, *214*, 116756. https://doi.org/10.1016/j.neuroimage.2020.116756

Stuss, D. T., & Alexander, M. P. (2007). Is there a dysexecutive syndrome? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 901–915. https://doi.org/10.1098/rstb.2007.2096

Suk, H.-I., & Shen, D. (2015). *Deep Learning in Diagnosis of Brain Disorders*. 203–213. https://doi.org/10.1007/978-94-017-7239-6_14

Sun, Y., Bi, Q., Wang, X., Hu, X., Li, H., Li, X., Ma, T., Lu, J., Chan, P., Shu, N., & Han, Y. (2019). Prediction of conversion from amnestic mild cognitive impairment to Alzheimer's disease based on the brain structural connectome. *Frontiers in Neurology*, *10*(JAN), 1178. https://doi.org/10.3389/fneur.2018.01178

Sun, Y., Wang, Y., Lu, J., Liu, R., Schwarz, C. G., Zhao, H., Zhang, Y., Xu, L., Zhu, B., Zhang, B., Liu, B., Wan, S., & Xu, Y. (2017). Disrupted functional connectivity between perirhinal and parahippocampal cortices with hippocampal subfields in patients with mild cognitive impairment and Alzheimer's disease. *Oncotarget*, *8*(58), 99112–99124. https://doi.org/10.18632/oncotarget.17944

Telesford, Q. K., Burdette, J. H., & Laurienti, P. J. (2013). An exploration of graph metric reproducibility in complex brain networks. *Frontiers in Neuroscience*, *0*(7 MAY), 67. https://doi.org/10.3389/fnins.2013.00067

Telesford, Q. K., Morgan, A. R., Hayasaka, S., Simpson, S. L., Barret, W., Kraft, R. A., Mozolic, J. L., & Laurienti, P. J. (2010). Reproducibility of Graph Metrics in fMRI Networks. *Frontiers in Neuroinformatics*, *4*, 117. https://doi.org/10.3389/fninf.2010.00117

Tibshirani, R., & Walther, G. (2005). Cluster validation by prediction strength. *Journal of Computational and Graphical Statistics*, *14*(3), 511–528. https://doi.org/10.1198/106186005X59243

Tononi, G., Sporns, O., & Edelman, G. M. (1999). Measures of degeneracy and redundancy in biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(6), 3257–3262. https://doi.org/10.1073/pnas.96.6.3257

Townley, R. A., Graff-Radford, J., Mantyh, W. G., Botha, H., Polsinelli, A. J., Przybelski, S. A., Machulda, M. M., Makhlouf, A. T., Senjem, M. L., Murray, M. E., Reichard, R. R., Savica, R., Boeve, B. F., Drubach, D. A., Josephs, K. A., Knopman, D. S., Lowe, V. J., Clifford R Jack, J., Petersen, R. C., & Jones, D. T. (2020). Progressive dysexecutive syndrome due to Alzheimer's disease: a description of 55 cases and comparison to other phenotypes. *Brain Communications*, *2*(1). https://doi.org/10.1093/BRAINCOMMS/FCAA068

Ungar, L., Altmann, A., & Greicius, M. D. (2014). Apolipoprotein E, gender, and Alzheimer's disease: An overlooked, but potent and promising interaction. *Brain Imaging and Behavior*, *8*(2), 262–273. https://doi.org/10.1007/s11682-013-9272-x

van den Heuvel, M. P., de Lange, S. C., Zalesky, A., Seguin, C., Yeo, B. T. T., & Schmidt, R. (2017). Proportional thresholding in resting-state fMRI functional connectivity networks and consequences for patient-control connectome studies: Issues and recommendations. *NeuroImage*, *152*, 437–449. https://doi.org/10.1016/j.neuroimage.2017.02.005

Váša, F., Bullmore, E. T., & Patel, A. X. (2018). Probabilistic thresholding of functional connectomes: Application to schizophrenia. *NeuroImage*, *172*, 326–340. https://doi.org/10.1016/j.neuroimage.2017.12.043

Visser, P. J., & Tijms, B. (2017). Brain Amyloid Pathology and Cognitive Function: Alzheimer

Disease Without Dementia? *JAMA*, *317*(22), 2285–2287. https://doi.org/10.1001/JAMA.2017.6895

Vogel, J. W., & Hansson, O. (2022). Subtypes of Alzheimer's disease: questions, controversy, and meaning. *Trends in Neurosciences*, *45*(5), 342–345. https://doi.org/10.1016/j.tins.2022.02.001

Voss, M. W., Wong, C. N., Baniqued, P. L., Burdette, J. H., Erickson, K. I., Prakash, R. S., McAuley, E., Laurienti, P. J., & Kramer, A. F. (2013). Aging brain from a network science perspective: Something to be positive about? *PLoS ONE*, *8*(11), e78345. https://doi.org/10.1371/journal.pone.0078345

Wang, J., Zuo, X., Dai, Z., Xia, M., Zhao, Z., Zhao, X., Jia, J., Han, Y., & He, Y. (2013). Disrupted functional brain connectome in individuals at risk for Alzheimer's disease. *Biological Psychiatry*, *73*(5), 472–481. https://doi.org/10.1016/j.biopsych.2012.03.026

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 440–442. https://doi.org/10.1038/30918

Way, T. W., Sahiner, B., Hadjiiski, L. M., & Chan, H. P. (2010). Effect of finite sample size on feature selection and classification: A simulation study. *Medical Physics*, *37*(2), 907. https://doi.org/10.1118/1.3284974

Wechsler, D. (1981). Manual for the Wechsler Adult Intelligence Scale - Revised. In *Psychological Corporation*. Psychological Corp. https://doi.org/Thesis_references-Converted #317

Wee, C. Y., Yap, P. T., & Shen, D. (2013). Prediction of Alzheimer's disease and mild cognitive impairment using cortical morphological patterns. *Human Brain Mapping*, *34*(12), 3411–3425. https://doi.org/10.1002/hbm.22156

Wee, C. Y., Yap, P. T., Zhang, D., Denny, K., Browndyke, J. N., Potter, G. G., Welsh-Bohmer, K. A., Wang, L., & Shen, D. (2012). Identification of MCI individuals using structural and functional connectivity networks. *NeuroImage*, *59*(3), 2045–2056. https://doi.org/10.1016/j.neuroimage.2011.10.015

Wei, R., Li, C., Fogelson, N., & Li, L. (2016). Prediction of conversion from mild cognitive impairment to Alzheimer's disease using MRI and structural network features. *Frontiers in Aging Neuroscience*, *8*(APR), 76. https://doi.org/10.3389/fnagi.2016.00076

Weigand, A. J., Bangen, K. J., Thomas, K. R., Delano-Wood, L., Gilbert, P. E., Brickman, A. M., & Bondi, M. W. (2020). Is tau in the absence of amyloid on the Alzheimer's continuum?: A study of discordant PET positivity. *Brain Communications*, *2*(1). https://doi.org/10.1093/braincomms/fcz046

Welchowski, T., Maloney, K. O., Mitchell, R., & Schmid, M. (2022). Techniques to Improve Ecological Interpretability of Black-Box Machine Learning Models: Case Study on Biological Health of Streams in the United States with Gradient Boosted Trees. *Journal of*

Agricultural, Biological, and Environmental Statistics, *27*(1), 175–197. https://doi.org/10.1007/s13253-021-00479-7

Whiteside, D. M., Kealey, T., Semla, M., Luu, H., Rice, L., Basso, M. R., & Roper, B. (2015). Verbal Fluency: Language or Executive Function Measure? *Http://Dx.Doi.Org/10.1080/23279095.2015.1004574*, *23*(1), 29–34. https://doi.org/10.1080/23279095.2015.1004574

Whitfield-Gabrieli, S., & Nieto-Castanon, A. (2012). Conn: A Functional Connectivity Toolbox for Correlated and Anticorrelated Brain Networks. *Brain Connectivity*, *2*(3), 125–141. https://doi.org/10.1089/brain.2012.0073

Wig, G. S. (2017). Segregated Systems of Human Brain Networks. *Trends in Cognitive Sciences*, *21*(12), 981–996. https://doi.org/10.1016/j.tics.2017.09.006

Wilson, R. S., Leurgans, S. E., Boyle, P. A., & Bennett, D. A. (2011). Cognitive Decline in Prodromal Alzheimer Disease and Mild Cognitive Impairment. *Archives of Neurology*, *68*(3), 351–356. https://doi.org/10.1001/ARCHNEUROL.2011.31

Xu, X., Li, W., Mei, J., Tao, M., Wang, X., Zhao, Q., Liang, X., Wu, W., Ding, D., & Wang, P. (2020). Feature Selection and Combination of Information in the Functional Brain Connectome for Discrimination of Mild Cognitive Impairment and Analyses of Altered Brain Patterns. *Frontiers in Aging Neuroscience*, *12*. https://doi.org/10.3389/fnagi.2020.00028

Zhang, L., Ni, H., Yu, Z., Wang, J., Qin, J., Hou, F., & Yang, A. (2020). Investigation on the Alteration of Brain Functional Network and Its Role in the Identification of Mild Cognitive Impairment. *Frontiers in Neuroscience*, *14*. https://doi.org/10.3389/fnins.2020.558434

Zhang, T., Liao, Q., Zhang, D., Zhang, C., Yan, J., Ngetich, R., Zhang, J., Jin, Z., & Li, L. (2021). Predicting MCI to AD Conversation Using Integrated sMRI and rs-fMRI: Machine Learning and Graph Theory Approach. *Frontiers in Aging Neuroscience*, *13*. https://doi.org/10.3389/fnagi.2021.688926

Zhao, X., Liu, Y., Wang, X., Liu, B., Xi, Q., Guo, Q., Jiang, H., Jiang, T., & Wang, P. (2012). Disrupted Small-World Brain Networks in Moderate Alzheimer's Disease: A Resting-State fMRI Study. *PLOS ONE*, *7*(3), e33540. https://doi.org/10.1371/JOURNAL.PONE.0033540

Ziegler, D. A., Janowich, J. R., & Gazzaley, A. (2018). Differential Impact of Interference on Internally- and Externally-Directed Attention. *Scientific Reports*, *8*(1), 2498. https://doi.org/10.1038/s41598-018-20498-8

Zilberter, Y., & Zilberter, M. (2017). The vicious circle of hypometabolism in neurodegenerative diseases: Ways and mechanisms of metabolic correction. In *Journal of Neuroscience Research* (Vol. 95, Issue 11, pp. 2217–2235). John Wiley and Sons Inc. https://doi.org/10.1002/jnr.24064

Zuo, X. N., He, Y., Betzel, R. F., Colcombe, S., Sporns, O., & Milham, M. P. (2017). Human

Connectomics across the Life Span. *Trends in Cognitive Sciences*, *21*(1), 32–45. https://doi.org/10.1016/j.tics.2016.10.005