Received: 22 November 2022 | Revised: 18 May 2023 | Accepted: 2 June 2023

DOI: 10.1002/alz.13412

Alzheimer's & Dementia®
THE JOURNAL OF THE ALZHEIMER'S ASSOCIATION

REVIEW ARTICLE

Artificial intelligence for diagnostic and prognostic neuroimaging in dementia: A systematic review

Robin J. Borchert[1,2] | Tiago Azevedo[3] | AmanPreet Badhwar[4,5] | Jose Bernal[6,7,8] | Matthew Betts[7,8,9] | Rose Bruffaerts[10,11] | Michael C. Burkhart[12] | Ilse Dewachter[11] | Helena M. Gellersen[8,12] | Audrey Low[13] | Ilianna Lourida[14] | Luiza Machado[15] | Christopher R. Madan[16] | Maura Malpetti[1] | Jhony Mejia[17] | Sofia Michopoulou[18] | Carlos Muñoz-Neira[19,20] | Jack Pepys[1,21] | Marion Peres[1] | Veronica Phillips[22] | Siddharth Ramanan[23] | Stefano Tamburin[24] | Hanz M. Tantiangco[25] | Lokendra Thakur[26,27,28] | Alessandro Tomassini[23] | Ashwati Vipin[29] | Eugene Tang[30] | Danielle Newby[31] | The Deep Dementia Phenotyping (DEMON) Network | Janice M. Ranson[14] | David J. Llewellyn[14,32] | Michele Veldsman[33] | Timothy Rittman[1]

Correspondence
Robin Borchert, Department of Clinical Neurosciences, University of Cambridge, Herchel Smith Building, Forvie Site, Robinson Way, Cambridge Biomedical Campus, Cambridge, CB2 0SZ, UK.
Email: rb729@medschl.cam.ac.uk

Funding information
Alzheimer's Research UK; National Institute for Health and Care Research (NIHR); National Institute for Health Research (NIHR); Alzheimer's Research UK and the Alan Turing Institute/Engineering and Physical Sciences Research Council, Grant/Award Number: EP/N510129/1; Medical Research Council, Grant/Award Number: MR/X005674/1; National Health and Medical Research Council (NHMRC); National Institute on Aging/National Institutes of Health, Grant/Award Number: RF1AG055654

Abstract

Introduction: Artificial intelligence (AI) and neuroimaging offer new opportunities for diagnosis and prognosis of dementia.

Methods: We systematically reviewed studies reporting AI for neuroimaging in diagnosis and/or prognosis of cognitive neurodegenerative diseases.

Results: A total of 255 studies were identified. Most studies relied on the Alzheimer's Disease Neuroimaging Initiative dataset. Algorithmic classifiers were the most commonly used AI method (48%) and discriminative models performed best for differentiating Alzheimer's disease from controls. The accuracy of algorithms varied with the patient cohort, imaging modalities, and stratifiers used. Few studies performed validation in an independent cohort.

Discussion: The literature has several methodological limitations including lack of sufficient algorithm development descriptions and standard definitions. We make recommendations to improve model validation including addressing key clinical questions, providing sufficient description of AI methods and validating findings in independent datasets. Collaborative approaches between experts

in AI and medicine will help achieve the promising potential of AI tools in practice.

KEYWORDS
artificial intelligence (AI), Alzheimer's disease, dementia, machine learning (ML), neurodegenerative diseases, neuroimaging

Highlights

- There has been a rapid expansion in the use of machine learning for diagnosis and prognosis in neurodegenerative disease
- Most studies (71%) relied on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset with no other individual dataset used more than five times
- There has been a recent rise in the use of more complex discriminative models (e.g., neural networks) that performed better than other classifiers for classification of AD vs healthy controls
- We make recommendations to address methodological considerations, addressing key clinical questions, and validation
- We also make recommendations for the field more broadly to standardize outcome measures, address gaps in the literature, and monitor sources of bias

1 | INTRODUCTION

There is a pressing need to improve diagnosis and prognosis for people with dementia. Up to 20% of people may receive the wrong diagnosis,[1] and differentiating between early symptoms in dementia based on clinical information and neuropsychological testing alone is subjective and prone to error. There is large geographic variability in the likelihood of receiving a diagnosis, even within a single country.[2] Diagnostic investigations such as neuroimaging and cerebrospinal fluid (CSF) tests can support clinical diagnosis; however it can take years to receive a diagnosis from the initial onset of symptoms.[3] Receiving a timely and accurate diagnosis is critical for people with dementia, their carers, and families:[4,5] it provides the opportunity for forward planning; and with the advent of disease modifying treatments an early accurate diagnosis will guide treatment selection, working toward a precision medicine approach.[6]

Neuroimaging is a non-invasive investigation used in routine clinical practice to support the diagnosis of dementia.[7,8] A range of neuroimaging methods are used in dementia and magnetic resonance imaging (MRI) is one of the most widely used to examine brain structure,[9,10] longitudinal patterns of atrophy,[11] and changes in brain function.[12–14] Positron emission tomography (PET) is available in specialist centers and is more expensive; it is used to measure metabolic activity, or using protein-specific ligands to identify underlying pathologies.[15–17]

Human clinical judgment has traditionally been used to interpret clinical neuroimaging.[9] Visual rating scales may support this assessment using features such as medial temporal lobe atrophy[18] and white matter hyperintensity load.[19,20] However, the development of more sophisticated approaches and richer data may mean that the most informative features are not amenable to human measurement or observation. For example, resting-state functional MRI can be used to derive a variety of connectivity metrics between 1000s of nodes that are amenable to machine learning (ML) approaches.[21] Deep learning methods have also demonstrated superiority to human neuroimaging interpretation.[22,23]

ML algorithms facilitate the automation of neuroimaging interpretation and have the potential to reduce bias and improve clinical decision making.[24–26] Neuroimaging data are particularly well-suited to analysis using ML, particularly deep learning, given its high dimensionality, non-linear nature and high covariance within the data. A large and growing number of ML studies have investigated how neuroimaging features can be used to predict cognitive diagnoses and conversion to dementia, fueled by the availability of large datasets, such as the Alzheimer's Disease Neuroimaging Initiative (ADNI).[27] However, uncertainty remains about which ML approaches have the greatest potential to inform clinical decision making and how their performance compares to human decision making.

We therefore conducted a systematic review to establish: (1) the extent to which ML approaches for neuroimaging have been used for the diagnosis and/or prognosis of neurodegenerative diseases; (2) how this field has progressed over time; (3) methodological challenges; and (4) the future directions to facilitate the translation of ML methods for patient benefit in dementia.

This review is part of a Special Issue on "Artificial Intelligence for Alzheimer's Disease and Related Dementias" published in *Alzheimer's & Dementia*. Together, this series provides a comprehensive overview of current applications of artificial intelligence (AI) to dementia, and future opportunities for innovation to accelerate research. Each review focuses on a different area of dementia research, including experimental models[28], drug discovery and trials optimization[29], genetics and omics[30], biomarkers[31], neuroimaging (this article), prevention[32], applied models and digital health[33], and methods optimization[34].

RESEARCH IN CONTEXT

1. **Systematic Review**: We conducted comprehensive searches of MEDLINE, Embase, Cochrane Library, BNI, PsycINFO, CINAHL, and Emcare to identify studies that examine the potential of artificial intelligence (AI) and machine learning methods applied to neuroimaging to inform clinical diagnosis and prognosis in dementia and other neurodegenerative diseases.
2. **Interpretation**: The use of AI in neuroimaging is expanding rapidly with the evidence base being dominated by studies conducted using the ADNI dataset, algorithmic classifiers, and structural MRI focusing on Alzheimer's disease. Improved diagnostic accuracy was observed when a combination of neuroimaging modalities was used, e.g., PET and structural MRI. Findings also suggest superior performance of discriminative models compared to algorithmic and generative classifiers for the classification of Alzheimer's disease vs healthy controls.
3. **Future Directions**: We highlight gaps in knowledge, current challenges, and issues to be addressed in future research around reproducibility and reporting, relevant clinical questions, and validation of results. We advocate wider collaboration between clinical, neuroimaging, and data science teams, and present recommendations to move toward clinically useful, machine learning methods applied to neuroimaging for dementia.

2 | METHODS

We conducted a systematic review to investigate the use of ML methods for diagnosis and/or prognosis in cognitive disorders including Alzheimer's disease (AD), mild cognitive impairment (MCI), Parkinson's disease (PD), vascular dementia, Lewy body dementia (LBD), frontotemporal dementia (FTD), progressive supranuclear palsy (PSP), Huntington's disease (HD) and corticobasal degeneration (CBD). The review is reported according to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines,[35] and the protocol was registered with PROSPERO (ID: CRD42021232249) prior to the screening of abstracts.

2.1 | Search strategy

The databases MEDLINE (via Ovid), Embase (via Ovid), Cochrane Library, BNI (via ProQuest), PsycINFO (via EBSCOhost), CINAHL (via EBSCOhost), and Emcare (via Ovid) were searched using the title, abstract, keyword, and MeSH term fields from inception to January 8, 2021, with the support of the Cambridge University Clinical School Library. Results were limited to English language studies. Full search terms for each database can be found in Supplementary Material 1. Studies which were known to the authors and met the inclusion/exclusion criteria of the review, but were not initially identified using the search strategy, were also included.

2.2 | PICOS framework

Outline of the parameters of this systematic review according to the PICOS framework:

- Participants: Patients with cognitive disorders due to neurodegenerative diseases.
- Index: Neuroimaging data assessed with ML for diagnosis and/or prognosis.
- Comparator: Traditional manual/subjective diagnostic/prognostic assessment.
- Outcome: Accuracy of diagnosis and/or prognosis.
- Study design: Controlled study.

2.3 | Inclusion & exclusion criteria

The inclusion and exclusion criteria used during the screening process to determine which studies would be included in the systematic review can be found below:

Inclusion criteria:

1. Primary research studies only.
2. Patient population consisting of AD, MCI, PD, vascular dementia, LBD, FTD, PSP, CBD, HD, and/or all-cause dementia.
3. Involving at least one of the following neuroimaging or neurophysiological modalities: structural or functional MRI, PET, single-photon computed tomography (SPECT), electroencephalogram (EEG), magnetoencephalography (MEG), or ultrasound.
4. Used ML methods to investigate diagnosis and/or prognosis of cognitive neurodegenerative disease(s).

Exclusion criteria:

1. Studies which did not include human participants.
2. Studies published in languages other than English.
3. Conference abstracts and book chapters.

4. Articles which did not include primary research, for example, reviews.
5. Studies where access to the full text was not available despite attempts from multiple individuals involved in the screening process.
6. Studies which did not use ML methods or only used simple logistic or linear regression methods for classification.
7. Studies which combined neuroimaging with other biomarkers, including CSF markers and/or genetics data, in the ML algorithms without reporting of model performance for neuroimaging features without these additional biomarkers.
8. Studies which focused on automated segmentation techniques which did not directly relate to diagnosis/prognosis of neurodegenerative diseases.
9. Studies which used AI methods for feature extraction but not classification.

2.4 | Study selection

The initial records were identified using the search criteria. These records underwent de-duplication using a Zotero (https://zotero.com) automation tool, which flagged possible duplicate studies, and were manually screened by a reviewer to merge genuine duplicates. Following de-duplication, all studies were screened across two stages. During the first stage, each abstract was independently reviewed by two reviewers to determine their eligibility for inclusion based on the outlined criteria using the screening tool Rayyan (https://www.rayyan.ai/). Once both reviewers screened their allocated abstracts, inclusion/exclusion decisions were unblinded. For abstracts where there was disagreement between screeners, a third independent reviewer assessed the abstract and made the final decision as to (1) progression to full-text screening stage or (2) exclusion.

The second stage involved full-text screening of all included studies by one reviewer per paper. For studies where the reviewer was unsure if the study met the outlined criteria, a second opinion was sought and a joint decision made after discussion with the second reviewer.

2.5 | Data extraction

One reviewer per paper manually collected data from each report independently into an Excel spreadsheet without the use of automation tools. The following data were extracted from the included studies:

1. Article information: First author, year, journal, country of first author's affiliated institution.
2. Study method: Patient population(s), neuroimaging modality, source of data. For studies using different datasets relating to a study, information regarding which specific dataset was extracted where possible. For example, for ADNI studies, the specific dataset used (ADNI-1, ADNI-2, ADNI-GO, J-ADNI) was identified and recorded where available.
3. ML methods, extracted neuroimaging features.
4. Receiver-operator curve (ROC) analysis results from the ML algorithm used to predict diagnosis/prognosis in the patient population, including accuracy (ACC), sensitivity (SEN), specificity (SPE), area under the curve (AUC), positive predictive value (PPV), and/or negative predictive value (NPV).

2.6 | Risk of bias assessment

Following the second stage of screening, all included studies were assessed for risk of bias by one reviewer using a hybrid version of the Joanna Briggs Institute (JBI) Critical Appraisal checklist covering the areas we deemed most relevant to this area of research.[36] The specific questions used for risk of bias assessment and their outcome for each study can be found in Supplementary Material 2. We only excluded studies exhibiting clear methodological concerns, such as lack of reporting of basic participant demographics, in order to accurately depict and identify current barriers in the literature limiting translation to clinical practice.

2.7 | Data synthesis and approaches to classification

We used descriptive statistics to determine the following characteristics of the extracted dataset: source of neuroimaging data, type of neuroimaging used, ML methods, focus on diagnosis and/or prognosis, accuracy of diagnostic/prognostic classifications, and global distribution of first authors' institutions. Studies using MRI were labeled according to the types of features used for the classification task including volumetric structural, non-volumetric structural, and functional MRI. Volumetric structural imaging was defined as MRI methods measuring the volume of specific regions using voxel-based segmentation techniques. Studies were classified as using non-volumetric structural MRI if the features used for classification were related to cortical thickness, texture, or surface area using T1- or T2-weighted images and/or diffusion tensor imaging (DTI) data. The type of AI algorithm used for the diagnostic/prognostic classification task was extracted. Studies which used AI methods for feature extraction but not classification were excluded.

Given a training set of labeled features, there are multiple ways to learn a classifier that can then be used to predict class membership for new, unlabeled instances. We categorized classifiers according to the object they seek to learn or model.

1. Generative classifiers learn the joint distribution of the features and labels.[37] Examples include naïve Bayes and linear/quadratic discriminant analysis. After training, it is possible to generate (hence the name) new pairs of features and labels by sampling from the learned joint distribution.
2. Discriminative classifiers learn the conditional distribution of the labels given the features.[38] Examples include logistic and

Gaussian process regression with potential regularization, k-nearest neighbors, and most ensemble methods (such as random forests).

3. Non-probabilistic, algorithmic classifiers directly learn the decision boundary in feature space.[39] Examples include maximum margin classifiers and support vector machines.

We note that some non-probabilistic classifiers can be reframed in a probabilistic light.[40,41] For this reason, some authors consider these methods to be discriminative in nature and draw less of a distinction between our types (2) and (3).

In order to determine how well a classifier generalizes to new data, models are typically evaluated using a validation set consisting of labeled data withheld from the training process. The model's predictions in the validation data can be compared to known labels using a variety of different metrics; precision, recall, accuracy, AUC, and F-scores are all estimated in this way. If a classifier performs much better on training data than on validation data, this can indicate overfitting. In such a case, the model may be refitted with regularization terms or priors that penalize model complexity.

Following data extraction, we conducted a meta-analysis. Considering the large number of studies from a single cohort and significant overlap of datasets, there is a risk of identifying spurious associations and false-positive findings when running a comprehensive meta-analysis.[42–44] We attempted to overcome these barriers by running a focused evaluation of the performance of ML algorithms, measured with AUC values, for a specific task: classification of AD versus healthy controls. This was achieved using a Stratified Weighted Random Method (SWRM) approach by assigning weights to the datasets and features (see further methodological details in Supplementary Material 1).

3 | RESULTS

The initial search strategy yielded 2709 studies, which underwent abstract screening following de-duplication. Three additional studies which were not picked up in the initial search strategy but met the inclusion criteria were identified by experts in the field and underwent full-text screening. The studies were consolidated to 255 studies after full-text screening (full list of references in Supplementary Material 3). A flow chart of the screening process reported according to the PRISMA 2020 guidelines[35] is shown in Figure 1. The publication time period ranged from 2005 to 2021. The included studies were classified by country based on the institutional affiliation of the first author. The most common countries included China (26%), USA (17%), Italy (7%), France (6%), and South Korea (6%).

Risk of bias assessment resulted in exclusion of three studies which exhibited clear methodological concerns, such as lack of reporting of basic participant demographics (supplementary material 2). The majority of studies used clearly defined inclusion criteria (95%) with detailed descriptions of participants and settings (91%). Only 41% of studies explicitly identified potential confounding factors.

3.1 | Datasets

Few studies used more than a single dataset, with 233 studies using one dataset, 18 used two datasets, and the remainder used three or more datasets. The most commonly used dataset was ADNI (see Figure 2). In the majority of the studies using data from ADNI, the specific cohort used (ADNI-1, ADNI-2, ADNI-GO, J-ADNI) was not stated (129 of 181) (Table 2 in Supplementary Material 1). Where the cohort was available ($n = 52$), 36 (69.2%) studies used a single cohort, 8 (15.4%) used two cohorts, and 8 (15.4%) used three cohorts. Of those that used ADNI-2 and ADNI-GO ($n = 11$), a majority ($n = 9$) also used ADNI-1. Apart from using the ADNI dataset alone, 19 studies used data from ADNI combined with other datasets including the UK Biobank and AIBL. The majority ($n = 11$) of these combination studies used a local dataset in addition to the ADNI dataset.

3.2 | AI methods

The classifier type most frequently used was a non-probabilistic algorithmic approach (48%), an example of which is support vector machines (SVM), followed by discriminative classifiers (32%) which includes most neural networks. Generative classifiers and "other" methods, mainly consisting of studies which combined multiple AI algorithms to generate novel or complex classification tools were difficult to categorize; each constituted 10% of the literature. Most of these studies focused heavily on computational methods which are not easily accessible to a clinical audience.

The number of studies which used algorithmic classifiers (mainly SVM) increased considerably between 2013 and 2015, after which its use stabilized. In contrast, there was a sharp rise in the number of studies using discriminative approaches (e.g., neural networks) starting in 2017, with discriminative studies outnumbering algorithmic studies for the first time in 2019 (Figure 3).

In order to unveil potential differences in performance between ML methods, we examined AUC values for classifying AD versus healthy controls across studies (Figure 4). Of note is that only 13% (11 of 84) of these studies reported a confidence interval for the AUC value. Of these 11 studies, 5 did not report the range of the confidence interval (e.g., 90% or 95%).

We employed a meta-analytic approach using the stratified weighted random method (SWRM) to weigh results based on the dataset, imaging modality, and type of ML method used (methodological details in the Supplementary Material 1). We found that for classification of AD versus healthy controls (i) discriminative models (SWRM = 3.39, RSD = 0.948, Heterogeneity = Considerable) performed better compared to algorithmic (SWRM = 2.42, RSD = 0.758, Heterogeneity = Substantial) and generative (SWRM = 2.14, RSD = 0.784, Heterogeneity = Moderate) classifiers; and (ii) each R table expected to have 49 rows but has in the range of 6-8, which indicates that most of the literature was limited to only few datasets and imaging modalities.

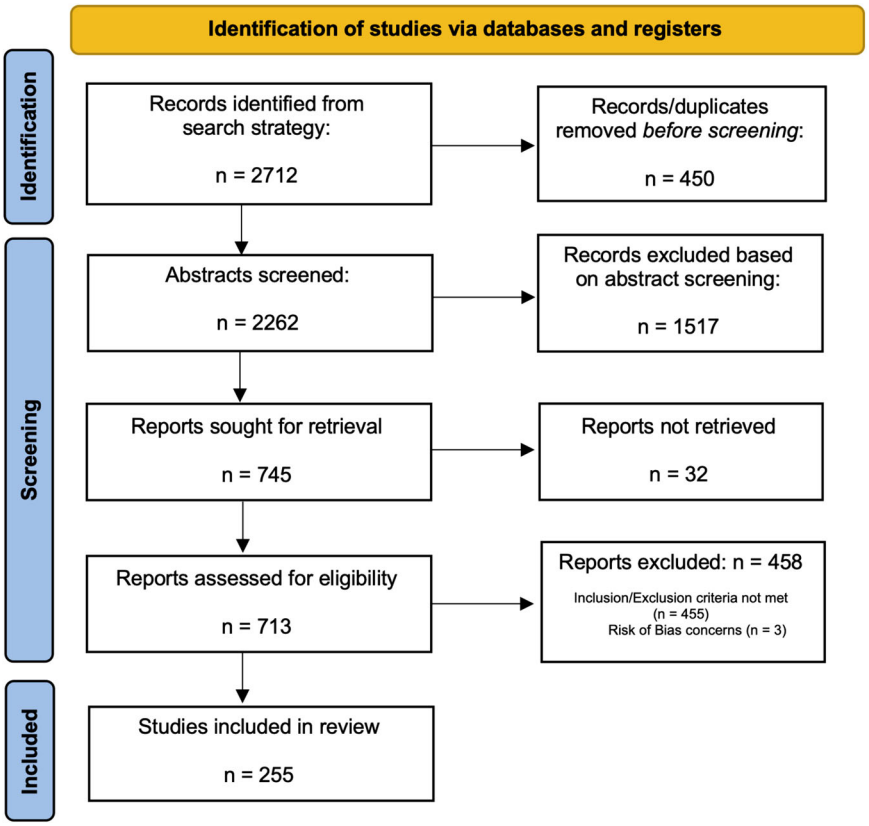

FIGURE 1 PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram for systematic review outlining the number of studies identified and excluded at each stage.

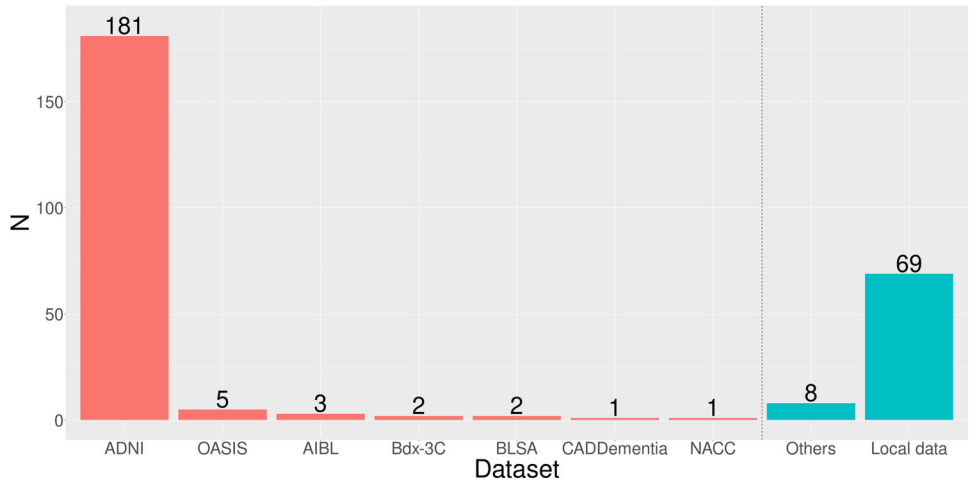

FIGURE 2 Datasets used across included studies. The majority of studies ($n = 181$, 71.0%) used the ADNI dataset alone or in combination with another dataset. Local data were used in 69 (27.1%) studies. Multiple studies used a combination of two datasets or more resulting in an overlap between the categories listed here. ADNI = Alzheimer's Disease Neuroimaging Initiative, OASIS = Open Access Series of Imaging Studies, AIBL = Australian Imaging, Biomarker & Lifestyle study of ageing, Bdx-3C = Bordeaux 3 Cities study, BLSA = Baltimore Longitudinal Study of Aging, CADDementia = Computer-Aided Diagnosis of Dementia challenge, NACC = National Alzheimer's Coordinating Center.

We identified four studies which used transfer learning for classification[45–48] which were trained on ImageNet[45] ADNI (normal controls and AD),[46] Human Connectome Project (HCP),[47] and generic images,[48] and were transferred to ADNI,[45] ADNI (stable and progressive MCI),[46] ADNI,[47] and ADNI (sMRI).[48] Transfer learning was typically used for fine tuning neural networks, particularly when the authors felt the dataset was not sufficiently large enough to properly train the neural network algorithm. Accuracy varied between these studies, including for the following classification tasks: AD versus healthy controls (90.4–99.1), MCI versus healthy

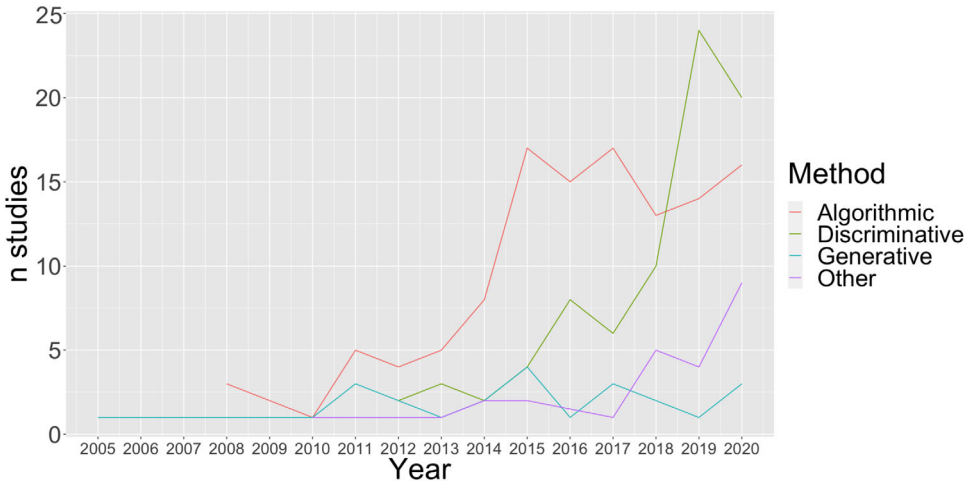

FIGURE 3 Changes in classification methods over time. This figure shows the rise in the use of discriminative classifiers in the last 4 years. The use of algorithmic classifiers increased up to 2015 and has remained steady since. The use of generative models has stayed relatively stable since its first use in 2005.

controls (83.2–99.2), and MCI converters versus non-converters (70.6–81.6).

3.3 | MRI

The number of imaging modalities used across the included studies can be found in Figure 5. Structural MRI and PET/SPECT were the most frequently used imaging modalities for diagnosis and prognosis of dementia, being used in approximately 71% and 25% of studies respectively. Around half of studies leveraged structural MRI alone (134 of 255) and those making use of multiple modalities (49 of 255) often used sMRI and PET (35 of 49) together. It is only since 2020 that studies incorporating three or more different modalities have begun to appear.[49–51]

In total, 68.6% (175 of 255) of studies relied on volumetric structural MRI measurements. In the few studies that tested traditional and AI approaches head-to-head, AI methods outperformed raw volumetric measurements, for example, against hippocampal volume for diagnosis[52,53] and for predicting conversion of MCI to AD.[54] The reported accuracy of AI methods for the diagnosis of AD varied between 60.2% and 99.3%. Of note, estimates in the lower range were found when using a multi-class classifier (i.e., AD vs. MCI vs. healthy controls, rather than AD vs. healthy controls)[55,56] or where an independent validation group was used.[57]

Contributing to heterogeneity, the aim of "diagnosis" differed between studies using structural MRI. For example, there were 17 studies specifically targeting early diagnosis in which "early" disease was variably defined by: MMSE score < 24[58–60]; CDR 0.5-1[48,61–63]; progression from MCI to AD within 18 months,[64,65] 2 years,[66] 3 years[67,68]; conversion more than 12 months after imaging[69]; or was not clearly defined.[70–72]

Studies using longitudinal structural MRI measures ($n = 6$)[69,73–77] suggest that multiple timepoints may be more accurate than baseline measures alone for the diagnosis of AD,[62] and were particularly useful when applied to the prediction of MCI to AD conversion.[69,75,77] Of interest, longitudinal changes in volumetric MRI may need to be considered in the context of baseline volumetry to be meaningful.[74]

Twenty-eight studies investigated the use of non-volumetric structural imaging features for diagnosis ($n = 24$) and/or prognosis/conversion ($n = 7$). The input consisted of T1- or T2-weighted images, DTI data, or a combination thereof, to estimate non-volumetric features such as cortical thickness, texture, and surface area. These studies focused on (i) optimization of image pre-processing techniques, (ii) investigation of feature selection methods, and (iii) optimization of classifiers and subsequent validation of the developed method. The accuracy for differentiating between AD patients and healthy controls ranged from 79.2% to 99.1%. Promising developments were noted for differential diagnosis (e.g., vascular dementia vs. AD)[78] and early diagnosis distinguishing MCI and healthy controls.[79–82] As expected, differentiating MCI subtypes and between MCI and AD cohorts was a more difficult task, which is also often the case in clinical practice. We found that performance was lower when predicting MCI conversion to AD, or conversion of stable MCI to progressive MCI.[83–85]

Twenty-six studies (the first published in 2012) used resting-state MRI (rsMRI); we did not identify any studies using task-based MRI. All but 4 studies[51,52,86,87] focused on diagnosis and the majority (20 of 26) used ADNI data, either as the primary dataset or as a replication dataset. Graph measures were often used to summarize network characteristics. Overall, the accuracy of discriminating between AD and controls ranged between 85% and 97%, but dropped when discriminating between MCI and controls (70-88%). Most studies reported the nodes which contribute most to discrimination between AD and controls: there was some heterogeneity, but most often components of the default mode network (DMN) were identified.[88–91]

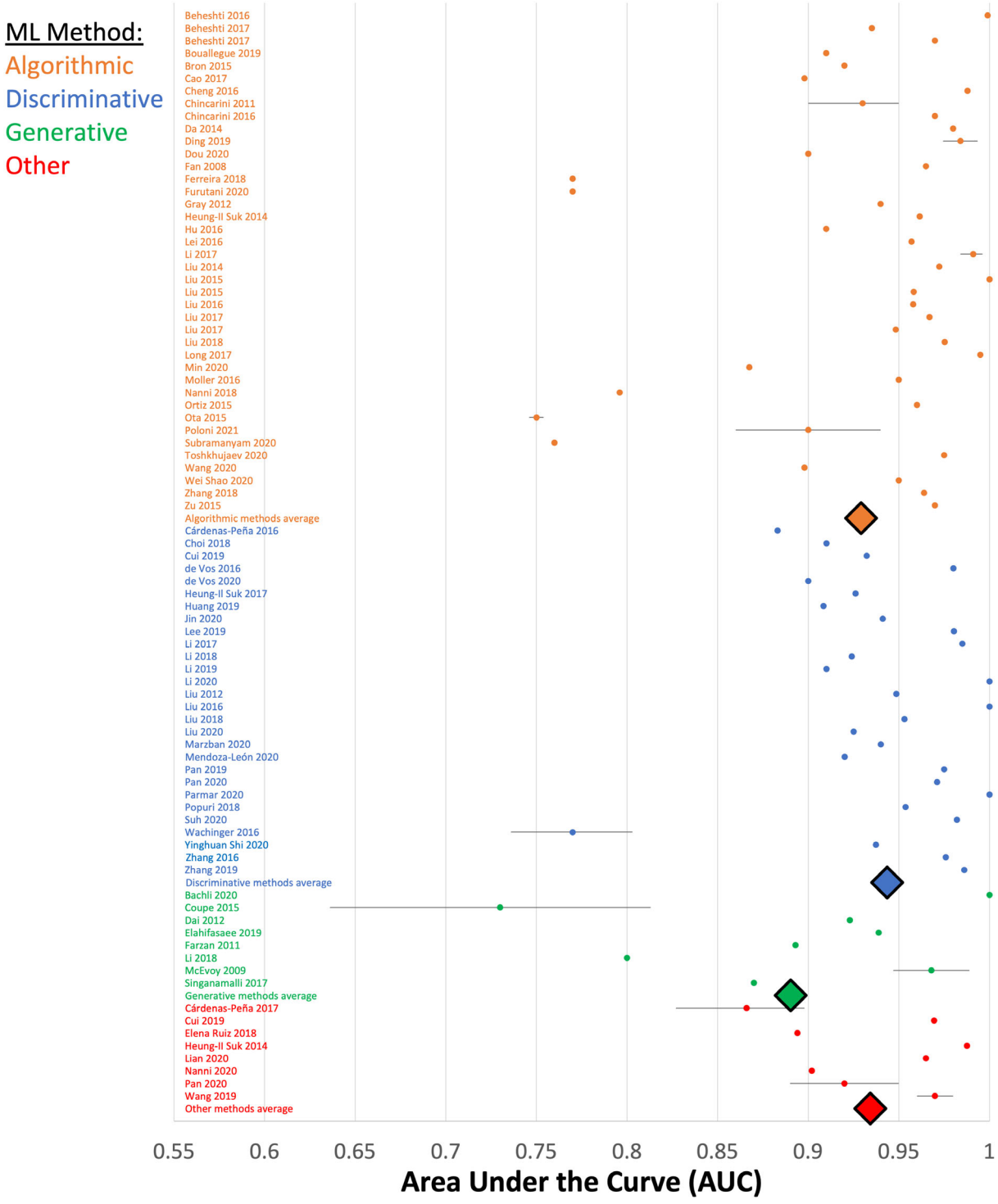

FIGURE 4 Forest plot depicting AUC values for classifications of AD patients versus healthy controls. Confidence intervals are shown where this was reported. Studies were stratified according to the type of machine learning method used including algorithmic (orange), discriminative (blue), generative (green) and other (red). Unweighted average AUC values for each type of machine learning method is depicted with a diamond.

3.4 | Neurophysiological imaging

We identified 24 studies which used neurophysiological imaging methods, only three of which investigated non-AD neurodegenerative diseases including PD and FTD.[92–94] The majority of the studies ($n = 21$) used quantitative EEG, while the remaining used either MEG,[95] event-related potential EEG[96], or combined EEG with SPECT.[97] Although half ($n = 12$) of these studies have been published since

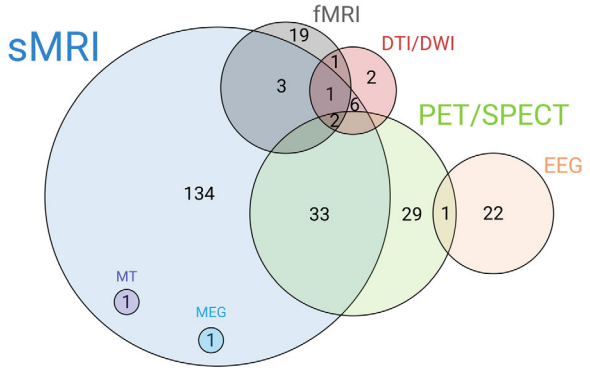

FIGURE 5 Imaging modalities used across included studies. fMRI = functional MRI, DTI = diffusion tensor imaging, DWI = Diffusion weighted imaging, EEG = Electroencephalography, MEG = magnetoencephalography MT = magnetization transfer, PET = positron emission tomography, sMRI = structural MRI, SPECT = single photon emission computed tomography.

2018, this cohort of publications also included some of the earliest studies identified in this review starting in 2005.[98,99] All neurophysiological studies used data from their local institution, the largest of which included EEG recordings from 272 participants,[100] although most studies (n = 13) included less than 50 participants. In a manner similar to other imaging modalities, SVM was the most common (n = 12) ML tool used and no other algorithm was used in more than three studies. Accuracy of discrimination between AD and healthy controls varied from 69% in the single MEG study[95] up to 100% in one study using four EEG features.[101]

3.5 | PET/SPECT imaging

Sixty-five studies were identified using PET imaging, aiming to improve early diagnosis (n = 46), prognosis (n = 13), or both (n = 6) using ML approaches. The most commonly used approach was SVM (n = 27), which when applied to FDG PET, demonstrated an accuracy of over 85% in studies for detecting AD hypometabolic patterns[102–104] and outperformed structural MRI when compared head-to-head.[105,106] Using SVM with FDG PET data distinguished AD (>86% accuracy) and MCI (>78.8% accuracy) from controls and predicted MCI conversion within 12 months and up to 5 years with accuracies ranging from 72% to 80%.[107–118] The same approach applied to amyloid PET also demonstrated accuracies of >85% for predicting MCI conversion and diagnosing AD.[115,117,119–121] Non-SVM approaches, such as convolutional neural networks and deep learning, on FDG PET and amyloid PET showed variable performance in predicting a final diagnosis of AD, cognitive decline, or MCI conversion,[46,122–132] with accuracy between 75% and 100%. Model accuracy in multicenter studies (>70% accuracy) was lower than that of those relying on local datasets (>78% accuracy).

Compared to ML methods which used PET alone, those which combined imaging modalities (i.e., FDG PET, amyloid PET, and/or MRI) were more accurate in terms of diagnosis of both MCI and AD (min accuracy: 56% for PET alone vs. 72% for PET and other modalities).[109,115–117,120,133–135] An additional approach used PET and structural MRI data in combination with other markers (i.e., apolipoprotein E4 [APOE4] status and cognitive scores) to train a classifier, then selected neuroimaging features for classification, showing better performance when neuroimaging data (gray matter density, amyloid burden, APOE4 status; r = −0.68) were used to predict individualized rate of cognitive decline in MCI, compared to cognitive predictors (depression, memory and executive function scores; r = −0.4).[136] Similarly, three studies showed that SPECT is able to classify MCI and AD, but its predictive value for MCI conversion improved when combined with other imaging modalities or cognitive assessments.[97,137,138]

3.6 | Approaches to prognosis in AD

Fifty-four studies investigated either prognosis or a combination of diagnosis and prognosis. The majority were retrospective designs (51 of 54). Of 54 studies, 47 (87%) looked at prognosis in terms of MCI to AD conversion. Of these studies, two approaches were used to evaluate the performance of prognostic predictions; some exclusively used baseline data (fixed), while others used multiple imaging time points (continuous) and related these to time to conversion.

MRI alone was the main imaging modality used (36 of 54 studies) with an additional six studies combining MRI and PET. Nine studies used only PET data,[102,111–113,116,122,126,139] one used SPECT,[137] and two used EEG data.[93,95] The main outcome measure for these studies was conversion to AD from MCI over a prespecified period of time (47 of 54 studies). A smaller proportion of studies (n = 4) used cognitive decline as an outcome measure. Similar to the diagnostic studies discussed in this review, the majority of the neuroimaging data came from the ADNI database (78%, 42 of 54 studies). An additional three studies combined local datasets with ADNI.

Thirty-eight studies used only baseline imaging data to predict a future diagnosis with a range of accuracy between 65% and 96% (mean AUC 0.79, standard deviation 0.09). Seven used multiple imaging timepoints to make predictions with accuracies between 73% and 92% (mean AUC 0.81, standard deviation 0.10). One paper found a substantial improvement with longitudinal data (AUC 0.93) compared to baseline data alone (AUC 0.54),[111] and a second paper achieved a high level of accuracy using baseline neuroimaging information with longitudinal cognitive scores (AUC = 0.90).[70]

Time to conversion was divided into two categories: conversion within a fixed timeframe (42 of 47), or a continuous measure of time of conversion (5 of 47). Of those that used a fixed timeframe, 5 studies considered conversion within 1 year (AUC range: 0.72-0.90), 8 studies within 18 months (AUC range: 0.68-0.79), 5 studies within 2 years (AUC range: 0.74-0.96), 17 studies within 3 years (AUC range: 0.65-0.93), and 7 studies predicted conversion over 3 years with a maximum of within 10 years (AUC range: 0.54-0.91).

The main outcome of the remaining studies that did not focus on MCI to AD progression (7 of 54) varied; 2 of 7 predicted cognitive scores (Alzheimer's Disease Assessment Scale—Cognitive Subscale [ADAS-Cog]) over time using longitudinal MRI,[140,141] while 2 other studies predicted both cognitive scores (Mini Mental State Examination [MMSE]) and MCI to AD conversion within 24 months.[77,142] Additionally, two of seven studies predicted conversion from cognitively normal to AD in 7[59] and 2 years.[64,143] Finally, only one paper examined prognosis in non-AD neurodegenerative diseases, namely PD and DLB[93] with an AUC of 0.87.

3.7 | Non-Alzheimer's dementias

The majority of studies that included patients with non-Alzheimer's dementia used neuroimaging features to improve the differential diagnosis between different dementia diagnoses. In total, 17 studies included a non-AD dementia group, 14 featured a non-AD dementia as the diagnosis of interest, with the remaining 3 using the non-AD groups as a control group. FTD or behavioral variant FTD (bvFTD) was the most commonly investigated non-AD dementia, with seven studies having FTD or bvFTD as their main focus.[92,93,144–148] These studies attempted differential diagnosis of FTD (from AD and/or LBD) most often using neuropsychological data and structural imaging (four of seven studies), with two studies using EEG[92,94] and one using structural MRI for classification based on post-mortem pathology.[147] Five studies used data routinely collected in clinics (for example, from memory clinics) to attempt differential diagnosis between patient groups based on imaging features and typically included FTD, LBD, PSP, CBD, PD dementia, and vascular dementia.[53,149–152]

Structural MRI was the most frequently used imaging modality (11 of 17 studies). Two studies focused on the differential diagnosis between PD and LBD,[93,153] and only two on vascular dementia.[78,154] The majority of studies used data from local hospitals or memory clinics (14 of 17 studies); one paper used local data combined with ADNI,[57] and three studies used multi-center or cohort data.[144,148,150] Since the majority of studies utilized prospective or retrospective data from local clinics, datasets were relatively small compared to multi-center studies like ADNI with most studies including 60 to 100 patients and some as low as 15 patients in a single diagnostic category.[78] The studies with larger patient numbers tended to come from multi-center studies[144,150] or used retrospective data over a long period of time.[147]

4 | DISCUSSION

In this systematic review, we examined 255 published studies using neuroimaging alone for the diagnosis or prognosis of neurodegenerative disease. The vast majority of studies (71%) used the ADNI dataset which primarily uses MRI and focuses on the conversion from MCI to AD. The dominance of ADNI means that this emphasis is reflected in the published literature, with the majority of studies using structural MRI alone or in combination with another MRI modality or PET, almost all of which focused on AD. The size of the ADNI data has led to a rapid rise since 2017 in the use of more complex discriminative AI methodologies, including deep learning models. These more complex models have in general outperformed simpler algorithmic and generative models, although comparison between studies is challenging given differences in diagnostic criteria and outcome measures. Most studies of diagnosis published ROC curve analysis results; however, there were marked differences between studies in definitions such as "early" dementia, and in the outcome measures used in prognostic studies. There remain significant gaps in the literature including non-Alzheimer's neurodegenerative diseases (most strikingly vascular dementia with only two studies), the limited application of promising neurophysiology methods, and validation in clinically relevant populations.

ML methods have been successfully applied to almost every aspect of neurodegenerative disease.[155] A previous review of ML for neuroimaging in dementia included studies up to 2016,[42] since when the field has expanded rapidly. Approximately 60% of the studies we included ($n = 152$) have been published since 2016. Some progress has been made on the concerns raised by Pellegrini and colleagues, including the overreliance on SVM classifiers and MRI. SVM was still the most frequently used classifier in our cohort which is unsurprising given that it was one of the first widely adopted methods. However, the overreliance on SVM classifiers has reduced, reflecting the rapid growth of this field and moving toward the use of a range of ML methodologies, as well as PET and/or multimodal approaches. However, despite this surge in studies, several barriers prevent the integration of these novel methods into everyday clinical practice. Below we discuss three critical issues identified from this systematic review: (1) reporting and reproducibility of methodology, (2) addressing clinically relevant questions, (3) validation of results.

4.1 | Methodological considerations

While it is encouraging to see a wide range of methods applied to neuroimaging data, the multiplicity of approaches creates a challenge in assessing the validity of each method, comparing between differing models, and independently reproducing the results. Although we did not systematically review reproducibility, in general we found limited descriptions of many models, and only a minority of studies reported the availability of code to enable replication.

Reproducibility and transparency in neuroimaging research is an increasingly prominent issue, most clearly outlined by Poldrack and colleagues.[156] The neuroimaging field has led the way in open science efforts, such as large data sharing platforms pioneered by the Human Connectome Project,[157] and introducing best practice for analysis and data sharing through the COBIDAS guidelines.[158,159] To increase the reliability of results, pre-registering analysis through platforms such as the Open Science Framework[160] has been advocated for in both neuroimaging studies[161] and ML methodologies.[162] More generally,

staged approaches to model validation in ML are available to improve confidence in model performance.[25]

We found that the combination of multiple imaging modalities, such as MRI and PET, improved the performance of ML models for classification tasks related to AD. We speculate that using features from multiple modalities enables the models to train on several different biomarkers which provide a more holistic representation of the underlying disease mechanisms, such as changes in structure (volumetric MRI), network-connectivity metrics (resting-state fMRI), and metabolic physiology (PET). Although the results suggest this approach may be beneficial, the limited number of studies identified here using this method means that it is difficult to suggest which combinations of modalities will be best at improving the performance of ML models.

4.2 | Addressing key clinical questions

Relevant clinical questions can be split into early diagnosis, differential diagnosis, prognosis and predicting response to treatment. There were no studies investigating the response to treatment, perhaps unsurprisingly given that the currently widely available treatments for dementia are symptomatic rather than disease modifying. The majority of studies considered the diagnosis of AD, or the prognostic prediction of MCI conversion to early AD. However, variability in definitions such as "early Alzheimer's disease" limited comparison between studies. This partly reflects the wider field where, for example, a clear definition of MCI has remained elusive despite recent efforts to reach such a consensus.[163]

We found no studies that assessed the common clinical challenge of differential diagnosis from among multiple (>2) possible diagnoses. This is a much harder problem to solve for ML algorithms because it requires a multi-class classifier which is computationally more challenging and typically yields lower accuracy than a binary classifier. The lack of appropriate multiclass data is a major limitation, particularly given the reliance on the ADNI dataset that consists almost exclusively of amnestic MCI or AD patients. The National Alzheimer's Coordinating Center dataset has Alzheimer's and non-Alzheimer's dementia patients from a real-world setting,[164] but is much more variable in scanning sequences (including MRI field strengths), and reports clinically defined diagnoses rather than research diagnostic criteria.

ROC curve analysis was widely used to characterize diagnostic classification performance. In particular, we found the AUC is often reported as the main measure of classification between groups, usually accompanied by the PPV and NPV. The PPV and NPV are more relevant to clinical practice, providing interpretation of the proportion of correct positive and negative results for a classification. The outcome measure for prognostic studies is more challenging. We found that studies predicting prognosis usually grouped outcomes and applied ROC curve analysis. This is particularly relevant for predicting MCI to AD conversion; however, it is not applicable to other situations, such as predicting the rate of cognitive decline in established dementia.

4.3 | Validation of results

We found that studies using an independent dataset for validation, as opposed to cross-validation or other similar methods, reported much lower accuracy, particularly when a community-based population was used. For instance, applying an SVM classifier trained on ADNI and applied to memory clinics found markedly reduced accuracy in the clinical setting (AUC = 0.76 for AD diagnosis) compared to that in the training dataset (AUC = 0.96).[57] A few recent studies have addressed the risk of overfitting by assessing generalizability in unseen independent research datasets,[104,165,166] collectively demonstrating the value of this approach in identifying methodological issues relevant to the overall model performance. Therefore, validation studies are critical, particularly those in a memory clinic setting where the tools are ultimately to be used.

The over-reliance on a single dataset such as ADNI introduces potential ethnic and socio-economic biases to models that may hamper generalization, an issue that has been specifically raised in the ADNI dataset.[167] Concerns have been raised more generally about bias in ML models,[168] including in the context of health applications.[169] This is of particular concern in marginalized ethnic groups who have poorer health indicators in general,[170] and who may miss out on access to health services due to socio-demographic, cultural, or religious beliefs,[171] including dementia services.[172,173] More representative datasets are critical for models to translate reliably to all parts of the population, to inform risk prediction models, and work toward closing gaps in health inequality related to dementia. Addressing bias in these collected datasets, and differences between genetic or ethnic groups in model performance, or applicability to different socio-economic populations, will be critical to address in ongoing data collection. It is unlikely that a single study or a single dataset can properly address these challenging issues, so collaboration between studies and between countries is required. This is happening to some extent in initiatives such as J-ADNI in Japan which is almost identical to the North American protocol and has been used to compare diagnosis and progression in dementia between both cohorts.[174] Other examples include the Longitudinal Aging Study in India (LASI-DAD)[175] and through initiatives such as the Genetic Frontotemporal dementia Initiative (GeNFI),[176] which recruits multi-nationally. Federated learning may also help address this issue by providing broader accessibility to datasets from diverse backgrounds and international sources.

A number of methodological approaches are available for measuring or mitigating bias.[177] Examples include the geometric solution to learn fair representations (He et al. 2020),[178] which removes correlations between the data and specified protected features, as well as IBM's AI Fairness 360 toolkit (Bellamy et al. 2019),[179] which provides an accessible set of fairness metrics for a model and accompanying explanations to help mitigate bias. We did not find the issue of bias to be discussed or addressed in the studies we reviewed.

BOX 1: Recommendations to move toward clinically useful, machine learning methods applied to neuroimaging for dementia

Recommendations for machine learning studies

Methodological considerations

- Provide sufficient description of the methods, with available code, to enable independent replication
- Use a staged approach to model validation
- Pre-register analysis
- Consider using multiple modalities

Addressing key clinical questions

- Clearly state the diagnostic criteria used
- For diagnosis, report performance in terms of ROC curve analysis, including PPV and NPV, and confidence intervals
- Clearly define measures of prognosis, and consider the use of odds ratios and survival analysis

Validation

- Independently validate models in at least one independent dataset
- Validate findings in a real-world dataset (e.g., memory clinics)

Recommendations for the field more broadly

- Work toward consensus on outcome measures for diagnosis and prognosis
- Establish large datasets of non-AD and/or multiple types of dementia
- Establish open datasets for EEG comparable to those with MRI
- Monitor ethnic and sociodemographic bias in data collection and encourage cross-study collaboration to address these biases

4.4 | Challenges for the field

Some of the issues we have highlighted can be addressed by individual researchers, but others require engagement from the neuroimaging, ML, and clinical communities more generally. This kind of collaboration has proven successful in initiatives such as ADNI. Although ADNI is a powerful dataset and has facilitated the use of more complex methodologies, similar collaborations for data collection and curation are required to help address ML for non-Alzheimer's neurodegenerative disease, and for EEG data.

Given the challenges of comparisons between studies using different methodologies and definitions, we suggest the field move toward consensus on outcome measures. Diagnostic criteria exist for the major neurodegenerative disorders, but better definitions of 'early' disease, and standard methods to assess prognosis would facilitate model selection. We outline our recommendations in Box 1.

In addition to overcoming these barriers related to transparency, establishing large, diverse datasets, external validation and consensus definitions, we will also need to address translational challenges more broadly to implement AI into real-world clinical settings.[180] Overcoming the technical obstacles of integrating AI will be required for different types of bias/artifacts when data are conglomerated from various sources/institutions[181] while ensuring the security and privacy of sensitive health records for storage and sharing.[182] Several factors currently limit the adoption of AI tools by clinicians including identity threat,[183,184] disruption of clinical workflow, and the uncertainty surrounding the basis of "black box" algorithms, particularly when the output disagrees with their own clinical judgement.[185] By improving interpretability, explainable AI may be the most amenable approach to building trust and understanding in the medical profession.[186] Furthermore, social and legal issues will require significant attention if implementation of AI into clinical practice is to be successful. For example, there remains uncertainty about which party is responsible when the use of AI tools result in harm from both legal[187] and patient[188] perspectives, while patients in general may prefer human supervision over AI.[189]

4.5 | Limitations

This systematic review has three main limitations. First, although we aimed to provide an informed and broad overview of the existing literature on this subject, our exclusion of reports not written in English and those where the full text was not available meant that some studies which would have otherwise met the inclusion criteria may not have been covered in this review. Two key additional exclusion criteria were the decisions not to include studies using linear regression for classification, and studies combining neuroimaging with other biomarkers without reporting the model performance for the neuroimaging features in isolation. Our motivation was to focus specifically on neuroimaging, and specifically on recognized ML methods, but it is possible we excluded studies with high clinical value and translational potential.

Second, the heterogeneity in classification tasks, ML methods used and statistical reporting across studies may have introduced bias when trying to decipher which tasks and results to extract. More specifically, this was an issue with the more technical studies which compared multiple (often > 5) ML methods across three or more classification groups introducing a large number of comparisons and results to consolidate and extract. For this reason, we decided to run our meta-analysis on a very specific task from which we could extract the AUC value for classifying AD versus healthy controls. This heterogeneity in AI methods, imaging modalities, and patient cohorts also meant that we were unable to provide insight into which features performed best for specific classification tasks. We do not address significant ethical issues in big data analysis of data security, consent to data sharing, and the acceptability of AI methods to clinicians and the general public.

Third, we employed a risk of bias screening tool that depended on a subjective judgment for each paper's inclusion or exclusion, and there

may have been heterogeneity in this assessment between screeners. We chose a low threshold for inclusion based on study quality in order to accurately depict and identify current barriers in the literature limiting translation to clinical practice. We only excluded studies exhibiting clear methodological concerns, such as lack of reporting of basic participant demographics. The screening tool had a binary outcome (inclusion/exclusion), and we were unable to investigate the potential relationship between study quality and ML performance.

5 | CONCLUSIONS

In this systematic review, we generate a number of recommendations to facilitate translation of ML methods for patient benefit in the diagnosis and prognosis of dementia. We highlight issues of methodological heterogeneity, clinical relevance of results, and validation/replication of findings. We offer a set of recommendations to address key gaps in the literature including the importance of addressing key clinical questions, providing sufficient details of AI methods, and validating findings in independent datasets which are clinically relevant. Looking forward, the field is likely to move toward the establishment of real-world datasets, multi-model imaging methods, and complex ML algorithms emphasizing the importance of providing sufficient methodological details to enable independent replication. We are optimistic that addressing these concerns will accelerate the translation of ML methods for patient benefit in neurodegenerative disease.

AFFILIATIONS

[1]Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK

[2]Department of Radiology, University of Cambridge, Cambridge, UK

[3]Department of Computer Science and Technology, University of Cambridge, Cambridge, UK

[4]Department of Pharmacology and Physiology, University of Montreal, Montreal, Canada

[5]Centre de recherche de l'Institut Universitaire de Gériatrie (CRIUGM), Montreal, Canada

[6]Centre for Clinical Brain Sciences, The University of Edinburgh, Edinburgh, UK

[7]Institute of Cognitive Neurology and Dementia Research, Otto-von-Guericke University Magdeburg, Magdeburg, Germany

[8]German Center for Neurodegenerative Diseases (DZNE), Magdeburg, Germany

[9]Center for Behavioral Brain Sciences, University of Magdeburg, Magdeburg, Germany

[10]Computational Neurology, Experimental Neurobiology Unit, Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium

[11]Biomedical Research Institute, Hasselt University, Diepenbeek, Belgium

[12]Department of Psychology, University of Cambridge, Cambridge, UK

[13]Department of Psychiatry, University of Cambridge, Cambridge, UK

[14]University of Exeter Medical School, Exeter, UK

[15]Department of Biochemistry, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

[16]School of Psychology, University of Nottingham, Nottingham, UK

[17]Department of Biomedical Engineering, Universidad de Los Andes, Bogotá, Colombia

[18]Imaging Physics, University Hospital Southampton NHS Foundation Trust, Southampton, UK

[19]Research into Memory, Brain sciences and dementia Group (ReMemBr Group), Translational Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

[20]Artificial Intelligence & Computational Neuroscience Group (AICN Group), Sheffield Institute for Translational Neuroscience (SITraN), Department of Neuroscience, University of Sheffield, Sheffield, UK

[21]Department of Biomedical Sciences, Humanitas University, Pieve Emanuele, Italy

[22]University of Cambridge Medical Library, Cambridge, UK

[23]Medical Research Council Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, UK

[24]Department of Neurosciences, Biomedicine and Movement Sciences, University of Verona, Verona, Italy

[25]Information School, University of Sheffield, Sheffield, UK

[26]Division of Genetics and Genomics, Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts, USA

[27]Broad Institute of MIT and Harvard, Cambridge, UK

[28]Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA

[29]Nanyang Technological University, Singapore

[30]Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, UK

[31]Department of Psychiatry, University of Oxford, Oxford, UK

[32]Alan Turing Institute, London, UK

[33]Department of Experimental Psychology, University of Oxford, Oxford, UK

AUTHOR CONTRIBUTIONS

Robin J. Borchert, Michele Veldsman, Timothy Rittman contributed to the conception of the work, drafting and revision of the manuscript for intellectual content. Robin J. Borchert, Michele Veldsman, Timothy Rittman, Jose Bernal, Eugene Tang contributed to the development of the protocol. Veronica Phillips conducted the literature search. Robin J. Borchert coordinated the screening process. Robin J. Borchert, Michele Veldsman, Timothy Rittman, Tiago Azevedo, AmanPreet Badhwar, Jose Bernal, Matthew Betts, Rose Bruffaerts, Helena M. Gellersen, Audrey Low, Christopher R. Madan, Maura Malpetti, Jhony Mejia, Sofia Michopoulou, Carlos Muñoz-Neira, Marion Peres, Siddharth Ramanan, Stefano Tamburin, Hanz M. Tantiangco, Lokendra Thakur, Alessandro Tomassini, Ashwati Vipin, Eugene Tang, Danielle Newby screened papers for inclusion in the review. Robin J. Borchert, Jose Bernal, Helena M. Gellersen, Audrey Low, Jhony Mejia, Carlos Muñoz-Neira, Marion Peres, Hanz M. Tantiangco extracted data from eligible papers. Robin J. Borchert, Michele Veldsman, Timothy Rittman, Jose Bernal, Lokendra Thakur contributed to analysis and interpretation of the data. Lokendra Thakur contributed to the meta-analytic approach. Robin J. Borchert, Michele Veldsman, Timothy Rittman, AmanPreet Badhwar, Jose Bernal, Matthew Betts, Rose Bruffaerts, Michael C. Burkhart, Ilse Dewachter, Audrey Low, Luiza Machado, Maura Malpetti, Jhony Mejia, Sofia Michopoulou, Jack Pepys, Stefano Tamburin, Lokendra Thakur, Ashwati Vipin contributed to the writing of the manuscript. Michele Veldsman and Timothy Rittman provided study supervision. Janice M. Ranson and

David J. Llewellyn conceived and organized the symposium from which this paper and others in the series originated, obtained funding, contributed to the conception of the work, revised the manuscript for intellectual content, and harmonized the manuscript with other papers in the series. Ilianna Lourida revised the manuscript for intellectual content and harmonized the manuscript with other papers in the series. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

With thanks to the Deep Dementia Phenotyping (DEMON) Network State of the Science symposium participants (in alphabetical order): Peter Bagshaw, Robin Borchert, Magda Bucholc, James Duce, Charlotte James, David Llewellyn, Donald Lyall, Sarah Marzi, Danielle Newby, Neil Oxtoby, Janice Ranson, Tim Rittman, Nathan Skene, Eugene Tang, Michele Veldsman, Laura Winchester, Zhi Yao. This paper was the product of a DEMON Network state of the science symposium entitled "Harnessing Data Science and AI in Dementia Research" funded by Alzheimer's Research UK. Race against Dementia Alzheimer's Research UK (ARUK-RADF2021A-010). Jose Bernal is supported by the MRC Doctoral Training Programme in Precision Medicine (Award Reference No. 2096671). Amanpreet Badhwar is supported by Fonds de recherche du Québec Santé—Chercheur boursiers Junior 1 and Fondation Courtois. Matthew Betts is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 425899996 – SFB 1436 Project A08 and by the German Federal Ministry of Education and Research (BMBF, funding code 01ED2102B) under the aegis of JPND. Eugene Tang, NIHR Clinical Lecturer, is funded by the National Institute for Health and Care Research (NIHR). The views expressed in this publication are those of the author(s) and not necessarily those of the NIHR, NHS or the UK Department of Health and Social Care. Sofia Michopoulou, NIHR Clinical Lecturer, was funded by the National Institute for Health Research (NIHR), the NIHR Applied Research Collaboration ARC Wessex, the Southampton Academy of Research and the Health Education England Topol Fellowship program. The views expressed in this publication are those of the author and not necessarily those of the funding bodies. Carlos Muñoz-Neira was supported by the Government of Chile through 'Becas Chile' and CONICYT—National Commission for Scientific and Technological Research [CONICYT—Comisión Nacional de Investigación Científica y Tecnológica], the University of Bristol (Grant Code G100030-150), and its Postdoctoral Research Associate position at the University of Sheffield. Janice Ranson and David Llewellyn are supported by Alzheimer's Research UK and the Alan Turing Institute/Engineering and Physical Sciences Research Council (EP/N510129/1). DJL also receives funding from the Medical Research Council (MR/X005674/1), National Institute for Health Research (NIHR) Applied Research Collaboration South West Peninsula, National Health and Medical Research Council (NHMRC), and National Institute on Aging/National Institutes of Health (RF1AG055654). Timothy Rittman is supported by the Cambridge Centre for Parkinson's Plus Disorders and the Cambridge Biomedical Research Centre.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflicts of interest. Author disclosures are available in the supporting information.

ORCID

Robin J. Borchert https://orcid.org/0000-0002-4673-9746
Tiago Azevedo https://orcid.org/0000-0002-2052-3832
AmanPreet Badhwar https://orcid.org/0000-0003-3414-3395
Jose Bernal https://orcid.org/0000-0003-3167-5134
Matthew Betts https://orcid.org/0000-0002-2840-4678
Rose Bruffaerts https://orcid.org/0000-0002-2631-9234
Michael C. Burkhart https://orcid.org/0000-0002-2772-5840
Ilse Dewachter https://orcid.org/0000-0001-7202-515X
Helena M. Gellersen https://orcid.org/0000-0001-7544-2311
Audrey Low https://orcid.org/0000-0002-2520-454X
Ilianna Lourida https://orcid.org/0000-0003-4439-2192
Luiza Machado https://orcid.org/0000-0001-8387-3832
Christopher R. Madan https://orcid.org/0000-0003-3228-6501
Maura Malpetti https://orcid.org/0000-0001-8923-9656
Jhony Mejia https://orcid.org/0000-0002-0313-6902
Sofia Michopoulou https://orcid.org/0000-0003-1974-8388
Carlos Muñoz-Neira https://orcid.org/0000-0002-4160-7875
Jack Pepys https://orcid.org/0000-0002-1441-0145
Marion Peres https://orcid.org/0000-0003-3111-6463
Veronica Phillips https://orcid.org/0000-0002-4383-9434
Siddharth Ramanan https://orcid.org/0000-0002-8591-042X
Stefano Tamburin https://orcid.org/0000-0002-1561-2187
Hanz M. Tantiangco https://orcid.org/0000-0002-8699-6579
Lokendra Thakur https://orcid.org/0000-0003-3505-9558
Alessandro Tomassini https://orcid.org/0000-0001-5645-6910
Ashwati Vipin https://orcid.org/0000-0002-3575-7287
Eugene Tang https://orcid.org/0000-0003-1030-9311
Danielle Newby https://orcid.org/0000-0002-3001-1478
Janice M. Ranson https://orcid.org/0000-0001-9491-3940
David J. Llewellyn https://orcid.org/0000-0002-2441-4246
Michele Veldsman https://orcid.org/0000-0003-2192-378X
Timothy Rittman https://orcid.org/0000-0003-1063-6937

REFERENCES

1. Fischer CE, Qian W, Schweizer TA, et al. Determining the impact of psychosis on rates of false-positive and false-negative diagnosis in Alzheimer's disease. *Alzheimers Dement Transl Res Clin Interv*. 2017;3:385-392. doi:10.1016/j.trci.2017.06.001
2. Cook LD, Nichol KE, Isaacs JD. The London memory service audit and quality improvement programme. *BJPsych Bull*. 2019;43:215-220. doi:10.1192/bjb.2019.18
3. Nedelec T, Couvy-Duchesne B, Monnet F, et al. Identifying health conditions associated with Alzheimer's disease up to 15 years before diagnosis: an agnostic study of French and British health records. *Lancet Digit Health*. 2022;4:e169-78. doi:10.1016/S2589-7500(21)00275-2
4. de Vugt ME, Verhey FRJ. The impact of early dementia diagnosis and intervention on informal caregivers. *Prog Neurobiol*. 2013;110:54-62. doi:10.1016/j.pneurobio.2013.04.005
5. Robinson L, Tang E, Taylor J-P. Dementia: timely diagnosis and early intervention. *BMJ*. 2015;350:h3029. doi:10.1136/bmj.h3029

6. Meco AD, Vassar R. Early detection and personalized medicine: future strategies against Alzheimer's disease. *Prog Mol Biol Transl Sci*. 2021;177:157-173. doi:10.1016/bs.pmbts.2020.10.002
7. Rittman T. Neurological update: neuroimaging in dementia. *J Neurol*. 2020;267:3429-3435. doi:10.1007/s00415-020-10040-0
8. Filippi M, Agosta F, Barkhof F, et al. EFNS task force: the use of neuroimaging in the diagnosis of dementia. *Eur J Neurol*. 2012;19:1487-1501. doi:10.1111/j.1468-1331.2012.03859.x
9. Harper L, Barkhof F, Scheltens P, Schott JM, Fox NC. An algorithmic approach to structural imaging in dementia. *J Neurol Neurosurg Psychiatry*. 2014;85:692-698. doi:10.1136/jnnp-2013-306285
10. Karas GB, Burton EJ, Rombouts SA, et al. A comprehensive study of gray matter loss in patients with Alzheimer's disease using optimized voxel-based morphometry. *NeuroImage*. 2003;18:895-907. doi:10.1016/s1053-8119(03)00041-7
11. Young AL, Marinescu RV, Oxtoby NP, et al. Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with Subtype and Stage Inference. *Nat Commun*. 2018;9:4273. doi:10.1038/s41467-018-05892-0
12. Pievani M, de Haan W, Wu T, Seeley WW, Frisoni GB. Functional network disruption in the degenerative dementias. *Lancet Neurol*. 2011;10:829-843. doi:10.1016/S1474-4422(11)70158-2
13. Greicius MD, Srivastava G, Reiss AL, Menon V. Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI. *Proc Natl Acad Sci*. 2004;101:4637-4642. doi:10.1073/pnas.0308627101
14. Badhwar A, Tam A, Dansereau C, Orban P, Hoffstaedter F, Bellec P. Resting-state network dysfunction in Alzheimer's disease: a systematic review and meta-analysis. *Alzheimers Dement Amst Neth*. 2017;8:73-85. doi:10.1016/j.dadm.2017.03.007
15. Klunk WE, Engler H, Nordberg A, et al. Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B. *Ann Neurol*. 2004;55:306-319. doi:10.1002/ana.20009
16. Lowe VJ, Curran G, Fang P, et al. An autoradiographic evaluation of AV-1451 Tau PET in dementia. *Acta Neuropathol Commun*. 2016;4. doi:10.1186/s40478-016-0315-6
17. Chételat G, Arbizu J, Barthel H, et al. Amyloid-PET and 18 F-FDG-PET in the diagnostic investigation of Alzheimer's disease and other dementias. *Lancet Neurol*. 2020;19:951-962. doi:10.1016/S1474-4422(20)30314-8
18. Scheltens P, Launer LJ, Barkhof F, Weinstein HC, van Gool WA. Visual assessment of medial temporal lobe atrophy on magnetic resonance imaging: interobserver reliability. *J Neurol*. 1995;242:557-560. doi:10.1007/BF00868807
19. Wahlund LO, Barkhof F, Fazekas F, et al. A new rating scale for age-related white matter changes applicable to MRI and CT. *Stroke*. 2001;32:1318-1322. doi:10.1161/01.str.32.6.1318
20. Fazekas F, Chawluk JB, Alavi A, Hurtig HI, Zimmerman RA. MR signal abnormalities at 1.5 T in Alzheimer's dementia and normal aging. *AJR Am J Roentgenol*. 1987;149:351-356. doi:10.2214/ajr.149.2.351
21. Wang J, Zuo X, He Y. Graph-based network analysis of resting-state functional MRI. *Front Syst Neurosci*. 2010;4. https://pubmed.ncbi.nlm.nih.gov/20589099/
22. Davatzikos C. Machine learning in neuroimaging: progress and challenges. *NeuroImage*. 2019;197:652-656. doi:10.1016/j.neuroimage.2018.10.003
23. Li X, Xiong H, Li X, et al. Interpretable deep learning: interpretation, interpretability, trustworthiness, and beyond. *Knowl Inf Syst*. 2022;64:3197-3234. doi:10.1007/s10115-022-01756-8
24. Hainc N, Federau C, Stieltjes B, Blatow M, Bink A, Stippich C. The bright, artificial intelligence-augmented future of neuroimaging reading. *Front Neurol*. 2017;8:489. doi:10.3389/fneur.2017.00489
25. Kohoutová L, Heo J, Cha S, et al. Toward a unified framework for interpreting machine-learning models in neuroimaging. *Nat Protoc*. 2020;15:1399-1435. doi:10.1038/s41596-019-0289-5
26. Nielsen AN, Barch DM, Petersen SE, Schlaggar BL, Greene DJ. Machine learning with neuroimaging: evaluating its applications in psychiatry. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2020;5:791-798. doi:10.1016/j.bpsc.2019.11.007
27. Mueller SG, Weiner MW, Thal LJ, et al. Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Alzheimers Dement J Alzheimers Assoc*. 2005;1:55-66. doi:10.1016/j.jalz.2005.06.003
28. Marzi SJ, Nott A, Sala Frigerio C, et al. Artificial intelligence for neurodegenerative experimental models. *Alzheimer's Dementia*. Submitted.
29. Doherty T, Yao Z, Al Khleifat A, et al. Artificial intelligence for dementia drug discovery and trials optimization. *Alzheimer's Dementia*.
30. Bettencourt C, Skene N, Bandres-Ciga S, et al. Artificial intelligence for dementia genetics and omics. *Alzheimer's Dementia*. Submitted.
31. Winchester LM, Harshfield EL, Shi L, et al. Artificial intelligence for alzheimer's disease and associated dementia biomarkers. *Alzheimer's Dementia*. Submitted.
32. Newby D, Orgeta V, Marshall CR, et al. Artificial intelligence for dementia prevention. *Alzheimer's Dementia*. Submitted.
33. Lyall DM, Kormilitzin A, Lancaster C, et al. Artificial intelligence for dementia applied models and digital health. *Alzheimer's Dementia*. Submitted.
34. Bucholc M, James C, Al Khleifat A, et al. Artificial intelligence for dementia research methods optimization. *Alzheimer's Dementia*. Submitted.
35. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. doi:10.1136/bmj.n71
36. Moola S, Munn Z, Tufanaru C, et al. *Chapter 7: systematic reviews of etiology and risk*. 2019. doi:10.46658/JBIRM-17-06
37. Bishop CM. *Probabilistic Generative Models (section 4.2). Pattern Recognit. Mach. Learn. Newer*. Springer-Verlag New York Inc.; 2007.
38. Bishop CM. *Probabilistic Discriminative Models (section 4.3). Pattern Recognit. Mach. Learn. Newer*. Springer-Verlag New York Inc.; 2007.
39. Bishop CM. *Sparse Kernel Machines Pattern Recognition and Machine Learning (Chapter 7). Pattern Recognit. Mach. Learn. Newer*. Springer-Verlag New York Inc.; 2007.
40. Franc V, Zien A, Schölkopf B, Support Vector Machines as Probabilistic Models. 2011.
41. Murphy K. *A probabilistic interpretation of SVMs. Mach. Learn. Probabilistic Perspect*. MIT Press; 2012.
42. Pellegrini E, Ballerini L, Hernandez M del CV, et al. Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: a systematic review. *Alzheimers Dement Diagn Assess Dis Monit*. 2018;10:519-535. doi:10.1016/j.dadm.2018.07.004
43. LeBlanc M, Zuber V, Thompson WK, et al. A correction for sample overlap in genome-wide association studies in a polygenic pleiotropy-informed framework. *BMC Genomics*. 2018;19:494. doi:10.1186/s12864-018-4859-7
44. Han B, Duong D, Sul JH, de Bakker PIW, Eskin E, Raychaudhuri S. A general framework for meta-analyzing dependent studies with overlapping subjects in association mapping. *Hum Mol Genet*. 2016;25:1857-1866. doi:10.1093/hmg/ddw049
45. Jain R, Jain N, Aggarwal A, Hemanth DJ. Convolutional neural network based Alzheimer's disease classification from magnetic resonance brain images. *Cogn Syst Res*. 2019;57:147-159. doi:10.1016/j.cogsys.2018.12.015
46. Lu D, Popuri K, Ding GW, Balachandar R, Beg MF. Alzheimer's Disease Neuroimaging I. Multiscale deep neural network based analysis of FDG-PET images for the early diagnosis of Alzheimer's disease. *Med Image Anal*. 2018;46:26-34.
47. Li W, Zhang L, Qiao L, Shen D. Toward a better estimation of functional brain network for mild cognitive impairment identification: a

transfer learning view. *IEEE J Biomed Health Inform*. 2020;24:1160-1168. doi:10.1109/JBHI.2019.2934230
48. Nanni L, Interlenghi M, Brahnam S, et al. Comparison of transfer learning and conventional machine learning applied to structural brain MRI for the early diagnosis and prognosis of Alzheimer's Disease. *Front Neurol*. 2020;11.
49. Li T-R, Wu Y, Jiang J-J, et al. Radiomics analysis of magnetic resonance imaging facilitates the identification of preclinical Alzheimer's Disease: an exploratory study. *Front Cell Dev Biol*. 2020;0. doi:10.3389/fcell.2020.605734
50. de Vos F, Schouten TM, Koini M, et al. Pre-trained MRI-based Alzheimer's disease classification models to classify memory clinic patients. *NeuroImage Clin*. 2020;27:102303. doi:10.1016/j.nicl.2020.102303
51. Rabin JS, Neal TE, Nierle HE, et al. Multiple markers contribute to risk of progression from normal to mild cognitive impairment. *NeuroImage Clin*. 2020;28:102400. doi:10.1016/j.nicl.2020.102400
52. Li F, Liu M. Alzheimer's disease neuroimaging initiative. A hybrid convolutional and recurrent neural network for hippocampus analysis in Alzheimer's disease. *J Neurosci Methods*. 2019;323:108-118. doi:10.1016/j.jneumeth.2019.05.006
53. Morin A, Samper-Gonzalez J, Bertrand A, et al. Accuracy of MRI classification algorithms in a tertiary memory center clinical routine cohort. *J Alzheimers Dis JAD*. 2020;74:1157-1166. doi:10.3233/JAD-190594
54. Costafreda SG, Dinov ID, Tu Z, et al. Automated hippocampal shape analysis predicts the onset of dementia in mild cognitive impairment. *NeuroImage*. 2011;56:212-219. doi:10.1016/j.neuroimage.2011.01.050
55. Guo Y, Zhang Z, Zhou B, et al. Grey-matter volume as a potential feature for the classification of Alzheimer's disease and mild cognitive impairment: an exploratory study. *Neurosci Bull*. 2014;30:477-489. doi:10.1007/s12264-013-1432-x
56. Cárdenas-Peña D, Collazos-Huertas D, Castellanos-Dominguez G. Enhanced data representation by kernel metric learning for dementia diagnosis. *Front Neurosci*. 2017;11:413. doi:10.3389/fnins.2017.00413
57. Klöppel S, Peter J, Ludl A, et al. Applying automated MR-based diagnostic methods to the memory clinic: a prospective study. *J Alzheimers Dis JAD*. 2015;47:939-954. doi:10.3233/JAD-150334
58. Cheng B, Liu M, Shen D, Li Z, Zhang D. Alzheimer's disease neuroimaging initiative. multi-domain transfer learning for early diagnosis of Alzheimer's disease. *Neuroinformatics*. 2017;15:115-132. doi:10.1007/s12021-016-9318-5
59. Coupé P, Fonov VS, Bernard C, et al. Detection of Alzheimer's disease signature in MR images seven years before conversion to dementia: toward an early individual prognosis. *Hum Brain Mapp*. 2015;36:4758-4770. doi:10.1002/hbm.22926
60. Gorji HT, Haddadnia J. A novel method for early diagnosis of Alzheimer's disease based on pseudo Zernike moment from structural MRI. *Neuroscience*. 2015;305:361-371. doi:10.1016/j.neuroscience.2015.08.013
61. Dai Z, Yan C, Wang Z, et al. Discriminative analysis of early Alzheimer's disease using multi-modal imaging and multi-level characterization with multi-classifier (M3). *NeuroImage*. 2012;59:2187-2195. doi:10.1016/j.neuroimage.2011.10.003
62. Hojjati SH, Ebrahimzadeh A, Babajani-Feremi A. Identification of the early stage of Alzheimer's disease using structural MRI and resting-state fMRI. *Front Neurol*. 2019;10:904. doi:10.3389/fneur.2019.00904
63. Khedher L, Ramírez J, Górriz JM, Brahim A, Segovia F. Early diagnosis of Alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images. *Neurocomputing*. 2015;151:139-150. doi:10.1016/j.neucom.2014.09.072
64. Pan D, Zeng A, Jia L, Huang Y, Frizzell T, Song X. Early detection of Alzheimer's disease using magnetic resonance imaging: a novel approach combining convolutional neural networks and ensemble learning. *Front Neurosci*. 2020;14.
65. Salvatore C, Cerasa A, Battista P, et al. Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: a machine learning approach. *Front Neurosci*. 2015;9:307. doi:10.3389/fnins.2015.00307
66. Chincarini A, Bosco P, Calvini P, et al. Local MRI analysis approach in the diagnosis of early and prodromal Alzheimer's disease. *NeuroImage*. 2011;58:469-480. doi:10.1016/j.neuroimage.2011.05.083
67. Hu K, Wang Y, Chen K, Hou L, Zhang X. Multi-scale features extraction from baseline structure MRI for MCI patient classification and AD early diagnosis. *Neurocomputing*. 2016;175:132-145. doi:10.1016/j.neucom.2015.10.043
68. Moradi E, Pepe A, Gaser C, Huttunen H, Tohka J. Alzheimer's Disease Neuroimaging Initiative. Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *NeuroImage*. 2015;104:398-412. doi:10.1016/j.neuroimage.2014.10.002
69. Zhu Y, Kim M, Zhu X, Kaufer D, Wu G. Alzheimer's disease neuroimaging initiative. Long range early diagnosis of Alzheimer's disease using longitudinal MR imaging data. *Med Image Anal*. 2021;67:101825. doi:10.1016/j.media.2020.101825
70. Li H, Fan Y. Early prediction of alzheimer's disease dementia based on baseline hippocampal mri and 1-year follow-up cognitive measures using deep recurrent neural networks. *Proc IEEE Int Symp Biomed Imaging*. 2019;2019:368-371. doi:10.1109/ISBI.2019.8759397
71. Lisowska A, Rekik I. Joint pairing and structured mapping of convolutional brain morphological multiplexes for early dementia diagnosis. *Brain Connect*. 2019;9:22-36. doi:10.1089/brain.2018.0578
72. Singanamalli A, Wang H, Madabhushi A. Cascaded multi-view canonical correlation (CaMCCo) for early diagnosis of Alzheimer's disease via fusion of clinical, imaging and omic features. *Sci Rep*. 2017;7:8137. doi:10.1038/s41598-017-03925-0
73. Davatzikos C, Xu F, An Y, Fan Y, Resnick SM. Longitudinal progression of Alzheimer's-like patterns of atrophy in normal older adults: the SPARE-AD index. *Brain J Neurol*. 2009;132:2026-2035. doi:10.1093/brain/awp091
74. Chincarini A, Sensi F, Rei L, et al. Integrating longitudinal information in hippocampal volume measurements for the early detection of Alzheimer's disease. *NeuroImage*. 2016;125:834-847. doi:10.1016/j.neuroimage.2015.10.065
75. Cui R, Liu M. Alzheimer's Disease Neuroimaging Initiative. RNN-based longitudinal analysis for diagnosis of Alzheimer's disease. *Comput Med Imaging Graph Off J Comput Med Imaging Soc*. 2019;73:1-10. doi:10.1016/j.compmedimag.2019.01.005
76. Farzan A, Mashohor S, Ramli R, Mahmud R. Discriminant analysis of intermediate brain atrophy rates in longitudinal diagnosis of alzheimer's disease. *Diagn Pathol*. 2011;6:105. doi:10.1186/1746-1596-6-105
77. Zhang D, Shen D, Initiative ADN. Predicting future clinical changes of MCI patients using longitudinal and multimodal biomarkers. *PLOS ONE*. 2012;7:e33182. doi:10.1371/journal.pone.0033182
78. Castellazzi G, Cuzzoni MG, Cotta Ramusino M, et al. A machine learning approach for the differential diagnosis of alzheimer and vascular dementia fed by MRI selected features. *Front Neuroinformatics*. 2020;0. doi:10.3389/fninf.2020.00025
79. Li Q, Wu X, Xu L, Chen K, Yao L. Classification of Alzheimer's disease, mild cognitive impairment, and cognitively unimpaired individuals using multi-feature kernel discriminant dictionary learning. *Front Comput Neurosci*. 2018;11:no pagination.
80. Jung WB, Lee YM, Kim YH, Mun CW. Automated classification to predict the progression of alzheimer's disease using whole-brain volumetry and DTI. *Psychiatry Investig*. 2015;12:92-102.

81. Ebadi A, Dalboni da Rocha JL, Nagaraju DB, et al. Ensemble classification of Alzheimer's disease and mild cognitive impairment based on complex graph measures from diffusion tensor images. *Front Neurosci*. 2017;11.
82. Kruthika KR, Rajeswari, Maheshappa HD. CBIR system using Capsule Networks and 3D CNN for Alzheimer's disease diagnosis. *Inform Med Unlocked*. 2019;14:59-68.
83. Gao F, Yoon H, Xu Y, et al. AD-NET: age-adjust neural network for improved MCI to AD conversion prediction. *NeuroImage Clin*. 2020;27:no pagination.
84. Hett K, Ta V-T, Manjón JV, Coupé P. Adaptive fusion of texture-based grading for Alzheimer's disease classification. *Comput Med Imaging Graph Off J Comput Med Imaging Soc*. 2018;70:8-16. doi:10.1016/j.compmedimag.2018.08.002
85. Eskildsen SF, Coupe P, Garcia-Lorenzo D, et al. Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning. *Neuroimage*. 2013;65:511-521.
86. Hojjati SH, Ebrahimzadeh A, Khazaee A, Babajani-Feremi A. Alzheimer's Disease Neuroimaging I. Predicting conversion from MCI to AD using resting-state fMRI, graph theoretical approach and SVM. *J Neurosci Methods*. 2017;282:69-80.
87. Hojjati SH, Ebrahimzadeh A, Khazaee A, Babajani-Feremi A. Predicting conversion from MCI to AD by integrating rs-fMRI and structural MRI. *Comput Biol Med*. 2018;102:30-39. doi:10.1016/j.compbiomed.2018.09.004
88. Li Y, Yang H, Lei B, Liu J, Wee CY. Novel effective connectivity inference using ultra-group constrained orthogonal forward regression and elastic multilayer perceptron classifier for MCI identification. *IEEE Trans Med Imaging*. 2019;38:1227-1239.
89. Nguyen DT, Ryu S, Qureshi MNI, Choi M, Lee KH, Lee B. Hybrid multivariate pattern analysis combined with extreme learning machine for Alzheimer's dementia diagnosis using multi-measure rs-fMRI spatial patterns. *PLoS ONE Electron Resour*. 2019;14:e0212582.
90. Jin D, Wang P, Zalesky A, et al. Grab-AD: generalizability and reproducibility of altered brain activity and diagnostic classification in Alzheimer's Disease. *Hum Brain Mapp*. 2020;41:3379-3391.
91. Wang M, Lian C, Yao D, Zhang D, Liu M, Shen D. Spatial-temporal dependency modeling and network hub detection for functional MRI analysis via convolutional-recurrent network. *IEEE Trans Biomed Eng*. 2020;67:2241-2252.
92. Garn H, Coronel C, Waser M, Caravias G, Ransmayr G. Differential diagnosis between patients with probable Alzheimer's disease, Parkinson's disease dementia, or dementia with Lewy bodies and frontotemporal dementia, behavioral variant, using quantitative electroencephalographic features. *J Neural Transm Vienna Austria*. 1996 2017;124:569-581. doi:10.1007/s00702-017-1699-6
93. Ruffini G, Ibañez D, Castellano M, et al. Deep learning with EEG spectrograms in rapid eye movement behavior disorder. *Front Neurol*. 2019;10.
94. Dottori M, Sedeño L, Martorell Caro M, et al. Towards affordable biomarkers of frontotemporal dementia: a classification study via network's information sharing. *Sci Rep*. 2017;7:3822. doi:10.1038/s41598-017-04204-8
95. Furutani N, Nariya Y, Takahashi T, et al. Decomposed temporal complexity analysis of neural oscillations and machine learning applied to Alzheimer's disease diagnosis. *Front Psychiatry*. 2020;11.
96. Chapman RM, McCrary JW, Gardner MN, et al. Brain ERP components predict which individuals progress to Alzheimer's disease and which do not. *Neurobiol Aging*. 2011;32:1742-1755. doi:10.1016/j.neurobiolaging.2009.11.010
97. Holler Y, Bathke AC, Uhl A, et al. Combining SPECT and quantitative EEG analysis for the automated differential diagnosis of disorders with amnestic symptoms. *Front Aging Neurosci*. 2017;9.
98. Dauwels J, Vialatte F, Musha T, Cichocki A. A comparative study of synchrony measures for the early diagnosis of Alzheimer's disease based on EEG. *NeuroImage*. 2010;49:668-693. doi:10.1016/j.neuroimage.2009.06.056
99. Cichocki A, Shishkin SL, Musha T, Leonowicz Z, Asada T, Kurachi T. EEG filtering based on blind source separation (BSS) for early detection of Alzheimer's disease. *Clin Neurophysiol Off J Int Fed Clin Neurophysiol*. 2005;116:729-737. doi:10.1016/j.clinph.2004.09.017
100. Buscema M, Vernieri F, Massini G, et al. An improved I-FAST system for the diagnosis of Alzheimer's disease from unprocessed electroencephalograms by using robust invariant features. *Artif Intell Med*. 2015;64:59-74. doi:10.1016/j.artmed.2015.03.003
101. Gallego-Jutglà E, Solé-Casals J, Vialatte F-B, Elgendi M, Cichocki A, Dauwels J. A hybrid feature selection approach for the early diagnosis of Alzheimer's disease. *J Neural Eng*. 2015;12:016018. doi:10.1088/1741-2560/12/1/016018
102. Toussaint P-J, Perlbarg V, Bellec P, et al. Resting state FDG-PET functional connectivity as an early biomarker of Alzheimer's disease using conjoint univariate and independent component analyses. *NeuroImage*. 2012;63:936-946. doi:10.1016/j.neuroimage.2012.03.091
103. Gray K, Wolz R, Heckemann R, Rueckert D, Hammers A. Structural differences in cognitively normal elderly individuals with abnormal amyloid biomarkers: detection using volumetric MRI in ADNI and AIBL. *Alzheimers Dement*. 2012(1):P337-8.
104. De Carli F, Nobili F, Pagani M, et al. Accuracy and generalization capability of an automatic method for the detection of typical brain hypometabolism in prodromal Alzheimer disease. *Eur J Nucl Med Mol Imaging*. 2019;46:334-347. doi:10.1007/s00259-018-4197-7
105. Ferreira LK, Rondina JM, Kubo R, et al. Support vector machine-based classification of neuroimages in Alzheimer's disease: direct comparison of FDG-PET, rCBF-SPECT and MRI data acquired from the same individuals. *Rev Bras Psiquiatr*. 2018;40:181-191.
106. Fan Y, Resnick SM, Wu X, Davatzikos C. Structural and functional biomarkers of prodromal Alzheimer's disease: a high-dimensional pattern classification study. *NeuroImage*. 2008;41:277-285. doi:10.1016/j.neuroimage.2008.02.043
107. Pardo JV, Lee JT, Kuskowski MA, et al. Fluorodeoxyglucose positron emission tomography of mild cognitive impairment with clinical follow-up at 3 years. *Alzheimers Dement*. 2010;6:326-333.
108. Pan X, Adel M, Fossati C, Gaidon T, Wojak J, Guedj E. Multiscale spatial gradient features for 18F-FDG PET image-guided diagnosis of Alzheimer's disease. *Comput Methods Programs Biomed*. 2019;180:no pagination.
109. Ortiz A, Munilla J, Alvarez-Illan I, Gorriz JM, Ramirez J. Exploratory graphical models of functional and structural connectivity patterns for Alzheimer's disease diagnosis. *Front Comput Neurosci*. 2015;9:1-18.
110. Li Y, Lu J, Jiang J, Zhang H, Zuo C. Radiomics: a novel feature extraction method for brain neuron degeneration disease using 18F-FDG PET imaging and its implementation for Alzheimer's disease and mild cognitive impairment. *Ther Adv Neurol Disord*. 2019;12.
111. Teng L, Li Y, Zhao Y, et al. Predicting MCI progression with FDG-PET and cognitive scores: a longitudinal study. *BMC Neurol*. 2020;20:148.
112. Cabral C, Morgado PM, Campos Costa D, Silveira M. Alzheimer's Disease Neuroimaging I. Predicting conversion from MCI to AD with FDG-PET brain images at different prodromal stages. *Comput Biol Med*. 2015;58:101-109.
113. Shen T, Jiang J, Lu J, et al. Predicting Alzheimer disease from mild cognitive impairment with a deep belief network based on 18F-FDG-PET images. *Mol Imaging*. 2019;18:1536012119877285.
114. Ota K, Oishi N, Ito K, Fukuyama H, SEAD-J Study Group. Alzheimer's Disease Neuroimaging Initiative. Effects of imaging modalities, brain atlases and feature selection on prediction of Alzheimer's disease.

J Neurosci Methods. 2015;256:168-183. doi:10.1016/j.jneumeth.2015.08.020
115. Zhan Y, Chen K, Wu X, et al. Identification of conversion from normal elderly cognition to Alzheimer's disease using multimodal support vector machine. *J Alzheimers Dis*. 2015;47:1057-1067.
116. Ben Bouallègue F, Mariano-Goulart D, Payoux P, Alzheimer's Disease Neuroimaging Initiative (ADNI). Joint Assessment of Quantitative 18F-Florbetapir and 18F-FDG Regional Uptake Using Baseline Data from the ADNI. *J Alzheimers Dis JAD*. 2018;62:399-408. doi:10.3233/JAD-170833
117. Yang BH, Chen JC, Chou WH, et al. Classification of Alzheimer's Disease from 18F-FDG and 11C-PiB PET imaging biomarkers using support vector machine. *J Med Biol Eng*. 2020;40:545-554.
118. Liu F, Wee C-Y, Chen H, Shen D. Inter-modality relationship constrained multi-modality multi-task feature selection for Alzheimer's Disease and mild cognitive impairment identification. *NeuroImage*. 2014;84:466-475. doi:10.1016/j.neuroimage.2013.09.015
119. El-Gamal FEA, Elmogy MM, Ghazal M, et al. A novel early diagnosis system for mild cognitive impairment based on local region analysis: a pilot study. *Front Hum Neurosci*. 2018;11:no pagination.
120. Xu L, Wu X, Chen K, Yao L. Multi-modality sparse representation-based classification for Alzheimer's disease and mild cognitive impairment. *Comput Methods Programs Biomed*. 2015;122:182-190.
121. Nozadi SH, Kadoury S. Classification of Alzheimer's and MCI patients from semantically parcelled PET images: a comparison between AV45 and FDG-PET. *Int J Biomed Imaging*. 2018;2018:no pagination.
122. Choi H, Jin KH. Alzheimer's Disease Neuroimaging I. Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging. *Behav Brain Res*. 2018;344:103-109.
123. Ding Y, Sohn JH, Kawczynski MG, et al. A deep learning model to predict a diagnosis of Alzheimer disease by using 18F-FDG PET of the Brain. *Radiology*. 2019;290:456-464.
124. Huang Y, Xu J, Zhou Y, Tong T, Zhuang X. Diagnosis of Alzheimer's disease via multi-modality 3D convolutional neural network. *Front Neurosci*. 2019;13.
125. Son HJ, Oh JS, Oh M, et al. The clinical feasibility of deep learning-based classification of amyloid PET images in visually equivocal cases. *Eur J Nucl Med Mol Imaging*. 2020;47:332-341.
126. Blazhenets G, Ma Y, Sorensen A, et al. Principal components analysis of brain metabolism predicts development of Alzheimer dementia. *J Nucl Med*. 2019;60:837-843.
127. Morgado P, Silveira M, Marques JS. Diagnosis of Alzheimer's disease using 3D local binary patterns. *Comput Methods Biomech Biomed Eng Imaging Vis*. 2013;1:2-12.
128. Popuri K, Balachandar R, Alpert K, et al. Development and validation of a novel dementia of Alzheimer's type (DAT) score based on metabolism FDG-PET imaging. *NeuroImage Clin*. 2018;18:802-813.
129. Lu D, Popuri K, Ding GW, Balachandar R, Beg MF. Alzheimer's disease neuroimaging I. Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer's disease using structural MR and FDG-PET images. *Sci Rep*. 2018;8:5697.
130. Liu M, Cheng D, Yan W. Alzheimer's disease neuroimaging initiative. classification of Alzheimer's disease by combination of convolutional and recurrent neural networks using FDG-PET images. *Front Neuroinformatics*. 2018;12.
131. Suk H-I, Lee S-W, Shen D. Alzheimer's Disease Neuroimaging Initiative. Latent feature representation with stacked auto-encoder for AD/MCI diagnosis. *Brain Struct Funct*. 2015;220:841-859. doi:10.1007/s00429-013-0687-3
132. Zhang F, Li Z, Zhang B, Du H, Wang B, Zhang X. Multi-modal deep learning model for auxiliary diagnosis of Alzheimer's disease. *Neurocomputing*. 2019;361:185-195. doi:10.1016/j.neucom.2019.04.093
133. Shao W, Peng Y, Zu C, Wang M, Zhang D. Hypergraph based multi-task feature selection for multimodal classification of Alzheimer's disease. *Comput Med Imaging Graph*. 2020;80:no pagination.
134. Zu C, Jie B, Liu M, et al. Label-aligned multi-task feature learning for multimodal classification of Alzheimer's disease and mild cognitive impairment. *Brain Imaging Behav*. 2016;10:1148-1159.
135. Liu L, Fu L, Zhang X, et al. Combination of dynamic (11)C-PIB PET and structural MRI improves diagnosis of Alzheimer's disease. *Psychiatry Res*. 2015;233:131-140. doi:10.1016/j.pscychresns.2015.05.014
136. Giorgio J, Landau SM, Jagust WJ, Tino P, Kourtzi Z. Modelling prognostic trajectories of cognitive decline due to Alzheimer's disease. *NeuroImage Clin*. 2020;26:102199. doi:10.1016/j.nicl.2020.102199
137. Borroni B, Anchisi D, Paghera B, et al. Combined 99mTc-ECD SPECT and neuropsychological studies in MCI for the assessment of conversion to AD. *Neurobiol Aging*. 2006;27:24-31. doi:10.1016/j.neurobiolaging.2004.12.010
138. Habert M-O, Horn J-F, Sarazin M, et al. Brain perfusion SPECT with an automated quantitative tool can identify prodromal Alzheimer's disease among patients with mild cognitive impairment. *Neurobiol Aging*. 2011;32:15-23. doi:10.1016/j.neurobiolaging.2009.01.013
139. Segovia F, Bastin C, Salmon E, Górriz JM, Ramírez J, Phillips C. Combining PET images and neuropsychological test data for automatic diagnosis of Alzheimer's disease. *PloS One*. 2014;9:e88687. doi:10.1371/journal.pone.0088687
140. Bhagwat N, Pipitone J, Voineskos AN, Chakravarty MM. Alzheimer's Disease Neuroimaging Initiative. An artificial neural network model for clinical score prediction in Alzheimer disease using structural neuroimaging measures. *J Psychiatry Neurosci JPN*. 2019;44:246-260. doi:10.1503/jpn.180016
141. Zhou J, Liu J, Narayan VA, Ye J. Modeling disease progression via multi-task learning. *NeuroImage*. 2013;78:233-248. doi:10.1016/j.neuroimage.2013.03.073
142. Lei B, Hou W, Zou W, Li X, Zhang C, Wang T. Longitudinal score prediction for Alzheimer's disease based on ensemble correntropy and spatial-temporal constraint. *Brain Imaging Behav*. 2019;13:126-137. doi:10.1007/s11682-018-9834-z
143. Battineni G, Chintalapudi N, Amenta F, Traini E. A comprehensive machine-learning model applied to magnetic resonance imaging (MRI) to predict Alzheimer's disease (AD) in older subjects. *J Clin Med*. 2020;9:2146. doi:10.3390/jcm9072146
144. Bachli MB, Sedeño L, Ochab JK, et al. Evaluating the reliability of neurocognitive biomarkers of neurodegenerative diseases across countries: a machine learning approach. *NeuroImage*. 2020;208:116456. doi:10.1016/j.neuroimage.2019.116456
145. Cajanus A, Hall A, Koikkalainen J, et al. Automatic MRI quantifying methods in behavioral-variant frontotemporal dementia diagnosis. *Dement Geriatr Cogn Disord Extra*. 2018;8:51-59. doi:10.1159/000486849
146. Möller C, Pijnenburg YAL, van der Flier WM, et al. Alzheimer disease and behavioral variant frontotemporal dementia: automatic classification based on cortical atrophy for single-subject diagnosis. *Radiology*. 2016;279:838-848. doi:10.1148/radiol.2015150220
147. Perry DC, Brown JA, Possin KL, et al. Clinicopathological correlations in behavioural variant frontotemporal dementia. *Brain J Neurol*. 2017;140:3329-3345. doi:10.1093/brain/awx254
148. Wang J, Redmond SJ, Bertoux M, Hodges JR, Hornberger M. A comparison of magnetic resonance imaging and neuropsychological examination in the diagnostic distinction of Alzheimer's disease and behavioral variant frontotemporal dementia. *Front Aging Neurosci*. 2016;8:119. doi:10.3389/fnagi.2016.00119
149. Kloppel S. Brain morphometry and functional imaging techniques in dementia: methods, findings and relevance in forensic neurology. *Curr Opin Neurol*. 2009;22:612-616.
150. Bruun M, Rhodius-Meester HFM, Koikkalainen J, et al. Evaluating combinations of diagnostic tests to discriminate different dementia types. *Alzheimers Dement Amst Neth*. 2018;10:509-518. doi:10.1016/j.dadm.2018.07.003

151. Houmani N, Vialatte F, Gallego-Jutglà E, et al. Diagnosis of Alzheimer's disease with electroencephalography in a differential framework. *PloS One*. 2018;13:e0193607. doi:10.1371/journal.pone.0193607
152. Koikkalainen J, Rhodius-Meester H, Tolonen A, et al. Differential diagnosis of neurodegenerative diseases using structural MRI data. *NeuroImage Clin*. 2016;11:435-449. doi:10.1016/j.nicl.2016.02.019
153. Oppedal K, Engan K, Eftestøl T, Beyer M, Aarsland D. Classifying Alzheimer's disease, Lewy body dementia, and normal controls using 3D texture analysis in magnetic resonance images. *Biomed Signal Process Control*. 2017;33:19-29. doi:10.1016/j.bspc.2016.10.007
154. Ritter K, Lange C, Weygandt M, et al. Combination of structural MRI and FDG-PET of the brain improves diagnostic accuracy in newly manifested cognitive impairment in geriatric inpatients. *J Alzheimers Dis JAD*. 2016;54:1319-1331. doi:10.3233/JAD-160380
155. Myszczynska MA, Ojamies PN, Lacoste AMB, et al. Applications of machine learning to diagnosis and treatment of neurodegenerative diseases. *Nat Rev Neurol*. 2020;16:440-456. doi:10.1038/s41582-020-0377-8
156. Poldrack RA, Baker CI, Durnez J, et al. Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nat Rev Neurosci*. 2017;18:115-126. doi:10.1038/nrn.2016.167
157. Van Essen DC, Smith SM, Barch DM, et al. The WU-Minn Human Connectome Project: an overview. *NeuroImage*. 2013;80:62-79. doi:10.1016/j.neuroimage.2013.05.041
158. Nichols TE, Das S, Eickhoff SB, et al. Best practices in data analysis and sharing in neuroimaging using MRI. *Nat Neurosci*. 2017;20:299-303. doi:10.1038/nn.4500
159. Pernet C, Garrido MI, Gramfort A, et al. Issues and recommendations from the OHBM COBIDAS MEEG committee for reproducible EEG and MEG research. *Nat Neurosci*. 2020;23:1473-1483. doi:10.1038/s41593-020-00709-0
160. Sullivan I, DeHaven A, Mellor D. Open and reproducible research on open science framework. *Curr Protoc Essent Lab Tech*. 2019;18:e32. doi:10.1002/cpet.32
161. Gentili C, Cecchetti L, Handjaras G, Lettieri G, Cristea IA. The case for preregistering all region of interest (ROI) analyses in neuroimaging research. *Eur J Neurosci*. 2021;53:357-361. doi:10.1111/ejn.14954
162. Hildebrandt M. Preregistration of machine learning research design. Against P-Hacking. *PROFILEDCOGITAS SUM COGITAS SUM 10 Years Profiling Eur. Citiz*. Amsterdam University Press; 2018:102-105. doi:10.1515/9789048550180-019
163. Dunne RA, Aarsland D, O'Brien JT, et al. Mild cognitive impairment: the manchester consensus. *Age Ageing*. 2021;50:72-80. doi:10.1093/ageing/afaa228
164. Beekly DL, Ramos EM, van Belle G, et al. The national Alzheimer's coordinating center (NACC) database: an Alzheimer disease database. *Alzheimer Dis Assoc Disord*. 2004;18:270-277.
165. Sørensen L, Nielsen M, Alzheimer's Disease Neuroimaging Initiative. Ensemble support vector machine classification of dementia using structural MRI and mini-mental state examination. *J Neurosci Methods*. 2018;302:66-74. doi:10.1016/j.jneumeth.2018.01.003
166. Qiu S, Joshi PS, Miller MI, et al. Development and validation of an interpretable deep learning framework for Alzheimer's disease classification. *Brain J Neurol*. 2020;143:1920-1933. doi:10.1093/brain/awaa137
167. Mendelson AF, Zuluaga MA, Lorenzi M, Hutton BF, Ourselin S. Selection bias in the reported performances of AD classification pipelines. *NeuroImage Clin*. 2017;14:400-416. doi:10.1016/j.nicl.2016.12.018
168. Sun W, Nasraoui O, Shafto P. Evolution and impact of bias in human and machine learning algorithm interaction. *PLOS ONE*. 2020;15:e0235502. doi:10.1371/journal.pone.0235502
169. Parikh RB, Teeple S, Navathe AS. Addressing bias in artificial intelligence in health care. *JAMA*. 2019;322:2377-2378. doi:10.1001/jama.2019.18058
170. Williams DR. Miles to go before we sleep: racial inequities in health. *J Health Soc Behav*. 2012;53:279-295. doi:10.1177/0022146512455804
171. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366:447-453. doi:10.1126/science.aax2342
172. Razai MS, Kankam HKN, Majeed A, Esmail A, Williams DR. Mitigating ethnic disparities in covid-19 and beyond. *BMJ*. 2021;372:m4921. doi:10.1136/bmj.m4921
173. Mukadam N, Cooper C, Livingston G. A systematic review of ethnicity and pathways to care in dementia. *Int J Geriatr Psychiatry*. 2011;26:12-20. doi:10.1002/gps.2484
174. Iwatsubo T, Iwata A, Suzuki K, et al. Japanese and North American Alzheimer's disease neuroimaging initiative studies: harmonization for international trials. *Alzheimers Dement*. 2018;14:1077-1087. doi:10.1016/j.jalz.2018.03.009
175. Lee J, Banerjee J, Khobragade PY, Angrisani M, Dey AB. LASI-DAD study: a protocol for a prospective cohort study of late-life cognition and dementia in India. *BMJ Open*. 2019;9:e030300. doi:10.1136/bmjopen-2019-030300
176. Rohrer JD, Nicholas JM, Cash DM, et al. Presymptomatic cognitive and neuroanatomical changes in genetic frontotemporal dementia in the Genetic Frontotemporal dementia Initiative (GENFI) study: a cross-sectional analysis. *Lancet Neurol*. 2015;14:253-262. doi:10.1016/S1474-4422(14)70324-2
177. Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A survey on bias and fairness in machine learning. *ACM Comput Surv*. 2021;54:115:1-115:35. doi:10.1145/3457607
178. A Geometric Solution to Fair Representations | Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society n.d. (accessed April 4, 2023). https://dl.acm.org/doi/abs/10.1145/3375627.3375864
179. Bellamy RKE, Dey K, Hind M, et al. AI Fairness 360: an extensible toolkit for detecting and mitigating algorithmic bias. *IBM J Res Dev*. 2019;63:4:1-4:15. doi:10.1147/JRD.2019.2942287
180. Bohr A, Memarzadeh K. The rise of artificial intelligence in healthcare applications. *Artif Intell Healthc*. 2020:25-60. doi:10.1016/B978-0-12-818438-7.00002-2
181. Obermeyer Z, Emanuel EJ. Predicting the future — Big Data, machine learning, and clinical medicine. *N Engl J Med*. 2016;375:1216-1219. doi:10.1056/NEJMp1606181
182. Yu K-H, Beam AL, Kohane IS. Artificial intelligence in healthcare. *Nat Biomed Eng*. 2018;2:719-731. doi:10.1038/s41551-018-0305-z
183. Lapointe L, Rivard S. A multilevel model of resistance to information technology implementation. *MIS Q*. 2005;29:461-491. doi:10.2307/25148692
184. Liberati EG, Ruggiero F, Galuppo L, et al. What hinders the uptake of computerized decision support systems in hospitals? A qualitative study and framework for implementation. *Implement Sci IS*. 2017;12:113. doi:10.1186/s13012-017-0644-2
185. Antoniadi AM, Du Y, Guendouz Y, et al. Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: a systematic review. *Appl Sci*. 2021;11:5088. doi:10.3390/app11115088
186. McDermid JA, Jia Y, Porter Z, Habli I. Artificial intelligence explainability: the technical and ethical dimensions. *Philos Transact A Math Phys Eng Sci*. 2021;379:20200363. doi:10.1098/rsta.2020.0363
187. Gerke S, Minssen T, Cohen G. Ethical and legal challenges of artificial intelligence-driven healthcare. *Artif Intell Healthc*. 2020:295-336. doi:10.1016/B978-0-12-818438-7.00012-5
188. Jamjoom AAB, Jamjoom AMA, Thomas JP, et al. Autonomous surgical robotic systems and the liability dilemma. *Front Surg*. 2022;9:1015367. doi:10.3389/fsurg.2022.1015367
189. Young AT, Amara D, Bhattacharya A, Wei ML. Patient and general public attitudes towards clinical artificial intelligence: a mixed

methods systematic review. *Lancet Digit Health*. 2021;3:e599-611. doi:10.1016/S2589-7500(21)00132-1

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Borchert RJ, Azevedo T, Badhwar AP, et al. Artificial intelligence for diagnostic and prognostic neuroimaging in dementia: A systematic review. *Alzheimer's Dement*. 2023;1-20. https://doi.org/10.1002/alz.13412