2019

The Characterization of Alzheimer's Disease and the Development of Early Detection Paradigms: Insights from Nosology, Biomarkers and Machine Learning

Isabel Milano

The Characterization of Alzheimer's Disease and the Development of Early Detection Paradigms: Insights from Nosology, Biomarkers and Machine Learning

A Thesis presented by

Isabel Rose Milano

To the Keck Science Department

Of Claremont McKenna, Pitzer, and Scripps Colleges

In partial fulfillment of

The degree of Bachelor of Arts

Senior Thesis in Neuroscience

April 29th, 2019

Table of Contents

Abstract

Alzheimer's Disease (AD) is the only condition in the top ten leading causes of death for which we do not have an effective treatment that prevents, slows, or stops its progression. Our ability to design useful interventions relies on (a) increasing our understanding of the pathological process of AD and (b) improving our ability for its early detection. These goals are impeded by our current reliance on the clinical symptoms of AD for its diagnosis. This characterizations of AD often falsely assumes a unified, underlying AD-specific pathology for similar presentations of dementia that leads to inconsistent diagnoses. It also hinges on postmortem verification, and so is not a helpful method for identifying patients and research subjects in the beginning phases of the pathophysiological process. Instead, a new biomarker-based approach provides a more biological understanding of the disease and can detect pathological changes up to 20 years before the clinical symptoms emerge. Subjects are assigned a profile according to their biomarker measures of amyloidosis (A), tauopathy (T) and neurodegeneration (N) that reflects their underlying pathology *in vivo*. AD is confirmed as the underlying pathology when subjects have abnormal values of both amyloid and tauopathy biomarkers, and so have a biomarker profile of A+T+(N)- or A+T+(N)+. This new biomarker based characterization of AD can be combined with machine learning techniques in multimodal classification studies to shed light on the elements of the AD pathological process and develop early detection paradigms. A guiding research framework is proposed for the development of reliable, biologically-valid and interpretable multimodal classification models.

Introduction

Alzheimer's Disease (AD), a condition affecting 50 million of people worldwide, is one of the top ten global causes of death (Patterson, 2018, p. 34). The number of diagnosed patients is expected to continue to grow exponentially as life expectancies increase and our ability to identify and diagnose the disorder improves. Stigma-reducing awareness campaigns, as well as improved training for caregivers and better diagnostic standards have all contributed to more accurate diagnoses worldwide. Unfortunately, AD is the only leading cause of death for which we still have no effective treatment that prevents, slows or stops the progression of the disease. Because of this, the main goal of the scientific field surrounding AD is the design of useful interventions, a task that depends heavily on (a) increasing our understanding of the pathophysiological process of AD and (b) improving our ability for its early detection.

Our understanding of AD has been somewhat muddling from the first time the term was used to describe a condition. In 1910, Emil Kraepelin, first used 'AD' to describe the findings his associate, Alois Alzheimer's, had made four years earlier, but many consider that in his description he might have downplayed the presence of vascular injuries that Alzheimer's originally found, in order to present a united, new condition entirely dependent on amyloid plaques and neurofibrillary tangles, now considered the hallmark lesions of AD (George et al., 2011, p. 420). In the influential 1997 'Nun Study,' Snowdon found some evidence of a vascular condition in many AD subjects, and, adding to the confusion, also found that there were several subjects that exhibited no clinical signs of

AD (no cognitive decline) but still had significant evidence of the pathological process of AD at autopsy (Snowdon et al., 1997). Although much research since then has focused on specifying the common ties between patients, the definition of AD is still unclear and many questions remain unanswered. This contributes to the stigma associated with the condition, and limits our ability for accurate diagnosis. Currently, standards for diagnosis rely on the clinical presentation of AD symptoms, even though these occur at least 10-20 years after the pathological process has begun and therefore do not allow for methods of early intervention which are crucial to advancing the field's research goals. Furthermore, a recent technological boom has paired our unclear characterization of AD with large amounts of data and data mining techniques that seem to advance without much concern for what we know about the neurobiological process of AD. At the same time, recent years have seen significant progress in the standardization and accessibility of AD biomarkers. The hope is that biomarkers, possibly in combination with advanced computational methods, will give us valuable insights into the otherwise undetectable pre-clinical process of AD and will allow us to develop methods for early detection and disease progression staging.

In the first section of this paper I explore the history and challenges of the diagnosis of AD, with particular emphasis on the discrepancies between the clinical and pathological characterizations of AD and on the role that biomarkers have in bridging the gap between them. In the second section, I turn to an examination of the neuropathological elements of AD, first with an analysis of current standard regions of interests and staging schemes derived from postmortem studies and then with an evaluation of the causal relationship

between amyloidosis and tauopathy in light of new evidence. Finally, my last section focuses on the application of our understanding of AD and AD biomarkers to AD classification studies for early detection of the disease through machine learning techniques.

Diagnosing Alzheimer's Disease

History and Standards of Alzheimer's Disease Nosology

The layperson's view of AD considers it a unified clinical-pathologic condition, of which age is the leading risk factor, that results in cognitive decline, significant memory loss and an increased likelihood of death. However, this view, upheld by most of the AD diagnostic standards, has led to a series of troubling results from postmortem verifications of the disease, suggesting that there are important limits to our diagnosis of the disease in everyday settings. For starters, 30-40% of cognitively unimpaired elders show evidence of the AD pathological process, and 10-30% of those diagnosed with AD show no evidence of the AD pathological process (Nelson et al., 2011; Bennett et al., 2006). Furthermore, mixed pathology, often involving other dementia-causing pathologies, is found in around 45% of elders with an AD diagnosis (Schneider et al., 2009). Analyzing the history of AD nosology and the current standards used for AD diagnosis is a valuable first step in understanding why these inconsistencies arise, and to evaluate whether our clinical-pathologic understanding of AD is enough to account for both the heterogeneity and specificity of the disease.

The most widely used set of formalized criteria for the diagnosis of Alzheimer's Disease (AD) is contained in the current version of the Diagnostic and Statistical Manual of Mental Disorders. The criteria outlined are used by insurance companies, social service agencies, and in court, as well as in clinical settings, to accurately classify AD, along with other

mental disorders, according to their psychiatric symptoms (George et al., 2011, p. 424) . Throughout its different editions, however, the DSM has showed profound changes in the way it categorizes and defines the progressive form of dementia that Kraepelin first called "Alzheimer's Disease" in 1910. The DSM-I and DSM-II used the term 'chronic brain syndrome' to refer to age-related progressive dementia, with ample interpretive room for identifying its causal factors (American Psychiatric Association (APA), 1952 and 1968)

In the DSM-III, the first version of which came out in 1980, the disorders are specified in much more detail (the manual triples in length) and there is a greater focus in categorizing according to the known or speculated biological substrates of the disorders. In this version they used the term Primary Degenerative Dementia to encompass progressive dementias, including AD, that were considered to only be distinguishable with access to histopathological data, and so represented subtypes that were not useful for purely clinical classification (George et al., 2011, 424; APA ,1980, p. 125). This changes by the 1987 version, DSM-III-R, in which Primary Degenerative Dementia of the Alzheimer's type is established as a type of 'Organic mental disorder,' a category was meant to include disorders involving cognitive deterioration due to physical brain pathology rather than psychiatric illness (APA, 1987). The manual emphasizes that AD should not be considered a mental disorder but rather a physical one, and so uses the variant Primary Degenerative Dementia of the Alzheimer's Type to refer to the most common clinical dementia syndrome, which arises from a discrete Alzheimer's pathology and which is characterized mainly by an insidious onset and progressive deterioration of symptoms.

In the DSM-IV, the focus on insidious onset and progressive deterioration in AD is kept, but the underlying categorization of disorders reflecting cognitive decline changes again, replacing the 'organic mental disorders' with a new group that includes 'Delirium, Dementia, and Amnestic and other cognitive disorders' (APA, 1994). The problem with the term 'organic mental disorders' is that it assumed that other mental disorders did not have a biological basis, an idea rooted in now out-dated dualist conceptions of the mind and body. Another issue with using the term, which is not acknowledged in the DSM-IV but which I will return to shortly is the lack of a discrete knowledge of the underlying pathologies, reflected in a lack of direct correlation between the pathological substrates and the clinical symptoms observed in many of the 'organic mental disorders' mentioned, particularly AD. The DSM-IV somewhat takes this into consideration by introducing a new diagnosis, 'Dementia due to Multiple Etiologies,' that accounted for cases of mixed pathology within dementia (APA, 1994, 155).

Nevertheless, the diagnosis of dementia is similar to what it had been in previous versions of the DSM. The cognitive decline is characterized as including definite memory impairment (in learning or recall) on one side, and either aphasia, apraxia, agnosia, or executive function disturbances on the other side. The cognitive deficits must be an impairment to daily functioning and a clear decline to classify as dementia. Dementia of the Alzheimer's Type is then characterized more precisely by the insidious onset and progressive deterioration of cognitive deficits. It's sub-categorized depending on whether there was an early or late onset of the condition, and whether it existed with or without behavioral disturbances.

Again, a radical change occurs with the publication of the fifth and most up-to-date version of the DSM (APA, 2013). The current criteria identifies a new category, Neurocognitive disorders (NCDs), encompassing delirium, dementias, amnestic and other cognitive disorders. Within this category, different criteria define whether a subject has delirium, Mild NCD or Major NCD, three separate diagnosis that mark the progression of the disorder. In the DSM-5 the term dementia is abandoned, although still accepted in some contexts, in order to account for cases of NCDs that are not considered deviations from healthy aging, which is what the term is habitually used for, such as NCDs caused by HIV or brain injury. The term dementia is replaced by Major NCD, and a Minor NCD is introduced so as to include less severe cases of cognitive dysfunction that can still be cared for in the clinical setting, and which might lead to Major NCD. The distinction between Major and Minor NCD hinges on whether the cognitive deficits interfere significantly with independence in daily functioning. This means that their distinction is not inherently discrete, and the disorders exist on a continuum.

Importantly, the DSM-5 also introduces a more in-depth explanation of cognitive domains that may or may not be involved in an NCD, along with examples of symptoms relevant to the particular domain, and psychiatric assessments that can be used to measure it. This explanation allows for more detailed specifications of the cognitive dysfunctions in each NCD's criteria, and the criteria for NCDs in general no longer requires memory dysfunction (which it did when talking about dementia in the DSM-III).

The neurocognitive disorders outlined in the DSM-5 include those due to AD, Frontotemporal Lobar Degeneration, Lewy Body Disease, Vascular Disease, traumatic brain injury, substance use, HIV infection, Prion disease, Parkinson's disease, Huntington's Disease, multiple etiologies or another medical condition. All of these share cognitive deficit symptoms that are acquired rather than developmental, and are considered separately from each other and from other mental disorders because they have a known, presumed or potentially discoverable underlying pathology.

The criteria for Major NCD involves, first, evidence of significant cognitive decline in at least one of the specified cognitive domains, as demonstrated by the subject's or a clinician's observation of cognitive decline, as well as with results from neuropsychological testing. These deficits must also interfere significantly with daily functioning, whereas for Minor NCD there must be evidence of more modest deficits that do not interfere with the subject's independence in everyday activities. While subjects with Minor NCD might report interference in daily functioning, with tasks requiring more effort or time, their ability to complete these tasks independently must still be preserved to be classified as Minor NCD. For both, the deficits cannot be explained by just delirium or any other mental disorder.

Once a subject is diagnosed with a Major or Minor NCD, the etiological subtype is determined by considering the criteria for each. Ideally, one of the set of criteria will apply while the others are used for differential diagnosis. For the diagnosis of NCDs due to neurodegenerative conditions, including Lewy Body Disease, FTLD or Parkinson's

Disease, once a subject is diagnosed with NCD they must also exhibit an insidious onset and gradual progression of symptoms, and these symptoms cannot be explained more adequately by another mental, neurological or systemic disorder. For NCD due to Alzheimer's particularly, the classification of Major or Minor NCD is followed by a classification of either Probable or Possible Alzheimer's Disease, depending on whether there is evidence of a causative AD genetic mutation from family history or genetic testing or whether particular criteria related to the symptoms' progression and type are met. In Major NCD, Probable AD is diagnosed if either there is evidence of a familial genetic mutation or if all three of the following criteria are met: (a) there is evidence of cognitive decline in the learning and memory domain and in another cognitive domain, (b) the symptoms progress steadily without extended plateaus and (c) there is no evidence of NCD due to mixed etiology (with particular emphasis on differential diagnosis of other neurodegenerative or cardiovascular diseases). If neither of these are met, then the subject should be diagnosed with Major NCD due to Possible Alzheimer's Disease. For Mild NCD, evidence of a causative AD genetic mutation is sufficient and required for diagnosing Probable AD, and Possible AD is diagnosed when (a-c) are met but there is no evidence of genetic mutation. In this case, however, (a) only requires deficits in the learning and memory domain.

Additional behavioral specifications are also outlined so as to support diagnosis, since 80% of individuals with major NCD due to AD exhibit behavioral and psychological symptoms **(**APA, 2013, 612**)**. For AD, these include possible depression and/or apathy at the mild level, and more psychotic manifestations, including agitation, irritability and wandering, at

more moderately severe levels. More extreme cases might also exhibit gait disturbance, dysphagia, incontinence, myoclonus and seizures.

Another key set of criteria for the diagnosis of AD comes from the work of the National Institute on Aging and the Alzheimer's Association in their 2011 revision of the criteria they had set forth in 1984 (McKhann et al., 2011). Although valuable for clinical diagnosis, their criteria differs form the DSM-5 in that it offers a more helpful approach for guiding research of AD. The main way they do this is by defining a new classification, 'AD dementia with evidence of the AD pathophysiological process,' that is meant solely for research purposes, since it can include evidence from autopsy.

Other crucial changes to the criteria set forth in 1984 had to be made to account for the 27 years of research on AD. Importantly, they distinguish between the AD-pathophysiological process and AD dementia, the former encompassing both the antemortem biological changes and the postmortem neuropathological substrate and the latter referring to the clinical syndrome that arises from such pathophysiological process. They expanded the criteria so that AD diagnosis would no longer require memory deficits, in order to account for cases of AD pathophysiology that present in nonamnestic forms. The DSM-5, published three years later, still requires memory impairment for the diagnosis of mild, major, probable and possible AD. The 2011 guidelines also incorporate more detailed accounts of other disorders that might often co-occur or be confused with AD, particularly of Dementia with Lewy bodies, vascular dementia, behavioral variant frontotemporal dementia, and primary progressive aphasia. Finally, the 2011 guidelines takes into

consideration research progress in the development of AD biomarkers and genetic mutations to give a more comprehensive account of what is known about AD.

The core clinical criteria for the diagnosis of AD are not too different from what we find in the DSM-5. For probable AD, there must be a clear decrease in cognitive function, with deficits demonstrating insidious onset and gradual progression, and no evidence of another condition such as cardiovascular disease or dementia due to Lewy Bodies. The nature of the cognitive deficits determines what kind of variant of AD the subject presents. It can either be amnestic, and require deficits in memory and recall, or non-amnestic, and involve the language, visuospatial, and/or executive function domains. For any of these variants, deficits in only one domain are not enough to diagnose AD, as they specify that at least two are required to be diagnosed with any type of dementia. If the subject continues to show decline in cognitive function, or if they are a carrier of one of the causative AD genetic mutations (in APP, PSEN1, or PSEN2), then the subject is diagnosed with Probable AD dementia with increased level of certainty. Possible AD dementia, according to the diagnostic guidelines, applies to subjects who meet most of the core criteria except that they either (a) show evidence of etiologically mixed presentation or (b) have symptoms that emerged suddenly (not insidiously) or for which there is not enough historical detail or neuropsychiatric evidence.

Finally, they specify criteria for Probable and Possible AD dementia with evidence of the AD pathological process. This diagnosis can be made when there is either evidence of AD pathology from autopsy or enough biomarker evidence to “increase the certainty that the

basis of the clinical dementia syndrome is the AD pathophysiological process," but the authors warn against using biomarker evidence routinely for diagnosis (McKhann et al., 2013, p. 6). The incorporation of biomarkers into our definition of AD is a crucial issue to consider, one discussed briefly in these guidelines, and one to which I will return later. For now, it is important to note that, in their consideration of biomarkers, the NIA/AA conclude that (in 2011) "the data are insufficient to recommend a scheme that arbitrates among all different biomarker combinations" and that more research is needed to standardize them and introduce them into the core criteria (McKhann et al., 2013, p. 8). For Possible AD dementia with evidence of the AD pathological process, the subject must have met the diagnostic criteria for a non-AD dementia.

Cognitive Domains in the Clinical Diagnosis of Alzheimer's Disease

In order to evaluate the validity and standardization of the criteria used to diagnose AD, and to come closer to a definition of AD that encompasses all variants of the disorder and that is based on what we know about the brain, it is important to consider what cognitive domains each set of criteria is based on. As we have seen, the NIA/AA guidelines and the DSM-5 criteria already show a crucial distinction in their evaluation of the cognitive domains affected by AD. The DSM-5 requires impairments to learning and memory whereas the NIA/AA guidelines allow for non-amnestic versions of AD that instead primarily impair other domains. At a lower level, however, the underlying categorization of cognitive function is a major factor in coming to a diagnostic conclusion, and when compared to a more robust research-based classification of cognitive domains can point to

how in line our clinical definition of AD is with current research standards in cognitive neuroscience. The National Institute of Mental Health's (NIMH) Research Domain Criteria (RDoC) offer us such a robust scheme. It is meant as a guiding framework for researchers, one that incorporates information from multiple domains (i.e. genomics, behavior, self-report) to approach a generalized and standardized understanding of mental health and illness that can be adapted to research findings (NIMH, 2013). While acknowledging the benefit of having reliable diagnostic standards, the NIMH criticized the lack of validity of the DSM's clinically based approach to pathopsyiology, and proposed these domains as a counter approach (Insel, 2013).

The RDoC define five constructs within cognitive systems: 'attention', 'perception', 'declarative memory', 'language', 'cognitive control' and 'working memory'. The only domain that demonstrates almost perfect overlap with the NIA/AA guidelines and DSM-5 criteria is the one pertaining to language. Let us first see how these categories compare to the main domain of interest when making AD diagnostic decisions in the DSM-5 and the NIA/AA guidelines, the one involving memory (although the NIA/AA allow for non-amnestic variants, they acknowledge that the amnestic version is the most prevalent).

In the DSM-5 this is the 'learning and memory' domain, and it involves assessment of immediate memory, free recall, cued recall and recognition memory (the last three of which are grouped under recent memory) (APA, 2013, p. 594). This fits reasonably well with the description of declarative memory in the RDoC, which emphasizes the processes of recall and recognition, but it also takes into consideration immediate memory, as

assessed in the subject's ability to repeat a list of numbers or words. Immediate memory impairments are in line with working memory impairments, which in the RDoC are encompassed by the 'working memory' domain, but in the DSM are recommended as a subprocess within executive functioning. The DSM's distinction between immediate memory and working memory seems to reflect a distinction between impairments in immediate recall for the former and impairments in recall and manipulation for the latter, but their inclusion in two different domains means that if a subject reflects impairments in both recall and manipulation of working memory, the subject would demonstrate deficits in two different domains, whereas when using the RDoC, the deficits would be subsumed only under the separate 'working memory' cognitive construct. When using the NIA/AA domains, memory impairments fit within the somewhat vague category encompassing deficits in acquiring and remembering, but there is no explicit mention of working memory impairments. Additionally, because none of the NIA/AA categories explicitly deal with attention processes, attention deficits might be included in this category when they affect the subject's ability to acquire information, or they might also be taken to reflect impairments in complex task handling, which is one aspect of another one of their domains that is more concerned with executive functioning. The distinction between deficits to executive functioning and memory updating can be difficult to make, since they both reflect aspects of cognition that often work together, along with attention processes, to guide behaviors. Because of this, it is important to have specific descriptions with non-overlapping components and precise assessments for each, which the RDoC do by defining a category for attention, declarative memory, cognitive control and working memory, separately, with separate recommended assessments.

The DSM-5 and NIA/AA guidelines also contain domains with processes that are not part of what the RDoC defines as 'cognitive functioning' processes, such as those subsumed under the 'social cognition' category in the DSM-5 and the 'personality/behavior' category of the NIA/AA. Both categories are best reflected in the RDoC criteria under the 'social processes' domain, which is distinct form the 'cognitive systems' domain involving cognitive processes. The lack of correlation between categories here raises two important issues. In the case of the NIA/AA, the category involves behavioral symptoms that in the DSM-5 are included only in the auxiliary behavioral specifications, which exist in roughly 80% of major NCD patients. This category is best reflected in the RDoC criteria under the 'social processes' domain, but also specifies behaviors such as compulsive or obsessive behaviors that do not perfectly fit within this category (in that they reflect inhibition deficits, for example). It raises the question of whether a category so loosely defined by a wide array of behavioral or personality-related symptoms, mostly involving social processes that the RDoC does not consider to reflect major cognitive processes, should be used as a domain of cognitive deficits for AD diagnosis.

The 'social cognition' category of the DSM-5 deals more specifically with deficits in emotion recognition and theory of mind, which are, again, subsumed under the 'social processes' category in the RDoC, the former reflected specifically in the 'social communication' subconstruct and the latter in the 'perception and understanding of self' subconstruct. Because of the more direct reference to emotion recognition deficits, this category brings up another significant issue when creating domains. Constructivist theories

of emotion, which have been gaining increased support in recent years for challenging classical theories of emotion, suggest that it is not possible to separate the emotional and cognitive processes à la Plato's chariot metaphor, where cognition directs and limits emotional processes, because emotional processes are not emotion-specific and so are involved in the formation of any cognitive process (Barrett, 2017). In the RDoC, both emotion recognition and emotional experiences are not considered within the cognitive domain. Emotion recognition is in the social processing domain, and emotional experiences are best accounted for in the arousal subdomain of the arousal and regulatory systems construct, if arousal is taken to be an inherent and basic property of emotional experiences (which according to most theories, it is). The issue becomes how to define emotion related disturbances in cognitive functioning. The DSM-5 does this by introducing emotion recognition deficits within the social cognition domain, but this does not correspond to the categories of the RDoC guidelines, and it still only reflects one aspect of emotional processing that might be affected.

Another category in the DSM-5 that does not correlate well with the RDoC constructs is the Perceptual-Motor domain, which encompasses deficits related to visual perception, visuo-construction, perceptual-motor integration, praxis and gnosis. This includes a wide array of deficits that deal with perception, movement and their integration through learning. The focus on perception is in line with the NIA/AA category involving visuospatial abilities and corresponds well to the visual perception subconstruct in the RDoC framework, although it specifically includes a subject's impaired ability to "orient clothing to the body" which in the DSM-5 is mentioned under the social cognition domain

(McKhann et al., 2013). The focus on movement and integration of perception with it fits best within the motor action construct in the RDoC domain for sensorimotor systems, which is defined separately from cognitive systems. The idea might be that subjects with AD will likely exhibit deficits in tasks that involve integrating perception and movement because of possible deficits in perception and the co-dependence of those systems in tasks that involve their integration. The RDoC motor systems construct does suggest that motivational processes from other domains will co-occur with motor actions (NIMH, 2013). However, they also say that motor systems explicitly includes the modulation of motor actions through learning, and so whether it reflects a unified set of cognitive deficits is not clear. Nevertheless, the category seems useful in carrying out cognitive assessments because it focuses on the interaction between perception and motor abilities, and is more detailed and inclusive than the visuospatial abilities category of the NIA.

Toward a Biomarker-Based Characterization of Alzheimer's Disease

The lack of correlation between cognitive domains is problematic when classification of deficits according to these domains makes up the core criteria for AD diagnosis. Having different diagnostic criteria logically leads to having differences in diagnostic outcomes. In a 2017 study, Dolci et al. compared the diagnosis of AD of 94 subjects using either criteria from the DSM-IV plus National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) or the 2011 NIA-AA guidelines, and found that 29% of the subjects

classified as demented according to the NIA-AA guidelines were not demented according to the DSM-IV criteria (Dolci et al., 2017).

It is necessary to come to an understanding of Alzheimer's Disease that allows for accurate diagnosis based on all available information, and that reflects a unified set of possible pathological changes that can lead to specific clinical symptoms. However, the widely used criteria outlined thus far portray a clinical-pathological definition of Alzheimer's Disease that treats clinical syndromes as reflective of AD pathological change without neuropathological evidence. Using a syndromal definition of AD is problematic, first of all, because it does not allow for early detection and intervention, since it cannot diagnose subjects with pathological AD but no symptoms. The prevalence of mixed pathologies that might affect cognitive function in old age also demonstrates the limits of such a definition, because in those cases it is more difficult to associate specific cognitive deficits with a single pathology. The DSM-5's incorporation of detailed cognitive domains is supposed to help in making such distinctions. For example, for cardiovascular disease the diagnosis relies partly on cognitive deficits within the 'complex attention' domain, whereas AD diagnosis, as we have seen, relies on impairments specific to the 'memory and learning' domain. Diagnosis of AD and its differential diagnosis hinges on the categorization of deficits according to cognitive domains, but our analysis suggests that these domains are not yet standardized, particularly when it comes to the categorization of deficits in working memory or executive function, social processes and emotional processes. Measuring cognitive decline in these different domains could be useful to see what processes are most targeted, but using them to include/exclude people from diagnosis is more problematic,

especially because the DSM-5 requires impairments in the learning and memory domain for AD diagnosis.

The reality is that the amnestic symptoms we often associate to AD are not specific to cases of AD neuropathology. 10-30% of patients diagnosed with AD show no evidence of AD pathology, and 30-40% of cognitively unimpaired elders (over age 80) reveal AD pathological changes at autopsy, suggesting that the error rate for AD diagnosis according to amnestic symptoms as has been presented in the DSM-5 can be of over 50% in the elderly (Jack et al., 2018, p. 552). As the NIA/AA acknowledges in its 2011 guidelines, there is enough evidence to suggest that nonamnestic syndromes, such as progressive aphasia, which is associated with language impairments, or posterior cortical atrophy, which is usually associated with visuoperceptual and spatial deficits, present the pathological changes associated with AD (Alladi et al., 2007, Rabinovici et al., 2008).

In order to improve our current understanding of AD, the NIA/AA recently published a new research framework for studying the condition that challenges the clinical-pathologic view of AD assumed in previous diagnostic criteria guidelines (Jack et al., 2018). They propose a biologically-based definition of AD in living persons that relies on neuropathological findings through biomarkers. The idea is that the AD pathological changes that are validated through postmortem examination can now be detected through a combination of multiple AD biomarkers that are capable of reflecting these changes *in vivo*. This new definition implies that, when designing studies of possible intervention

methods for AD, criterion validity is established if the intervention modifies both biomarkers and cognitive symptoms.

The NIA/AA establish three categories of AD biomarkers which measure distinct aspects of neuropathological change. Through consideration of biomarkers in each category, a subject can is given a 'biomarker profile' that defines where they place on the (new) Alzheimer's Continuum. Category A measures β-amyloidosis, or the deposition of β-amyloid (Aβ) in the brain, through either cortical amyloid PET or cerebral spinal fluid (CSF) measures of Aβ42. Category T contains biomarkers that measure tauopathy, or the accumulation of fibrillary tau which gives rise to neurofibrillary tangles, and includes CSF measures of phosphorylated tau and cortical tau PET. Category (N) refers to biomarkers of neurodegeneration, and includes CSF measures of T-tau, FDG-PET, and MRI atrophy. This last category is put in parenthesis because, as opposed to A and T, it measures changes that are not specific to AD, as neurodegeneration can occur in non-AD conditions, especially with cases of mixed etiologies in the elderly (Jack et al., 2018, 539). Because of this, it is not considered in their definition of Alzheimer's disease, but it does affect a subject's placement in the Alzheimer's continuum, and, together with clinical symptoms, is important to their definition of AD stages, which is introduced later.

Their proposed definition, therefore, is based only on the biomarkers in A and T. More specifically, if a subject has abnormal values of A, they are placed in the Alzheimer's continuum. The subject will then be placed in the Alzheimer's Disease category if they also exhibit abnormal values in T, regardless of the results of N. If, instead, no abnormal

values in T are detected, the subject is placed in either the 'Alzheimer's pathological change' category, if results from N are normal, or in the 'Alzheimer's and concomitant suspected non Alzheimer's pathologic change' category, if there is evidence of neurodegeneration. The biomarker profiles and their respective category are outlined in Table 1. Whether results in each biomarker category are normal or indicative of Alzheimer's depends on a cut-point. Because AD pathology is defined as a continuous process and not a binary, they also suggest that investigators take a 'three range approach' in which two cut points separate the results as either clearly normal, in the intermediate range, or clearly abnormal. The more lenient cut point can be used in studies focused on early detection of pathological change, and the more conservative one can be used in studies were diagnostic accuracy is critical, an approach used in understanding biomarker results for other diseases (Jack et al., 2018).

They then designed disease staging matrices that incorporate cognitive symptoms but keep them separate from the biological biomarker-based definition of Alzheimer's by considering the subject's biomarker profile on one axis, and the severity of cognitive deficits on the other. They divide the cognitive deficit continuum in cognitively unimpaired, Mild Cognitive Impairment (MCI) and Dementia. The distinction between the two last categories is the same as the distinction between major and mild Neurocognitive Disorder in the DSM-5, and hinges on the impact of the symptoms to independence in daily functioning. Table 2 shows their application biomarker profiles to clinical staging.

Table 1: Possible biomarker profiles with their respective diagnostic category as suggested by the NIA/AA. Blue cells corresponds to profiles associated with the Alzheimer's Continuum, with darker cells corresponding to profiles that are AD-specific. Adapted from

Biomarker Profile	A	T	(N)	Diagnostic Category
A- T- (N)-	-	-	-	Normal AD biomarkers
A+ T- (N)-	+	-	-	Alzheimer's Pathologic Change
A+ T+ (N)-	+	+	-	Alzheimer's Disease
A+ T+ (N)+	+	+	+	Alzheimer's Disease
A+ T- (N)+	+	-	+	Alzheimer's and concomitant suspected non Alzheimer's pathologic change
A- T+ (N)-	-	+	-	Non-AD pathologic change
A- T- (N)+	-	+	+	Non-AD pathologic change
A- T+ (N)+				Non-AD pathologic change

Table 2: Syndromal staging within the AD specific biomarker profiles. Adapted from Jack et al. (2018)

Biomarker Profile	Clinical Diagnosis		
	Cognitively Unimpaired	Minor Neurocognitive Disorder	Major Neurocognitive Disorder
A+ T+ (N)-	Preclinical AD	Prodromal AD	Major ND due to AD
A+ T+ (N)+			

Increased risk of quick short term clinical progression →

This model for understanding AD might prove to be very useful in the research domain, and with better standardization of AD biomarker approaches and accessibility, it might also change the way we diagnose AD in everyday settings, but it certainly raises some issues. One one hand, it is an attractive shift in the literature because it makes use of new, standardized biomarker information and incorporates it into our model of the disease, therefore making efficient use of available information. Furthermore, it is a more quantifiable and biological-based definition, meaning that it is less likely to result in the

clinical-pathologic inconsistencies mentioned at the start of this section. A more biological approach also lessens the effect of discrepancies between the cognitive domain categories outlined by different institutions, and specifically allows for identifying and accounting for non-amnestic presentations of the disease. Importantly, it no longer makes use of the clinical presentation of AD as a starting point. By not assuming that a worse clinical presentation of AD corresponds to increased AD pathology, it leaves room to identify non-pathologic factors that might contribute to a worsening clinical condition in individual subjects and that might help us understand more about AD.

On the other hand, however, we must consider whether we know enough about the pathological process of AD as a discrete entity to define AD strictly in biological terms. Are biomarkers capable of detecting AD-specific pathological changes? This is particularly important in the design of future studies that may rely on biomarker information to give insights into AD pathology, and even more important if these studies make use machine learning algorithms to extract AD relevant meaning from biomarker data. Studies that attempt to derive accurate classification methods for the early detection of AD using machine learning need to be able to learn AD-specific differences in biomarkers, which can only be achieved with an understanding of what characterizes the pathological process of AD and how or when each biomarker reflects this process. We need an robust apprehension of the causal relationship between these biomarkers and the neuropathological processes they relate to in order to evaluate marker results as well as the prioritization scheme suggested by the NIA/AA. Does the proposed discriminating role of amyloidosis biomarkers and non-specific role of neurodegeneration markers fit with our

understanding of the pathological process? The next section on analyzing what we know about AD pathology in order to answer these questions by considering current pathological verification and staging standards in addition to dominant theories and research literature pertaining to the causal role of the different pathological elements.

Alzheimer's Disease Pathology

Major Components of Alzheimer's Pathology and the Amyloid Cascade Hypothesis

Pathological studies performing postmortem evaluations dating back to Alzheimer's study in the early 1900s have identified two main hallmark lesions of AD; Aβ plaques and neurofibrillary tangles (NFTs). These lesions are found in the brains of AD subjects along with neurodegeneration, which reflects the neuronal and synaptic loss thought to be caused by the lesions, and which best correlates with the rate and nature of cognitive decline in AD (Serrano-Pozo et al., 2011). The main in-vivo biomarkers for neurodegeneration includes detection of area-specific atrophy in MRI and FDG PET, the latter of which uses fluorodeoxyglucose, an analog of glucose, to measure changes in synaptic and neuritic functioning (which are largely glucose-dependent) (Jack et al., 2018).

NFTs result from the aggregation of hyperphosphorilated molecules of the microtubule associated protein (MAP) tau in the form of paired helical filaments in cell bodies. A hyperphosphorilated form of tau is more negative than its regular form, and as a result it prefers to bind to other tau molecules than to the microtubules it is associated with, which leads to the disintegration of the microtubule **(**Luo, 2015, p. 469**)**. When the misfolding and aggregation of tau occurs in the axons or dendrites, the result is a more fibrous and usually smaller structure labeled neuropil threads (NTs**)** that are also revealing of AD pathology (Dening & Thomas, 2013, p. 88). Biomarkers that reflect tau pathology include elevated CSF phosphorylated tau (p-tau), which measures the levels of phosphorylated tau

in cerebral spinal fluid, and tau PET, which uses relatively recently developed tau-specific tracers to localize tau deposits in the brain (Jack et al., 2018).

Aβ plaques are extracellular aggregations of the Aβ peptide that result from an overproduction of Aβ through abnormal β-secretase and γ-secretase cleavage of the Amyloid Precursor Protein (APP). APP cleavage can produce molecules of Aβ40 or Aβ42. Aβ40 constitutes 95% of cerebral Aβ and is more soluble than its variant form (Serrano-Pozo et al., 2011, p. 9). Aβ42 is more likely to aggregate, and therefore thought to lead to the formation of toxic plaques characteristic of AD. Biomarkers that measure levels of fibrillary Aβ deposition include high ligand retention on amyloid PET or low CSF Aβ42 (Jack et al., 2018).

Because Aβ plaques do not correlate with cognitive impairment in AD subjects, and because plaques are often found in autopsies of cognitive unimpaired subjects, morphological characterization of types of plaques is extremely important in order to distinguish toxic, AD-related plaques to others that might be reflective of normal aging. Unfortunately, the literature does not converge on comprehensive plaque nomenclature, and terms are often used interchangeably. The main (and apparently most useful) morphological distinction is between diffuse and dense-core plaques.

Diffuse plaques are large, have ill-defined borders (which means they are sometimes not even referred to as plaques, just deposits) and are often found in cognitively unimpaired elders. Dense-core plaques, also known as senile, local or neuritic (although ‘neuritic’

should be used to entail more than just dense-core), are smaller and more focal and, importantly, are generally associated with negative effects to the surrounding neuropil, which might include neuronal or synaptic loss, abnormal activation of microglial cells, reactive astrocytes or distended axonal or dendritic processes referred to as dystrophic neurites (Serrano-Pozo et al., 2011). The mechanism through which Aβ deposits can lead to neuronal or synaptic loss and whether microglial cells and reactive astrocytes have any causal role in the pathophysiological process are heavily disputed. In addition, the plaques most associated with AD pathology contain levels of hyperphosphorilated tau in dystrophic neurites. The term 'neuritic plaques' is usually used to refer to these more AD-specific plaques, which are important because they are areas where tauopathy and amyloidosis apparently integrate, and so are critical to understanding the causality of the pathological processes. A central question in AD is whether neuritic plaques are necessary and sufficient for the development of tau pathology, because if this were the case it would show a downstream causal role from amyloidosis to tauopathy. Whether diffuse plaques, also found in the brains of subjects of AD, have an effect on the subsequent pathological process also needs to be considered to answer this question.

Understanding AD, therefore, requires uncovering the causal relationship between tau and Aβ aggregates, and modeling their connection to subsequent neurodegeneration and cognitive decline. The prevailing view of the pathological stream in the literature is the amyloid cascade hypothesis, which was suggested first by Hardy and Higgins in 1992 (Hardy & Higgins, 1992). It hypothesizes that AD pathology begins with amyloid accumulation in limbic and association cortices, which leads to the formation of Aβ

plaques. Neuritic plaques affect synaptic functioning and lead to microglial and astocytic activation, which eventually alters kinase and phosphotase activities enough to result in the formation and spread of NFTs. This is followed by widespread neuronal dysfunction and selective neuronal loss (neurodegeneration). In terms of biomarkers, this hypothesis suggests that AD pathological changes will first be detected by CSF measures of Aβ42, followed by Amyloid PET, CSF levels of phosphorylated tau, tau PET and finally by neurodegeneration biomarkers, including MRI atrophy measures and FDG PET.

Evidence for the amyloid cascade hypothesis comes largely from genetic studies of populations that demonstrate a higher risk for AD. Studies on the familial variant of AD (FAD) showed that a missense mutation in the Aβ section of the APP gene increases the likelihood of AD (Goate et al., 1991). In humans with trisomy 21 (Down Syndrome) an extra copy of the APP gene corresponds to higher rates of the disorder and earlier onset (Luo, 2015, p. 471). Mutations in presenilin 1 and 2, two major components of the Aβ production enzyme γ-secretase, lead to higher rates of Aβ42 and early disease onset (Selkoe & Hardy, 2016). Subjects with allele ApoE4 (instead of ApoE3), through increased binding to Aβ, demonstrate a lower rate of Aβ clearance and higher rates of AD (Selkoe & Hardy, 2016). In addition to genetic evidence, recent studies that have injected rodents with human Aβ oligomers have suggested that that Aβ accumulation can lead to synaptic dysfunction, tau hyperphosphorylation and neuritic dystrophy (Jin et al., 2011). All of these are considered evidence for the crucial and causal role of Aβ accumulation in AD.

However, more evidence is needed to prove the causal role of Aβ in AD, as some inconsistencies remain. One issue is that the amyloid cascade hypothesis does not explain the shocking observation that Snowdon first made in the 'Nun Study,' and that pathological studies since have corroborated, that cognitively unimpaired elders often show high levels of Aβ plaques in the brain (Snowdon et al., 1997; Nelson et al., 2011). This had led many to believe that Aβ plaques themselves, although typical of AD, might not have a downstream causal role, since they often appear without the tauopathy, neurodegeneration and cognitive dysfunction of AD.

Furthermore, the amyloid cascade hypothesis does not provide us with a well-supported model of how Aβ itself becomes toxic and how this toxicity causes the onset of the disease. Part of the issue is that studies focused on uncovering the mechanisms through which Aβ becomes toxic are based on FAD or other cases where genetic mutations are the cause of changes in Aβ production. Although they might point to the centrality of Aβ in AD, they do not tell us how Aβ production can become maladaptive in mutation-free subjects. The toxicity of Aβ in these cases might be a result of something else entirely. A recent study found that people with trisomy 21 had increased Aβ deposition regardless of whether they had an extra copy of the APP gene, suggesting that the mechanism that causes amyloid deposition is much less clear than what was thought to be, and might include genes that are not specific to amyloid (Wiseman et al., 2018) In addition, evidence of Aβ leading to dystrophic neurites and tauopathy has come from tissue culture or mice studies that produce limited results with unrealistic setups (Drachman, 2014). Studies focusing on the possible causal role of Aβ in situations where it is highly overexpressed run the risk of

ignoring the role of tau in bringing about these changes. In fact, reducing the amount of tau in mice that overexpress Aβ can prevent the negative downstream effects of synaptic dysfunction and even cognitive decline (Roberson et al., 2011). Furthermore, active immunization studies have been able to decrease the amount of Aβ in subject's brains but have not resulted in decreases in the rate of cognitive decline, even when plaques were removed (Nicoll et al., 2016).

Another issue involves the localization and spread of the disease. Aβ levels or the number of Aβ plaques do not correlate with the rate of neurodegeneration and cognitive decline in AD subjects, which occur much later in the disease and are best correlated with the amount of neurofibrillary tangles. The amyloid cascade hypothesis postulates that Aβ deposition only initiates the pathological process, through which it causes neural and synaptic loss that is logically more correlated with the level of cognitive function. This hypothesis would fit well with findings that amyloid deposits exist primarily in areas that are more prone to develop neurofibrillary tangles and neurodegeneration in the initial stages of disease progression, but as we will see later, this is not the case. The amyloid cascade hypothesis needs an explanation of how widespread amyloid burden, or plaques in certain regions, can lead to tauopathy or neurodegeneration in different, localized areas. This requires first and foremost a better understanding of the localization and spread of amyloidosis, tauopathy and neurodegeneration in AD, and how these relate to the presentation of the clinical symptoms of the disease (whether their amnestic, or dealing more with executive function, etc.)

In order to address these issues, and evaluate the contributions of new findings to the amyloid cascade hypothesis, I will explore the current standards for identifying AD pathology in postmortem brains as presented in 2012 NIA/AA guidelines and the standardized staging models these criteria are based on: the Braak staging for tauopathy and the Thal staging for Aβ deposition. Autopsy verification of AD is considered the most dependable (albeit impractical) way of identifying the condition, and so best reflects what we consider to be AD. Autopsy verification relies heavily on the Braak and Thal staging schemes, and so these are also examined below. The goal is to get a more in depth understanding of what AD is so that we can evaluate whether the amyloid cascade hypothesis and the new biomarker based framework for are adequate forms of characterizing AD.

Standards in the Postmortem Verification and Staging of Alzheimer's Disease

In 2012, the NIA/AA outlined an ABC scoring system that should be used in the evaluation of postmortem brains to identify AD (Hyman et al., 2012). This scoring is based on the semiquantative evaluation of the number of Aβ plaques according to the Thal Staging system (A), NFTs according to the Braak staging system (B), and neuritic plaques according to the The Consortium to Establish a Registry for Alzheimer's Disease (CERAD) guidelines (C). The separate assessment of general Aβ plaques and neuritic plaques was decided on to account for the range of Aβ deposits that are present in the brain (in A) while at the same time quantifying how many of those reflect surrounding neuritic or synaptic damage (in C). Except for in the late stages of neuropathology, there is no clear

relation between the semiquantitative assessment of neuritic plaques according to CERAD and the topographical distribution of Aβ aggregates described by Thal, suggesting that these categories should be evaluated separately (Hyman et al., 2012). The standards are adapted to each fit a four point scale (0-3), and final reporting includes reference to all parameters (i.e Alzheimer Disease Neuropathologic Changes: A3, B2, C2) and can be converted to a four point scale of AD neuropathalogic level (Not, Low, Intermediate, High). According to this conversion, 'Not' can only be applied to cases with an A0 and C0, where there is no evidence of Aβ deposits. Once a subject has at least some Aβ deposits (A1 and above), if they do not have a score of B2 or B3 (Braak stages III-VI), then they are assigned to the 'Low' group despite the amount of Aβ or neuritic plaques. 'High' is reserved for subjects with high amounts of Aβ (A3) and neuritic (C2 or C3) plaques and an advanced Braak Stage (B3).

Importantly, the NIA/AA suggest brain regions of interest for each parameter that should be evaluated hierarchically for Aβ deposits. First considerations include the middle frontal gyrus, the superior and middle temporal gyri and the inferior parietal lobule, all of which should be evaluated for all three parameters. If these regions test positive for Aβ deposits, then the next regions to be evaluated are the hippocampus and entorhinal cortex, where C can be considered but only A and B score should be used for scoring, followed by the basal ganglia at the level of the anterior commissure with basal nucleus of Meynert, where B can be considered but only A defines the score. Then, if the evaluation is positive in these regions, the pathologist should evaluate the levels of A in the midbrain including the Substantia Nigra and in the cerebral cortex and dentate nucleus area. The occipital cortex,

particularly Brodmann Areas 17 and 18 should be stained for NFTs and scored on B, and A and C should be considered in these areas. These guidelines also suggest differential diagnosis of Lewy Body Disease, Vascular Brain Injury and Hippocampal Sclerosis through staining in particular regions. These regions of interest are derived from the findings of Braak et al., Thal et al. and CERAD.

With regards to the distribution and progression of NFTs, Braak outlines six stages that define the typical case of AD tauopathy (Braak et al., 2006). Their results, obtained through immunocytochemistry and the Gallyas silver staining technique, suggested that tauopathy begins in a specific area in the perirhinal cortex, (which they refer to as the transentorhinal cortex to capture its structural similarity and proximity to the entorhinal cortex) a region highly associated with memory, and progresses along the neocortex and inwards towards the hippocampus. More particularly, NFTs first appear in the tranentorhinal cortex (Stage I), and then progress to the entorhinal cortex and CA1 and CA2 regions of the hippocampus proper (Stage II), followed by extension toward sensory areas of the temporal neocortex, particularly affecting the neocortex of the fusiform and lingual gyri (Stage III). Stage IV shows involvement of most of the occipital-temporal gyrus as well as new involvement of part of the insular cortex and a wider involvement of neocortical association areas. In Stage V, NFTs spread to the superior temporal gyrus and mildly affect the high order association areas of the frontal, parietal, and occipital neocortex. By Stage VI, tauopathy reaches the secondary and primary neocortical areas and in the occipital lobe has spread to the primary visual cortex (affecting the striate and

parastriate areas of the occipital neocortex, mostly in Brodmann areas 17 and 18, respectively).

Thal et al. detected Aβ deposits using both the Campbell-Switzer silver technique, best suited for detecting deposits witih dystrophic neurites, and by immunohistochemistry, which detects all accumulations of Aβ (Thal et al., 2002). Their results suggested that Aβ deposits begin to appear in the neocortex and progress inwards toward the midbrain. In phase 1, deposits are found in areas of the neocortex, usually focused in the temporal and occipital regions. Phase 2 shows involvement of the entorhinal cortex, area CA1 of the hippocampus, and the insular cortex primarily, but also sometimes extending into the amygdala or cingulate gyrus. Phase 3 involves deposits in a number of subcortical regions, including the caudate nucleus, the putamen, the claustrum, the basal forebrain nuclei, the substantia immobilata (which includes the nucleus basalis of Meynert) the thalamus and the hypothalamus. In addition, deposits may extend further into the white matter, the hippocampus and into some areas of the central gray in the midbrain. In Phase 4 of β-amyloidosis deposits appear in the brainstem's medulla oblongata and in the substantia nigra of the basal ganglia, and are increased in the central gray in the midbrain and area CA4 of the hippocampus. Phase 5 is characterized by extension into a series of brainstem nuclei, including the locus coeruleus in the pons, the tegmental nuclei and the Raphe nuclei as well as the granular layers of the cerebellum, leaving the dentate nucleus intact.

The standardized progressions suggest differential pathways for amyloidosis and tauopathy in the brain. Amyloid accumulation begins and progresses more widely, covering most of

the cortical areas and extending rather unselectively towards areas of the midbrain. Tau aggregation begins in the upper layers of the transentorhinal cortex, spreading to the rest of the transentohrinal cortex, the entohrinal cortex, some subcortical nuclei and part of the hippocampus and then accumulating in neocortical association areas. An important aspect of the spreading pattern of tau as outlined by Braak and Braak and replicated more recently by Cho et al. is that the spread of NFTs is correlated with the severity and duration of cognitive decline in AD (Cho et al., 2016, Huber et al., 2018). The first part of tau progression, Stages 1 to 2 are associated with prodromal Alzheimer's disease, the later stages 3 to 4 with mild cognitive impairment or mild neurocognitive disorder, and the final stages 5 to 6 with Alzheimer's disease (Spillantini et al., 2013). Genetic studies have also suggested that tau dysfunction is sufficient to cause neurodegeneration and dementia (Goedert, 2004). Furthermore, tau aggregates spread in a more step-wise manner than amyloid deposits, as does regional volume atrophy (Cho et al., 2016). This suggests that tau propagation relies on connectivity by spreading through white matter tracts rather than through proximity, which might be a better fit for amyloid deposit spread. This distinction may be due to the fact that NFTs are intracellular accumulations whereas Aβ plaques appear extracellularly. Further evidence for the centrality of connectivity to tau spreading comes from tracking infusion-related tau pathology in mice and observing rapid and robust propagation of tauopathy through particular tracts (Ahmed et al., 2013).

Updating our Understanding of AD Pathology: Implications for the Amyloid Cascade Hypothesis and the NIA/AA Biomarker-Based framework

The connectivity-dependency of the propagation of tau might be one of the reasons why recent work has brought to light possible errors in the seemingly generalizable standard Braak stages that question what we consider to be the beginning of the pathophysiological process of AD. To come to a better understanding of the beginning of AD, Braak conducted a study in 2011 that involved examining the brains of AD subjects under 30 years old. Results demonstrated that rather than starting in the transenthorinal cortex, tau accumulation begins in subcortical areas with diffuse connections to the cortex before the appearance of NFTs, particularly in the locus coeruleus, a significant catecholaminergic nuclei (Braak et al., 2011). Importantly, Braak presented this as evidence against the amyloid cascade hypothesis, because of all the subjects involved, only 1/42 had amyloid deposits whereas 38/42 had tau aggregates, meaning that tau aggregates before amyloid, and so refuting the causal role of amyloid.

More recent studies have been able to use advanced connectivity measures, as well as improved and more standardized AD biomarkers to shed additional light on the progression of tauopathy and amyloidosis. A 2016 longitudinal voxel-based morphometry study demonstrated that degeneration in the nucleus basalis of Meynert, another catecholaminergic nuclei with diffuse cortical connections, precedes and predicts degeneration in the entorhinal cortex (Schmitz et al. 2016). Moreover, they found that memory impairments typical of AD consistently appeared only after the neurodegeneration had spread from the nucleus basalis of Meynert to the entorhinal cortex, suggesting a specific pathway that when damaged produces the well-known clinical symptoms.

Incorporation of non-amnestic presentations of AD would be a valuable next step in localizing this pathway and finding out how and why nonamnestic presentations differ.

Other attempts have been made at linking the progression of AD pathological markers to functionally significant processes in the brain to potentially help explain the relationship between pathology and clinical symptoms in AD. Some have pointed out that the regions heavily involved in AD tauopathy, particularly the entorhinal cortex, the hippocampus and the association cortices, together with the parahippocampal cortex, are also the key components of what is known as the posterior Default Mode Network (DMN), a network that appears in resting state subjects during fMRI and that has been sometimes associated with processing of information regarding the self as well as with other cognitive processes (Cho et al., 2016).

In an impressive multimodal connectivity study, Jones et al. demonstrated that the connectivity of the posterior DMN decreases throughout the the course of AD in a manner consistent with the known spatial involvement of pathological AD markers and that its cascading network failure begins before the formation of amyloid plaques (Jones et al., 2015). This latter finding suggests that tau-related loss in connectivity precedes the causal stream of amyloid accumulation, and so provides evidence against the amyloid cascade hypothesis. Interestingly, they also found that the connectivity within the DMN, between the posterior and ventral parts of the system, actually shows an increase as the disease progresses, after the initial deterioration of the posterior DSM. This increased connectivity was found to correlate with elevated amyloid levels and declining hippocampal volume,

key markers of AD progression. They hypothesize that as a result of failures in the posterior DMN, a transient compensatory mechanism is activated to increase connectivity between the posterior and ventral networks, and that the metabolic demands associated with this increased connectivity could be the phenomenon that triggers downstream pathological processes. This is supported by animal studies that suggest that Aβ secretion and deposition is enhanced by neuronal activity (Li et al., 2013).

Adding to these discoveries, Palmvquist conducted a study to uncover regions of early accumulations of Aβ, which they suspected would be more localized than the neocortical spread outlined in Thal's Phase 1 because of the difficulties of precise early localization resulting from the time lag between the beginning of amyloidosis and the onset of clinical symptoms (Palmvquist et al., 2017). To do this they compared non-accumulators (with normal levels of amyloid) with early Aβ accumulators, who have abnormally low levels of CSF Aβ42 and normal levels of overall Aβ (detected by PET) and whose rate of accumulation of Aβ fibrils matches those with both abnormal CSF Aβ42 and PET Aβ levels. They found that Aβ fibrils accumulate first in the posterior cingulate cortex, the precuneus and the medial orbitofrontal cortex, regions centrally involved in the DMN, before extending across the neocortex. Furthermore, they found that early accumulators exhibited hypoconnectivity within the DSM and between the DSM and the frontoparietal network, but that even earlier accumulators (with less abnormal levels of CSF Aβ42) exhibited hyperconnectivity in these areas, potentially supporting the hypothesis of Jones et al. (2016).

Rather than suggest that the amyloid cascade hypothesis is incorrect, these studies point to a more complex interaction between tau and Aβ aggregates that causes the downstream neuropathological process of AD. More specifically, they point to a need for clarification on how tau pathology spreads and what role Aβ deposition has in the intensity and direction of this process. In 2018, Jacobs et al. conducted a study that combined a multitude of tau, amyloid, hippocampal and diffusion tract imaging methods with an intricate design to test the hypothesis that "tau deposition is associated with aberrant structural connectivity under the influence of increased amyloidosis" (Jacobs et al., 2018).

They found, first, that lower baseline hippocampal volume was associated with increased mean diffusivity of the hippocampal cingulum bundle (HCB), a tract that connects the hippocampus to the posterior cingulate cortex (as well as other areas) and that forms part of a network subserving memory. Then they demonstrated that they could predict changes in overall levels of tau aggregates in the posterior cingulate cortex from baseline levels of diffusivity of the HCB. They obtained this result even when normalizing for hippocampal volume, although hippocampal volume itself was not able to predict changes in PCC tau levels, suggesting that the relationship between HCB diffusivity and PCC tau is stronger than between hippocampal volume and HCB diffusivity. This result supports the view that tau accumulates in a tract specific manner, where abnormally diffusive tracts lead to an increase in accumulation of tau in downstream regions. Importantly, this association between PCC tau and HCB diffusivity was found only in subjects with abnormal levels of Aβ and not in those with regular Aβ levels. This leads to their conclusion that the relationship between HCB diffusivity and PCC tau is stronger when the subject exhibits

high levels of amyloidosis. Finally, they were able to connect these processes to cognitive impairment by predicting memory impairments from HCB baseline diffusivity and, again, they found that the connection was stronger when the subject had abnormal levels of Aβ than when they did not.

The tract specific propagation of tau, Jones et al.'s finding of connectivity failures preceding cortical plaque accumulations and Jacobs et al.'s findings of the dependence of downstream tau related pathology on the presence of abnormal levels of Aβ, as well as some of the other findings outlined thus far, might suggest an alternative to the amyloid cascade hypothesis in which tau related changes cause downstream pathology and modulation of the toxicity of Aβ in subjects with already abnormal levels of Aβ. The differential spread of tauopathy and amyloidosis, according to which tauopathy begins in central areas in the brain and spreads towards cortical areas and amyloidosis does the opposite, is more in line with this understanding of the relationship between tau and amyloid where each component might modulate the spread and toxicity of the other. Amyloid might act as a gatekeeper for the pathological process of AD, but amyloid accumulations alone are not the central cause of the process as suggested by the amyloid cascade hypothesis. Availability of alternative theories is important for the field to not be narrowed to an amyloid-centric view.

Let us return to the A/T/(N) biomarker approach of the NIA/AA. Although they specify that their framework does not assume the amyloid cascade hypothesis, and they provide ideas for alternative hypothesis testing, the dependency on A+ markers to define the

Alzheimer's Continuum seems to indicate at least some reliance on the belief that amyloid is more central than other markers in characterizing AD.

Burke et al. conducted an autopsy based study that evaluated the presence and incidence of dementia in each NIA/AA proposed biomarker profile (Burke et al., 2018) and that provides insights into the practical applications of these biomarkers. They found that the two biomarker profiles with A+ and T+ values, corresponding to the AD specific profiles, were significantly more correlated with both prevalence and incidence of dementia, with similar rates for N+ and N- subjects. This support the NIA/AA's consensus that a combination of abnormal levels of tau and amyloid account for the traditional AD profile. However, they also found that the lowest rate of dementia prevalence and incidence occurred in subjects with only abnormal amyloid biomarkers (A+ T- (N)-), and that this prevalence was almost the same for subjects with normal biomarker levels (A- T- (N)-). Amyloid alone is therefore not enough to cause downstream cognitive decline, and not specific enough to cases of AD, but the NIA/AA does consider it enough for identification on the Alzheimer's continuum. The importance of this biomarker profile within the NIA/AA framework is likely do to an interest in identifying early pathologic change, but its preference over A-T+(N)+ profiles, which have higher rates of prevalence and incidence of dementia, seems like it is due, at least partly, to an assumption of the centrality of amyloid in line with the amyloid cascade hypothesis. If we make use of this aspect of their proposed framework, a biologically based approach might still not be able to explain how amyloid plaque burden is often found in cognitively unimpaired elders, and

might bias future studies against investigations of possible amyloid-independent tauopathy spread and progression.

Despite the possible overvaluing of A biomarkers in defining the Alzheimer's continuum, the A/T/(N) approach is largely in line with our understanding of the pathological elements of AD, particularly in their requirement of A+T+ to characterize the condition. Even though they do not account directly for non-AT(N) pathology, the A/T/(N) profiles provide a framework against which other pathological elements can be evaluated. Furthermore, analysis of the intricate relationship between amyloidosis and tauopathy presented in this section is highly suggestive of the need for a standardized system for classifying biomarkers that is interpretable and AD-specific and that facilitates the design of studies that can adequately shed light on the relationship between pathological elements of AD. The causality of the pathological elements of AD is still unresolved, but the NIA/AA biomarker approach is a useful tool for identifying different combinations of these elements and evaluating their differential effects on pathological progression, cognitive decline or other biomarkers.

Multimodal Integration of Biomarkers for Early Detection of Alzheimer's Disease Using Machine Learning

Challenges and Advantages of Multimodal Machine Learning Approaches

Machine learning has been of interest to the literature concerning AD primarily because of the possibilities it offers for early detection of the disease. Early detection of AD is crucial to our understanding of AD, for it would allow us to identify and investigate AD subjects early on in the AD neuropathological process. More importantly, early detection can contribute to the development of intervention methods that might alleviate the subsequent symptoms and neuropathologic progression of AD, and allows for patients to have earlier access to care settings where they can be treated according to the best methods available. Currently, these interventions, principally in the form of antidementia drugs and caregiver interventions, are limited, but they are supported by evidence that shows they can help by improving cognitive function, treating depression, improving caregiver mood, and delaying institutionalisation (Prince et al., 2011). In addition, because of the prevalence of multiple neurodegerative conditions in patients far into the AD process, early detection also facilitates the isolation of AD, which might lead to better diagnostic accuracy by helping us avoid the confounding of neurodegenerative disorders that often occurs in later diagnoses of AD.

The development of machine learning algorithms for improved classification of AD relies on the establishment of useful biomarkers of AD progression. A standardized

understanding of AD biomarkers did not exist until the proposed NIA/AA framework, and so the literature of machine learning for AD has progressed rather blindly without standardized understanding of the progression of biomarkers and how these relate to both the clinical presentation and the neuropathologic process of AD. In the 2018 framework, the NIA/AA define AD as the presentation of two possible biomarker profiles, A+/T+/N- and A+/T+/N-. The implication behind this definition is that biomarkers of neurodegeneration, which include total tau CSF measures (t-tau), FDG PET and MRI, are not specific enough to AD and so their normality or abnormality cannot be used to diagnose a subject as having AD (or as being in the Alzheimer's continuum, a term that includes the biomarker profiles of AD as well as A+ profiles with T-). This is supported by an NIA/AA study from 2016 where they found that the correlation between biomarkers of neurodegeneration (particularly between t-tau and hippocampal volume as assessed by MRI) was minimal (59%) and much smaller than the correlation between A biomarkers (Amyloid PET and CSF AB42) (Vos et al., 2016). A recent study by Ekman et al. also found inconsistency within the N domain of biomarkers in subjects with MCI (Ekman et al., 2018).

Nevertheless, the literature on ML for AD classification has relied heavily on imaging modalities for neurodegeneration, particularly on MRI but also on FDG PET (See Spasov et al., 2019, Nguyen et al., 2019, Sorensen et al., 2017 or Lu et al., 2018 for examples). Even a large proportion of multimodal studies, which incorporate multiple biomarkers in a learning algorithm, have limited their scope to combining biomarkers within the N domain to accurately classify AD.

In order to improve and validate current neurodegeneration biomarkers, as well as newer and less standardized markers such as those involving plasma levels, EEG signals or fMRI connectivity, MRI, FDG-PET and t-tau need to be studied in relation to A and T biomarkers, especially when the desired outcome is accurate AD classification. Neurodegeneration is the pathological element of AD most correlated with clinical symptoms associated with the disease, which is why using MRI and FDG-PET might result in high accuracy rates for predicting clinically diagnosed AD, but they do not reflect AD-specific processes, and biomarkers within this domain show high variability (i.e. are not standardized), so they should not be used in isolation for accurate AD diagnosis. Even if N domain biomarkers are capable of achieving high accuracy rates in differentiating between AD and healthy controls or between prodromal stages of AD and AD, the algorithm is unlikely to be learning AD-specific changes, and its ability to differentiate between conditions is somewhat unsurprising considering that disease progression once a subject has already been diagnosed with Mild or Major NCD due to AD is highly correlated with neurodegeneration rates. Because of this, studies involving only N domain biomarkers are unlikely to be useful in identifying AD-specific patterns long before the appearance of cognitive systems, and this is the primary goal of early detection paradigms. Instead, I propose that future research focus explicitly on developing multimodal machine learning algorithms that combine biomarkers from each of the NIA/AA defined categories. These models are advantageous in that they can provide us with classification methods that make use of large amounts (or all) of available information, can allow us to see the relative importance of different modalities and their features, can provide us with information about the relationship between AD neuropathological elements, can help us evaluate non-

standardized biomarkers with relation to standardized ones, and can contribute to the identification of more robust biomarker profiles that account for variation among the AD population.

Literature Review

In recent literature, multimodal ML methods that combine A/T/N biomarkers have been quite successful in the classification of AD and its prodromal stages. The following review demonstrates their benefits and highlights some of the main issues that need to be considered in the development of such methods. One major complication of multimodal techniques is that they can be more prone to overfitting as a larger amount of weights (for the modalities and the features of each modality) need to be derived. Another difficulty in the development of such models is accounting for missing values and biased datasets, where there is significantly more information about one modality or demographic than there is about others. Furthermore, the correlation between features between modalities, depending on what modalities were chosen, might need to be evaluated and fixed for efficient model building. Different ML methods, as well as different methods for feature extraction, feature selection and model validation are used in the following papers to tackle these issues, to varying degrees of success.

Just this year, Lee et al. assessed the conversion of MCI (similar to Mild Neurodegenerative Disorder) to probable AD with a deep learning recurring neural net variant that incorporated multimodal information from Aβ42, t-tau and p-tau CSF

measures, MRI hippocampal volume and entorhinal cortical thickness, cognitive performance through neuropsychiatric assessment of executive function and memory, and demographic information for 1618 subjects taken from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (307 MCIc, 558 MCInc, 415 controls, 338 AD) (Lee et al., 2019). An interesting and relatively unique aspect of this experiment was that the CSF levels and cognitive performance data were longitudinal, collected at multiple time points, and so the model was both multimodal and longitudinal, having to account for large amounts of (possibly correlated) information. The central question of the study was whether longitudinal and multimodal algorithms perform better when trying to classify MCIc (converted) from MCInc (non-converting). Because training in RNNs with long inputs, as the ones needed for longitudinal analysis, can sometimes be difficult, they used Gated Recurrent Units (GRU) to extract feature vectors for each modality and then concatenated the vectors. Final prediction of the MCI conversion rate for each subject was then performed using L1 regularized logistic regression. They compared the results of their multimodal model to a single modal method based on cognitive performance (which, out of all the markers, gave the best single modal performance) and to a baseline model with all modalities but where longitudinal data were not included. Their findings indicated that multimodal and longitudinal information increased the accuracy of their prediction. Their proposed model was able to predict MCI conversion in 6 months with an accuracy of 0.81, a sensitivity of 0.84 and a specificity of 0.80. Furthermore, they concluded that their GRU-based model was advantageous as compared to other models

(particularly kernel based SVMs) in that it was able to handle irregular longitudinal data and incomplete samples (which they had a lot of) and in that it can be more easily expanded to incorporate other modalities (by simply adding another GRU).

Deep Learning has also been used in other AD classification studies in recent years, but their necessity for the problem has been questioned, as many others have preferred to use simpler, more interpretable methods and have not necessarily had worse results. Of particular interest are the studies that use decision trees to get reliable and interpretable predictions.

In 2013, Gray et al. considered the three-way problem of distinguishing AD from healthy controls, MCI from healthy controls, and MCIc from MCInc using a random forest method and 149 ADNI subjects (37 AD, 75 MCI, and 35 healthy controls) (Gray et al., 2013). In random forests, the final predicted class of a subject is obtained by combining the predictions of a series of individual decision trees. It uses the method of bootstrap aggregating (bagging) as well as random feature selection in the construction of the trees to reduce variance and avoid overfitting. They combine random forests with a manifold learning technique, through which they derive supervised similarity measures for the different modalities. By delivering consistent pairwise similarity measures, their method is relatively interpretable and allows for a combination of multimodal and unbalanced features that can readily be extended to multi-class classification (more easily than SVMs). The modalities used in this experiment included MRI regional volume measures, FDG-PET regional intensities, Aβ42, t-tau and p-tau CSF levels, and APoE genotyping. They

compared the performance of their random forest when the information from each modality was simply concatenated (as in Lee et al. 2019) to when it was embedded using a joint similarity matrix based on the pairwise similarity measures. They found that when the features where embedded, classification rates for all three cases was higher, suggesting that there is correlated and complementary information between modalities that when incorporated into the model can improve the performance of the classifier. For AC-healthy control classification, they achieved an accuracy rate of 89% with 87.9% sensitivity and 90% specificity. For the other two problems, performance was expectedly worse, but still above chance (MCI-Healthy controls: acc=74.6%, sen=77.5%, spec=67.9%; MCIc-MCInc: acc = 58.0, sens = 57.1, spec = 58.7).

Another Random Forest based classification scheme was presented by Dauwan et al. in a study focused on differential classification of AD and Lewy Body disease (Dauwan et al., 2016). Although more oriented towards differential diagnosis than early detection, the study is interesting because it combined an especially wide array of modalities, including clinical symptoms, CSF measures and MRI as well as resting state EEG signals for 198 subjects from the Amsterdam Dementia Cohort (66 subjects for each of the three conditions). To quantify clinical symptoms, they used the standard MMSE (Mini Mental State Examination) for general assessment of cognitive dysfunction, the trail making test part A (TMT-A) for assessing motor speed, the visual association test (VAT) for assessing episodic memory and a Digit span test to assess attention. This collection of neuropsychiatric tests provide a better measure of all the possible cognitive dysfunctions of AD (as seen in our discussion of cognitive domains in section 1) than does an assessment

of MMSE alone, which is more common in the literature. For CSF, they used all the AD-related CSF levels and included additional features that reflected the ratio of tau and Aβ42, in order to account for the idea that the ratio of tau to Aβ42 rather than the magnitudes of each might be what modulates the pathological process of AD (**SOURCE**). MRI measures included assessment of medial temporal lobe atrophy, cortical atrophy and white matter hyperintensities. Another interesting feature of this study is that they derived a Variable Importance Score (VIMP) for each feature so that they could better assess the contributions of each feature and determine performances for various combinations of them. Their analysis of feature importance revealed the EEG was particularly important for the differentiation of LBD with AD and healthy controls, whereas it did not have much of an effect on the classification of AD from controls, for which MMSE was the best discriminant. This latter finding is unsurprising since they did not deal specifically with early AD patients where clinical symptoms are not yet presented fully but pathology exists. Nevertheless, the study found that incorporating the CSF and MRI biomarkers increased the accuracy of their model. Unfortunately, they did not assess the relative values of CSF and MRI explicitly in the paper, but their design could allow for such comparisons. For the discrimination of AD from healthy controls (in which we are most interested), they achieved a high accuracy of 91%, a sensitivity of 92% and a specificity of 91% when combining all available features.

A final, very recent, multimodal decision tree based model classifying AD, MCI and subjects with subjective cognitive decline (SCD, used as control) was proposed by Mofrad et al. (2019). The study was oriented toward validating the interpretation of CSF

biomarkers proposed by the NIA/AA, and so their decision to use a decision tree algorithm was largely based on their desire for and belief in increased intuitive interpretability for diagnostic purposes. Using a decision tree model that also incorporated regressively learning cutoff values (rather than using standardized ones) also allowed them to identify interesting subgroups within the sampled population, which would not have been possible using solely regression models for which the class labels need to be known a priori. Their supervised and nonparametric Classification and Regression Tree (CART) model incorporated data from all three categories of CSF as well as from demographics, ApoE genotyping and MMSE measures of cognitive decline for 1446 subjects diagnosed with the NIA/AA criteria (1004 AD, 442 healthy controls). Their analysis revealed that the best predictive model relied on two cutoff points for CSF Aβ42 (the most discriminative feature) and one cutoff point for CSF t-tau, and was not improved by incorporating age, p-tau levels, sex or ApoE status. The model resulted in an overall accuracy of 90% (sensitivity of 93%, specificity of 88%) and 76% (sensitivity of 84% and specificity of 70%) for distinguishing AD from SCD controls and MCIc from MCInc, respectively. To evaluate the generalizability of their model, they tested its ability to distinguish AD from SCD on an independent cohort, and were able to achieve similar, if somewhat lower, predictive performances.

The combination of cutoff points allowed them to identify two AD subgroups of CSF profiles, one typical profile with low levels of Aβ42 and high levels of tau, and one atypical one with even lower levels of Aβ42 but normal levels of tau, both of which showed similar patterns of MRI atrophy and cognitive decline. Because p-tau and t-tau

were correlated, the atypical subgroup likely corresponds to the two A+/T- biomarkers within the Alzheimer's continuum but outside of the Alzheimer's Disease label (Jack et al., 2018). When combined with the markers of neurodegeneration and cognitive decline, the NIA/AA label is Alzheimer's and concomitant suspected non Alzheimer's pathologic change with dementia. Furthermore, they found that most misclassified subjects (n=50 of 89.56%) within the MCIc-MCInc comparison were subjects with abnormal biomarker profiles who had not clinically progressed to AD, and so represented subjects that might still convert to AD in the future. The identification of these subgroups, made possible by their choice of design, reveals the discrepancy between the clinical diagnosis according to pre-2018 NIA standards and the biomarker profile, but it also reflects the promising ability of machine learning techniques to identify and further analyze these groups through interpretation of their biomarkers.

Despite the simplicity and intuitive interpretability of decision trees, the more sophisticated Support Vector Machines (SVM), seem to be the preferred or most common machine learning technique applied to AD classification,

An influential study from 2011 used a linear SVM to combine multimodal information from 239 ADNI subjects for the prediction of MCI conversion to AD (MCIc-MCInc comparison using 69 MCIc and 170 MCInc). Davatzikos et al. combined CSF measures, clinical information, demographics and ApoE status with an intricate score of brain atrophy, called SPARE AD (Spatial Pattern of Abnormalities for Recognition of Early AD), derived from pattern recognition of MRI scans as proposed by Fan et al. (2008).

When using only the SPARE-AD score in their linear SVM, they were able to correctly classify MCIc subjects 94.7% of the time, but at the cost of classifying MCInc correctly only 37.8% of the time. This reflects the non-specificity of neurodegeneration markers like MRI to AD even when using a high dimensional pattern recognition method like SPARE AD for the identification of relevant features. The classification rate and the specificity of the model increased slightly with the addition of biomarkers, with maximum overall rates reached when using t-tau (classification rate of 61.7%, sensitivity of 84.2%, specificity of 51.2%). However, combining Aβ42, t-tau and Aβ42, p-tau and tau/Aβ42 ratios each individually with SPARE-AD showed similar results. They attributed their low classification rates at least partly to the fact that many misclassified MCInc are likely to convert to MCIc or to another neurodegenerative disorder in the future, and so focused more on the sensitivity and specificity metrics. Overall their results suggested that multimodal methods using MRI and CSF measures provided the most reliable models of MCI to AD conversion. They further concluded that, because SPARE AD scores were associated with faster decline in MMSE scores and because of the high sensitivity of their SPARE AD SVM, brain atrophy measures might be critical in the identification of MCIc. However, this only indicates that cognitive decline is more correlated with brain atrophy than other biomarkers, and cannot help in the identification of MCIc in individual subjects where their cognitive decline and atrophy might be due to another neurodegenerative condition.

Another influential study in 2011 distinguished AD and MCI subjects from healthy controls using an SVM that incorporated multimodal information from 202 ADNI subjects (51 AD, 43 MCIc, 56 MCInc, 52 healthy controls) using a kernel combination method (Zhang et al., 2011). This mixed kernel approach allows the learning of each modality to rely on a kernel that best reflects the distribution of its features rather than assuming the same distribution with a single kernel on a concatenated feature vector, and was selected for effective fusion of multimodal data. They compared their model to a single modal SVM and a multimodal SVM with direct feature concatenation, and found that their model consistently outperformed the alternatives, achieving a high classification rate of 93.2% (sensitivity of 93.0% and specificity of 93.3%) for AD-HC, with an AUC of 0.98 and of 76.4% (sensitivity of 91.8% and specificity of 66.0%) for MCI-HC, with an AUC of 0.81.

Despite the relatively high performance scores of the mixed kernel SVM of Zhang et al., kernel combination methods have a higher risk of overfitting, and so would require additional evaluations of the model's generalizability to ensure its predictive value. Unfortunately, Zhang et al. do not include this in their study, and so their conclusions still require further validation. This year, Varatharajah et al. published a study with a robust design that included an evaluation of their final proposed multimodal model's generalizability as well as comparisons of different standard machine learning approaches (Varatharajah et al., 2019). They combined demographic information with genetic factors (ApoE genotyping as well as genotyping for nine other AD related genes), MMSE measures of cognitive decline, cognitive resilience measures, CSF biomarkers (all), MRI measures, FDG-PET and F-florbetapir (amyloid) PET for 135 ADNI MCI subjects (39

MCIc, 96 MCInc). They derived a cognitive resilience score by combining the subjects's years of education with their responses to the American National Adult Reading Test (ANART) which gives an estimate of pre-morbid verbal IQ. The MRI features included volume and thickness measures of AD-related ROIs as well as hippocampal volume. Overall, they considered a wide range of biomarkers and multiple machine learning methods, which makes this study particularly interesting. Their model choice was a linear kernel SVC, but they started off comparing it to an RBF (nonlinear) kernel SVC in order to assess whether the problem of AD classification is linearly separable, and whether more complicated nonlinear methods need to be used to tackle it. Then, they compared the goodness of fit of the linear SVC with a multi kernel learning algorithm and a General Linear Model (GLM) using elastic net regularization (to avoid overfitting). For feature extraction, they used an information based approach (rather than a statistical correlation-based approach) that relies on a join mutual information (JMI) metric for random variables. This approach starts with an empty feature set and adds attributes sequentially by choosing the attribute that maximized the JMI between the attributes and the outcome (Varatharajah et al., 2019, p. 5).

The RBF kernel and linear kernel SVC had similar ROC curves and classification accuracy rates, suggesting that the classification of MCIc and MCIc is a linearly separable problem that does not necessitate nonlinear models. To evaluate the generalizability of the different models compared, they calculated a cross-validated AUC score for increasing amounts of features incorporated into each of the models, and compared their change over time for the training and testing data. This telling and applicable method revealed that even though all

methods showed increased AUC as features were incorporated in the training dataset, in the testing dataset the mixed kernel method's AUC decreased significantly once enough features were added. This suggests that mixed kernel methods are more likely to overfit, and are therefore less generalizable than the fairly consistent GLM and SVC methods. Their performance evaluation showed that the linear SVC outperformed both the MLK and the GLM, achieving a maximum AUC of 93%, sensitivity of 93% and specificity of 77% with 65 features. The GLM and mixed kernel classifier performed similarly (but not as well) when using an optimum number of 25 and 5 features, respectively. Finally, they looked at the role of individual modalities in the linear SVC and found that CSF measures were the highest predictor of AD with a cross validated AUC of about 0.9, which is in line with the NIA/AA's interpretation of CSF biomarkers. Interestingly, they also found that when removing CSF measures from the model, the AUC of the linear SVC did not reduce by much. The implication of this is that modalities are likely to be correlated, and so practices such as the JMI paradigm used above are essential to deriving efficient multimodal classifiers.

Proposed Research Framework for the Development of Early Diagnosis Paradigms

These studies demonstrate the validity and efficacy of multimodal approaches for AD classification in the development of early detection paradigms. A summary of the key studies considered is presented in Table 3. Relatively high performance scores are achieved throughout, and some studies specifically report better performances for multimodal approaches when compared to single modal approaches (Zhang et al., (2011),

for instance). These models allow for the inclusion of most available biomarker information, which is necessary in order to be certain that the methods are learning meaningful AD-specific changes and that they are accounting for all elements that characterize the pathology. In line with the proposed NIA/AA biomarker framework, at least A and T domain biomarkers need to be included for a valid assessment of AD, and additional domains should be included to explore their relationship to amyloid and tau. As the analysis of AD pathology revealed, many questions about the AD process are still somewhat unresolved, particularly concerning the causal relationship between markers of the disease. Machine learning based studies of AD classification, therefore, should avoid assumptions of a unified understanding and should instead be oriented towards providing highly interpretable models that can provide insights into the relationship between biomarkers. I suggest that a key way for multimodal studies to increase their interpretability and ability to elucidate more about the AD process is by using the A/T/(N) biomarker profiles to evaluate the contributions of individual subjects and to relate the pathological elements to the clinical presentation of AD. Although this approach was not adopted by any of the studies considered in my literature review, likely because it was so recently proposed, an analysis of their methodologies and results still provide us with useful guidelines for the design of future multimodal machine learning AD classification studies guided by the A/T/(N) framework.

When it comes to choosing an appropriate machine learning algorithm, I propose that future studies apply Decision Tree-based algorithms or linear SVCs and avoid neural nets, nonlinear methods or mixed kernel methods. The Decision Tree based algorithms proposed

Table 3: Summary of key multimodal AD classification studies that integrate A, T and N biomarkers

KEY STUDIES	ML method	Modalities included	Feature importance estimation
Lee et al. (2019)	RNN	CSF (all), MRI	Single modal feature extraction and concatenation-based integration
Gray et al. (2013)	Random Forest	CSF (all), MRI, FDG-PET, ApoE	Embedding of features using joint similarity measures of decision tree agreement
Dauwan et al. (2016)	Random Forest	CSF (all), MRI, EEG	Each feature has a Variable Importance Score (VIMP)
Mofrad et al. (2019)	Classification and Regression Tree (CART)	CSF (all), ApoE genotyping	Regression tree allowed for learning (rather than setting) of cut-off values that established the relative importance of the markers
Zhang et al. (2011)	Mixed kernel SVM	CSF (all), MRI, FDG-PET, ApoE	Atlas warping algorithm extracts ROIs from MRI
Davatzikos et al. (2011)	SVM	CSF (all), MRI, ApoE	Complex SPARE AD pattern recognition algorithm extracts MRI features
Varatharajah et al. (2019)	Linear SVC (+ RBF kernel SVC, mixed kernel SVC, GLM)	CSF (all), MRI, FDG-PET, Amyloid PET, ApoE (+ 9 more genes)	Information based feature extraction method using joint mutual information scores and different amounts of extracted features.

by Gray et al. (2013), Dauwan et al. (2016) and Mofrad et al. (2019) all achieve relatively high performance scores (with a maximum accuracy rate of 92% achieved by Dauwan et al. in differentiating AD subjects from healthy controls), generally avoid overfitting, and have the advantage of being highly interpretable. Mofrad et al.'s study, for example, was able to provide valuable insights into the relationship between tau and amyloid CSF biomarkers that is in line with an understanding of AD as involving a two way interaction between the key pathological elements (rather than the unidirectional understanding suggested by the amyloid cascade hypothesis). Varatharajah et al. demonstrate that linear SVCs can achieve even higher performances (AUC of 93%, sensitivity of 93% and

specificity of 77%) without necessarily reducing interpretability. The linear SVC outperformed the RBF kernel SVC, suggesting that linear approaches are sufficient for AD classification paradigms. The linear SVC also outperformed the mixed kernel approach. Although in Zhang et al. (2011) the mixed kernel approach achieved good performances, because the results were not corroborated in the more robust study by Varatharajah et al. (2019), and because mixed kernel methods have the additional disadvantage of not being able to easily incorporate additional modalities, I suggest that they should not be preferred over linear SVCs. Neural nets, as the one proposed by Lee et al. (2019), are robust to irregular data, but are not very interpretable and might not exhibit as high performances as other methods when considering data with large amounts of features, as is the case with multimodal studies.

Because of the large amounts of data, multimodal studies might be particularly prone to overfitting. As a result, I propose that all multimodal classification studies include some method for evaluating the generalizability of their model. Generalizability evaluations are particularly important when a large percentage of studies use data from the same database, and even more so when this database is known to be non-representative of the general population. Only 5% of the subjects included in the ADNI database, used in 5 of the 7 key studies in this review, identify as African-American, a percentage that is too low for a reliable examination of whether African-Americans differ from Caucasians, and one that is likely to be even lower when inclusion criteria involves information from multiple modalities, since social and economic factors heavily influence access to these modalities, and African-Americans are one of the groups most marginalized by the effects of these

factors. Studies continue to use these biased datasets and assume their generalizability despite the higher prevalence of AD for African-Americans, and findings of significantly lower CSF tau levels that might suggest heterogeneity of the AD pathological process not accounted for in the ADNI dataset (Shin et al., 2016; Howell et al., 2017).

As an urgent starting point, the Alzheimer's scientific field must invest in exclusively African-American cohorts, as well as cohorts specific to other marginalized groups, in order to counter data collection biases and include more heterogeneity of the disease for increased understanding of the AD process, especially if we are to embrace a biomarker based framework. Another necessary step is to always include generalizability assessments that demonstrate the model's overfitting avoidance. One acceptable method for doing this is to simply test the model on an independent cohort, as is done by Mofrad et al. (2019). An interesting approach might be to train a model on an ADNI dataset and then evaluate its performance on a cohort made up of subjects not well represented in the ADNI, such as African Americans. Another interesting method is proposed by Varatharajah et al. (2019) and involves sequentially increasing the amount of considered features and comparing AUC scores of the model separately on the testing and training datasets.

Lastly, I propose that increased interpretability will be achieved by future multimodal classification studies that use an appropriate method for estimating feature importance. Feature importance evaluations allow for informed multimodal integration that not only helps us visualize the relationship among biomarkers, but also lead to more robust models that account for correlation between variables. Dauwan et al. (2016) decide on a simple

approach that gives a variable importance score (VIMP) to each feature and allows for manually incorporating or deleting selected features in the model. For more intricate models, I suggest using information-based approaches, as suggested by Gray et al. (2013) or Varatharajah et al. (2019). Both successfully use joint similarity estimates for feature extraction, thus reducing the amount of correlated data incorporate into the model, and demonstrate that an information-based approach is more successful than simple concatenation of features or mixed kernel methods, the latter of which will likely also reduce the interpretability of the model.

In summary, I propose that future multimodal studies using machine learning for the development of early diagnostic paradigms (a) classify their subjects according to the A/T/(N) biomarker profiles, (b) rely on linear SVCs or decision tree based algorithms, (c) make efforts to incorporate a diverse population to account for the heterogeneity of AD, (d) include an assessment of generalizability, (e) assess and visualize relative feature importance and (f) make use of information-based methods for multimodal integration.

Conclusion

The discrepancy between the clinically-based characterization of AD and the pathological process of AD requires a movement toward a more biologically based definition that makes use of advances in the standardization and accessibility of biomarkers and is practical for both diagnostic and research purposes. An A/T/(N) biomarker profile as proposed by the NIA/AA, combined with clinical information, serves as a useful bridge between our clinical and pathological standards, and should be considered in the design of experiments that require a valid assessment of AD. According to this framework, biomarkers of both amyloidosis and tauopathy must be abnormal for a subject to be classified as AD. Their requirement of multimodal biomarker information for the accurate diagnosis of AD is largely supported by pathological evidence of an intricate relationship between tauopathy and amyloidosis in which they mediate each other's function, and a secondary, but highly correlated relationship between clinical status and neurodegeneration.

An important limit to this approach is its reliance on hard to access modalities for accurate assessment of AD. CSF measures are expensive and require lumbar punctures, and PET or MRI imaging modalities are, in many settings, difficult to access. However, this framework will significantly increase the validity of diagnoses and is absolutely crucial to developing research based interventions for early detected AD subjects. As we embrace this approach and begin to rely more on biomarker data, therefore, we must put significant effort into not further marginalizing already underrepresented communities. This is critical

considering that the incidence of dementia is is growing at a faster rate in ostensively developing countries. The 10/66 Dementia Research Group is an organization that is helping spearhead changes in the scientific community by conducting population-based research in Latin America, the Caribbean, India, Russia, China and South East Asia (Patterson, 2018). They are so named because when they begun 10% of global research was being conducted in low/middle income countries even though they contain 66% of people with dementia (Patterson, 2018, p. 14).

If we continue to collect diverse biomarker data, our framework will become more robust, and we will be able to develop better performing and more valid machine learning models. Multimodal AD classification studies of AD that combined A, T and N biomarkers have made use of machine learning algorithms like decision trees, support vector machines and neural networks, along with a variety of techniques for feature extraction and evaluation, overfitting avoidance and generalizability estimation to successfully advance our ability for early detection of AD.

Literature Cited

Ahmed, Z., Cooper, J., Murray, T. K., Garn, K., McNaughton, E., Clarke, H., ... & Bose, S. (2014). A novel in vivo model of tau propagation with rapid and progressive neurofibrillary tangle pathology: the pattern of spread is determined by connectivity, not proximity. Acta neuropathologica, 127(5), 667-683.

Alladi, S., Xuereb, J., Bak, T., Nestor, P., Knibb, J., Patterson, K., & Hodges, J. R. (2007). Focal cortical presentations of Alzheimer'sdisease. Brain, 130(10), 2636-2645

American Psychiatric Association, Committee on Nomenclature and Statistics, American Psychiatric Association, & Committee onNomenclature and Statistics. (1952). Diagnostic and statistical manual of mental disorders. American Psychiatric Assn., Mental Hospital Service.

American Psychiatric Association. (1968). DSM-II. Diagnostic and statistical manual of mental disorders, 2nd edition.

American Psychiatric Association. (1980). DSM-III. Diagnostic and statistical manual of mental disorders, 3rd edition.

American Psychiatric Association. (1987). DSM-III-R. Diagnostic and statistical manual of mental disorders, Revised 3rd edition.

American Psychiatric Association. (1994). DSM-IV. Diagnostic and statistical manual of mental disorders, 4th edition.

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (DSM-5®). American Psychiatric Pub.

Barrett, L. F. (2017). The theory of constructed emotion: an active inference account of interoception and categorization. Social cognitive and affective neuroscience, 12(1), 1-23.

Bennett, D. A., Schneider, J. A., Arvanitakis, Z., Kelly, J. F., Aggarwal, N. T., Shah, R. C., & Wilson, R. S. (2006). Neuropathology of older persons without cognitive impairment from two community-based studies. Neurology, 66(12), 1837-1844.

Braak, H., & Del Tredici, K. (2011). The pathological process underlying Alzheimer's disease in individuals under thirty. Acta neuropathologica, 121(2), 171-181.

Braak, H., Alafuzoff, I., Arzberger, T., Kretzschmar, H., & Del Tredici, K. (2006). Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. Acta neuropathologica, 112(4), 389-404.

Burke, B. T., Latimer, C. S., Keene, D. C., Montine, T. J., Sonnen, J., McCormick, W., ... & Crane, P. K. (2018). AUTOPSY FINDINGS OF CEREBRAL β-AMYLOIDOSIS, TAUOPATHY, AND NEURODEGENERATION (A, T AND N) IN A COMMUNITY-BASED SAMPLE: IMPLICATIONS FOR DEMENTIA MITIGATION STRATEGIES. Alzheimer's & Dementia: The Journal of the Alzheimer's Association, 14(7), P1593-P1595.

Cho, H., Choi, J. Y., Hwang, M. S., Kim, Y. J., Lee, H. M., Lee, H. S., ... & Lyoo, C. H. (2016). In vivo cortical spreading pattern of tau and amyloid in the Alzheimer disease spectrum. Annals of neurology, 80(2), 247-258.

Dauwan, M., van der Zande, J. J., van Dellen, E., Sommer, I. E., Scheltens, P., Lemstra, A. W., & Stam, C. J. (2016). Random forest to differentiate dementia with Lewy

bodies from Alzheimer's disease. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 4, 99-106.

Davatzikos, C., Bhatt, P., Shaw, L. M., Batmanghelich, K. N., & Trojanowski, J. Q. (2011). Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. Neurobiology of aging, 32(12), 2322-e19.

Dening, T., & Thomas, A. (Eds.). (2013). Oxford textbook of old age psychiatry. Oxford University Press.

Dolci, G. A., Damanti, S., Scortichini, V., Galli, A., Rossi, P. D., Abbate, C., ... & Scarpini, E. (2017). Alzheimer's Disease Diagnosis: Discrepancy between Clinical, Neuroimaging, and Cerebrospinal Fluid Biomarkers Criteria in an Italian Cohort of Geriatric Outpatients: A Retrospective Cross-sectional Study. Frontiers in Medicine, 4, 203.

Drachman, D. A. (2014). The amyloid hypothesis, time to move on: Amyloid is the downstream result, not cause, of Alzheimer's disease. Alzheimer's & Dementia, 10(3), 372-380.

Ekman, U., Ferreira, D., & Westman, E. (2018). The A/T/N biomarker scheme and patterns of brain atrophy assessed in mild cognitive impairment. Scientific reports, 8(1), 8431.

George, D. R., Whitehouse, P. J., & Ballenger, J. (2011). The evolving classification of dementia: placing the DSM-V in a meaningful historical and cultural context and pondering the future of "Alzheimer's". Culture, Medicine, and Psychiatry, 35(3), 417-435.

Goate, A., Chartier-Harlin, M. C., Mullan, M., Brown, J., Crawford, F., Fidani, L., ... & Mant, R. (1991). Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. Nature, 349(6311), 704.

Goedert, M. (2004, February). Tau protein and neurodegeneration. In Seminars in cell & developmental biology (Vol. 15, No. 1, pp. 45-49). Academic Press.

Gray, K. R., Aljabar, P., Heckemann, R. A., Hammers, A., Rueckert, D., & Alzheimer's Disease Neuroimaging Initiative. (2013). Random forestbased similarity measures for multi-modal classification of Alzheimer's disease. NeuroImage, 65, 167-175.

Hardy, J. A., & Higgins, G. A. (1992). Alzheimer's disease: the amyloid cascade hypothesis. Science, 256(5054), 184-186.

Howell, J. C., Watts, K. D., Parker, M. W., Wu, J., Kollhoff, A., Wingo, T. S., ... & Hu, W. T. (2017). Race modifies the relationship between cognition and Alzheimer's disease cerebrospinal fluid biomarkers. Alzheimer's research & therapy, 9(1), 88.

Huber, C. M., Yee, C., May, T., Dhanala, A., & Mitchell, C. S. (2018). Cognitive decline in preclinical Alzheimer's disease: amyloid-beta versus tauopathy. Journal of Alzheimer's Disease, 61(1), 265-281.

Hyman, B. T., Phelps, C. H., Beach, T. G., Bigio, E. H., Cairns, N. J., Carrillo, M. C., ... & Mirra, S. S. (2012). National Institute on Aging– Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease. Alzheimer's & dementia, 8(1), 1-13.

Insel, T., & National Institute of Mental Health. (2013, April 29). Post by Former NIMH Director Thomas Insel: Transforming Diagnosis. Retrieved from

https://www.nimh.nih.gov/about/directors/thomas-insel/blog/2013/transforming-diagnosis.shtml

Jack, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., ... & Liu, E. (2018). NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. Alzheimer's & Dementia, 14(4), 535-562.

Jacobs, H. I., Hedden, T., Schultz, A. P., Sepulcre, J., Perea, R. D., Amariglio, R. E., ... & Johnson, K. A. (2018). Structural tract alterations predict downstream tau accumulation in amyloid-positive older individuals. Nature neuroscience, 21(3), 424.

Jin, M., Shepardson, N., Yang, T., Chen, G., Walsh, D., & Selkoe, D. J. (2011). Soluble amyloid β-protein dimers isolated from Alzheimer cortex directly induce Tau hyperphosphorylation and neuritic degeneration. Proceedings of the National Academy of Sciences, 108(14), 5819-5824.

Jones, D. T., Knopman, D. S., Gunter, J. L., Graff-Radford, J., Vemuri, P., Boeve, B. F., ... & Jack Jr, C. R. (2015). Cascading network failure across the Alzheimer's disease spectrum. Brain, 139(2), 547-562.

Lee, G., Nho, K., Kang, B., Sohn, K. A., & Kim, D. (2019). Predicting Alzheimer's disease progression using multi-modal deep learning approach. Scientific reports, 9(1), 1952.

Li, X., Uemura, K., Hashimoto, T., Nasser-Ghodsi, N., Arimon, M., Lill, C. M., ... & Berezovska, O. (2013). Neuronal activity and secreted amyloid β lead to altered amyloid β precursor protein and presenilin 1 interactions. Neurobiology of disease, 50, 127-134.

Luo, L. (2015). Principles of neurobiology. Garland Science.

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack Jr, C. R., Kawas, C. H., ... & Mohs, R. C. (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's & dementia, 7(3), 263-269.

Mofrad, R. B., Schoonenboom, N. S., Tijms, B. M., Scheltens, P., Visser, P. J., van der Flier, W. M., & Teunissen, C. E. (2019). Decision tree supports the interpretation of CSF biomarkers in Alzheimer's disease. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 11, 1-9.

National Institute of Mental Health. Definitions of the RDoC Domains and Constructs. (2013). Retrieved from https://www.nimh.nih.gov/research/research-funded-by-nimh/rdoc/definitions-of-the-rdoc-domains-and-constructs.shtml#part_154187

Nelson, P. T., Head, E., Schmitt, F. A., Davis, P. R., Neltner, J. H., Jicha, G. A., ... & Scheff, S. W. (2011). Alzheimer's disease is not "brain aging": neuropathological, genetic, and epidemiological human studies. Acta neuropathologica, 121(5), 571-587.

Nicoll, J. A., Barton, E., Boche, D., Neal, J. W., Ferrer, I., Thompson, P., ... & Seubert, P. (2006). Aβ species removal after Aβ42 immunization. Journal of Neuropathology & Experimental Neurology, 65(11), 1040-1048.

Palmqvist, S., Schöll, M., Strandberg, O., Mattsson, N., Stomrud, E., Zetterberg, H., ... & Hansson, O. (2017). Earliest accumulation of β- amyloid occurs within the default-mode network and concurrently affects brain connectivity. Nature communications, 8(1), 1214.

Patterson, C. (2018). World Alzheimer Report 2018—The state of the art of dementia research: New frontiers. Alzheimer's Disease International (ADI): London, UK.

Prince, M., Bryce, R., & Ferri, C. (2018). World Alzheimer Report 2011: The benefits of early diagnosis and intervention.

Rabinovici, G. D., Jagust, W. J., Furst, A. J., Ogar, J. M., Racine, C. A., Mormino, E. C., ... & Gorno-Tempini, M. L. (2008). Aβ amyloid and glucose metabolism in three variants of primary progressive aphasia. Annals of neurology, 64(4), 388-401.

Roberson, E. D., Halabisky, B., Yoo, J. W., Yao, J., Chin, J., Yan, F., ... & Palop, J. J. (2011). Amyloid-β/Fyn–induced synaptic, network, and cognitive impairments depend on tau levels in multiple mouse models of Alzheimer's disease. Journal of Neuroscience, 31(2), 700-711.

Schneider, J. A., Arvanitakis, Z., Leurgans, S. E., & Bennett, D. A. (2009). The neuropathology of probable Alzheimer disease and mild cognitive impairment. Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society, 66(2),
200-208.

Selkoe, D. J., & Hardy, J. (2016). The amyloid hypothesis of Alzheimer's disease at 25 years. EMBO molecular medicine, 8(6), 595-608.

Serrano-Pozo, A., Frosch, M. P., Masliah, E., & Hyman, B. T. (2011). Neuropathological alterations in Alzheimer disease. Cold Spring Harbor perspectives in medicine, 1(1), a006189.

Shin, J., & Doraiswamy, P. M. (2016). Underrepresentation of African-Americans in Alzheimer's Trials: A Call for Affirmative Action. Frontiers in Aging Neuroscience, 8, 123.

Snowdon, D. A., Greiner, L. H., Mortimer, J. A., Riley, K. P., Greiner, P. A., & Markesbery, W. R. (1997). Brain infarction and the clinical expression of Alzheimer disease: the Nun Study. Jama, 277(10), 813-817.

Spillantini, M. G., & Goedert, M. (2013). Tau pathology and neurodegeneration. The Lancet Neurology, 12(6), 609-622.

Thal, D. R., Rüb, U., Orantes, M., & Braak, H. (2002). Phases of Aβ-deposition in the human brain and its relevance for the development of AD. Neurology, 58(12), 1791-1800.

Varatharajah, Y., Ramanan, V. K., Iyer, R., & Vemuri, P. (2019). Predicting Short-term MCI-to-AD Progression Using Imaging, CSF, Genetic Factors, Cognitive Resilience, and Demographics. Scientific reports, 9(1), 2235.

Vos, S. J., Gordon, B. A., Su, Y., Visser, P. J., Holtzman, D. M., Morris, J. C., ... & Benzinger, T. L. (2016). NIA-AA staging of preclinical Alzheimer disease: discordance and concordance of CSF and imaging biomarkers. Neurobiology of aging, 44, 1-8.

Wiseman, F. K., Pulford, L. J., Barkus, C., Liao, F., Portelius, E., Webb, R., ... & Collins, T. (2018). Trisomy of human chromosome 21 enhances amyloid-β deposition independently of an extra copy of APP. Brain, 141(8), 2457-2474.

Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D., & Alzheimer's Disease Neuroimaging Initiative. (2011). Multimodal classification of Alzheimer's disease and mild cognitive impairment. Neuroimage, 55(3), 856-867.

Acknowledgements

I am extremely grateful for the guidance, support and helpful insights of my primary advisor, Michael Spezio, in researching and writing my senior thesis. A big thanks as well to Claremont Mckenna College and the Keck Science Department, for furthering my love for science and encouraging this kind of work. Thanks to Sebastian Naze, my secondary advisor, for supervising my work and giving me such helpful feedback. I could not have done any of this without the endless support of my family and friends.

Appendix: Thesis Poster

The Characterization of Alzheimer's Disease and the Development of Early Detection Paradigms: Insights from Nosology, Biomarkers and Machine Learning

Isabel Rose Milano

Claremont McKenna College (April 2019)

Abstract

Alzheimer's Disease is the only condition in the top ten leading causes of death for which we don't have an effective treatment that prevents, slows, or stops the progression of the disease. Our ability to design useful interventions relies on (a) increasing our understanding of the pathological process of AD and (b) improving our ability for its early detection. These goals are impeded by our current reliance on the clinical symptoms of AD for its diagnosis. Instead, a new biomarker-based approach provides a more biological understanding of the disease that does not hinge on postmortem verification and that can detect pathological changes before the clinical symptoms emerge. This new biomarker based characterization of AD can then be combined with machine learning techniques in multimodal classification studies to shed light on the elements of the AD pathological process and develop early detection paradigms.

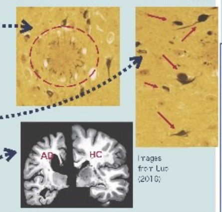
Introduction

Alzheimer's Disease (AD) is a condition affecting 50 million of people worldwide and one of the top ten global causes of death (Petterson, 2018). The number of diagnosed patients is expected to continue to grow exponentially as life expectancies increase and our ability to identify and diagnose the disorder improves. Stigma-reducing awareness campaigns, as well as improved training for caregivers and better diagnostic standards, have all contributed to more accurate diagnoses worldwide. Unfortunately, AD is the only leading cause of death for which we still have no effective treatment that prevents, slows or stops the progression of the disease. Because of this, the main goal of the scientific field surrounding AD is the design of useful interventions, a task that depends heavily on (a) increasing our understanding of the pathophysiological process of AD and (b) improving our ability for its early detection. Our understanding of AD has been somewhat nebulous from the first time the term was used to describe a condition. Although much research since then has focused on specifying the common ties between AD patients, many questions remain unanswered. This contributes to the stigma associated with the condition, and limits our ability for accurate diagnosis. Currently, standards for diagnosis rely on the clinical presentation of AD symptoms, even though these occur at least 10-20 years after the pathological process has begun and therefore do not allow for methods of early intervention. Furthermore, a recent technological boom has paired our unclear characterization of AD with large amounts of data and data mining techniques that seem to advance without much concern for what we know about the neurobiological process. At the same time, recent years have seen significant progress in the standardization and accessibility of AD biomarkers. The hope is that biomarkers, in combination with advanced computational methods, will give us valuable insights into the otherwise undetectable pre-clinical stages of AD and will allow us to develop methods for early detection and disease progression staging.

Diagnosing AD	The Neuropathological Process of AD	Multimodal ML techniques for early detection of AD
• History and challenges of current clinical standards for AD Diagnosis • Progressing toward a biomarker-based approach	• What characterizes the AD neuropathological process? • Current standards for the pathological assessment of AD • Reassessing the Amyloid Cascade Hypothesis • Implications for biomarker-based approaches	• Value and challenges of Multimodal machine learning classification for AD • Review of recent literature • Proposed research framework for future multimodal studies

Toward a biomarker-based understanding of Alzheimer's Disease

A	Biomarkers of Amyloidosis Measures of the aggregation of Aβ	CSF $A\beta_{42}$ CSF $A\beta_{42}/A\beta_{40}$ ratio Amyloid PET
T	Biomarkers of Tauopathy Measures of the aggregation of phosphorylated tau (in neurofibrillary tangles)	CSF phosphorylated tau (P-tau) Tau PET
(N)	Biomarkers of Neurodegeneration Measures of cerebral atrophy or neuronal injury	MRI (cortical thickness, regional atrophy, hippocampal volume, etc). FDG PET



Adapted from Jack et al. (2018)

Biomarker Profile	A	T	(N)	Diagnostic Category
A- T- (N)-	-	-	-	Normal AD biomarkers
A+ T- (N)-	+	-	-	Alzheimer's Pathologic Change
A+ T+ (N)-	+	+	-	Alzheimer's Disease
A+ T+ (N)+	+	+	+	Alzheimer's Disease
A+ T- (N)+	+	-	+	Alzheimer's and concomitant suspected non Alzheimer's pathologic change
A- T+ (N)-	-	+	-	Non-AD pathologic change
A- T- (N)+	-	+	+	Non-AD pathologic change
A- T+ (N)+				Non-AD pathologic change

☐ Alzheimer's Continuum ■ Alzheimer's Disease

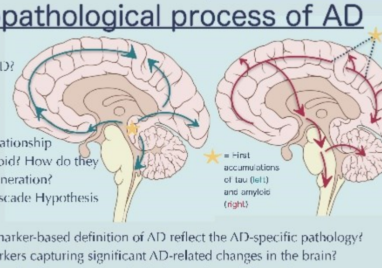
Biomarker Profile	Clinical Diagnosis: Cognitively Unimpaired	Minor Neurocognitive Disorder	Major Neurocognitive Disorder
A+ T+ (N)- A+ T+ (N)+	Preclinical AD	Prodromal AD	Major ND due to AD

Increased risk of rapid short term clinical progression →

- **30-40%** of cog. unimpaired elders show evidence of AD pathological process at autopsy (Jack et al. 2018)
- **10-30%** of AD diagnosed show no evidence of the AD pathological process at autopsy (Jack et al. 2018).
- **45%** of elders with a Probable AD diagnosis exhibit mixed pathology at autopsy (Schneider et al. 2009)
- Claims-based beneficiary classification for AD can reduce the prevalence of its diagnosis by up to **50%** (Newcomer et al. 1999).
- **22–64%** of patients with Early-Onset AD exhibit non-amnestic presentations of AD (Mendez 2012)

The neuropathological process of AD

- What pathological profile characterizes AD?
- Where does the pathology begin? how does it progress?
- What is the causal relationship between tau and amyloid? How do they bring about neurodegeneration?
- Does the Amyloid Cascade Hypothesis still stand?



- Does our revised biomarker-based definition of AD reflect the AD-specific pathology?
- Are A, T and N biomarkers capturing significant AD-related changes in the brain?
- How can we use A, T and N biomarkers to shed light on the pathology of AD?

Multimodal integration of biomarkers for early detection using Machine Learning

Challenges in the design of multimodal AD classification studies:

- Integrating biomarker information from multiple modalities
- Overfitting, as there is an increase in the amount of weight parameters needed
- Accounting for the correlation between data from different modalities
- Handling imbalances in the data due to (a) missing biomarker data or (b) underrepresentation of racial minorities in available data

To accurately assess whether a subject has AD or to find where they fall on the Alzheimer's continuum, multimodal classification must include A, T and N biomarkers

Advantages of multimodal AD classification studies:

- Can shed light on the relationship between the neuropathological elements of AD
- Can make use of most (or all) available information
- Can contribute to the evaluation of non-standardized biomarkers by comparing them to standardized ones
- Can identify more robust biomarker profiles to account for the heterogeneity of AD.

KEY STUDIES	ML method	Modalities included	Feature importance estimation
Lee et al. (2019)	RNN	CSF (all), MRI	Single model feature extraction and concatenation-based integration
Gray et al. (2013)	Random Forest	CSF (all), MRI, FDG-PET, ApoE	Embedding of features using joint similarity measures of decision tree agreement
Dauwan et al. (2016)	Random Forest	CSF (all), MRI, EEG	Each feature has a Variable Importance Score (VIMP)
Mofrad et al. (2019)	Classification and Regression Tree (CART)	CSF (all), ApoE genotyping	Regression tree allowed for learning (rather than setting) of cut-off values that established the relative importance of the markers
Zhang et al. (2011)	Mixed kernel SVM	CSF (all), MRI, FDG-PET, ApoE	Atlas warping algorithm extracts ROIs from MRI
Davatzikos et al. (2011)	SVM	CSF (all), MRI, ApoE	Complex SPARE AD pattern recognition algorithm extracts MRI features
Varatharajah et al. (2019)	Linear SVC (+ RBF kernel SVC, mixed kernel SVC, GLM)	CSF (all), MRI, FDG-PET, Amyloid PET, ApoE (+ 9 more genes)	Information based feature extraction method using joint mutual information scores and different amounts of extracted features.

Conclusion

The discrepancy between the clinically-based characterization of AD and the pathological process of AD requires a movement toward a more biologically-based definition practical for both diagnostic and research purposes. An A/T/(N) biomarker profile, proposed by the NIA/AA, combined with clinical information, serves as a useful bridge between our clinical and pathological standards, and should be considered in the design of experiments that require a valid assessment of AD. Their requirement of multimodal biomarker information for the accurate diagnosis of AD is largely supported by pathological evidence of an intricate relationship between tauopathy and amyloidosis, and a secondary, but highly correlated relationship between clinical status and neurodegeneration. Multimodal AD classification studies of AD that combined A, T and N biomarkers have made use of machine learning algorithms like decision trees, support vector machines and neural networks, along with a variety of techniques for feature extraction, overfitting avoidance and generalizability estimation to advance our ability for early detection of AD.

Acknowledgements. I am extremely grateful for the help and support of my primary advisor, Michael Spezio in researching and writing my senior thesis. A big thanks as well to Claremont McKenna College and the Keck Science Department for furthering my love for science and encouraging this kind of work. Thanks to Sebastian Naze, my secondary advisor, for agreeing to supervise and offering their support. I could not have done any of this without the endless support of my family and friends.

References

1. Dauwan, M., van der Zande, J. J., van Dellen, E., Sommer, I. E., Scheltens, P., Lemstra, A. W., & Stam, C. J. (2016). Random forest to differentiate dementia with Lewy bodies from Alzheimer's disease. Alzheimer's & Dementia, 4, 99-106.
2. Gray, K. R., Aljabar, P., Heckemann, R. A., Hammers, A., Rueckert, D. (2013). Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. NeuroImage, 65, 167-175.
3. Jack Jr, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., ... & Liu, E. (2018). NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. Alzheimer's & Dementia, 14(4), 535-562.
4. Lee, Garam, et al. "Predicting Alzheimer's disease progression using multi-modal deep learning approach." Scientific reports 9.1 (2019): 1952.
5. Luo, L. (2016). Principles of neurobiology. Garland Science.
6. Mendez, M. F. (2012). Early-onset Alzheimer's disease: nonamnestic subtypes and type 2 AD. Archives of medical research, 43(8), 677-685.
7. Mofrad, R. B., et al. (2019). [illegible] Alzheimer's disease. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, [illegible]
8. Newcomer, R., Clay, T., Luxenberg, J. S., & Miller, R. H. (1999). Misclassification and selection bias when identifying Alzheimer's disease solely from Medicare claims records. Journal of the American Geriatrics Society, 47(2), 215-219.
9. Petterson, C. World Alzheimer Report 2018 (Rep.). London: Alzheimer's Disease International.
10. Schneider, J. A., Arvanitakis, Z., Leurgans, S. E., & Bennett, D. A. (2009). The neuropathology of probable Alzheimer disease and mild cognitive impairment. Annals of Neurology, 66(2), 200-208.
11. Varatharajah, Y., Ramanan, V. K., Iyer, R., & Vemuri, P. (2019). Predicting Short-term MCI-to-AD Progression Using Imaging, CSF, Genetic Factors, Cognitive Resilience, and Demographics. Scientific reports, 9(1), 2235.