Philosophy of Medicine
Original Research

Reconfiguring Boundaries Between Healthy Aging and Future Dementia: Ground-Truthing Comparisons in Neuroimaging Research on Predictive ML-Based Diagnosis of Alzheimer's Disease

Paula Muhr[1]

1 Department of Design, Brand University of Applied Sciences, Hamburg, Germany. Email: paula.muhr@brand-university.de

Abstract

This paper addresses emerging research on machine learning (ML) algorithms for predictive diagnosis of Alzheimer's disease (AD) in individuals without clear-cut clinical symptoms using structural neuroimaging data. By analyzing published studies, it examines how predictive diagnostic separability between healthy subjects and future AD patients is constructed at the level of neuroimaging data during algorithm training and evaluation. It argues that researchers' choices about how to organize comparisons of referential neuroimaging datasets—called ground-truths—during algorithm training and evaluation reconfigure boundaries between healthy aging and future dementia in this research in conceptually inconsistent ways. Due to the current focus on algorithmic performance metrics in the research context, the implications of these choices for the evolving production of biomedical knowledge about AD, which are discussed here, have heretofore been overlooked.

1. Introduction

Over the past fifteen years, there has been a surge in medical research focused on applying machine learning (ML), a subfield of artificial intelligence (AI), to the predictive diagnosis of diverse neurological disorders, particularly by using different modalities of neuroimaging data (Arbabshirani et al. 2017; Segato et al. 2020).[1] This research aims to develop ML algorithms—which "infer calculation rules from sets of data" (Jaton 2023, 788) and generalize these rules to unseen data—to enable automated diagnoses of complex

[1] The most common modalities of neuroimaging data used in this research include structural magnetic resonance imaging (MRI), functional magnetic resonance imaging (fMRI), positron emission tomography (PET), computed tomography (CT), functional near-infrared spectroscopy (fNIRS), and electroencephalography (EEG) (Segato et al. 2020, 4–5).

Pitt Open Library Publishing (cc) BY

This work is published by Pitt Open Library Publishing and is licensed under a Creative Commons Attribution 4.0 International License. © The Author(s).

ISSN 2692-3963 | Date Received 03-01-2025 | Date Accepted 03-11-2026 | Date First Published 23-11-2026
Philosophy of Medicine | DOI 10.5195/pom.2026.239 | Volume 7 | Issue 1 | pp. 1–30

disorders.[2] A substantial portion of this emerging research centers on developing diagnostic ML models for dementia and especially for Alzheimer's disease (AD).

Dementia is an umbrella term for symptoms such as memory loss, language problems, cognitive impairments, and behavioral changes that severely impact affected individuals' daily lives, limiting their autonomy. Currently, an estimated 60 million people worldwide have dementia, with the number forecast to almost triple by 2050 (Nichols et al. 2022). Up to 80% of dementia cases are attributed to AD (Alzheimer's Association 2024). This neurodegenerative disease is marked by progressive neural damage of still unknown causes and neurophysiological mechanisms. Despite recently approved disease-modifying drugs (Doran and Sawyer 2024; Kang 2024), there is no cure, and AD inevitably ends in death.

Although age is considered the biggest risk factor, AD is not an unavoidable part of aging. However, especially in its early stages, when treatments might be more effective, AD is challenging to distinguish clinically from healthy aging (Suemoto and Leite 2023).[3] Since the 2000s, to counter this problem, the agenda of AD research has been increasingly informed by the twofold "preventive logic" (Leibing 2014, 222; see also Leibing 2018; Leibing and Schicktanz 2021). Primary prevention involves addressing fourteen modifiable risk factors, including diabetes, smoking, hearing loss, education, and social isolation (Livingston et al. 2024).[4] Secondary prevention entails searching for biomarkers of early-stage AD—biological measurements that indicate underlying pathological processes. It is in the context of secondary prevention that emerging research on ML algorithms for predictively diagnosing AD from neuroimaging data before clear-cut clinical symptoms manifest is increasingly being viewed as a potential future solution to the currently unresolvable problem of early clinical diagnosis (Frizzell et al. 2022).

A growing number of humanities scholars have critically discussed the benefits and pitfalls of the broader preventive turn in AD research, focusing primarily on its ethical and sociopolitical implications, and on "what is at stake for aging individuals" (Leibing 2018, 705). On a positive note, Peter J. Whitehouse and Daniel R. George emphasize the productiveness of the preventive turn in downplaying unrealistic promises of ever finding a single cure, and in shifting attention instead to "maintaining brain and cognitive health" (2021, 243–244). Other, more critical accounts highlight the epistemic ambiguities of technological innovations in early AD diagnosis, questioning whether such innovations are socially and ethically responsible and under which conditions (Boenink et al. 2016; Leibing 2018; Leibing and Schicktanz 2021). Several authors have critiqued the fact that research, and even more so public policies, too often frame dementia prevention as the responsibility of the individual, disregarding the need for broader societal changes (Daly 2024; Leibing et al. 2025; Petersen et al. 2025). Moreover, Elizabeth Ford et al. (2023) have explored various

[2] For a succinct and accessible introduction to ML, see Alpaydin (2016). For a critical analysis of ML, see Mackenzie (2015). As Mackenzie pertinently notes, the "techniques of machine learning nearly all pivot around ways of transforming, constructing or imposing some kind of shape on the data and using that shape to discover, decide, classify, rank, cluster, recommend, label or predict what is happening or what will happen" (2015, 432).
[3] For philosophical debates on whether aging should be understood as a normal biological process, a disease, or a modifiable trajectory, and for discussions of the normative underpinnings of the notion of "healthy aging," see Blasimme (2021), Cardona (2008), De Winter (2015), Green and Hillersdal (2021), Pfaller and Schweda (2019), and Saborido and García-Barranquero (2022). For a broader philosophical discussion on how to conceptualize health and disease, see Schermer and Binney (2024).
[4] The additional modifiable risk factors include hypertension, obesity, depression, physical inactivity, excessive alcohol consumption, traumatic brain injury, air pollution, untreated vision loss, and high LDL cholesterol (Livingston et al. 2024).

ethical concerns, including unfairness and potential psychosocial harm, which might arise if digital and AI-based tools were to be used for population screening of early dementia.[5]

However, existing analyses have largely overlooked, or only marginally engaged with,[6] emerging research on ML algorithms for predictive AD diagnosis in individuals without clear-cut clinical symptoms using neuroimaging data, which is at the center of this paper. One reason for this might be that this research belongs to basic science. It is, therefore, still removed from directly shaping current clinical practice or public health policies (Li et al. 2022), although this might change in the future. Regardless of whether ML algorithms for predictive diagnosis of AD ever find clinical application, I argue that this research is operationally reconfiguring boundaries between healthy aging and future dementia during the process of algorithm development, whose details and broader implications I discuss here. Even if currently situated in the research context, this reconfiguring deserves attention as it is performed across a growing body of studies, which are published in biomedical journals and thus partake in the production of evolving medical knowledge about AD.

This paper examines the ongoing reconfiguring of the boundary between healthy aging and future dementia in the current research on ML algorithms for predictive AD diagnosis using neuroimaging data by focusing on the concrete epistemic processes that underpin the development of such algorithms in the research context. In doing so, the paper is informed by and contributes to recent work in science and technology studies (STS) and philosophy of science that "approach[es] science as performative" and, therefore, analyzes specific operations, choices, and contingencies through which scientific knowledge is produced in "science writings and settings" (Green and Hillersdal 2021, 3; see also Mol 2002). More specifically, I build on the emerging STS scholarship that explores AI "in the making" (Jaton and Sormani 2023, 630) by attending to how data practices and semantic negotiations are performed during ML algorithm development in specific medical research settings (Henriksen and Bechmann 2020; Henriksen and Olesen 2021; Jaton 2023; Muhr 2023, 2025; Winter and Carusi 2022). Importantly, instead of foregrounding the purely technical aspects of AI, this scholarship investigates "the human dimension of data work" and the epistemic consequences of human decisions and judgments "along the [ML] model's production chains" (Wirth et al. 2025, 3), asking: "Where do actors make choices about how machine learners should predict and classify the world?" (Lee 2025, 6136). My paper aligns with this orientation.

Much of this scholarship has focused on what, in the context of ML systems, is referred to as ground-truthing practice—the use of "benchmarked repositories called ground-truth datasets" (Jaton 2021, 1) during algorithm training and evaluation. Such benchmark datasets are deployed in the process of algorithm development for a twofold purpose. First, ground-truth data are used to define what an algorithm needs to learn during training and, second, to provide benchmarks for evaluating how well a new algorithm performs the task

[5] On the controversies surrounding population-based screening of dementia and translating early-intervention approaches into clinical practice, as well as their potentially detrimental effects on affected individuals, see Blasimme et al. (2021), Gunnarson et al. (2021), Swallow (2017), and Zeiler et al. (2023). For broader critical discussions of preventive medicine and the resulting overmedicalization, expansion of patienthood, and overdiagnosis, see Giroux (2024), Timmermans and Buchbinder (2010), and Tinland (2022).

[6] Ford et al. (2023) discuss this research only briefly, drawing on existing systematic reviews, and focusing mainly on technical aspects, such as the types of algorithms used, before turning to the ethical implications of the broader category of digital biomarkers.

for which it has been designed.[7] So far, scholars have primarily examined how ground-truth datasets are constructed when developing new medical algorithms, including what data are needed, how they are collected, and how they are labeled (Henriksen and Bechmann 2020; Henriksen and Olesen 2021; Jaton 2023). Most of these studies draw on ethnographic fieldwork examining algorithm development within a single research group or research site.

My analysis expands this scholarship by investigating how ground-truth datasets are operatively deployed across multiple studies when developing predictive algorithms for AD diagnosis using neuroimaging data. In other words, I shift the focus from how researchers construct ground-truth datasets to how they work with them during algorithm training and evaluation. I am not looking at the development of a particular algorithm, or a specific research site but instead analyze methods sections in selected published papers. This allows me to examine the different choices researcher teams make about how to organize comparisons of ground-truth data during algorithm training and evaluation with the aim of establishing AI-based predictive diagnostic separability between healthy subjects and future AD patients at the level of neuroimaging data. As I will show, these choices determine what is compared and how, including whether to compare neuroimaging data from healthy subjects and AD patients, or from individuals within a medical category called mild cognitive impairment (MCI), as well as whether to deploy direct or multistage comparisons.

The theoretical framework of my analysis is provided by recent literature that theorizes comparison, when used in scientific contexts, not as a neutral instrument that enables researchers to obtain "objective insights" but as a context-specific and historically situated epistemic practice in its own right (Epple and Erhart 2020, 22; see also Heintz 2016). Specifically, I draw on the work of Angelika Epple and Walter Erhart, who approach comparison in distinctly operative terms, as a "'relationing' activity" that goes way beyond stating mere differences" and is "always bound up with actors and agencies that perform the comparisons and connect them with their purposes and possible outcomes—intended or not" (2020, 17). Comparisons, as they show, are not found readily at hand but need to be constructed as part of a practice that becomes "established through repetition and routines, cultural habits, and historical patterns" (2020, 18). What is particularly pertinent for my analysis is that, once established, practices of comparison do not simply reveal preexisting differences between the entities being compared but actively produce new relations between them, resulting in novel semantic effects. As Bettina Heintz (2016) has pointed out, while the semantic effects may be unintended, or even remain unnoticed, by the actors who perform the comparison, they often extend beyond the context in which comparisons are performed, shaping broader epistemic understandings of the phenomena under investigation.

It is this understanding of comparison that I draw on in my close reading of studies on ML-based predictive AD diagnosis using neuroimaging data published between 2010 and 2025. Based on my analysis, I argue that different choices about how to organize comparisons during ground-truthing are reconfiguring boundaries between healthy aging and future dementia in conceptually inconsistent ways across the studies examined. I discuss the inconsistencies at two different levels—at the level of five dominant approaches

[7] As shown by Jaton (2021, 2), computer scientists rely on "ground-truth datasets while shaping and implementing supervised *and* unsupervised ML algorithms." For a distinction between supervised and unsupervised ML algorithms, both of which are developed in the context of research on predictive AD diagnosis, see Alpaydin (2016).

to organizing comparison that I identified in the field, and also within each approach, at the level of individual studies. I further argue that, while the resulting inconsistencies are partly informed by existing heterogeneity in how the boundaries between health and disease in relation to AD are currently drawn in the medical context (Daly and Keuck 2024; Schermer 2023), the comparison practices used in training and evaluating ML models do not simply reflect this heterogeneity but further compound it.

With this aim in mind, the paper is structured as follows: I first discuss the current ambiguities of early diagnosis of AD, focusing on biomarker-informed aspects of clinical diagnosis and the conceptual debates about defining AD in the medical literature.[8] I then provide a brief overview of the segment of the emerging research into ML models for predictively diagnosing AD from neuroimaging data to contextualize my subsequent analysis of select studies. Next, based on a close reading of the selected studies, I identify and analyze five distinct, coexisting approaches to how researchers organize comparisons of neuroimaging data during ground-truthing to establish predictive diagnostic separability between healthy individuals and future AD patients. In the final section, I examine the inconsistent ways in which these approaches draw boundaries between health and disease by establishing different relations between healthy controls, MCI, and AD, and discuss the resulting implications. I conclude by summarizing the main findings of my analysis.

2. Clinical and Conceptual Ambiguities of Early-Stage AD and MCI

Despite decades of intensifying research, AD remains "an equivocal, poorly delineated [medical] phenomenon" (Boenink et al. 2016, 3; see also Daly 2025; Daly and Keuck 2024; Schermer 2023; Van der Molen et al. 2025). Providing a detailed account of its multiple ambiguities, which are historically rooted,[9] differently distributed across clinical, research, and public-health contexts, and arising from "different knowledge-driven values and practical aims" (Daly and Mastroleo 2024, 160) is beyond the scope of this paper. Instead, this section broadly sketches some of the ambiguities of clinically diagnosing early-stage AD that the emerging research on ML-based predictive AD diagnosis using neuroimaging data promises to solve, as well as the current conceptual confusion in the medical literature as to how to draw the boundary between health and disease in relation to AD (Schermer 2023).

Currently, clinical diagnosis relies on a battery of cognitive, behavioral, and neurological tests that assess cognitive impairment (that is, clinical phenotype) and is further supported by diverse in vivo neuroimaging and cerebrospinal fluid (CSF) biomarkers that detect the associated neuropathology (Porsteinsson et al. 2021).[10] Noninvasive structural neuroimaging, primarily magnetic resonance imaging (MRI), is deployed to evaluate and localize underlying anatomical brain damage and exclude other conditions (Domínguez-Fernández et al. 2023). Invasive molecular neuroimaging—such as various forms of positron emission

[8] I focus on the biomarker-informed aspects of clinical diagnosis rather than symptom-based ones because, as noted above, the ML research examined here is situated within what Leibing (2014) designates as secondary prevention, which aims to identify biomarkers of AD before clear clinical symptoms emerge.

[9] For an insightful historical analysis of how, since its inception in the early twentieth century to the present day, the medical definition of AD has been based on conceptually shaky, uncertain ground, see Keuck (2021).

[10] For an overview of diverse clinical diagnostic tests used to assess cognitive deficits (for example, the Mini-Mental State Examination [MMSE] and Montreal Cognitive Assessment [MoCA]), changes in day-to-day functioning (for example, the Functional Activities Questionnaire [FAQ]), and behavioral symptoms such as anxiety or apathy, see Porsteinsson et al. (2021, 377–379).

tomography (PET)—and different CSF tests are implemented to identify AD's two early-stage "neuropathological hallmarks" (Porsteinsson et al. 2021, 371): beta-amyloid plaques and neurofibrillary tau tangles.[11] This diagnostic use of neuroimaging biomarkers is grounded in the currently dominant, although not universally accepted, ATN framework, where ATN stands for *a*myloid plaques, *t*au tangles, and *n*eurodegeneration.[12] This framework posits that the pathophysiological changes start with the accumulation of amyloid plaques, followed by tau tangles, and lead to neurodegeneration that results in the progressive loss of brain function and structure (Jack et al. 2016).

Importantly, no single cognitive test or biomarker is sufficient for clinical diagnosis, and none has a standardized threshold for defining the positive outcome (Dubois et al. 2021; Janeiro et al. 2020). Therefore, when diagnosing patients, doctors must interpret complex, multimodal data using clinical judgment, a process criticized in the medical context for its "subjectivity" (Borchert et al. 2023). The concern is that, in the absence of standardized thresholds, different clinicians may interpret the same combination of test results and biomarkers differently, making their diagnostic decisions mutually inconsistent and difficult to reproduce, and increasing the likelihood that more ambiguous AD cases are missed by some clinicians. What further complicates clinical diagnosis, especially in early stages, is that the symptoms of cognitive decline are often insidious. Subjective memory complaints do not necessarily correlate with performance on cognitive tests (Anderson 2019), and many older patients have multiple comorbidities, including non-AD neuro-degenerative diseases (Alzheimer's Association 2024). Due to these ambiguities, many individuals remain undiagnosed or are diagnosed at advanced stages of dementia when the underlying brain damage is untreatable.

So far, my discussion has focused on the pragmatic challenges of clinical diagnosis, especially in the early stages of AD, as these are the challenges that the research on ML algorithms for predictively diagnosing AD using neuroimaging data promises to solve. However, in addition to the diagnostic difficulties in clinical context, there are also broader conceptual disparities among medical experts about how to draw the boundary between health and disease in relation to AD, particularly when positive biomarkers are present in individuals who manifest no detectable clinical symptoms (Daly and Keuck 2024; Schermer 2023). These disparities are closely related to the fact that the exact role of plaques and tangles in the neuropathology of AD remains debated. Studies have shown that amyloid plaques, tau tangles, and the associated brain damage occur decades before cognitive symptoms manifest (Lloret et al. 2019). On this basis, the Alzheimer's Association Expert Working Group has proposed that abnormal core biomarkers alone are sufficient to diagnose AD, even in individuals with no clinical signs of cognitive impairment (Jack et al. 2024). According to this deterministic "biological definition of AD," positive tau and amyloid biomarkers mean that the disease already exists "at this initially detectable stage and will eventually produce symptoms if the individual lives long enough" (2024, 5148).

[11] Beta-amyloid plaques are clumps of protein fragments that accumulate between brain cells. Tau tangles are abnormal aggregations of tau-protein inside brain cells (Jack et al. 2016). For an overview of different types of PET biomarkers, see Domínguez-Fernández et al. (2023). Both CSF tests and molecular neuroimaging, the latter of which involves injecting small amounts of radioactive tracers, are invasive methods.

[12] Notably, the Lancet Commissions (Livingston et al. 2024) shift attention away from biomarker-based definitions of AD toward a broader clinical and public-health approach to dementia prevention. As this paper focuses on biomarker-based research, I do not discuss this "biomarker-independent" conceptualization of AD (Daly and Keuck 2024, 1471). For more details, see Daly and Keuck (2024).

This definition has been criticized as driving an increasing "biomarkerization of the AD concept which may have substantial effects on AD research and the dementia field as a whole" (Van der Molen et al. 2025, 9). By contrast, the International Working Group draws the boundary differently, arguing that the diagnosis of AD requires both positive biomarkers and clinical signs of cognitive impairment. On this view, asymptomatic biomarker-positive individuals "should be considered only at-risk for progression to Alzheimer's disease" (Dubois et al. 2021, 484) but not as yet having a disease.

The latter "clinicobiological" definition of AD (Daly and Mastroleo 2024, 160) draws on empirical findings showing that the presence of positive biomarkers does not necessarily lead to AD. To begin with, "all stages of Alzheimer's disease brain lesions (including amyloid β and tau lesions) are found in two-thirds of individuals aged at least 70 years in systematic post-mortem examination, regardless of clinical status" (Dubois et al. 2021, 488). Moreover, in patients diagnosed with AD, the extent and distribution of brain pathologies and the trajectories of cognitive decline vary significantly. While this does not mean that the biomarkers lack clinical utility, it does mean that there is no straightforward correlation between the current models of AD neuropathology—which underpin the use of biomarkers—and its clinical manifestations (see also Schermer 2023). This, in turn, implies that the predictive value of tau and amyloid biomarkers in presymptomatic individuals remains contested and limited (Daly and Keuck 2024, 1470), which is exactly the area of application in which the biomarkers are seen as particularly needed.

It is against this background that the aforementioned secondary preventive logic has gained traction in medical research, emphasizing the need for developing methods of reliable early clinical detection of AD before the affected individuals exhibit fully established symptoms of dementia. In addition to the intensified search for new biomarkers, the introduction of the medical construct of MCI has been central to this turn (Leibing 2018). Since the early 2000s, MCI has been "extensively used worldwide, both in clinical and in research settings, to define the grey area between intact cognitive functioning and clinical dementia" (Petersen et al. 2014, 214). MCI is also crucial to my discussion because of its key role in the emerging research on ML algorithms for predictive AD diagnosis, which I discuss later.

By definition, MCI is characterized primarily by mild memory deficits in individuals whose overall cognitive functioning remains preserved. Among other reasons, MCI was introduced to aid the discovery of "signs and symptoms that could be used as reliable predictive markers of [Alzheimer's] disease development" (Petersen et al. 2014, 214). But its introduction has also led to several thus far unresolved challenges. The diagnosis relies on a clinical interview, aimed at assessing subjective complaints and various cognitive tests to measure the level of cognitive deficits, and can be extended through CSF and neuro-imaging biomarkers (Anderson 2019; Bradfield 2023). Operationally, MCI is even more challenging to diagnose than AD because of the mildness of its symptomatic manifestations and the ambiguity of neuropathological correlates. Challenges include considerable inconsistencies in "how a cognitive 'impairment' is operationalized" concerning differences in the cut-off point on cognitive tests (Anderson 2019, 82). Conceptually, MCI is a highly ambiguous transitional category between healthy aging and clinical AD, and its medical status lacks consensus. Some define it as a prodromal stage marking the beginning of AD (Pike and Kinsella 2019). Others use it to designate healthy individuals at risk for AD

(Golomb et al. 2004). This lack of consensus about the status of MCI further complicates the already contested boundary between health and disease in relation to AD.

Furthermore, the prognostic utility of MCI is debatable. According to various studies, the average conversion rate from MCI to AD ranges from 8% to 15% per year (Petersen 2016). However, one meta-analysis concluded that "most people with MCI will not progress to dementia even after 10 years of follow-up" (Mitchell and Shiri-Feshki 2009, 252). While MCI remains stable in some individuals for many years, a significant proportion of subjects (16–26%) spontaneously recover, reverting to normal cognitive functioning (Canevelli et al. 2016; Koepsell and Monsell 2012). And even in those individuals whose cognitive impairments worsen, this might be due to neurodegenerative conditions other than AD (Bradfield 2023). Hence, besides its clinical and conceptual ambiguity, MCI is also a highly heterogeneous diagnostic category that incorporates individuals with both progressive and nonprogressive mild cognitive deficits of different underlying causes.

Despite its fuzzy boundaries, MCI is used in clinical and research practices as an intermediary category that underpins efforts for early AD detection. Its predictive potential, especially in the research context, is increasingly tied to finding ways to operatively distinguish MCI individuals who are likely to develop dementia from those who will remain stable (Alzheimer's Association 2024; Pandya et al. 2016, 60). Much of the emerging ML research into predictive AD diagnosis using neuroimaging data addresses precisely this question, with the explicit aim of developing a purportedly less "subjective" alternative to the current clinical diagnosis (Borchert et al. 2023), whose practical and conceptual ambiguities were sketched earlier in this section. However, as I will argue, while this research promises to overcome some of the challenges of early clinical diagnosis, it instead intensifies existing ambiguities by redefining boundaries between health and disease in conceptually inconsistent ways, thereby compounding the already unstable and partly conflicting boundaries that currently coexist in the medical context. Before turning to this argument, I will briefly introduce the research on ML-based predictive AD diagnosis using neuroimaging data.

3. Situating Current Research on ML-Based Predictive AD Diagnosis

Initially, the emerging research into ML models for AD diagnosis focused on designing algorithms to differentiate AD patients from healthy individuals or patients with other types of dementia. It did so by using primarily neuroimaging data and, to a lesser extent, other data types like cognitive test scores, blood-based and CSF biomarkers, and genetic information (Li et al. 2022). More recently, the focus has shifted toward developing ML models for predictive AD diagnosis in terms of likely future conversion from MCI (Frizzell et al. 2022; Grueso and Viejo-Sobera 2021; Leandrou et al. 2018; Singh et al. 2024). Most studies have developed predictive ML models for neuroimaging data, predominantly one particular modality, structural MRI, on which I focus here. Thus, the studies discussed in this paper have aimed to create ML algorithms that could be deployed clinically to predictively diagnose AD through automated analysis of structural MRI scans of healthy subjects or those tentatively diagnosed with MCI.

Several aspects of this research need to be introduced to provide a conceptual basis for my subsequent analysis. First, the development of predictive ML algorithms is closely tied to the establishment of large-scale open-access digital databases since the mid-2000s

(Borchert et al. 2023; Singh et al. 2024). When constructing ground-truth datasets for training and evaluating ML algorithms for predictive AD diagnosis, most studies have utilized the publicly available research datasets by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (Weiner et al. 2015).[13] Other less prominently used but nevertheless important databases include the Open Access Series of Imaging Studies (OASIS) and the Australian Imaging Biomarkers and Lifestyle Study of Aging (AIBL).[14] These databases were established to enable research on AD progression by collecting longitudinal neuroimaging (MRI and PET), clinical, genetic, and biochemical biomarker data from individuals diagnostically classified as cognitively unimpaired, MCI subjects, and AD patients. The neuroimaging data and the diagnostic categories assigned to them within these open-access databases provide the source for ground-truth datasets on which predictive algorithms for AD diagnosis are trained and evaluated. However, as my analysis in the subsequent section will show, researchers make important interpretive decisions when constructing and working with their ground-truth datasets, thereby establishing new semantic relations between the diagnostic categories of healthy control, MCI, and AD, which often go beyond those used in the open-access datasets.

Second, ML-driven predictive diagnosis using MRI images differs from conventional group-level statistical analysis. Instead of identifying statistically significant group differences in brain features between patient and control cohorts, ML algorithms make predictions about individual subjects by "automatically classify[ing] each subject into one of the groups in the study (e.g., healthy vs. patient)" based on MRI data (Arbabshirani et al. 2017, 138). Simply put, the ML-driven prediction of a future AD diagnosis is treated as an image classification problem—the algorithm is trained to assign classificatory labels to MRI images according to predefined diagnostic categories (for example, "AD patient"). Importantly, image classification in this context is informed by two related assumptions. The first assumption is the existence of unknown visual patterns of disease-related tissue pathology in MRI scans from MCI subjects who will progress to AD ("MCI converters"), which are presumed to be absent in MRI scans from MCI subjects who will remain stable ("MCI non-converters"). The second assumption is that although invisible to humans, these latent pathological data patterns, which are expected to be distributed throughout the three-dimensional MRI data volumes (Nanni et al. 2020, 2), can be identified by adequately trained ML algorithms. This means that the classification problem underlying the production of ML algorithms for predictive AD diagnosis is conceptually framed as an early detection of some already present pathological changes in brain structure, which are assumed to reliably indicate future progression to AD. In other words, it is the identification of such patterns in MRI data that serves as the basis for establishing diagnostic separability between healthy aging and future AD in this research.

Third, ML algorithms learn to perform such image classifications without being explicitly programmed with predetermined rules. Instead, they learn through the process of ground-truthing, during which researchers define the problem to be solved computationally by selecting and organizing reference datasets used for algorithm training and evaluation (Jaton 2021). The process starts with exposing ML algorithms to training data, which, in this context, consist of labeled MRI images, where each image corresponds

[13] See https://adni.loni.usc.edu/.
[14] See http://www.oasis-brains.org and http://www.aibl.csiro.au.

to a fixed diagnostic category (for example, "healthy control," "MCI converter," or "AD patient"). The algorithm's task is to identify differential patterns (structural features of MRI scans) between the predefined categories of the training data and infer generalizable patterns that can be applied to unseen cases, enabling accurate predictions. A trained predictive model should be capable of accurately assigning diagnostic labels to new, previously unseen MRI data, based on the patterns it has learned from the training data. Researchers evaluate the trained algorithm's generalization ability—its capacity to make accurate future predictions on unseen data—by quantifying its performance on test data, using metrics such as prediction accuracy.

Discussions and comparisons of prediction accuracies and related quantitative performance metrics dominate the medical discourse (Leandrou et al. 2018). However, my analysis here disregards performance metrics because they approach ML algorithms as finished products and place the focus on their technical aspects. Instead, I focus on the interpretative decisions researchers make when organizing comparisons between MRI data from different groups of subjects during ground-truthing. Such decisions include, for example, whether MRI data from MCI converters are compared directly with those from MCI non-converters or whether the converter/non-converter comparison is combined with comparisons involving healthy controls and AD patients, and if so, how such complex comparisons are organized. They also include choices on how to define MCI converters and non-converters and over what time frame to perform the comparisons. As I show in the following sections, these decisions are far from merely technical because they shape how relations between healthy aging, MCI, and future AD dementia are established at the level of neuroimaging data and, in doing so, participate in reconfiguring the boundaries between health and disease within the research context.

To make this argument, in what follows, I analyze the methods sections of 36 studies on ML-based predictive AD diagnosis published between 2010 and 2025. In selecting the studies that comprise my empirical sample, I have employed a two-stage procedure. First, I have identified published studies on ML algorithms for predictive AD diagnosis that use MRI data by consulting multiple systematic reviews (Borchert et al. 2023; Frizzell et al. 2022; Grueso et al. 2021; Li et al. 2022; Pellegrini et al. 2018; Singh et al. 2024; Valizadeh et al. 2025; Vermeulen et al. 2025; Zhao et al. 2023) and by searching PubMed, an online database of biomedical literature. Then, based on a close reading of the studies' methods sections, I have selected studies that exemplify the diversity of ways in which researchers organize comparisons during algorithm training and evaluation.[15] My aim is not to provide an exhaustive survey of the field but to identify and analyze distinct comparative approaches used by researchers to establish predictive diagnostic separability between healthy individuals and future AD patients at the level of MRI data. Consequently, other aspects of researchers' methodological choices, such as whether they used the ADNI or alternative

[15] These studies are: Agostinho et al. (2024), Aguilar et al. (2013), Bae et al. (2021), Bron et al. (2015), Casanova et al. (2018), Cheng et al. (2015), Costafreda et al. (2011), Cuingnet et al. (2011), Duchesne and Mouiha (2011), Ezzati et al. (2019), Franciotti et al. (2023), Ghaffari et al. (2022), Gupta et al. (2019), Hoang et al. (2023), Hojjati et al. (2018), Khojaste-Sarakhsi et al. (2025), Lebedev et al. (2014), Li et al. (2025), Lin et al. (2020), Liu et al. (2019), Lu et al. (2018), Mofrad, Lundervold, and Lundervold (2021), Mofrad et al. (2021), Moscoso et al. (2019), Nanni et al. (2020), Ning et al. (2018), Plant et al. (2010), Ren et al. (2023), Spasov et al. (2019), Suk et al. (2016), Sun et al. (2017), Tong et al. (2017), Xi et al. (2025), Yu et al. (2024), Zhang et al. (2012), and Zheng et al. (2025).

open-access databases or which types of ML algorithms they employed, were not of primary interest in my selection.

4. Reconfiguring the Boundary Between Disease and Health Through Different Comparisons

As discussed in the previous section, ML algorithms for predictive AD diagnosis learn during ground-truthing to identify what, in this research field, are presumed to be latent abnormalities in brain anatomy indicative of incipient dementia. They do so by comparing predefined classes of MRI data to identify differential visual patterns between these image classes. If we adopt Epple and Erhart's (2020, 15) description of the activity of comparing as an "operation that puts into perspective" two or more objects of comparison; that is, *comparata,* "in respect to a *tertium comparationis,*" with the aim of establishing similarities and differences between *comparata*, then we can say the following: During ground-truthing in the studies comprising my empirical sample, classes of MRI images, each labeled with a particular diagnostic category, serve as *comparata,* whereas the *tertium comparationis* consists of neuroanatomical structures as represented in MRI images. The aim of comparison is to establish differences between the *comparata* with respect to this *tertium comparationis,* which are regarded as diagnostically relevant for distinguishing between healthy individuals and those with future dementia.

Two aspects are important for the analysis that follows. First, what the previous paragraph should already make apparent is that diagnostic differentiation between healthy individuals and those with future dementia at the level of MRI data is defined in purely relational terms as it depends on the choice of classes of images—and their corresponding diagnostic labels—included in the comparison, a point that I develop further in this section. Second, as Heintz (2016) has argued, while comparisons are typically performed in order to identify differences of interest between the objects of comparison based on a *tertium comparationis,* their semantic effects are not limited to the production of such differences. Instead, comparisons should be seen as "relationing instruments" because the identification of differences requires the establishment of new semantic relations between the objects of comparison (2016, 307). In other words, although comparisons typically rely on preexisting categories that have prior meanings, the activity of comparing recontextualizes the relations between these categories. It thus produces shifts in their meanings, in ways that may be unintended or overlooked by those who perform the comparisons. As I argue in the subsequent section, such unintended semantic shifts are occurring in current research on algorithms for predictive AD diagnosis.

In the remainder of this paper, I examine researchers' decisions about which classes of MRI images an algorithm is made to compare during training and testing by approaching ground-truthing as a comparative practice that has not yet conceptually stabilized in this particular field. Instead, this practice consists of multiple approaches, which, as I will show, are reconfiguring the boundary between healthy aging and future dementia at the level of neuroimaging data in conceptually inconsistent ways by establishing different comparative arrangements during ground-truthing. Based on a close reading of the studies comprising my empirical sample, I identified five such widespread approaches. It is important to note that not all approaches discussed here are mutually exclusive; some can and have been combined within a single study. However, prying them apart analytically allows me to

examine what I claim are their distinctive semantic effects. Instead of providing an exhaustive overview of all possible approaches to defining comparative arrangements during ground-truthing in this research, I focus on the five approaches I identified in order to make apparent their semantic disparities in reconfiguring the boundaries between health and disease.

4.1 Assuming Equivalence

The first approach used during the ground-truthing of ML algorithms for predictive neuroimaging-based AD diagnosis is the simplest. While it was more prevalent in early studies (Aguilar et al. 2013; Costafreda et al. 2011; Lebedev et al. 2014; Plant et al. 2010), it nevertheless continues to be applied in more recent research (Ezzati et al. 2019). In this approach, a predictive ML algorithm is trained on MRI data to distinguish AD patients from healthy controls without using MRI data from MCI subjects. Once trained, the algorithm is then tested on MRI data from MCI subjects to predict which will convert to AD by classifying these data into MCI converters or MCI non-converters.

The assumption underpinning this approach is that the binary comparison between healthy controls and AD patients is operationally equivalent to the binary comparison between MCI converters and MCI non-converters, the former being a "simplified version" of the latter (Tong et al. 2017, 158). On this assumption, an algorithm that accurately separates MRI scans from healthy controls and AD patients is expected to differentiate MRI scans from MCI converters and MCI non-converters with similar accuracy. As a relationing activity, this comparison thus implicitly equates MCI non-converters with healthy subjects, on the one hand, and MCI converters with AD patients, on the other hand. Although these different diagnostic categories remain formally distinct, they are relationally organized in ways that allow differences identified between healthy controls and AD patients to stand in for differences between MCI non-converters and MCI converters.

4.2 Defining Conversion Through Cut-Off Periods

Another approach that has been particularly widely deployed in the research community involves training and evaluating predictive ML algorithms using only baseline MRI scans from MCI subjects, who are retrospectively categorized as converters or non-converters based on whether they developed AD within a specific cut-off period.[16] Unlike in the previous approach, the comparison between AD patients and healthy controls is here entirely neglected during ground-truthing. Instead, diagnostically relevant relations at the level of MRI data are established only within the category of MCI. Therefore, the semantically critical decision—and the crucial comparative operation—entails retrospectively dividing the otherwise undifferentiated group of MCI subjects into converters and non-converters.

This decision is operationally dependent on follow-up examinations because open-access databases such as ADNI do not label subjects as MCI converters and MCI non-converters. Instead, at each point in the process, participants are classified as healthy

[16] A baseline scan is taken at the first examination.

controls, MCI subjects, or AD patients.[17] To assign the labels of MCI converters or MCI non-converters, researchers must determine whether an individual remained in the MCI category or progressed to AD after a specific period following the baseline examination. Importantly, while labeling subjects as MCI converters or MCI non-converters depends on their follow-up examinations, the MRI data used for algorithm ground-truthing are taken from each subject's baseline examination. The converter/non-converter distinction is thus produced by relating baseline MRI scans to a temporally distant diagnostic outcome and then projecting this relation back onto the baseline scans as if it were already contained within these images.

The decisive comparative choice within this approach concerns determining the cut-off period used for distinguishing converters from non-converters, because the cut-off period provides the basis for defining the *comparata* on which the algorithm is trained and tested to identify diagnostic differences. Problematically, different studies have used different cut-off periods. These have ranged from 18 months (Bron et al. 2015; Cuingnet et al. 2011) to 48 months (Duchesne and Mouiha 2011) to 3 years (Casanova et al. 2018; Hoang et al. 2023; Hojjati et al. 2018). However, researchers have rarely justified their choices, and no consensus exists in the community regarding an optimal cut-off period, reflecting the ambiguity of current diagnostic boundaries, as discussed in section 2. Consequently, the operative distinction between MCI converters and MCI non-converters used during ground-truthing is highly ambivalent as it is produced through researchers' inconsistent decisions about how to define conversion in terms of the cut-off period after baseline.

Notably, Alexis Moscoso et al. found that extending the cut-off period from two to five years affected the ML algorithms' predictive power because "most of the MCI patients that remained stable over short periods and subsequently progressed to AD dementia had evident atrophies at baseline" (2019, 1). Based on this finding, Moscoso et al. conclude that "machine learning models that use these patients to learn non-disease patterns" have "reduced performance" and, we may add, ambiguous biological interpretability.[18] For the present discussion, however, what is particularly significant is that extending the temporal cut-off changes how the conversion, as the basis of ground-truthing comparison, is defined. Individuals classified as MCI non-converters in studies with shorter cut-offs might be reclassified as MCI converters in studies using longer cut-offs, even if the data stems from the same open-access database. Yet the result is not simply variability in labeling but, even more importantly, variability in the conceptual framing of conversion, since different cut-off periods establish different relations between present MCI status and future dementia, and thus draw a different boundary between health and disease.

4.3 Modeling Conversion Trajectories

Unlike the previous approaches, which use MRI data from a single point in time (that is, the baseline) during ground-truthing, the third classificatory comparison adopts a distinctly

[17] Already this example shows that ground-truthing involves more than simply adopting the diagnostic categories provided by the ADNI and other open-access databases. By retrospectively reclassifying MCI subjects as converters and non-converters on the basis of later follow-up examinations, researchers establish new relations within the existing diagnostic category of MCI.

[18] Nevertheless, studies that develop predictive algorithms based on significantly shorter cut-off periods continue to be published (see, for example, Hoang et al. 2023).

longitudinal perspective.[19] This comparison deploys MRI data repeatedly collected from the same individuals over an extended period. Here, the diagnostically relevant differences are configured by comparing temporal trajectories that trace the conversion from MCI to AD based on the changes between baseline and multiple follow-up MRI scans. The assumption underpinning this comparison is that the conversion from health to disease can be understood as a gradual process whose reliable prediction requires linking multiple points in time to model this trajectory. Conversion is thus defined as a relation across sequential points in time, rather than the presence of a pattern detectable within any single scan.

The crucial comparative operation in this approach consists in relating multiple scans of the same individual to one another in ways that make patterns of change over time comparable across subjects. However, there are inconsistencies in how such modeling of conversion trajectories is performed across studies. Researchers make different decisions regarding the period covered by the longitudinal measurements, the number of serial scans considered sufficient per subject, and the intervals between these scans. For example, Zhuo Sun et al. (2017, 3) deployed longitudinal MRI data from the ADNI database for a three-year period and included only MCI subjects who had "a baseline scan, together with follow-up scans at 6, 12, 24, and 36 months." Yanyan Lin et al. (2020) also used the ADNI database but covered two years, selecting MCI subjects scanned at baseline and two follow-ups, each at an interval of six months. Samaneh A. Mofrad et al. (2021, 3) combined the ADNI and AIBL databases, using "longitudinal brain MRI data from subjects scanned multiple times (at least twice) over a period of 15 years." In a recent study, Guojiang Zheng et al. (2025) based their longitudinal modeling only on two scans, separated by a fixed two-year interval.

None of these different decisions was justified by the researchers. Yet, importantly, these differences do not merely affect the granularity of longitudinal modeling but also how conversion is conceptually constituted as a temporal relation that structures the comparison. A trajectory defined through two scans collected over six months establishes a different relation between MCI and AD than one defined through multiple scans collected over a decade. In the first case, conversion is implicitly understood as a relatively quick and easily identifiable process, whereas in the other as a slower and more insidious one. Consequently, different studies construct different versions of conversion through which the boundary between present health and future dementia is drawn in conceptually inconsistent ways.

4.4 Constructing Complex Comparative Arrangements

The fourth approach, which has become increasingly prevalent in recent studies, displaces a single binary comparison with more complex comparative arrangements. The underlying assumption is that differences between MRI data from MCI non-converters and MCI converters may be too subtle to be identified through their direct comparison alone and that additional information can be gained by combining multiple categories or several related comparisons into arrangements through which diagnostic differences are produced. Various ways of combining categories and comparisons are being explored within the research community.

[19] While the second approach also incorporates a temporal dimension through the retrospective classification of MCI subjects as converters and non-converters, the MRI data used for training and evaluation remain restricted to baseline scans.

To predict conversion from MCI to AD, some researchers have deployed a method called multitask learning. In this method, an algorithm learns several mutually related classification tasks concurrently to draw on their commonalities and differences. For example, Simeon Spasov et al. (2019) trained and tested their deep-learning algorithm to simultaneously distinguish MCI non-converters from MCI converters and AD patients from healthy controls. As a relationing activity, the comparison between MCI non-converters and MCI converters is here not performed independently but entangled with the comparison between healthy controls and AD patients and thus shaped by it.

Another widely used method within this approach is transfer learning. Whereas multitask learning involves solving several tasks concurrently, in transfer learning, tasks are learned sequentially. In some cases, an algorithm is pretrained "using visual information similar to that used in the target task" to acquire generic knowledge that will aid in target classification (Bae et al. 2021, 53). The algorithm is then retrained using data that are directly relevant to the target task. In multiple studies (Bae et al. 2021; Cheng et al. 2015; Ren et al. 2023; Tong et al. 2017; Yu et al. 2024; Li et al. 2025), algorithms for predicting future MCI conversion were first pretrained to classify MRI scans from AD patients and healthy controls and then retrained to distinguish MRI scans from MCI converters and MCI non-converters. Relationally, this approach treats the comparison between MCI converters and MCI non-converters and the comparison between AD patients and healthy controls as neither equivalent nor entirely separate. Instead, it establishes a relation of partial similarity between them and assumes that distinctions learned in one comparison can refine the other.

Recently, several studies have deployed a different version of transfer learning in which various types of deep-learning architectures that had been pretrained on vast databases of generic color images (for example, ImageNet) were then fine-tuned on grayscale structural MRI images (Ghaffari et al. 2022; Hoang et al. 2023; Nanni et al. 2020). For example, Loris Nanni et al. (2020) performed fine-tuning of different pretrained convolutional neural network (CNN) models for the binary comparison between MCI non-converters and MCI converters, as well as between healthy controls and AD patients. The underlying assumption here is that distinctions acquired through generic computer-vision tasks, such as visual object recognition, can be transferred to the identification of disease-related patterns of cerebral atrophy in MRI scans.

One further method that is increasingly used in this research is multiclass classification, in which more than two *comparata* are mutually related within a single comparative arrangement (Khojaste-Sarakhsi et al. 2025; Liu et al. 2019; Suk et al. 2016). Here, an ML algorithm is trained and evaluated to simultaneously differentiate among four categories: healthy controls, MCI non-converters, MCI converters, and AD patients. Unlike multitask or transfer learning, the relationing activity underpinning this comparison aims to establish diagnostic distinctions across all four categories at once, as each category is differentiated in relation to three others, rather than against only one opposing category.

Across these different methods, the distinctive feature of this approach is that boundaries between healthy aging and future dementia are configured in potentially more nuanced ways, as the different arrangements through which comparisons are organized allow researchers to establish more complex comparative relations across multiple *comparata*.

4.5 Integrating Heterogeneous Data Modalities

The fifth approach entails comparisons organized around multiple data modalities. Unlike the previous approaches that relied on MRI data alone, multimodal comparisons combine MRI data with other neuroimaging and non-imaging modalities when training and evaluating ML algorithms for predictive AD diagnosis. This approach is characterized by a pronounced heterogeneity in the choice and number of incorporated data modalities. For example, some researchers have developed predictive algorithms for identifying MCI converters by integrating structural MRI—which focuses on neuroanatomical changes—with fMRI, a functional MRI modality that provides complementary information about subjects' brain activity (Hojjati et al. 2018). Others tested alternative binary combinations, integrating MRI data with PET scans (Agostinho et al. 2024; Lu et al. 2018), cognitive tests (Mofrad et al. 2021; Zhang et al. 2012), or genetic data (Ning et al. 2018; Xi et al. 2025). Some studies have moved beyond such binary combinations and have started to simultaneously incorporate multiple data modalities such as MRI, CSF levels, PET scans, genetic data, cognitive scores, and demographic information (Franciotti et al. 2023; Gupta et al. 2019; Yu et al. 2024).

In multimodal comparisons, the difference between MCI converters and MCI non-converters is thus configured through relations established between heterogeneous forms of data that are assumed to capture different, potentially complementary aspects of the disease. However, medically interpreting the algorithmically derived differences can be conceptually challenging, as it is unclear how algorithms integrate complementary information from different data modalities. Moreover, the boundaries between healthy aging and future dementia are configured differently across studies, depending on which modalities are incorporated into the comparison and how relations between these modalities are organized.

4.6 Summary

Taken together, the five approaches analyzed in this section demonstrate that ground-truthing in neuroimaging research on ML algorithms for predictive AD diagnosis is not a homogeneous comparative practice but instead operates with diverse comparative arrangements. The inconsistencies within each approach reveal how individual studies configure diagnostic relations differently. However, the differences across the five approaches raise broader questions about how the boundaries between health and dementia are conceptually produced through ML-based comparisons, a topic that I address in the next section.

5. Conceptual Inconsistencies Across the Approaches and Their Implications

The above overview of the five approaches to defining the conditions of comparing MRI data, which underpin the development of ML algorithms for predictive AD diagnosis, has aimed to highlight the diversity of these approaches. We have seen that despite the shared aim of achieving statistically significant diagnostic separability between healthy aging and future dementia at the level of MRI data, the approaches differ both in the categories of subjects included in the comparison and in how these categories are related to one another

during ground-truthing. What is particularly important to emphasize is that researchers' decisions on how to define the task to be learned and how to organize the comparisons of different categories of MRI data during ground-truthing are primarily motivated by pragmatic reasons. This is because "individual algorithms may perform better than the others for a given task" (Nanni et al. 2020, 3) and for a particular classificatory comparison. Within the research community, the adequacy of a particular approach aimed at establishing the algorithmic separability between MCI converters and MCI non-converters at the level of MRI images is evaluated based on optimal performance metrics, such as accuracy (Pellegrini et al. 2018). In other words, the comparisons are not primarily designed with medical interpretability or conceptual stringency in mind but to increase the statistically defined efficacy of automated image classification. Thus, the focus is on solving the problem of differentiating between MCI converters and MCI non-converters in distinctly technical terms. Sometimes, this includes testing different comparisons within a single study to determine which delivers higher classification accuracy (Nanni et al. 2020; Sun et al. 2017).

However, in this section, I will draw on Heintz (2016) to show that the different approaches to defining classification tasks and the comparisons on which they rely are not semantically neutral. They do more than simply produce differences among the categories being compared in terms of potential predictive diagnosis of AD. Depending on how MRI images—alone or sometimes also in combination with non-imaging data—are organized into comparisons within each approach, these approaches establish distinctly different relations between MCI, AD, and healthy aging, thus implicitly reconfiguring boundaries between disease and health in conceptually inconsistent ways. Due to the focus on technical feasibility in the current ML-based research, the resulting conceptual inconsistencies among these different approaches, which admittedly are not self-evident, have so far been overlooked.

The first approach, which conflates MCI converters with AD patients and MCI non-converters with healthy subjects, operates with simplified binary categories of health and disease. It implies that health and disease can be clearly delineated through algorithmic analysis of MRI data, treating MRI differences as sufficient on their own for identifying future dementia in non-symptomatic individuals or those with vague cognitive difficulties. In this sense, it aligns with the logic of the biological model mentioned previously, which equates disease with the presence of detectable abnormalities on what Clifford R. Jack et al. (2024) call core biomarkers. However, the first approach discussed here could be said not only to inherit this logic but also to further amplify it. Instead of relying on the combined interpretation of multiple biomarkers and their associated quantitative thresholds, it diagnostically frames future dementia solely through the relationally constituted neuroanatomical differences derived from unconstrained comparisons between MRI data from AD patients and healthy controls. By operationally collapsing the diagnostic distinction between already manifest AD and future conversion to AD at the level of MRI data, this framing ignores the semantic significance of clinical symptoms in defining disease. In doing so, it implicitly medicalizes presymptomatic states whose eventual clinical manifestation may, in fact, remain uncertain. The boundary between healthy aging and future dementia, as configured within this approach, thus appears more biologically deterministic than in either the biological or clinicobiological models discussed in the medical literature (Schermer 2023) and could be seen as a further intensification of the

ongoing "biomarkerization" of AD (Van der Molen et al. 2025, 1). My point here is not to assess whether this approach is clinically justified and ethically defensible (cf. Blasimme et al. 2021; Swallow 2017) but to flag that an implicit conceptual change emerges from particular methodological decisions made during algorithm ground-truthing.

The second and third approaches arguably complicate the establishment of a biologically deterministic boundary between health and disease by explicitly operationalizing the temporal dimension of the transition between these two extremes but do so in distinctly different ways and with different semantic effects.

The second approach defines the boundary by first imposing a quantitatively defined cut-off period—that is, *x* months or years from baseline—by which time the disease, if present, has become symptomatically manifest, and then using the information about the diagnostic status at follow-up to project the resulting distinction back onto a time when the clinically relevant symptoms of AD were still absent. By paying attention to the diagnostic shift to AD after the cut-off period used to delimit the time frame within which conversion is meant to take place, this approach implicitly acknowledges the diagnostic importance of symptoms, as these are necessarily entailed in the clinical diagnosis of AD. In this sense, it departs from a purely biological framing, even though the difference between future dementia and healthy aging is ultimately negotiated at the level of MRI data. However, it nevertheless retains a binary distinction between health and disease because it operates with fixed cut-off periods that impose a predefined temporal boundary and, problematically, differ across studies. Moreover, these cut-off periods appear to be arbitrary, rather than grounded either in empirical data or in a clear conceptual framework. This raises a more fundamental conceptual problem: If the boundary between health and disease changes with temporal parameters chosen during ground-truthing without any explicit justification, this boundary cannot be claimed to refer to an underlying biological reality. Instead, the boundary is constructed through the comparative arrangements and operationally imposed on predictive ML algorithms that are then trained to identify it.

Conversely, the third, longitudinal approach conceptualizes conversion from health to disease not as being limited to a predefined time window but as developing across varied trajectories empirically derived from patients' sequential data. By treating conversion as a process that can be dynamically modeled over time and attending to the existence of different individual trajectories, this approach operates with a boundary between healthy aging and future dementia that is not conceived as a fixed dividing line but as a gradual transition. Health and disease are configured here as shifting positions within a continuous spectrum that connects presymptomatic stages to those with full-blown symptom manifestations. While not identical to it, this understanding appears more closely aligned with the clinicobiological model discussed in the medical literature, which likewise avoids equating detectable pathology with the immediate presence of disease (Dubois et al. 2024). Conceptually, however, the longitudinal approach to conversion appears implicitly to challenge a more pragmatic notion that health and disease need to be separated by a diagnostically clearly identifiable boundary. This approach, therefore, introduces a tension between a processual understanding of the transition from health to disease and the discrete diagnostic categories through which this progression is made classifiable and medically interpretable.

It is worth noting, however, that, despite their differences, both the second and third approaches operate with the tacit approximation that the progression from health to AD

follows a relatively linear, one-way development via MCI. In both approaches, MCI subjects who experience one or more reversions of symptoms, or exhibit other less straightforward developments, are typically excluded from the ground-truthing comparative arrangements through which conversion is operationalized (Sun et al. 2017).[20] Consequently, although spontaneous remission, as mentioned earlier, is a clinical reality for a significant proportion of MCI subjects, it is treated as noise in these comparative arrangements, leading to an implicit homogenization of MCI that becomes inscribed into the algorithms.

The fourth approach, which is characterized by the most pronounced methodological differences within the approach, implicitly acknowledges a certain level of semantic overlap between different diagnostic categories as well as the limitations of simple binary comparisons by the very fact that it operationally combines multiple categories or multiple comparisons. Yet how the boundary between health and disease is configured differs considerably across the diverse methods and the particular comparative arrangements they set up, thus establishing different relations between healthy controls, AD patients, and MCI subjects. What the methods share, however, is that, through their reliance on comparative arrangements involving multiple categories or mutually interconnected comparisons, they create complex, interconnected relations across multiple diagnostic categories. Therefore, across the methods comprising the fourth approach, the boundary between disease and health is not defined in simple binary terms, as in the first approach. Yet, unlike the second and third approaches, this complexity does not arise from incorporating the temporal dimension into the comparison but from the interconnected relations that these more complex comparative arrangements establish among different diagnostic categories.

Finally, the fifth approach, by virtue of performing comparisons that combine MRI data with other neuroimaging modalities as well as non-imaging data—including genetic information and demographic data—shifts the negotiation of the boundary between healthy aging and future AD beyond the single data modality. The boundary is thus implicitly conceptualized in more complex terms, tacitly acknowledging the multifactorial nature of the disease, which is no longer diagnostically related to a single neurophysiological difference but to a complex network of mutually interacting factors, including genetic and demographic influences. Moreover, in this approach, a number of studies explicitly incorporate symptom measures such as cognitive test scores, thus aiming to link AD pathophysiology to clinical phenotypes as the basis for predictive AD diagnosis.

In sum, my analysis of these five approaches has aimed to demonstrate that current ML research on predictive AD diagnosis does not merely inherit the conceptual ambiguities surrounding AD and MCI that already exist in the medical literature (Schermer 2023). While these ambiguities provide the background conditions, the ways in which researchers organize comparisons during ground-truthing result in comparative arrangements that reconfigure the relations between healthy aging, MCI, and future dementia in novel, yet conceptually inconsistent ways. Depending on how comparisons are organized, the boundary between health and disease appears as a sharp biological distinction, a retrospectively constituted threshold, a gradual temporal transition, a product of multiple interconnected comparisons, or an effect emerging from the integration of heterogeneous data modalities. Consequently, rather than simply providing a straightforward technical

[20] Such exclusions, which remain without explicit justification in the published studies, might be motivated by the lack of data on individuals who experience remissions in open-access databases or because wayward trajectories are more challenging to model and interpret. For one recent exception, see Dang et al. (2023).

solution to the current clinical ambiguities of the AD diagnosis, the research on predictive ML-based diagnosis using neuroimaging data multiplies the ways in which the boundary between health and disease can be drawn, thereby introducing new conceptual inconsistencies into an already heterogeneous field (Daly 2025).

An important implication of these findings is that approaches that draw the boundaries between healthy aging and future dementia in different ways should not be assessed only through common performance metrics. While such metrics may indicate the technical success of a particular approach, they do not address the conceptual differences between the alternative ways in which the relations between healthy aging, MCI, and future dementia are configured during ground-truthing. This means that the researchers' competitive focus on producing predictive algorithms that achieve increasingly better results on quantitative performance metrics overlooks the epistemic contingencies arising from the conceptual inconsistencies in how the boundary between AD and healthy aging is defined during ground-truthing across different comparisons.

While this has no influence on how well an ML algorithm for predictive AD diagnosis performs the task for which it was designed, it does have implications for present-day and future medical understanding of AD. Whether the boundary between health and disease in this context is conceptualized in purely biological or clinicobiological terms, as well as whether it is understood in binary terms or as a dynamic continuum, should not be determined by which comparative arrangement happens to accidentally produce the highest predictive accuracy. My intention is not to argue that any of the current approaches, or the implicit boundary between health and disease inscribed into them, is inherently either better or worse. Instead, I argue that questions concerning how the boundary between health and disease in AD ought to be defined should first be addressed explicitly in conceptual terms. The design of ground-truthing procedures should then be developed in accordance with such decisions, rather than allowing these questions to be settled implicitly through the selection of whichever comparative arrangement happens to achieve the highest predictive accuracy.

6. Conclusion

This paper has examined current research on ML algorithms for the predictive diagnosis of AD before the clinical onset of dementia using neuroimaging data. By moving beyond quantitative performance metrics of algorithms as finished products, I have attended to the human decisions that shape the ground-truthing operations through which predictive ML algorithms are developed. Importantly, I have not treated ground-truthing as a purely technical procedure for training and evaluating predictive models. Instead, I approached it as a comparative practice through which relations between healthy aging, MCI, and future dementia are established at the level of neuroimaging data with the aim of constructing predictive diagnostic separability between healthy aging and future dementia.

Based on a close reading of 36 studies published between 2010 and 2025, I analyzed researchers' decisions about how to organize comparisons during ground-truthing. In doing so, I identified five coexisting approaches that differ in how they define the classificatory task to be learned and in how they relate different diagnostic categories to one another. Researchers employ the different classificatory comparisons pragmatically, exploring which ones deliver better performance metrics with a particular ML model. However, I have

argued that, depending on how comparisons are organized within each approach, these approaches configure the boundary between health and disease in conceptually inconsistent ways: as a sharply delineated biological distinction or as a multifactorial phenomenon; as a binary division between health and disease or as a gradual transition between them; as a relation constituted retrospectively through diagnostic outcomes at follow-ups or through dynamically modeled trajectories over time. In this sense, current ML research on predictive AD diagnosis does not merely inherit existing conceptual ambiguities that dominate the broader medical literature on AD (Daly 2025; Daly and Keuck 2024; Schermer 2023) but contributes to their further multiplication.

The significance of this finding lies in the fact that these different comparative arrangements are primarily evaluated through performance metrics such as predictive accuracy. Yet accuracy can establish whether an algorithm successfully performs the task for which it was designed, not whether the underlying configuration of the boundary between healthy aging and future dementia is conceptually preferable, medically more appropriate, or ethically justifiable. In fact, predictive algorithms with similar performance metrics may draw very different boundaries, depending on the comparisons that were used during ground-truthing. Consequently, questions concerning how the boundary between health and disease in AD ought to be drawn remain implicit within researchers' decisions about how ground-truth comparisons are organized. Recognizing these decisions as epistemically consequential does not mean that predictive ML research on AD should be dismissed nor that its promissory potential to offer new ways of dealing with long-standing diagnostic challenges should be negated. It does, however, challenge the framing of this research in opposition to the purportedly "subjective" expert-based clinical diagnosis, thus presenting it as a straightforward technological fix. It also suggests that discussions about the conceptual foundations of predictive diagnosis deserve at least as much attention as the discussions about algorithmic performance, as the former shape the biomedical understanding of AD currently being produced in the research context.

Acknowledgments

Earlier versions of this paper were presented at the conference Digital Futures in the Making: Imaginaries, Politics, and Materialities (University of Hamburg, 2022), the conference Preventia: Optimising Bios Through Technologies (Hanse-Wissenschaftskolleg in Delmenhorst, 2022), the 2023 virtual workshop of the Sociology of Science and Technology Network (SSTNET), and the lecture series of the Phenomenology of Dementia and Aging Network (University of Cologne, 2025). I thank the organizers and participants for their comments and discussions. I am particularly grateful to Silke Schicktanz, Ayo Wahlberg, Melissa Park, Klaus Høyer, Erik Norman Dzwiza-Ohlsen, Ragna Winniewski, Janna Hastings, and Nicolas Baya-Laffite for their valuable feedback. I also thank the three anonymous reviewers for their constructive and insightful comments, and Alison Lockhart for her exceptional assistance throughout the editorial process. I am especially grateful to Annette Leibing for her generous support.

Disclosure Statement

No competing interest was reported by the author.

References
a) Empirical Sample

Agostinho, Daniel, Marco Simões, Miguel Castelo-Branco, and Alzheimer's Disease Neuroimaging Initiative (ADNI). 2024. "Predicting Conversion from Mild Cognitive Impairment to Alzheimer's Disease: A Multimodal Approach." *Brain Communications* 6, no. 4, fcae208. https://doi.org/10.1093/braincomms/fcae208.

Aguilar, Carlos, Eric Westman, J-Sebastian Muehlboeck, et al. 2013. "Different Multivariate Techniques for Automated Classification of MRI Data in Alzheimer's Disease and Mild Cognitive Impairment." *Psychiatry Research: Neuroimaging* 212, no. 2: 89–98. https://doi.org/10.1016/j.pscychresns.2012.11.005.

Bae, Jinhyeong, Jane Stocks, Ashley Heywood, et al. 2021. "Transfer Learning for Predicting Conversion from Mild Cognitive Impairment to Dementia of Alzheimer's Type Based on a Three-Dimensional Convolutional Neural Network." *Neurobiology of Aging* 99: 53–64. https://doi.org/10.1016/j.neurobiolaging.2020.12.005.

Bron, Esther E., Marion Smits, Wiro J. Niessen, and Stefan Klein. 2015. "Feature Selection Based on the SVM Weight Vector for Classification of Dementia." *IEEE Journal of Biomedical and Health Informatics* 19, no. 5: 1617–1626. https://doi.org/10.1109/JBHI.2015.2432832.

Casanova, Ramon, Ryan T. Barnard, Sarah A. Gaussoin, et al. 2018. "Using High-Dimensional Machine Learning Methods to Estimate an Anatomical Risk Factor for Alzheimer's Disease Across Imaging Databases." *NeuroImage* 183: 401–411. https://doi.org/10.1016/j.neuroimage.2018.08.040.

Cheng, Bo, Mingxi Liu, Daoqiang Zhang, Brent C. Munsell, and Dinggang Shen. 2015. "Domain Transfer Learning for MCI Conversion Prediction." *IEEE Transactions on Bio-Medical Engineering* 62, no. 7: 1805–1817. https://doi.org/10.1109/TBME.2015.2404809.

Costafreda, Sergi G., Ivo D. Dinov, Zhuowen Tu, et al. 2011. "Automated Hippocampal Shape Analysis Predicts the Onset of Dementia in Mild Cognitive Impairment." *NeuroImage* 56, no. 1: 212–219. https://doi.org/10.1016/j.neuroimage.2011.01.050.

Cuingnet, Rémi, Emilie Gerardin, Jérôme Tessieras, et al. 2011. "Automatic Classification of Patients with Alzheimer's Disease from Structural MRI: A Comparison of Ten Methods Using the ADNI Database." *NeuroImage* 56, no. 2: 766–781. https://doi.org/10.1016/j.neuroimage.2010.06.013.

Duchesne, Simon, and Abderazzak Mouiha. 2011. "Morphological Factor Estimation via High-Dimensional Reduction: Prediction of MCI Conversion to Probable AD." *International Journal of Alzheimer's Disease,* 914085. https://doi.org/10.4061/2011/914085.

Ezzati, Ali, Andrea R. Zammit, Danielle J. Harvey, et al. 2019. "Optimizing Machine Learning Methods to Improve Predictive Models of Alzheimer's Disease." *Journal of Alzheimer's Disease* 71, no. 3: 1027–1036. https://doi.org/10.3233/jad-190262.

Franciotti, Raffaella, Davide Nardini, Mirella Russo, et al. 2023. "Comparison of Machine Learning-Based Approaches to Predict the Conversion to Alzheimer's Disease from Mild Cognitive Impairment." *Neuroscience* 514: 143–152. https://doi.org/10.1016/j.neuroscience.2023.01.029.

Ghaffari, Hamed, Hassan Tavakoli, and Gila Pirzad Jahromi. 2022. "Deep Transfer Learning–Based Fully Automated Detection and Classification of Alzheimer's Disease on Brain MRI." *British Journal of Radiology* 95, no. 1136, 20211253. https://doi.org/10.1259/bjr.20211253.

Gupta, Yubraj, Ramesh K. Lama, Goo-Rak Kwon, and Alzheimer's Disease Neuroimaging Initiative (ADNI). 2019. "Prediction and Classification of Alzheimer's Disease Based on Combined Features from Apolipoprotein-E Genotype, Cerebrospinal Fluid, MR, and FDG-PET Imaging Biomarkers." *Frontiers in Computational Neuroscience* 13, 72. https://doi.org/10.3389/fncom.2019.00072.

Hoang, Gia M., Ue-Hwan Kim, and Jae G. Kim. 2023. "Vision Transformers for the Prediction of Mild Cognitive Impairment to Alzheimer's Disease Progression Using Mid-Sagittal sMRI." *Frontiers in Aging Neuroscience* 15, 1102869. https://doi.org/10.3389/fnagi.2023.1102869.

Hojjati, Seyed H., Ata Ebrahimzadeh, Ali Khazaee, Abbas Babajani-Feremi, and Alzheimer's Disease Neuroimaging Initiative (ADNI). 2018. "Predicting Conversion from MCI to AD by Integrating rs-fMRI and Structural MRI." *Computers in Biology and Medicine* 102: 30–39. https://doi.org/10.1016/j.compbiomed.2018.09.004.

Khojaste-Sarakhsi, M., Seyedhamidreza S. Haghighi, and S.M.T. Fatemi Ghomi. 2025. "A 3D Multi-Modal Multi-Scale End-to-End Classifier for Alzheimer's Disease Diagnosis." *Medical Engineering & Physics* 143, no. 1, 104382. https://doi.org/10.1016/j.medengphy.2025.104382.

Lebedev, A.V., E. Westman, G.J.P. Van Westen, et al. 2014. "Random Forest Ensembles for Detection and Prediction of Alzheimer's Disease with a Good Between-Cohort Robustness." *NeuroImage: Clinical* 6: 115–25. https://doi.org/10.1016/j.nicl.2014.08.023.

Li, Yuhan, Donghao Niu, Keying Qi, Dong Liang, and Xiaojing Long. 2025 "An Imaging and Genetic-Based Deep Learning Network for Alzheimer's Disease Diagnosis." *Frontiers in Aging Neuroscience* 17, 1532470. https://doi.org/10.3389/fnagi.2025.1532470.

Lin, Yanyan, Kexin Huang, Hanxiao Xu, et al. 2020. "Predicting the Progression of Mild Cognitive Impairment to Alzheimer's Disease by Longitudinal Magnetic Resonance Imaging-Based Dictionary Learning." *Clinical Neurophysiology* 131, no. 10: 2429–2439. https://doi.org/10.1016/j.clinph.2020.07.016.

Liu, Mingxia, Jun Zhang, Ehsan Adeli, and Dinggang Shen. 2019. "Joint Classification and Regression via Deep Multi-Task Multi-Channel Learning for Alzheimer's Disease Diagnosis." *IEEE Transactions on Biomedical Engineering* 66, no. 5, 1195–1206. https://doi.org/10.1109/tbme.2018.2869989.

Lu, Donghuan, Karteek Popuri, Gavin Weiguang Ding, Rakesh Balachandar, Mirza Faisal Beg, and Alzheimer's Disease Neuroimaging Initiative (ADNI). 2018. "Multimodal and Multiscale Deep Neural Networks for the Early Diagnosis of Alzheimer's Disease Using Structural MR and FDG-PET Images." *Scientific Reports* 8, 5697. https://doi.org/10.1038/s41598-018-22871-z.

Mofrad, Samaneh A., Arvid Lundervold, and Alexander S. Lundervold. 2021. "A Predictive Framework Based on Brain Volume Trajectories Enabling Early Detection of Alzheimer's Disease." *Computerized Medical Imaging and Graphics* 90, 101910. https://doi.org/10.1016/j.compmedimag.2021.101910.

Mofrad, Samaneh A., Astri J. Lundervold, Alexandra Vik, and Alexander S. Lundervold. 2021. "Cognitive and MRI Trajectories for Prediction of Alzheimer's Disease." *Scientific Reports* 11, 2122. https://doi.org/10.1038/s41598-020-78095-7.

Moscoso, Alexis, Jesús Silva-Rodríguez, Jose M. Aldrey, et al. 2019. "Prediction of Alzheimer's Disease Dementia with MRI Beyond the Short-Term: Implications for the Design of Predictive Models." *NeuroImage: Clinical* 23, 101837. https://doi.org/10.1016/j.nicl.2019.101837.

Nanni, Loris, Matteo Interlenghi, Sheryl Brahnam, et al. 2020. "Comparison of Transfer Learning and Conventional Machine Learning Applied to Structural Brain MRI for the Early Diagnosis and

Prognosis of Alzheimer's Disease." *Frontiers in Neurology* 11, 576194. https://doi.org/10.3389/fneur.2020.576194.

Ning, Kaida, Bo Chen, Fengzhu Sun, et al. 2018. "Classifying Alzheimer's Disease with Brain Imaging and Genetic Data Using a Neural Network Framework." *Neurobiology of Aging* 68: 151–158. https://doi.org/10.1016/j.neurobiolaging.2018.04.009.

Plant, Claudia, Stefan J. Teipel, Annahita Oswald, et al. 2010. "Automated Detection of Brain Atrophy Patterns Based on MRI for the Prediction of Alzheimer's Disease." *NeuroImage* 50, no. 1: 162–174. https://doi.org/10.1016/j.neuroimage.2009.11.046.

Ren, Fujia, Chenhui Yang, and Y.A. Nanehkaran. 2023. "MRI-Based Model for MCI Conversion Using Deep Zero-Shot Transfer Learning." *Journal of Supercomputing* 79: 1182–1200. https://doi.org/10.1007/s11227-022-04668-0.

Spasov, Simeon, Luca Passamonti, Andrea Duggento, Pietro Liò, Nicola Toschi, and Alzheimer's Disease Neuroimaging Initiative (ADNI). 2019. "A Parameter-Efficient Deep Learning Approach to Predict Conversion from Mild Cognitive Impairment to Alzheimer's Disease." *NeuroImage* 189: 276–287. https://doi.org/10.1016/j.neuroimage.2019.01.031.

Suk, Heung-Il, Seong-Whan Lee, Dinggang Shen, and Alzheimer's Disease Neuroimaging Initiative (ADNI). 2016. "Deep Sparse Multi-Task Learning for Feature Selection in Alzheimer's Disease Diagnosis." *Brain Structure and Function* 221, no. 5: 2569–2587. https://doi.org/10.1007/s00429-015-1059-y.

Sun, Zhuo, Martijn van de Giessen, Boudewijn P. Lelieveldt, and Marius Staring. 2017. "Detection of Conversion from Mild Cognitive Impairment to Alzheimer's Disease Using Longitudinal Brain MRI." *Frontiers in Neuroinformatics* 11, 16. https://doi.org/10.3389/fninf.2017.00016.

Tong, Tong, Qinquan Gao, Ricardo Guerrero, et al. 2017. "A Novel Grading Biomarker for the Prediction of Conversion from Mild Cognitive Impairment to Alzheimer's Disease." *IEEE Transactions on Biomedical Engineering* 64, no. 1: 155–165. https://doi.org/10.1109/tbme.2016.2549363.

Xi, Yang, Qian Wang, Chenxue Wu, Lu Zhang, Ying Chen, and Zhu Lan. 2025. "Predicting Conversion of Alzheimer's Disease Based on Multi-Modal Fusion of Neuroimaging and Genetic Data." *Complex & Intelligent Systems* 11, 58. https://doi.org/10.1007/s40747-024-01680-0.

Yu, Qi, Qian Ma, Lijuan Da, et al. 2024. "A Transformer-Based Unified Multimodal Framework for Alzheimer's Disease Assessment." *Computers in Biology and Medicine* 180, 108979. https://doi.org/10.1016/j.compbiomed.2024.108979.

Zhang, Daoqiang, Dinggang Shen, and Alzheimer's Disease Neuroimaging Initiative (ADNI). 2012. "Predicting Future Clinical Changes of MCI Patients Using Longitudinal and Multimodal Biomarkers." *PLoS ONE* 7, e33182. https://doi.org/10.1371/journal.pone.0033182.

Zheng, Guojiang, Yang Lu, and Hui Chen. 2025. "Deep Learning-Based Framework for Predicting Mild Cognitive Impairment Progression in Neurology Using Longitudinal MRI." *IEEE Access* 13, 68903–68919. https://doi.org/10.1109/ACCESS.2025.3562432.

b) General References

Alpaydin, Ethem. 2016. *Machine Learning: The New AI*. MIT Press.

Alzheimer's Association. 2024. "2024 Alzheimer's Disease Facts and Figures." *Alzheimer's & Dementia* 20, no. 5: 3708–3821. https://doi.org/10.1002/alz.13809.

Anderson, Nicole D. 2019. "State of the Science on Mild Cognitive Impairment (MCI). *CNS Spectrums* 24, no. 1: 78–87. https://doi.org/10.1017/S1092852918001347.

Arbabshirani, Mohammad R., Sergey Plis, Jing Sui, and Vince D. Calhoun. 2017. "Single Subject Prediction of Brain Disorders in Neuroimaging: Promises and Pitfalls." *NeuroImage* 145, Part B: 137–165. https://doi.org/10.1016/j.neuroimage.2016.02.079.

Blasimme, Alessandro. 2021. "The Plasticity of Ageing and the Rediscovery of Ground-State Prevention." *History and Philosophy of the Life Sciences* 43, no. 2, 67. https://doi.org/10.1007/s40656-021-00414-6.

Blasimme, Alessandro, Marco Canevelli, and Matteo Cesari. 2021. "Shifting the Focus of Dementia Prevention: Ethical Considerations." *AJOB Neuroscience* 12, no. 4: 240–242. https://doi.org/10.1080/21507740.2021.1941405.

Boenink, Marianne, Harro van Lente, and Ellen Moors, eds. 2016. *Emerging Technologies for Diagnosing Alzheimer's Disease: Innovating with Care*. Palgrave Macmillan.

Borchert, Robin J., Tiago Azevedo, AmanPreet Badhwar, et al. 2023. "Artificial Intelligence for Diagnostic and Prognostic Neuroimaging in Dementia: A Systematic Review." *Alzheimer's & Dementia* 19, no. 12: 5885–5904. https://doi.org/10.1002/alz.13412.

Bradfield, Nicholas I. 2023. "Mild Cognitive Impairment: Diagnosis and Subtypes." *Clinical EEG and Neuroscience* 54, no. 1: 4–11. https://doi.org/10.1177/15500594211042708.

Canevelli, Marco, Giulia Grande, Eleonora Lacorte, et al. 2016. "Spontaneous Reversion of Mild Cognitive Impairment to Normal Cognition: A Systematic Review of Literature and Meta-Analysis." *Journal of the American Medical Directors Association* 17, no. 10: 943–948. https://doi.org/10.1016/j.jamda.2016.06.020.

Cardona, Beatriz. 2008. "'Healthy Ageing' Policies and Anti-Ageing Ideologies and Practices: On the Exercise of Responsibility." *Medicine, Health Care and Philosophy* 11, no. 4: 475–483. https://doi.org/10.1007/s11019-008-9129-z.

Daly, Timothy. 2024. "Dementia Prevention Guidelines Should Explicitly Mention Deprivation." *AJOB Neuroscience* 15, no. 1: 73–76. https://doi.org/10.1080/21507740.2023.2225461.

———. 2025. "The Meanings of Alzheimer's Disease." *Philosophy of Medicine* 6, no. 1: 1–11. https://doi.org/10.5195/pom.2025.251.

Daly, Timothy and Lara Keuck. 2024. "Alzheimer's Disease: Engaging with an Unstable Category." In *Handbook of the Philosophy of Medicine,* edited by Thomas Schramme and Mary Walker. Springer. https://doi.org/10.1007/978-94-017-8706-2_113-1.

Daly, Timothy, and Ignacio Mastroleo. 2024. "Reducing Confusion Surrounding Expert Conceptions of Alzheimer's and Dementia: A Practical Analysis." *Journal of Neuropsychology* 19, no. 2: 158–164. https://doi.org/10.1111/jnp.12398.

Dang, Mingxi, Caishui Yang, Kewei Chen, Peng Lu, He Li, and Zhanjun Zhang. 2023. "Hippocampus-Centred Grey Matter Covariance Networks Predict the Development and Reversion of Mild Cognitive Impairment." *Alzheimer's Research & Therapy* 15, 27. https://doi.org/10.1186/s13195-023-01167-z.

De Winter, Gunnar. 2015. "Aging as Disease." *Medicine, Health Care and Philosophy* 18, no. 2: 237–243. https://doi.org/10.1007/s11019-014-9600-y.

Domínguez-Fernández, Celtia, June Egiguren-Ortiz, Jone Razquin, et al. 2023. "Review of Technological Challenges in Personalised Medicine and Early Diagnosis of Neurodegenerative Disorders." *International Journal of Molecular Sciences* 24, no. 4, 3321. https://doi.org/10.3390/ijms24043321.

Doran, Sarah J., and Russell P. Sawyer. 2024. "Risk Factors in Developing Amyloid Related Imaging Abnormalities (ARIA) and Clinical Implications." *Frontiers in Neuroscience* 18, 1326784. https://doi.org/10.3389/fnins.2024.1326784.

Dubois, Bruno, Nicolas Villain, Giovanni B. Frisoni, et al. 2021. "Clinical Diagnosis of Alzheimer's Disease: Recommendations of the International Working Group." *Lancet Neurology* 20, no. 6: 484–496. https://doi.org/10.1016/S1474-4422(21)00066-1.

Dubois, Bruno, Nicolas Villain, Lon Schneider, et al. 2024. "Alzheimer Disease as a Clinical-Biological Construct: An International Working Group Recommendation." *JAMA Neurology* 81, no. 12: 1304–1311. https://doi.org/10.1001/jamaneurol.2024.3770.

Epple, Angelika, and Walter Erhart. 2020. "Practices of Comparing: A New Research Agenda Between Typological and Historical Approaches". In *Practices of Comparing,* edited by Angelika Epple, Walter Erhart and Johannes Grave. Transcript Verlag. https://doi.org/10.14361/9783839451663-002.

Ford, Elizabeth, Richard Milne, and Keegan Curlewis. 2023. "Ethical Issues When Using Digital Biomarkers and Artificial Intelligence for the Early Detection of Dementia." *WIREs Data Mining and Knowledge Discovery* 13, no. 3, e1492. https://doi.org/10.1002/widm.1492.

Frizzell, Tory O., Margit Glashutter, Careesa C. Liu, et al. 2022. "Artificial Intelligence in Brain MRI Analysis of Alzheimer's Disease over the Past 12 Years: A Systematic Review." *Ageing Research Reviews* 77, 101614. https://doi.org/10.1016/j.arr.2022.101614.

Giroux, Élodie. 2024. "Risk and Disease: Two Alternative Ways of Modelling Health Phenomena." In *A Pragmatic Approach to Conceptualization of Health and Disease*, edited by Maartje Schermer and Nicholas Binney. Springer. https://doi.org/10.1007/978-3-031-62241-0_13.

Golomb, James, Alan Kluger and Steven H. Ferris. 2004. "Mild Cognitive Impairment: Historical Development and Summary of Research." *Dialogues in Clinical Neuroscience* 6: 351–367. https://doi.org/10.31887/DCNS.2004.6.4/jgolomb.

Green, Sara, and Line Hillersdal. 2021. "Aging Biomarkers and the Measurement of Health and Risk." *History and Philosophy of the Life Sciences* 43, no. 1, 28. https://doi.org/10.1007/s40656-021-00367-w.

Grueso, Sergio, and Raquel Viejo-Sobera. 2021. "Machine Learning Methods for Predicting Progression from Mild Cognitive Impairment to Alzheimer's Disease Dementia: A Systematic Review." *Alzheimer's Research & Therapy* 13, no. 1, 162. https://doi.org/10.1186/s13195-021-00900-w.

Gunnarson, Martin, Alexandra Kapeller, and Kristin Zeiler. 2021. "Ethico-Political Aspects of Conceptualizing Screening: The Case of Dementia." *Health Care Analysis* 29, no. 4: 343–359. https://doi.org/10.1007/s10728-021-00431-3.

Heintz, Bettina. 2016. "Wir leben im Zeitalter der Vergleichung: Perspektiven einer Soziologie des Vergleichs." *Zeitschrift für Soziologie* 45, no. 5: 305–323. https://doi.org/10.1515/zfsoz-2015-1018.

Henriksen, Anne, and Anja Bechmann. 2020. "Building Truths in AI: Making Predictive Algorithms Doable in Healthcare." *Information, Communication & Society* 23, no. 6: 802–816. https://doi.org/10.1080/1369118x.2020.1751866.

Henriksen, Anne, and Finn Olesen. 2021. "Experimenting on the Enactment of Predictive AI: The Quest for a Future Proactive Healthcare Sector." *STS Encounters* 12, no. 1. https://doi.org/10.7146/stse.v12i1.135404.

Jack, Clifford R., J.S. Andrews, Thomas G. Beach, et al. 2024. "Revised Criteria for Diagnosis and Staging of Alzheimer's Disease: Alzheimer's Association Workgroup." *Alzheimer's & Dementia* 20, no. 8: 5143–5169. https://doi.org/10.1002/alz.13859.

Jack, Clifford R., David A. Bennett, Kaj Blennow, et al. 2016. "A/T/N: An Unbiased Descriptive Classification Scheme for Alzheimer Disease Biomarkers." *Neurology* 87, no. 5: 539–547. https://doi.org/10.1212/wnl.0000000000002923.

Janeiro, Manuel H., Carlos G. Ardanaz, Noemí Sola-Sevilla et al. 2020. "Biomarkers in Alzheimer's Disease." *Advances in Laboratory Medicine* 2, no. 1: 27–50. https://doi.org/10.1515/almed-2020-0090.

Jaton, Florian. 2021. "Assessing Biases, Relaxing Moralism: On Ground-Truthing Practices in Machine Learning Design and Application." *Big Data & Society* 8, no. 1: 1–15. https://doi.org/10.1177/20539517211013569.

———. 2023. "Groundwork for AI: Enforcing a Benchmark for Neoantigen Prediction in Personalized Cancer Immunotherapy.'" *Social Studies of Science* 53, no. 5: 787–810. https://doi.org/10.1177/03063127231192857.

Jaton, Florian, and Philippe Sormani. 2023. "Enabling 'AI'? The Situated Production of Commensurabilities." *Social Studies of Science* 53, no. 5: 625–634. https://doi.org/10.1177/03063127231194591.

Kang, Connie. 2024. "Donanemab: First Approval." *Drugs* 84: 1313–1318. https://doi.org/10.1007/s40265-024-02087-4.

Keuck, Lara. 2021. "A Window to Act? Revisiting the Conceptual Foundations of Alzheimer's Disease in Dementia Prevention." In *Preventing Dementia? Critical Perspectives on a New Paradigm of Preparing for Old Age,* edited by Annette Leibing and Silke Schicktanz. Berghahn Books.

Koepsell, Thomas D., and Sarah E. Monsell. 2012. "Reversion from Mild Cognitive Impairment to Normal or Near-Normal Cognition." *Neurology* 79, no. 15: 1591–1598. https://doi.org/10.1212/wnl.0b013e31826e26b7.

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

Lee, Francis. 2025. "The Practices and Politics of Machine Learning: A Field Guide for Analyzing Artificial Intelligence." *AI & Society* 40, no. 8: 6135–6148. https://doi.org/10.1007/s00146-025-02430-7.

Leibing, Annette. 2014. "The Earlier the Better: Alzheimer's Prevention, Early Detection, and the Quest for Pharmacological Interventions." *Culture, Medicine and Psychiatry* 38, no. 2: 217–236. https://doi.org/10.1007/s11013-014-9370-2.

———. 2018. "Situated Prevention: Framing the 'New Dementia'." *Journal of Law, Medicine & Ethics* 46, no. 3: 704–716. https://doi.org/10.1177/1073110518804232.

Leibing, Annette, and Silke Schicktanz, eds. 2021. *Preventing Dementia? Critical Perspectives on a New Paradigm of Preparing for Old Age*. Berghahn Books.

Leibing, Annette, Silke Schicktanz, and Alessandro Blasimme. 2025. "Social, Ethical, and Epistemological Aspects of Dementia Prevention: The Three-Country BEAD Study." *Geriatrics Gerontology and Aging* 19, e0000339. https://doi.org/10.53886/gga.e0000339_en.

Li, Renjie, Xinyi Wang, Katherine Lawler, Saurabh Garg, Quan Bai, and Jane Alty. 2022. "Applications of Artificial Intelligence to Aid Early Detection of Dementia: A Scoping Review on Current Capabilities and Future Directions." *Journal of Biomedical Informatics* 127, 104030. https://doi.org/10.1016/j.jbi.2022.104030.

Livingston, Gill, Jonathan Huntley, Kathy Y. Liu, et al. 2024. "Dementia Prevention, Intervention, and Care: 2024 Report of the Lancet standing Commission." *The Lancet* 404, no. 10452: 572–628. https://doi.org/10.1016/S0140-6736(24)01296-0.

Lloret, Ana, Daniel Esteve, Maria-Angeles Lloret, et al. 2019, "When Does Alzheimer's Disease Really Start? The Role of Biomarkers." *International Journal of Molecular Sciences* 20, no. 22, 5536. https://doi.org/10.3390/ijms20225536.

Mackenzie, Adrian. 2015. "The Production of Prediction: What Does Machine Learning Want?" *European Journal of Cultural Studies* 18, no. 4–5: 429–445. https://doi.org/10.1177/1367549415577384.

Mitchell, A.J. and M. Shiri-Feshki. 2009. "Rate of Progression of Mild Cognitive Impairment to Dementia: Meta-Analysis of 41 Robust Inception Cohort Studies." *Acta Psychiatrica Scandinavica* 119: 252–265. https://doi.org/10.1111/j.1600-0447.2008.01326.x.

Mol, Annemarie. 2002. *The Body Multiple: Ontology in Medical Practice*. Duke University Press.

Muhr, Paula. 2023. "The Unobserved Anatomy: Negotiating the Plausibility of AI-Based Reconstructions of Missing Brain Structures in Clinical MRI Scans." In *Plausibilisierung und Evidenz: Dynamiken und Praktiken von der Antike bis zur Gegenwart,* edited by Antje Flüchter, Birte Förster, Britta Hochkirchen, and Silke Schwandt. Bielefeld University Press.

———. 2025. "Search for Elusive Neuroimaging Biomarkers: Machine Learning, Resting-State fMRI, and the Reconfiguration of Diagnosis in Functional Neurological Disorder." *Digital Society* 4, no. 3, 83. https://doi.org/10.1007/s44206-025-00238-y.

Nichols, Emma, Jaimie D. Steinmetz, Stein Emil Vollset, et al. 2022. "Estimation of the Global Prevalence of Dementia in 2019 and Forecasted Prevalence in 2050: An Analysis for the Global Burden of Disease Study 2019." *The Lancet: Public Health* 7, no. 2: e105–e125. https://doi.org/10.1016/S2468-2667(21)00249-8.

Pandya, Seema Y., Matthew A. Clem, Lynette M. Silva, and Fu L. Woon. 2016. "Does Mild Cognitive Impairment Always Lead to Dementia? A Review." *Journal of the Neurological Sciences* 369: 57–62. https://doi.org/10.1016/j.jns.2016.07.055.

Pellegrini, Enrico, Lucia Ballerini, Maria D. Hernandez, et al. 2018. "Machine Learning of Neuroimaging for Assisted Diagnosis of Cognitive Impairment and Dementia: A Systematic Review." *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* 10, no. 1: 519–535. https://doi.org/10.1016/j.dadm.2018.07.004.

Petersen, Niklas, Mattia Andreoletti, Alessandro Blasimme, Cynthia Lazzaroni, Annette Leibing, and Silke Schicktanz. 2025. "Disentangling Responsibility: Perspectives on Dementia Prevention from Stakeholders in Canada, Germany and Switzerland." *Sociology of Health & Illness* 48, no. 1, e70126. https://doi.org/10.1111/1467-9566.70126.

Petersen, Ronald C. 2016. “Mild Cognitive Impairment.” *Continuum* 22, no. 2: 404–418. https://doi.org/10.1212/CON.0000000000000313.

Petersen, Ronald C., B. Caracciolo, C. Brayne, S. Gauthier, V. Jelic, and L. Fratiglioni. 2014. “Mild Cognitive Impairment: A Concept in Evolution.” *Journal of Internal Medicine* 275, no. 3: 214–228. https://doi.org/10.1111/joim.12190.

Pfaller, Larissa, and Mark Schweda. 2019. “Excluded from the Good Life? An Ethical Approach to Conceptions of Active Ageing.” *Social Inclusion* 7, no. 3, 44–53. https://doi.org/10.17645/si.v7i3.1918.

Pike, Kerryn and Glynda Kinsella. 2019. “Alzheimer’s Disease: Prodromal Stages and Dementia.” In *Degenerative Disorders of the Brain,* edited by Darren Hocking, John Bradshaw and Joanne Fielding. Routledge.

Porsteinsson, A.P., R.S. Isaacson, S. Knox, M.N. Sabbagh, and I. Rubino. 2021. “Diagnosis of Early Alzheimer’s Disease: Clinical Practice in 2021.” *Journal of Prevention of Alzheimer’s Disease* 8, no. 3: 371–386. https://doi.org/10.14283/jpad.2021.23.

Saborido, Cristian, and Pablo García-Barranquero. 2022. “Is Aging a Disease? The Theoretical Definition of Aging in the Light of the Philosophy of Medicine.” *The Journal of Medicine and Philosophy* 47, no. 6: 770–783. https://doi.org/10.1093/jmp/jhac030.

Schermer, Maartje H. 2023. “Preclinical Disease or Risk Factor? Alzheimer’s Disease as a Case Study of Changing Conceptualizations of Disease.” *The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine* 48, no. 4: 322–334. https://doi.org/10.1093/jmp/jhad009.

Schermer, Maartje, and Nicholas Binney, eds. 2024. *A Pragmatic Approach to Conceptualization of Health and Disease*. Springer. https://doi.org/10.1007/978-3-031-62241-0.

Segato, Alice, Aldo Marzullo, Francesco Calimeri, and Elena De Momi. 2020. “Artificial Intelligence for Brain Diseases: A Systematic Review.” *APL Bioengineering* 4, no. 4, 041503. https://doi.org/10.1063/5.0011697.

Singh, Soraisam G., Dulumani Das, Utpal Barman, and Manob J. Saikia. 2024. “Early Alzheimer’s Disease Detection: A Review of Machine Learning Techniques for Forecasting Transition from Mild Cognitive Impairment.” *Diagnostics* 14, no. 16, 1759. https://doi.org/10.3390/diagnostics14161759.

Suemoto, Claudia K., and Renata E.P. Leite. 2023. “Autopsy Studies Are Key to Identifying Dementia Cause.” *Lancet Healthy Longevity* 4, no. 3: e94–e95. https://doi.org/10.1016/S2666-7568(23)00022-3.

Swallow, Julia. 2017. “Expectant Futures and an Early Diagnosis of Alzheimer’s Disease: Knowing and Its Consequences.” *Social Science & Medicine* 184: 57–64. https://doi.org/10.1016/j.socscimed.2017.05.017.

Timmermans, Stefan, and Mara Buchbinder. 2010. “Patients-in-Waiting: Living Between Sickness and Health in the Genomics Era.” *Journal of Health and Social Behavior* 51, no. 4: 408–423. https://doi.org/10.1177/0022146510386794.

Tinland, Julia. 2022. “Personalised Prevention: Increasing or Decreasing Over-Medicalisation, Overdiagnosis and Overtreatment?” In *Personalized Medicine in the Making: Philosophical Perspectives from Biology to Healthcare,* edited by Chiara Beneduce and Marta Bertolaso. Springer.

Valizadeh, Gelareh, Reza Elahi, Zahra Hasankhani, Hamidreza S. Rad, and Ahmad Shalbaf. 2025. "Deep Learning Approaches for Early Prediction of Conversion from MCI to AD Using MRI and Clinical Data: A Systematic Review." *Archives of Computational Methods in Engineering* 32, no. 2: 1229–1298. https://doi.org/10.1007/s11831-024-10176-6.

Van der Molen, Lennart H., Marianne Boenink, Harro van Lente, and Edo Richard. 2025. "Changing Definitions of Disease: Transformations in the Diagnostic Criteria for Alzheimer's Disease." *Alzheimer's & Dementia* 21, no. 4, e70133. https://doi.org/10.1002/alz.70133.

Vermeulen, Robin J., Vebjørn Andersson, Jimmy Banken, et al. 2025. "Limited Generalizability and High Risk of Bias in Multivariable Models Predicting Conversion Risk from Mild Cognitive Impairment to Dementia: A Systematic Review." *Alzheimer's & Dementia* 21, no. 4, e70069. https://doi.org/10.1002/alz.70069.

Weiner, Michael, Dallas Veitch, Paul Aisen, et al. 2015. "Impact of the Alzheimer's Disease Neuroimaging Initiative, 2004 to 2014." *Alzheimer's & Dementia* 11, no. 7: 865–884. https://doi.org/10.1016/j.jalz.2015.04.005.

Whitehouse, Peter J., and Daniel R. George. 2021. "Afterword. Looking Forward." In *Preventing Dementia? Critical Perspectives on a New Paradigm of Preparing for Old Age*, edited by Annette Leibing and Silke Schicktanz. Berghahn Books.

Winter, Peter, and Annamaria Carusi. 2022. "'If You're Going to Trust the Machine, Then That Trust Has Got to Be Based on Something': Validation and the Co-Constitution of Trust in Developing Artificial Intelligence (AI) for the Early Diagnosis of Pulmonary Hypertension (PH)." *Science & Technology Studies* 35, no. 4: 58–77. https://doi.org/10.23987/sts.102198.

Wirth, Assia, Clément L. Ludec, Camille Girard-Chanudet, and Maxime Cornet. 2025. "Who Decides When Artificial Intelligence Works? Algorithmic Production as a Continuous Evaluation Process." *Digital Society* 4, no. 3, 65. https://doi.org/10.1007/s44206-025-00224-4.

Zeiler, Kristin, A. Segernäs, and M. Gunnarson. 2023."Entering the Grey Zone of Aging Between Health and Disease: A Critical Phenomenological Account." *Continental Philosophy Review* 57, no. 4: 659–676. https://doi.org/10.1007/s11007-023-09618-y.

Zhao, Zhen, Joon H. Chuah, Khin W. Lai, et al. 2023. "Conventional Machine Learning and Deep Learning in Alzheimer's Disease Diagnosis Using Neuroimaging: A Review." *Frontiers in Computational Neuroscience* 17, 1038636. https://doi.org/10.3389/fncom.2023.1038636.