

# Machine Learning Methods for Predicting Conversion from Mild Cognitive Impairment to Alzheimer's Disease. A Systematic Review.

Sergio Grueso (✉ [sgrueso@uoc.edu](mailto:sgrueso@uoc.edu))

Universitat Oberta de Catalunya <https://orcid.org/0000-0002-3888-912X>

Raquel Viejo-Sobera

Universitat Oberta de Catalunya

---

## Research

**Keywords:** Alzheimer's disease, Conversion, Machine Learning, Magnetic Resonance, Mild Cognitive Impairment, PRISMA, Positron Emission Tomography, Prediction

**Posted Date:** March 15th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-288564/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

# Abstract

**Background:** Increase in life-span in our society is a double-edged sword that entails a growing number of patients with neurocognitive disorders, Alzheimer's disease being the most prevalent. Advances in medical imaging and computational power, enable new methods for early detection of neurocognitive disorders with the goal of preventing or reducing cognitive decline. Computer-aided image analysis and early detection of changes in cognition is a promising approach for patients with mild cognitive impairment, sometimes a prodromal stage of Alzheimer's disease.

**Methods:** We conducted a systematic review following PRISMA guidelines of studies where Machine Learning was applied to neuroimaging data in order to predict the progression from Mild Cognitive Impairment to Alzheimer's disease. After removing duplicates, we screened 159 studies and selected 47 for a qualitative analysis.

**Results:** Most studies used Magnetic Resonance Image and Positron Emission Tomography data but also Magnetoencephalography. The datasets were mainly extracted from the Alzheimer's disease Neuroimage Initiative (ADNI) database with some exceptions. Regarding the algorithms used, the most common were support vector machines, but more complex models such as Deep Learning, combined with multimodal and multidimensional data (neuroimaging, clinical, cognitive, biological, and behavioral) achieved the best performance.

**Conclusions:** Although performance of the different models still has room for improvement, the results are promising and this methodology has a great potential as a support tool for clinicians and healthcare professionals.

## 1 Background

The increase in life-span experienced in Western societies has largely been driven by medical and technological advances (1), however, this improvement has resulted in an increasing number of people diagnosed with neurocognitive disorders. In 2010, dementia was associated with \$604 billion dollars of healthcare expenses in EEUU (2). Every year, ten million new cases of dementia are registered and by 2050 it is estimated that 135 million people will have some degree of dementia (3). Age is the main risk factor for dementia; the prevalence is 1–2% at the age of 65 but reaches 30% at the age of 85. From all neurodegenerative disorders, about 60–90% are characterized as Alzheimer's Disease (AD) subtype (depending on the diagnostic criteria used) (4), for which there is yet no cure.

Patients are typically diagnosed when the symptoms of cognitive decline have already manifested. In such cases, the diagnosis was determined too late, failing to implement preventive protocols to reduce cognitive decline. Pharmacological and non-pharmacological treatments have proven to be effective in reducing cognitive and behavioural symptoms in early stages of the disease (5). In light of these treatments, recent studies have focused on detecting patients with cognitive impairment that have not reached dementia in order to delay or prevent its development. The last edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) includes a specific category for this type of patients called Mild Neurocognitive Disorder, analogous to the Mild Cognitive Impairment (MCI) whose main characteristic is having minor memory impairment (4) (throughout this review, MCI will be used instead of Mild Neurocognitive Disorder as it is more frequent in the scientific literature). MCI can, in some cases, be a prodromal stage of dementia, especially for AD (6).

In late stages, AD is easily detectable with neuroimaging techniques and cerebrospinal fluid evaluations for the presence of neurofibrillary clumps, beta-amyloid and tau proteins (7), and temporal cortex atrophy (4). In early stages, however, when these biomarkers have not clearly emerged, early detection of the disease or its progression from MCI to AD remains challenging. Conventional neuroimaging techniques as Magnetic Resonance Image (MRI) or Positron Emission Tomography (PET) have had limited utility so far in early AD detection (8, 9). However, the scientific community now has access to thousands of neuroimaging longitudinal datasets from healthy, MCI, and AD subjects along with other variables (i.e. demographic, biological, and cognitive measurements, etc.) stored in public databases such as the Alzheimer Disease Neuroimaging Initiative (ADNI) (<http://adni.loni.usc.edu>). These datasets can be compared and analysed to perform classification and automatic detection of AD and MCI progression (10, 11) using newly developed computer-aided techniques like Machine Learning (ML) algorithms.

The ML paradigm consists of training an algorithm with a dataset; in this case, neuroimaging results together with other clinical variables, to extract common factors that help classify subjects according to a variable of interest. In the case of an early diagnosis of AD and distinction from a stable MCI condition, for example, the algorithm learns to classify the data according to the specific diagnosis and extracts which factors have been the most relevant for the differentiation between the target groups. Subsequently, the trained algorithm can be used to classify a specific individual for which we do not know the diagnosis and thus manage to assist in the therapeutic approach (12–14). This technique can be applied to any disease that occurs with morphological changes or with characteristic neural patterns. See Arbabshirani, Plis, Sui, & Calhoun (15) for a review of the same objective and methodology but applied to autism, attention deficit disorder, and schizophrenia.

Recent work has demonstrated that ML algorithms are able to classify images from AD, MCI, and healthy participants with very high accuracy levels (16, 17). Although such classification has provided valuable information about AD biomarkers, for this technology to have more substantial clinical impact by empowering a clinician to administer a customized treatment protocol, it is necessary to determine and predict whether a MCI patient will progress to AD or remain stable. The goal of this systematic review is to analyze the existing classification methods based in ML algorithms applied to neuroimaging data in combination with other variables for predicting MCI to AD conversion.

## 2 Methods

To perform this systematic review we followed the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines (18, 19). A systematic search was done to find studies that included ML methods to predict MCI to AD conversion using neuroimaging techniques. Only articles written in English and published between 2010 and September 2020 (included) were selected. Articles published before 2010 were not included because the technological (e.g., computational power offered, graphical processing units) and methodological (e.g., ML and deep learning algorithm development) gap between those studies and the current standards makes them hardly comparable. We performed an advanced search concatenating terms with Boolean operators in PubMed, PsycINFO, ProQuest, Google Scholar and Web of Science databases as follows: ("*computational neuroscience*" OR "*artificial intelligence*" OR "*AI*" OR "*machine learning*") AND ("*neuroimaging technique*" OR "*neuroimage*") AND ("*Alzheimer*" OR "*Alzheimer's*") AND ("*mild cognitive impairment*" OR "*MCI*") AND ("*conversion*" OR "*prediction*" OR "*predicting*" OR "*follow-up*").

After removing duplicates, the eligibility criteria were applied by two independent reviewers (SG and RVS) to select only the articles that included: 1) prediction of MCI to AD conversion; 2) use of neuroimaging data; 3) classification methods based on ML algorithms; and 4) accuracy results.

Once the selection of studies was concluded, the following data was extracted for each study: 1) first author and year of publication; 2) groups; 3) sample size and mean age; 4) database; 5) neuroimaging technique used and variables selected; 6) classification method; 7) validation method; 8) accuracy achieved and 9) Area Under the ROC Curve (Table 3).

We also analyzed the risk of bias of the selected studies. The aspects considered in the analysis of bias were based on the Cochrane guidelines for systematic reviews (20) but the exact criteria were adapted by taking into account the particular methodology and goals of the studies, focused on creating and validating a classification model in large datasets. The criteria used are detailed in Table 1.

Table 1  
*Risk of bias analysis criteria*

Risk of bias	Score	Criteria
Database	Low (0)	Use of validated and widely used dataset for the study of biomarkers of Alzheimer's Disease (AD) including several years of follow-up with information of stable and progressive MCI patients
	Medium (1)	Use of similar database with less widespread usage
	High (2)	The participants were selected by the authors and no validated database was used
Validation of the classification method	Low (0)	The study validates the classification method with a test sample and/or an independent sample
	Medium (1)	It uses a different validation method
	High (2)	There is no validation of the classification method
Mathematical development of the algorithms	Low (0)	Explanation of their theoretical basis or architecture for Neural Networks
	Medium (1)	The authors refer to literature but do not develop their mathematical notation or architecture
	High (2)	No information about the model

We also performed an interpretability analysis based on the framework proposed by Kohoutová et al. (21). These authors developed three levels of assessment for the interpretability of ML models in neuroimaging based on the model, the feature, and the biology; also, each level has several sublevels. *Model-level* assessment consists on evaluating the model as a whole and testing it in different contexts and conditions. The sublevels include sensitivity and specificity, generalizability, behavioural analysis, representational analysis, and analysis of confounds. *Feature-level* assessment consists on evaluating the significance of individual features used in prediction, including stability, feature importance, and visualization. Finally, the *biology-level* assessment is a validation of the model based on its neurobiological plausibility and it

has two sublevels: literature (relationship with the model with previous literature) and invasive studies (the possibility of using more invasive methods).

We assessed whether the studies included in the review complied with each of the sublevels but we did not include behavioural analysis, representational analysis, and invasive studies sub-levels. *Behaviour analysis* sub-level was not considered because the only “behaviour” of the model is to classify subjects, and the behaviour is measured as accuracy, which is included in the sensitivity and specificity sublevels. *Representational analysis* compares the model with other models, other brain regions or other experimental settings; in our review, the main goal of almost all studies was to find neural patterns that predict AD and therefore it is common to use the whole brain as a feature. Also, there is only one experimental setting aimed to find maximum classification accuracy so it cannot be compared to similar experiments in the same study, only with similar literature (which represents another sub-level). Finally, the *invasive studies* sub-level is not applicable because the long-term objective of these investigations is to find a non-invasive method of predicting AD as soon as possible.

### 3 Results

As shown in Fig. 1, the workflow followed for the article selection included the four phases (identification, screening, eligibility, and inclusion) proposed by the PRISMA guidelines (18, 19). The 159 articles remaining after eliminating duplicates were screened and, after applying the exclusion criteria, 47 articles were selected for the review.

The risk of bias analysis is shown in Fig. 2 and Table 2. The overall risk of bias of all the studies was considered low. From the 48 articles selected at the eligibility stage, only one study (22) was not included in the qualitative analysis because of the high risk of bias. The sample size in this study was seven subjects, and it did not include any validation method. Therefore, the final number of studies included in the qualitative analysis was 47.

Table 2

*Risk of bias analysis for individual studies*

Author (year)	Algorithm	Validation	Database	Total bias of the study
Plant et al. (2010) (23)	0	0	2	2
Costafreda et al. (2011) (24)	1	0	1	2
Chincarini et al. (2011) (25)	0	0	0	0
Filipovych et al. (2011) (26)	0	0	0	0
Hinrichs et al. (2011) (8)	0	0	0	0
Westman et al. (2011) (27)	0	0	1	1
Zhang et al. (2011) (28)	0	0	0	0
Cho et al. (2012) (29)	0	0	0	0
Gray et al. (2012) (30)	0	0	0	0
Toussaint et al. (2012) (31)	1	0	0	1
Zhang et al. (2012) (9)	0	0	0	0
Casanova et al. (2013) (32)	0	0	0	0
Liu, X. et al. (2013) (33)	0	0	0	0
Wee et al. (2013) (34)	0	0	0	0
Young et al. (2013) (35)	1	0	0	1
Guerrero et al. (2014) (36)	0	0	0	0
Lebedev et al. (2014) (37)	0	0	0	0
Liu, M. et al. (2014) (38)	0	0	0	0
Liu, F. et al. (2014) (39)	0	0	0	0
Min et al. (2014) (40)	0	0	0	0
Suk et al. (2014) (41)	0	0	0	0
Cheng et al. (2015) (42)	0	0	0	0
Moradi et al. (2015) (43)	0	0	0	0
Ritter et al. (2015) (44)	1	0	0	1
Salvatore et al. (2015) (45)	0	0	0	0
Collij et al. (2016) (46)	0	2	2	4
Li et al. (2016) (47)	0	0	2	2
López et al. (2016) (48)	2	0	0	2
Thung et al. (2016) (49)	0	0	0	0
Long et al. (2017) (50)	0	0	0	0
Donnelly-Kehoe et al. (2018) (51)	0	0	0	0
Gao et al. (2018) (52)	1	0	0	1
Khanna et al. (2018) (53)	1	0	0	1
Popuri et al. (2018) (54)	0	0	0	0
Gupta et al. (2019) (55)	2	0	0	2
Lee et al. (2019) (56)	0	0	0	0
Pusil et al. (2019) (57)	0	0	0	0

*Note.* This table shows the results of the bias analysis performed based on Higgins et al. (20) with the punctuations specified in Table 1.

Author (year)	Algorithm	Validation	Database	Total bias of the study
Spasov et al. (2019) (58)	0	0	2	2
Wee et al. (2019) (59)	0	0	0	0
Abrol et al. (2020) (60)	0	0	0	0
Gao et al. (2020) (61)	0	0	0	0
Giorgio et al. (2020) (62)	0	0	0	0
Lin et al. (2020) (63)	1	0	0	1
Pan et al. (2020) (64)	0	0	0	0
Ramon-Julvez et al. (2020) (65)	0	0	0	0
Xiao et al. (2020) (66)	1	0	0	1
Xu et al. (2020) (67)	0	0	0	0
	0	0	2	2
	12/96	2/96	12/96	27/288

*Note.* This table shows the results of the bias analysis performed based on Higgins et al. (20) with the puntuations specified in Table 1.

The studies selected for the qualitative analysis are presented in Table 3 following the structure explained in the data extraction section (study, cohort, sample [mean age], database, features and neuroimaging technique, classification method, validation method, results [% accuracy], and AUC ROC).

Magnetic Resonance was the most common kind of neuroimaging used (in 28 out of 47 studies), followed by Positron Emission Tomography (PET, in three out of 47 studies), 13 studies included data from both techniques (MRI and PET), two studies used Magnetoencephalography (MEG) data and one study used MRI and MEG data.

Regarding the source of the datasets, 40 out of 47 studies used the ADNI database in any of its versions (ADNI-1, 2, 3 or GO) to obtain samples of healthy, MCI and AD subjects. Of the remaining seven studies, two used data from AddNeuroMed (<https://consortiapedia.fastercures.org/consortia/anm/>) database (24, 27) and five collected their own data (23, 46, 48, 57, 67).

Although almost all studies used the same database, the cohorts selected varied across them. Most articles (28 out of 47 studies) divided their participants in four groups: healthy controls, stable MCI patients (sMCI), progressive MCI patients (pMCI), and AD patients. Ten articles selected three cohorts formed by MCI, AD, and healthy subjects, although in order to predict the conversion to AD, they had to distinguish between pMCI and sMCI patients. The remaining nine studies used different groups of participants: only sMCI and pMCI (44, 48, 49, 57), only healthy controls and MCI (53, 61, 62, 67), or the distinction between early and late MCI (59).

The sample size also varied across studies, Pusil et al. (57) has the smallest sample with 54 subjects and Popuri et al. (54) has the largest sample with 1,294 subjects. The sample size follows an ascendant trend across years, which may be attributed to the increased data availability in the ADNI database. Mean age ranged from 62 to 79 years old. Although eight studies did not include the mean age of the sample, they used an ADNI database and therefore the age range might be similar to the rest of the studies. The variations in age between studies may be due to differences in participant selection and the moment when the study was conducted (since the ADNI database has been incorporating more data over the years).

As for feature selection, the most common were whole brain volumes, selected in 24 articles, and intensity measurements of glucose metabolism, selected in 13 PET studies, also nine studies included biological features (as APOE4 genotype). Other selected features were neuropsychological test results (seven out of 47 studies) and demographic variables as age (six out of 47 studies). 19 studies only used one type of feature such as 3D MRI data or whole brain grey matter volumes and 28 studies selected two or more different types of features.

Regarding the ML methods used to classify the patients and detect probable MCI to AD progression, the most popular were those based in Support Vector Machine (SVM). SVM was used in eight out of the 47 studies, and also in five in combination with other methods such as a Neural Network for feature selection (56) or Locally-Linear Embedding for dimensionality reduction (Liu et al., 2013). SVM is a supervised ML algorithm that has demonstrated its utility in neuroimaging-based applications, especially in classification of future clinical outcomes (68). This method takes every measurement from every subject as a single point in a multidimensional space, with the number of dimensions being the total number of features of that particular dataset (for example, 93 grey matter volumes from regions of interest). The algorithm then finds the maximal margin separating hyperplane that optimally differentiates groups of data points representing different classes (e.g. pMCI vs.

sMCI, or AD vs. HC). The data instances closest to the group boundaries are the support vectors and are, by definition, the ones that determine the position of the hyperplane. The mapping into a higher dimensional space is done by a kernel function, usually polynomial or Gaussian (24). The SVM algorithm is trained with labelled data (indicating whether the data belongs to a healthy person, sMCI, pMCI, or AD patient, for example) to generate this multidimensional space. Once the model has been trained we can introduce a new subject with MCI and it will be classified in the multidimensional space into the boundaries of one of the previously defined groups (i.e. sMCI, pMCI, AD, etc.). For example, if the new patient is classified as belonging to the AD group, we can infer that this subject is more likely to develop a future AD due to being more similar to subjects in that group. The different groups for classification will depend on the specific methodology of each study.

The combination of SVM with other methods allows better feature selection and to avoid overfitting of data, this will facilitate the generalization of the model (i.e. achieving high accuracy when applied to different datasets). For example, Thung et al. (49) used SVM with multiple kernels (linear, Gaussian, and polynomial) after feature selection with least squares and logistic elastic net regressions and also matrix completion with label-guided low-rank matrix completion method. On the other hand, Toussaint et al. (31) used non-linear SVM with Gaussian Radial Basis Function kernel but only after a two-sample t-test and a spatial independent component analysis, performed for the detection of glucose metabolism and characteristic region patterns of AD patients. Other classification methods used were Random Forest (37, 43, 51) or Neural Networks that can have different architectures but the most commonly used for image classification tasks were Convolutional Neural Networks (58, 59, 64, 65).

As for validation methods, Cross-Validation was selected in 27 studies, with different numbers of folds and/or iterations. Cross-Validation consists in dividing the sample in two parts, one to train the algorithm (training set) and another one for validation (testing set). This partition can be done several times changing the train/test split of the data, and the accuracy of each iteration can be averaged to obtain a more robust quantification of the model performance instead of just validating the model on one test sample.

Another validation method is the Leave-One-Out Cross Validation, selected in five studies. In this case, the model is trained with all the data except for one data point, then it tries to classify the data point left out and does the same with the rest of the sample in subsequent iterations. The train/test method was selected in ten studies with different percentages of data partitions. Westman et al. (27) validated the model on an independent test set of 51 subjects and Popuri et al. (54) also performed the validation with an independent sample.

The results of ML classification algorithms can be assessed based on their sensitivity (percentage of correctly detected pMCI patients or true positive ratio) and specificity (percentage of healthy or sMCI subjects correctly identified or true negative ratio), or accuracy (percentage of correctly classified subjects). By changing the decision threshold of the classifier we can compensate the ratio between true positive/true negative and generate a graphic representation of that ratio, or what is known as the Receiver Operating Characteristic (ROC) curve (69). The calculation of the area under the ROC curve (AUC ROC) represents a good quantitative index for the comparison of classification models, since it indicates the ability of the model to predict both the presence and non-presence of disease, or in this case, the progression or lack of progression from MCI to AD (70). An AUC ROC of one implies a perfect classification of every subject in the sample. The maximum accuracies achieved by every study in the prediction of AD conversion from MCI patients or the accuracy of the method in discriminating between a progressive/stable MCI are shown in the "Results" column of Table 3; the AUC coefficient is presented when available. The best results with the SVM algorithm were obtained by Pusil et al. (2019) with 100% accuracy, but using a small sample of 56, making the model hardly generalizable. In studies with bigger samples, Guerrero et al. (36) had the highest accuracy results (97.1% with 511 subjects) followed by Long et al. (2017; 96.5% and 427 subjects), Gupta et al. (2019; 93.6%, 158 subjects) and Wee et al. (2019; 92.5%, 1083 subjects). Finally, the highest AUC ROC value was 0.99 in Long et al. (2017) followed by 0.96 in Xu et al. (2020) and 0.95 in Gupta et al. (2019).

Table 3

*Studies selected following PRISMA guidelines*

Author (year)	Groups	Sample size (mean age)	Database	Neuroimaging technique and features	Classification method	Validation method	Results (% accuracy)	AUC ROC
Plant et al. (2010) (23)	HS AD MCI	18 (64.8) 32 (68.8) 24 (69.7)	Sample collected for the study	MRI: Whole-brain volume measures	SVM Bayes VFI	Train/test method: AD + HS as train set and MCI as test set.	SVM/Bayes/VFI accuracy for pMCI vs sMCI:  SVM: 50 Bayes: 58.3 VFI: 75	NA
Costafreda et al. (2011) (24)	HS AD MCI	88 (73.6) 71 (74.9) 103 (74.1)	AddNeuroMed	MRI: 3D hippocampal morphometric measures	nl-SVM-RBFk	4-fold Cross Validation	pMCI vs sMCI: 80	NA
Chincarini et al. (2011) (25)	HS AD sMCI pMCI	189 (76.6) 144 (75.5) 166 (75.7) 136 (75.1)	ADNI-1	MRI: GM volumes	SVM	20-fold Cross Validation	NA	0.74
Filipovych et al. (2011) (26)	HS AD sMCI pMCI	63 (75.2) 54 (77.4) 174 (74.5) 68 (76.2)	ADNI-1	MRI: Whole-brain GM density	Semi-supervised SVM	Leave-one-out Cross Validation	pMCI: 79.4 sMCI: 51.7	0.69
Hinrichs et al. (2011) (8)	HS AD MCI	66 (76.2) 58 (76.6) 119 (75.1)	ADNI-1	MRI and PET: scan data, APOE4 genotype, CSF assays and cognitive tests results	MK-SVM	Train/test method: AD + HS as train set and MCI as test set	pMCI vs sMCI: NA	0.79
Westman et al. (2011) (37)	HS AD MCI	112 (73) 117 (76) 122 (75)	AddNeuroMed	MRI: whole-brain volume, age and education	OPLS	Train/test method: sample of 51 subjects	pMCI vs sMCI: 73	NA

*Note.* AUC = Area Under the Curve; AD = Alzheimer's Disease; HS = Healthy Subjects; MCI = Mild Cognitive Impairment; IMCI = late MCI; eMCI = early MCI; pMCI = progressive MCI; sMCI = stable MCI; WM = White Matter; GM = Grey Matter; CNN = Convolutional Neural Network; rDNN = randomized Deep Neural Network; FCS = Functional Connectivity Strength; RF = Random Forest; SVM = Support Vector Machine; EN = Elastic Nets; AB = Ada-Boost; nl-SVM-RBFk = non-linear SVM with Radial Basis Function kernel; SNP = Single Nucleotide Polymorphisms; GPR = Gaussian Process Regression; PLS = Partial Least Squares; OPLS = Orthogonal Partial Least Squares; MMSE = Mini Mental State Examination; VFI = Voting Feature Intervals; NN = Neural Network; AD-NET = Age-Adjust Neural Network; Res-Net = Deep Residual Neural Network; SNN = Spiking Neural Network; EL = Ensemble Learning; RLR = Regularized Logistic Regression; ; F-FDG = Fluorine 18 fluorodesoxyglucose; DAT = Dementia Alzheimer Type; IDAT = late DAT; eDAT = early DAT; FPDS = FDG-PET DAT Score; ICA = Independent Component Analysis; GMB = Gradient Boosting Model; M2TL = Multimodal manifold-regularized transfer learning; ss = sample selection; M3TL = Multi-Modal Multi-Task Learning; DBM = Deep Boltzmann Machine; ELM = Extreme Learning Machine; MG2G = Multiple Graph2Gauss; MCFS = Multi-Cluster Feature Selection; HLR = Hierarchical Logistic Regression; NA = Not Applicable.

Author (year)	Groups	Sample size (mean age)	Database	Neuroimaging technique and features	Classification method	Validation method	Results (% accuracy)	AUC ROC
Zhang et al. (2011) (28)	HS  AD sMCI pMCI	52 (75.3) 51 (75.2) 56 (75.3) 43 (75.3)	ADNI-1	MRI and PET: Volume, intensity and CSF ( $A\beta_{42}$ , t-tau y p-tau) measurements	SVM	10-fold Cross Validation	pMCI: 91.5  sMCI: 73.4	NA
Cho et al. (2012) (29)	HS AD sMCI pMCI	160 (76.2) 128 (76.0) 131 (74.8) 72 (74.8)	ADNI-1	MRI: Cortical thickness	LDA	Train/test method: 50/50 partition	pMCI vs sMCI: 70	NA
Gray et al. (2012) (30)	HS AD sMCI pMCI	54 (NA) 50 (NA) 64 (NA) 53 (NA)	ADNI-1	PET: Signal intensity and relative change over 12 month	NI-SVM-RBFk	Train/test method: 75/25 partition with 1000 iterations	sMCI vs pMCI: 63.1	0.66
Toussaint et al. (2012) (31)	HS AD sMCI pMCI	80 (76.4) 80 (76.0) 40 (76.4) 40 (76.4)	ADNI-1	PET: Glucose metabolic signal and clinical measures	Two-sample t-test + spatial ICA + nl-SVM-RBFk	Leave-one-out Cross Validation	pMCI vs sMCI: 80	NA

*Note.* AUC = Area Under the Curve; AD = Alzheimer's Disease; HS = Healthy Subjects; MCI = Mild Cognitive Impairment; IMCI = late MCI; eMCI = early MCI; pMCI = progressive MCI; sMCI = stable MCI; WM = White Matter; GM = Grey Matter; CNN = Convolutional Neural Network; rDNN = randomized Deep Neural Network; FCS = Functional Connectivity Strength; RF = Random Forest; SVM = Support Vector Machine; EN = Elastic Nets; AB = Ada-Boost; nl-SVM-RBFk = non-linear SVM with Radial Basis Function kernel; SNP = Single Nucleotide Polymorphisms; GPR = Gaussian Process Regression; PLS = Partial Least Squares; OPLS = Orthogonal Partial Least Squares; MMSE = Mini Mental State Examination; VFI = Voting Feature Intervals; NN = Neural Network; AD-NET = Age-Adjust Neural Network; Res-Net = Deep Residual Neural Network; SNN = Spiking Neural Network; EL = Ensemble Learning; RLR = Regularized Logistic Regression; ; F-FDG = Fluorine 18 fluorodesoxyglucose; DAT = Dementia Alzheimer Type; IDAT = late DAT; eDAT = early DAT; FPDS = FDG-PET DAT Score; ICA = Independent Component Analysis; GMB = Gradient Boosting Model; M2TL = Multimodal manifold-regularized transfer learning; ss = sample selection; M3TL = Multi-Modal Multi-Task Learning; DBM = Deep Boltzmann Machine; ELM = Extreme Learning Machine; MG2G = Multiple Graph2Gauss; MCFS = Multi-Cluster Feature Selection; HLR = Hierarchical Logistic Regression; NA = Not Applicable.

Author (year)	Groups	Sample size (mean age)	Database	Neuroimaging technique and features	Classification method	Validation method	Results (% accuracy)	AUC ROC
Zhang et al. (2012) (9)	HS AD sMCI pMCI	47 (NA) 40 (NA) 42 (NA) 38 (NA)	ADNI-1	MRI and PET: Volume, intensity and CSF ( $A\beta_{42}$ , t-tau y p-tau) measurements	M3TL	10-fold Cross Validation	pMCI vs sMCI: 73.9	0.79
Casanova et al. (2013) (32)	HS AD sMCI pMCI	188 (75.9) 171 (75.5) 182 (75.2) 153 (75.0)	ADNI-1	MRI: GM volume	RLR	10-fold Cross Validation	pMCI vs sMCI: 61.5	NA
Liu, X. et al. (2013) (33)	HS AD sMCI pMCI	138 (76) 86 (75) 93 (75) 97 (75)	ADNI-1	MRI: Volume and cortical thickness	SVM EN LDA	<i>Leave-one-out</i> Cross Validation	pMCI vs sMCI: SVM: 66 EN:68 LDA: 68	0.53 0.61 0.68
Wee et al. (2013) (34)	HS AD sMCI pMCI	200 (75.8) 198 (75.7) 111 (75.3) 89 (74.8)	ADNI-1	MRI: Cortical thickness and correlation of cortical thickness between pairs of ROIs	Mk-SVM	10-fold Cross Validation	pMCI vs sMCI: 75.05	0.84
Young et al. (2013) (35)	HS AD sMCI pMCI	73 (75.9) 63 (75.2) 96 (75.6) 47 (74.5)	ADNI-1	MRI and PET: Volume, intensity, APOE4 genotype and CSF ( $A\beta_{42}$ , t-tau y p-tau) measurements	Gaussian Process	<i>Leave-one-out</i> Cross Validation	sMCI vs pMCI: 74.1	0.79

*Note.* AUC = Area Under the Curve; AD = Alzheimer's Disease; HS = Healthy Subjects; MCI = Mild Cognitive Impairment; IMCI = late MCI; eMCI = early MCI; pMCI = progressive MCI; sMCI = stable MCI; WM = White Matter; GM = Grey Matter; CNN = Convolutional Neural Network; rDNN = randomized Deep Neural Network; FCS = Functional Connectivity Strength; RF = Random Forest; SVM = Support Vector Machine; EN = Elastic Nets; AB = Ada-Boost; nl-SVM-RBFk = non-linear SVM with Radial Basis Function kernel; SNP = Single Nucleotide Polymorphisms; GPR = Gaussian Process Regression; PLS = Partial Least Squares; OPLS = Orthogonal Partial Least Squares; MMSE = Mini Mental State Examination; VFI = Voting Feature Intervals; NN = Neural Network; AD-NET = Age-Adjust Neural Network; Res-Net = Deep Residual Neural Network; SNN = Spiking Neural Network; EL = Ensemble Learning; RLR = Regularized Logistic Regression; ; F-FDG = Fluorine 18 fluorodesoxyglucose; DAT = Dementia Alzheimer Type; IDAT = late DAT; eDAT = early DAT; FPDS = FDG-PET DAT Score; ICA = Independent Component Analysis; GMB = Gradient Boosting Model; M2TL = Multimodal manifold-regularized transfer learning; ss = sample selection; M3TL = Multi-Modal Multi-Task Learning; DBM = Deep Boltzmann Machine; ELM = Extreme Learning Machine; MG2G = Multiple Graph2Gauss; MCFS = Multi-Cluster Feature Selection; HLR = Hierarchical Logistic Regression; NA = Not Applicable.

Author (year)	Groups	Sample size (mean age)	Database	Neuroimaging technique and features	Classification method	Validation method	Results (% accuracy)	AUC ROC
Guerrero et al. (2014) (36)	HS AD sMCI pMCI	175 (76.3) 106 (75.4) 114 (75.1) 116 (74.7)	ADNI-1 ADNI-GO	3D MRI data	SVM	Train/test	pMCI vs sMCI: 97.1	NA
Lebedev et al. (2014) (37)	HS AD MCI	225 (75.9) 185 (75.2) 165 (75.5)	ADNI-1	MRI: Cortical thickness, demographic variables and APOE4 genotype	RF	Independent sample	pMCI vs sMCI: 82.3	0.83
Liu, M. et al. (2014) (38)	HS AD sMCI pMCI	229 (76.0) 198 (75.7) 236 (74.9) 167 (74.9)	ADNI-1	MRI. Whole-brain GM density	SVM	10-fold Cross Validation	pMCI vs sMCI: 70.7	NA
Liu, F. et al. (2014) (39)	HS AD MCI	52 (75.3) 51 (75.2) 99 (75.3)	ADNI-1	MRI and PET: Volume and intensity measurements	Mk-SVM	10-fold Cross Validation	sMCI vs pMCI: 67.8	0.69
Min et al. (2014) (40)	HS AD sMCI pMCI	128 (76.1) 97 (75.9) 117 (75.0) 117 (75.2)	ADNI-1	MRI: Multi-atlas GM volume measurements	SVM	10-fold Cross Validation	pMCI vs sMCI: 72.4	0.67
Suk et al. (2014) (41)	HS AD MCI	101 (75.9) 93 (75.5) 204 (74.9)	ADNI-1	MRI and PET: Volume and intensity measurements	DBM	10-fold Cross Validation	pMCI vs sMCI: 75.9	0.74

*Note.* AUC = Area Under the Curve; AD = Alzheimer's Disease; HS = Healthy Subjects; MCI = Mild Cognitive Impairment; IMCI = late MCI; eMCI = early MCI; pMCI = progressive MCI; sMCI = stable MCI; WM = White Matter; GM = Grey Matter; CNN = Convolutional Neural Network; rDNN = randomized Deep Neural Network; FCS = Functional Connectivity Strength; RF = Random Forest; SVM = Support Vector Machine; EN = Elastic Nets; AB = Ada-Boost; nl-SVM-RBFk = non-linear SVM with Radial Basis Function kernel; SNP = Single Nucleotide Polymorphisms; GPR = Gaussian Process Regression; PLS = Partial Least Squares; OPLS = Orthogonal Partial Least Squares; MMSE = Mini Mental State Examination; VFI = Voting Feature Intervals; NN = Neural Network; AD-NET = Age-Adjust Neural Network; Res-Net = Deep Residual Neural Network; SNN = Spiking Neural Network; EL = Ensemble Learning; RLR = Regularized Logistic Regression; ; F-FDG = Fluorine 18 fluorodesoxyglucose; DAT = Dementia Alzheimer Type; IDAT = late DAT; eDAT = early DAT; FPDS = FDG-PET DAT Score; ICA = Independent Component Analysis; GMB = Gradient Boosting Model; M2TL = Multimodal manifold-regularized transfer learning; ss = sample selection; M3TL = Multi-Modal Multi-Task Learning; DBM = Deep Boltzmann Machine; ELM = Extreme Learning Machine; MG2G = Multiple Graph2Gauss; MCFS = Multi-Cluster Feature Selection; HLR = Hierarchical Logistic Regression; NA = Not Applicable.

Author (year)	Groups	Sample size (mean age)	Database	Neuroimaging technique and features	Classification method	Validation method	Results (% accuracy)	AUC ROC
Cheng et al. (2015) (42)	HS AD sMCI pMCI	52 (NA) 51 (NA) 56 (NA) 53 (NA)	ADNI-1	MRI and PET: Volume, intensity and CSF ( $A\beta_{42}$ , t-tau y p-tau) measurements	M2TL	10-fold Cross Validation	sMCI vs pMCI:80.1	0.85
Moradi et al. (2015) (43)	HS AD sMCI pMCI	231 (NA) 200 (NA) 100 (NA) 164 (NA)	ADNI-1	MRI: GM volumes, age and cognitive measures	RF	10-fold Cross Validation	pMCI vs sMCI: 82	0.90
Ritter et al. (2015) (44)	sMCI pMCI	151 (74.1) 86 (74.6)	ADNI-1	MRI and PET: Neuropsychological test, clinical variables, cortical thickness, demographic data and intensity measurements	SVM with RBFk Classification tree RF	30 iterations of 10-fold Cross Validation	sMCI vs pMCI: SVM: 66.5 Classification Tree: 66.1 RF: 63.1	NA
Salvatore et al. (2015) (45)	HS AD sMCI pMCI	162 (76.3) 137 (76.0) 134 (74.5) 76 (74.8)	ADNI-1	MRI: GM and WM volumes	SVM	20-fold Cross Validation	pMCI vs sMCI: 66	NA
Collij et al. (2016) (46)	HS AD MCI	100 (66.7) 100 (63.2) 60 (62.8)	Sample collected for the study	MRI: Cortical thickness	SVM	Train/test method: 50/50 partition	pMCI vs sMCI: 70.8	0.77
Li et al. (2016) (47)	HS AD sMCI pMCI	42 (65.6) 25 (69.4) 10 (66.5) 21 (68.6)	ADNI-1	MRI: GM whole-brain FCS and functional data	SVM	Leave One Out Cross Validation	pMCI vs SMCI: 80.6	NA

*Note.* AUC = Area Under the Curve; AD = Alzheimer's Disease; HS = Healthy Subjects; MCI = Mild Cognitive Impairment; IMCI = late MCI; eMCI = early MCI; pMCI = progressive MCI; sMCI = stable MCI; WM = White Matter; GM = Grey Matter; CNN = Convolutional Neural Network; rDNN = randomized Deep Neural Network; FCS = Functional Connectivity Strength; RF = Random Forest; SVM = Support Vector Machine; EN = Elastic Nets; AB = Ada-Boost; nl-SVM-RBFk = non-linear SVM with Radial Basis Function kernel; SNP = Single Nucleotide Polymorphisms; GPR = Gaussian Process Regression; PLS = Partial Least Squares; OPLS = Orthogonal Partial Least Squares; MMSE = Mini Mental State Examination; VFI = Voting Feature Intervals; NN = Neural Network; AD-NET = Age-Adjust Neural Network; Res-Net = Deep Residual Neural Network; SNN = Spiking Neural Network; EL = Ensemble Learning; RLR = Regularized Logistic Regression; ; F-FDG = Fluorine 18 fluorodesoxyglucose; DAT = Dementia Alzheimer Type; IDAT = late DAT; eDAT = early DAT; FPDS = FDG-PET DAT Score; ICA = Independent Component Analysis; GMB = Gradient Boosting Model; M2TL = Multimodal manifold-regularized transfer learning; ss = sample selection; M3TL = Multi-Modal Multi-Task Learning; DBM = Deep Boltzmann Machine; ELM = Extreme Learning Machine; MG2G = Multiple Graph2Gauss; MCFS = Multi-Cluster Feature Selection; HLR = Hierarchical Logistic Regression; NA = Not Applicable.

Author (year)	Groups	Sample size (mean age)	Database	Neuroimaging technique and features	Classification method	Validation method	Results (% accuracy)	AUC ROC
López et al. (2016) (48)	sMCI  pMCI	21 (72.7) 12 (75.7)	Sample collected for the study	MRI and MEG: Cognitive reserve; APOE genotype; hippocampal volumes; 3D MRI data; MEG recordings; neuropsychological tests	HLR	Train/test method: 75/25 partition	pMCI vs sMCI: 95.5	0.97
Thung et al. (2016) (49)	sMCI pMCI	53 (75.7) 60 (75.2)	ADNI-1	MRI: Whole-brain GM volume measures and changes in 4 years of follow-up	Mk-SVM	10-fold Cross Validation	pMCI vs sMCI: 78.2	0.84
Long et al. (2017) (50)	HS AD sMCI pMCI	135 (76.2) 65 (75.6) 132 (75.2) 95 (75.1)	ADNI-1	MRI: Whole-brain GM and Whole-brain WM	SVM	10-fold Cross Validation	pMCI vs sMCI: 96.5 with GM: 96.5 with WM: 96.0	GM: 0.99 WM: 0.99
Donnelly-Kehoe et al. (2018) (51)	HS AD sMCI pMCI	100 (NA) 100 (NA) 100 (NA) 100 (NA)	ADNI-1	MRI: Demographic, Morphometric and MMSE	RF SVM AB	Train/test method: 75/25 partition	NA	0.75 0.76 0.62
Gao et al. (2018) (52)	HS AD MCI	94 (76.3) 58 (74.2) 147 (74.8)	ADNI-1	MRI and PET: Hippocampus measurement, Medical history, Neuropsychological tests and Volume-based morphometry	GPR PLS	Train/test method: AD + HS as train set and MCI as test set + follow-up	sMCI vs pMCI GPR:82.2 PLS:85.5	NA

*Note.* AUC = Area Under the Curve; AD = Alzheimer's Disease; HS = Healthy Subjects; MCI = Mild Cognitive Impairment; IMCI = late MCI; eMCI = early MCI; pMCI = progressive MCI; sMCI = stable MCI; WM = White Matter; GM = Grey Matter; CNN = Convolutional Neural Network; rDNN = randomized Deep Neural Network; FCS = Functional Connectivity Strength; RF = Random Forest; SVM = Support Vector Machine; EN = Elastic Nets; AB = Ada-Boost; nl-SVM-RBFk = non-linear SVM with Radial Basis Function kernel; SNP = Single Nucleotide Polymorphisms; GPR = Gaussian Process Regression; PLS = Partial Least Squares; OPLS = Orthogonal Partial Least Squares; MMSE = Mini Mental State Examination; VFI = Voting Feature Intervals; NN = Neural Network; AD-NET = Age-Adjust Neural Network; Res-Net = Deep Residual Neural Network; SNN = Spiking Neural Network; EL = Ensemble Learning; RLR = Regularized Logistic Regression; ; F-FDG = Fluorine 18 fluorodesoxyglucose; DAT = Dementia Alzheimer Type; IDAT = late DAT; eDAT = early DAT; FPDS = FDG-PET DAT Score; ICA = Independent Component Analysis; GMB = Gradient Boosting Model; M2TL = Multimodal manifold-regularized transfer learning; ss = sample selection; M3TL = Multi-Modal Multi-Task Learning; DBM = Deep Boltzmann Machine; ELM = Extreme Learning Machine; MG2G = Multiple Graph2Gauss; MCFS = Multi-Cluster Feature Selection; HLR = Hierarchical Logistic Regression; NA = Not Applicable.

Author (year)	Groups	Sample size (mean age)	Database	Neuroimaging technique and features	Classification method	Validation method	Results (% accuracy)	AUC ROC
Khanna et al. (2018) (53)	HS MCI	315 (NA) 609 (NA)	ADNI-1	MRI and PET: Volume, clinical and SNP measures	GBM	10 iterations of a 10-fold Cross Validation	C-index (it's a generalization of the AUC ROC calculation for binary classification): 0.86	0.95
Popuri et al. (2018) (54)	sHS uHS pSH pMCI sMCI eDAT IDAT	753 (75.4) 110 (78.9) 58 (78.2) 486 (74.8) 881 (75.0) 232 (76.6) 464 (75.8)	ADNI-1	PET: Glucose metabolic signal	FPDS	Independent group	Classification of DAT+/DAT-: sMCI = 70.4 pMCI = 67.9	sMCI vs pMCI at 2, 3 and 5 years: 0.80 0.79 0.77
Gupta et al. (2019) (55)	HS AD sMCI pMCI	38 (76.7) 38 (77.1) 36 (74.2) 46 (76.1)	ADNI-1	MRI and PET: Volume, intensity and CSF ( $A\beta_{42}$ , t-tau y p-tau) measurements	Mk-SVM	10-fold Cross Validation	pMCI vs sMCI: 93.6	0.95
Lee et al. (2019) (56)	HS AD sMCI pMCI	229 (75.9) 198 (75.3) 214 (75.0) 160 (74.9)	ADNI-1	MRI: GM volumes	rDNN + SVM	10-fold Cross Validation	pMCI vs sMCI: 88.5	0.95
Pusil et al. (2019) (57)	sMCI pMCI	27 (71.2) 27 (74.8)	Sample collected for the study	MEG: Brain connectivity matrix	MCFS + SVM with RBF kernel	Train /test method: 80/20 partition	pMCI vs sMCI: 100	NA

*Note.* AUC = Area Under the Curve; AD = Alzheimer's Disease; HS = Healthy Subjects; MCI = Mild Cognitive Impairment; IMCI = late MCI; eMCI = early MCI; pMCI = progressive MCI; sMCI = stable MCI; WM = White Matter; GM = Grey Matter; CNN = Convolutional Neural Network; rDNN = randomized Deep Neural Network; FCS = Functional Connectivity Strength; RF = Random Forest; SVM = Support Vector Machine; EN = Elastic Nets; AB = Ada-Boost; nl-SVM-RBFk = non-linear SVM with Radial Basis Function kernel; SNP = Single Nucleotide Polymorphisms; GPR = Gaussian Process Regression; PLS = Partial Least Squares; OPLS = Orthogonal Partial Least Squares; MMSE = Mini Mental State Examination; VFI = Voting Feature Intervals; NN = Neural Network; AD-NET = Age-Adjust Neural Network; Res-Net = Deep Residual Neural Network; SNN = Spiking Neural Network; EL = Ensemble Learning; RLR = Regularized Logistic Regression; ; F-FDG = Fluorine 18 fluorodesoxyglucose; DAT = Dementia Alzheimer Type; IDAT = late DAT; eDAT = early DAT; FPDS = FDG-PET DAT Score; ICA = Independent Component Analysis; GMB = Gradient Boosting Model; M2TL = Multimodal manifold-regularized transfer learning; ss = sample selection; M3TL = Multi-Modal Multi-Task Learning; DBM = Deep Boltzmann Machine; ELM = Extreme Learning Machine; MG2G = Multiple Graph2Gauss; MCFS = Multi-Cluster Feature Selection; HLR = Hierarchical Logistic Regression; NA = Not Applicable.

Author (year)	Groups	Sample size (mean age)	Database	Neuroimaging technique and features	Classification method	Validation method	Results (% accuracy)	AUC ROC
Spasov et al. (2019) (58)	HS AD	184 (74.6) 192 (75.6) 181 (73.7) 228 (72.2)	ADNI-1	MRI: 3D data, demographic, neuropsychological, and biological (APOE4) measures	CNN	Train/test method: 90/10 partition	pMCI vs sMCI: 86	0.92
	sMCI pMCI							
Wee et al. (2019) (59)	HS AD MCI eMCI IMCI	ADNI-1/ADNI-2: 242/300 (76.9/75.6) 355/261 (76.3/75.3) 415/NA (75.9) NA/314 (72.9) NA/208 (73.7)		Cortical thickness	Graph NN	10-fold Cross Validation	Conversion from: IMCI to AD: 75 eMCI to AD: 92	NA
Abrol et al. (2020) (60)	HS AD sMCI pMCI	237 (74.3) 157 (75.1) 245 (72.1) 189 (74.2)	ADNI-1 ADNI-2 ADNI-3 ADNI-GO	3D MRI data	ResNET	Train/test method: 80/20 partition	pMCI vs sMCI: 77.8	0.78
Gao et al. (2020) (61)	HS sMCI pMCI	847 (56.9) 129 (74.8) 168 (74.8)	ADNI-1	3D MRI data	Age prediction + AD-NET	5-fold Cross Validation	pMCI vs sMCI; 76	0.81

*Note.* AUC = Area Under the Curve; AD = Alzheimer's Disease; HS = Healthy Subjects; MCI = Mild Cognitive Impairment; IMCI = late MCI; eMCI = early MCI; pMCI = progressive MCI; sMCI = stable MCI; WM = White Matter; GM = Grey Matter; CNN = Convolutional Neural Network; rDNN = randomized Deep Neural Network; FCS = Functional Connectivity Strength; RF = Random Forest; SVM = Support Vector Machine; EN = Elastic Nets; AB = Ada-Boost; nl-SVM-RBFk = non-linear SVM with Radial Basis Function kernel; SNP = Single Nucleotide Polymorphisms; GPR = Gaussian Process Regression; PLS = Partial Least Squares; OPLS = Orthogonal Partial Least Squares; MMSE = Mini Mental State Examination; VFI = Voting Feature Intervals; NN = Neural Network; AD-NET = Age-Adjust Neural Network; Res-Net = Deep Residual Neural Network; SNN = Spiking Neural Network; EL = Ensemble Learning; RLR = Regularized Logistic Regression; ; F-FDG = Fluorine 18 fluorodesoxyglucose; DAT = Dementia Alzheimer Type; IDAT = late DAT; eDAT = early DAT; FPDS = FDG-PET DAT Score; ICA = Independent Component Analysis; GMB = Gradient Boosting Model; M2TL = Multimodal manifold-regularized transfer learning; ss = sample selection; M3TL = Multi-Modal Multi-Task Learning; DBM = Deep Boltzmann Machine; ELM = Extreme Learning Machine; MG2G = Multiple Graph2Gauss; MCFS = Multi-Cluster Feature Selection; HLR = Hierarchical Logistic Regression; NA = Not Applicable.

Author (year)	Groups	Sample size (mean age)	Database	Neuroimaging technique and features	Classification method	Validation method	Results (% accuracy)	AUC ROC
Giorgio et al. (2020) (62)	HS MCI	167 (NA) 167 (NA)	ADNI-1	MRI and PET: GM density;  Biological and cognitive measurements	GMLVQ	10-fold Cross Validation	pMCI vs sMCI: 81.4	NA
Lin et al. (2020) (63)	HS AD sMCI pMCI	200 (73.9) 102 (75.7) 205 (71.8) 110 (73.9)	ADNI-1	MRI and PET: Volume, cortical thickness, intensity measurements, APOE4 presence and levels of Aβ42, T-tau and P-tau in CSF	LASSO + ELM with Gaussian kernel	10-fold Cross Validation	sMCI vs pMCI: 84.7	0.88
Pan et al. (2020) (64)	HS AD sMCI pMCI	262 (74.5) 237 (76.0) 175 (74.5) 115 (74.8)	ADNI-1	2D MRI data	CNN + EL	5-fold Cross Validation on independent sample	pMCI vs sMCI: 62	0.59
Ramon-Julvez et al. (2020) (65)	HS AD sMCI pMCI	181 (NA) 191 (NA) 227 (NA) 179 (NA)	ADNI-1	MRI data and Jacobian determinant of diffeomorphic transformations	CNN	10-fold Cross Validation	sMCI vs pMCI: 89	0.94
Xiao et al. (2020) (66)	HS AD sMCI pMCI	50 (77.8) 51 (75.8) 45 (71.9) 51 (72.5)	ADNI-1	MRI: GM volumes	Logistic Regression	10-fold Cross Validation	pMCI vs sMCI: 72.9	NA
Xu et al. (2020) (67)	HS MCI	53 (69.6) 76 (73.7)	Sample collected for the study	MEG: Brain connectivity matrix	MG2G Embedding model	Train/validation/test method: 85/10/5 partition	HS vs pMCI vs sMCI: 82 pMCI vs sMCI: 87	0.75–0.96
<p><i>Note.</i> AUC = Area Under the Curve; AD = Alzheimer's Disease; HS = Healthy Subjects; MCI = Mild Cognitive Impairment; IMCI = late MCI; eMCI = early MCI; pMCI = progressive MCI; sMCI = stable MCI; WM = White Matter; GM = Grey Matter; CNN = Convolutional Neural Network; rDNN = randomized Deep Neural Network; FCS = Functional Connectivity Strength; RF = Random Forest; SVM = Support Vector Machine; EN = Elastic Nets; AB = Ada-Boost; nl-SVM-RBFk = non-linear SVM with Radial Basis Function kernel; SNP = Single Nucleotide Polymorphisms; GPR = Gaussian Process Regression; PLS = Partial Least Squares; OPLS = Orthogonal Partial Least Squares; MMSE = Mini Mental State Examination; VFI = Voting Feature Intervals; NN = Neural Network; AD-NET = Age-Adjust Neural Network; Res-Net = Deep Residual Neural Network; SNN = Spiking Neural Network; EL = Ensemble Learning; RLR = Regularized Logistic Regression; ; F-FDG = Fluorine 18 fluorodesoxyglucose; DAT = Dementia Alzheimer Type; IDAT = late DAT; eDAT = early DAT; FPDS = FDG-PET DAT Score; ICA = Independent Component Analysis; GMB = Gradient Boosting Model; M2TL = Multimodal manifold-regularized transfer learning; ss = sample selection; M3TL = Multi-Modal Multi-Task Learning; DBM = Deep Boltzmann Machine; ELM = Extreme Learning Machine; MG2G = Multiple Graph2Gauss; MCFS = Multi-Cluster Feature Selection; HLR = Hierarchical Logistic Regression; NA = Not Applicable.</p>								

Finally, regarding the interpretability analysis, Table 4 shows that most of the studies presented results of specificity and sensitivity (44 out of 47 studies), all the studies performed a stability measurement of their model, only four studies did not compared their results with the existing

literature, and only seven did not specify which features were the most important for the classification task. On the other hand, only 19 studies presented their results along with some kind of visualization of the most relevant brain areas for the prediction of MCI conversion. Finally, only 17 articles made an analysis of confounds and Lebedev et al. (37) were the only group that complied with the generalizability sublevel, testing their model in different cohorts.

Table 4

*Analysis of the interpretability based on Kohoutová et al. (21)*

Author (year)	Model			Feature			Biography
	SEN/SPE	GN	AC	ST	IMP	VIS	LIT
Plant et al. (2010) (23)	√	-	√	√	√	-	√
Costafreda et al. (2011) (24)	√	-	√	√	√	√	√
Chincarini et al. (2011) (25)	√	-	√	√	√	-	√
Filipovych et al. (2011) (26)	√	-	-	√	√	√	√
Hinrichs et al. (2011) (8)	√	-	-	√	√	√	√
Westman et al. (2011) (37)	√	-	-	√	√	√	√
Zhang et al. (2011) (28)	-	-	-	√	√	√	√
Cho et al. (2012) (29)	√	-	-	√	√	√	√
Gray et al. (2012) (30)	√	-	√	√	√	√	√
Toussaint et al. (2012) (31)	√	-	√	√	√	√	√
Zhang et al. (2012) (9)	√	-	-	√	√	√	√
Casanova et al. (2013) (32)	√	-	-	√	√	-	√
Liu, X. et al. (2013) (33)	√	-	-	√	√	√	√
Wee et al. (2013) (34)	√	-	-	√	√	-	√
Young et al. (2013) (35)	√	-	-	√	-	-	-
Guerrero et al. (2014) (36)	√	-	√	√	√	-	√
Lebedev et al. (2014) (37)	√	√	-	√	√	√	√
Liu, M. et al. (2014) (38)	√	-	-	√	√	√	√
Liu, F. et al. (2014) (39)	√	-	-	√	-	-	-
Min et al. (2014) (40)	√	-	√	√	-	-	-
Suk et al. (2014) (41)	√	-	√	√	√	-	√
Cheng et al. (2015) (42)	√	-	√	√	-	-	-
Moradi et al. (2015) (43)	√	-	-	√	√	-	√
Ritter et al. (2015) (44)	√	-	-	√	√	-	√
Salvatore et al. (2015) (45)	√	-	-	√	√	√	√
Collij et al. (2016) (46)	√	-	√	√	√	√	√
Li et al. (2016) (47)	√	-	√	√	√	√	√
López et al. (2016) (48)	√	-	-	√	√	-	√
Thung et al. (2016) (49)	√	-	√	√	√	-	√
Long et al. (2017) (50)	√	-	-	√	√	-	√

*Note.* This table shows an interpretability analysis performed for each study selected in our review following the framework proposed in Kohoutová et al. (2020). Presence (√) or absence (-) of the different sublevels assessments. Behavioural Analysis, Representational Analysis and Invasive studies sub-levels are not applicable to this type of studies by its definition. SEN = Sensitivity; SPE = Specificity; GN = Generalizability; AC = Analysis of confounds; ST = Stability; IMP = Importance; VIS = Visualization; LIT = Literature.

	Model			Feature			Biography
Donnelly-Kehoe et al. (2018) (51)	√	-	√	√	√	-	√
Gao et al. (2018) (52)	√	-	-	√	√	-	√
Khanna et al. (2018) (53)	-	-	√	√	√	-	√
Popuri et al. (2018) (54)	√	-	√	√	√	-	√
Gupta et al. (2019) (55)	√	-	-	√	√	-	√
Lee et al. (2019) (56)	√	-	-	√	√	√	√
Pusil et al. (2019) (57)	√	-	-	√	√	-	√
Spasov et al. (2019) (58)	√	-	-	√	√	-	√
Wee et al. (2019) (59)	√	-	√	√	√	√	√
Abrol et al. (2020) (60)	√	-	√	√	√	√	√
Gao et al. (2020) (61)	√	-	-	√	-	-	√
Giorgio et al. (2020) (62)	√	-	-	√	√	-	√
Lin et al. (2020) (63)	√	-	-	√	√	-	√
Pan et al. (2020) (64)	-	-	-	√	√	-	√
Ramon-Julvez et al. (2020) (65)	√	-	-	√	-	-	√
Xiao et al. (2020) (66)	√	-	-	√	-	-	√
Xu et al. (2020) (67)	√	-	-	√	√	√	√
Total (48)	44	1	17	47	40	19	43
<i>Note.</i> This table shows an interpretability analysis performed for each study selected in our review following the framework proposed in Kohoutová et al. (2020). Presence (√) or absence (-) of the different sublevels assessments. Behavioural Analysis, Representational Analysis and Invasive studies sub-levels are not applicable to this type of studies by its definition. SEN = Sensitivity; SPE = Specificity; GN = Generalizability; AC = Analysis of confounds; ST = Stability; IMP = Importance; VIS = Visualization; LIT = Literature.							

## 4 Discussion

In this systematic review, we analysed 47 studies, conducted over the last ten years, which used neuroimaging data to predict conversion to AD from MCI using ML algorithms.

The complexity of neuroimaging results and the amplitude of the deterioration and symptoms present in multiple areas and functions in AD, make its prediction practically impossible in patients with MCI by simply visualizing an image or result of MRI, PET, or other neuroimaging techniques. Nevertheless, using the publicly available neuroimaging data collected over the last decades, together with the newly developed ML algorithms, researchers can not only distinguish the brains of AD patients and healthy people with high accuracy, but can also predict MCI patient's disease progression (i.e., whether a MCI patient will progress to AD or remain stable in the future). This information is highly valuable for clinicians in order to achieve more accurate diagnoses and therefore set treatment plans that can slow down the progression of the disease and prevent higher degrees of cognitive impairment.

The 47 studies analysed reached different levels of accuracy using classification methods based on ML algorithms. Only seven studies (44, 49, 53, 57, 61, 62, 67) focused exclusively on predicting MCI conversion, most studies also tried to find the main differences between healthy controls and AD patients. The specific search for AD biomarkers is much more abundant in the literature than predicting progression from a MCI or even from healthy subjects (71). In any case, in the studies that carried out both tasks and in studies that focused on the prediction of AD conversion, the distinction between controls and AD was always more accurate than the distinction between pMCI vs. sMCI, showing the difficulty of finding biomarkers before the characteristic symptoms of AD-related neurodegeneration appear.

One of the main challenges of this review was to compare studies with highly variable methodologies including different samples, preprocessing techniques, types of neuroimaging data, and also different classification and validation methods. Still, studies that achieved higher levels of accuracy have in common the use of multimodal and multidimensional data combined with increasingly complex classification methods. Easy-to-implement algorithms, such as those based on SVM, are leaving room to more complex algorithms based on Deep learning paradigms such as Neural Networks, capable of identifying dementia-associated subtle changes of brain morphology in a way

able to increase the number of correctly classified subjects. All methods seemed to benefit from the inclusion of demographic variables and cognitive measurements, and even genetic variables if these were available. Nevertheless, in order for these techniques to be able to help clinicians in their everyday practice, a balance is needed between the most advanced data and algorithms that achieve the higher performance, and the data and methods that might be available in the clinical practice. In this sense, future studies might need to focus more on achieving high performance using large datasets with more essential (and easily obtainable) data such as structural MRI, demographic, and screening cognitive results.

Regarding the sample, most studies use the public available data from the ADNI database. This database is still incorporating new data and the most recent studies even use ADNI-2 and ADNI-3 (59, 60). The main problem of the studies performed ten years ago, is their smaller sample size. Furthermore, even if two studies report similar accuracies, a study with a bigger sample size will have results that are more generalizable. For example, Plant et al. (23) and Popuri et al. (54) obtained similar accuracies of 75% and 79% of correct classifications respectively, but Popuri et al. (54) used a sample about 30 times larger.

Upon review of the recent developments in the field, it is apparent that as these patient populations continue to age and their disease progression matures and is clinically diagnosed, the inclusion of their ongoing neuroimaging data will enrich these public databases thereby enabling improved validation of classification methods.

Nevertheless, it should be noted that a follow-up of two or three years, might not be long enough to detect progression to AD. Therefore, subjects considered as stable MCI or even as healthy subjects might, in fact, progress to AD in the long term. This problem will always be present with the inclusion of new cases in the ADNI database, but the follow-ups recorded will be increasingly long-lasting, thus being more useful. On the other hand, in order for these methods to be clinically useful, the models have to be tested, not only in big samples, but also in more variable and diverse groups of people, other than the ADNI sample. This approach is highlighted in Lebedev et al. (37), which applied the model to the ADNI and then to the AddNeuroMed cohort, achieving similar performance and accuracy results in both, making it more robust for future clinical implementation.

The brain areas selected as important for the discrimination of AD patients from healthy subjects or sMCI have been mainly located in the temporal lobe such as the hippocampus, amygdala, entorhinal cortex, or cingulate gyrus, some parietal areas such as the precuneus, and the rostral and caudal areas of the medial frontal lobe. These regions have been widely validated by the scientific literature as relevant in the progression to AD (8, 29, 30, 72–76). This coincidence between the literature and the algorithm results supports the notion that the classification methods can detect differences between groups based on relevant neuroimaging features.

In terms of accuracy, although the algorithms are useful and able to discriminate the brain characteristics of AD, the performance of the algorithms are far from being specific enough to leave complete diagnosis in the hands of automated methods, so the judgment of a clinical professional will remain crucial in the near future. Nevertheless, the automated methods discussed above present a low-cost approach that can be useful as a first approximation, a method to discriminate ambiguous cases, and as a support tool for large datasets.

Clinical research is moving towards a broader and more open context where professionals from very different disciplines might be interested in these types of studies. As such, it is important to present the results from complex neuroimaging classification studies as clearly as possible. The framework to interpret ML models provided by Kohoutová et al. (21) is a helpful starting point for this purpose. Most or all of the studies reviewed here included information about the specificity and sensitivity (model level), the stability of the models, and the most important features selected (feature level), along with a comparison with the previous literature (biology level). However, there are some important issues that should be addressed in future studies such as the inclusion of visualizations of the most relevant brain areas to predict MCI conversion, an adequate analysis of confounds, and generalization methods. These specific improvements would provide more comprehensive and comparable studies.

Regarding the limitations of the review, it is worth mentioning that we did not include methodological details such as the preprocessing methods to obtain the neuroimaging results, or the mathematical development of the algorithms. This information could have provided a better understanding of each model performance and how the data is classified to differentiate between groups, but these deep methodological analyses were out of the scope of this review given its more clinically-oriented focus

## 5 Conclusions

The recent trend in research to find diagnostic automation methods presents great potential in the early detection of neurodegenerative diseases. Since structural changes appear before the clinical symptoms manifest, there is a valuable period of time in which the morphological and functional changes in the brain can be detected and, therefore, used to predict and provide clinical treatment to slow down the future development of a neurological disease.

Research in this field is still rapidly advancing, new increasingly complex algorithms continue to be developed, and access to higher levels of computational capacity is also increasing, as well as the precision and resolution of neuroimaging techniques. In the future, we can expect faster, more precise, and more efficient classification methods that may be directly incorporated into the neuroimaging techniques themselves that enable the generation of a diagnostic hypothesis with a simple scan of a patient's brain. However, the challenge to translate this knowledge to daily practice remains. This challenge will be overcome on one hand by increasing the generalizability of the classification methods as they are applied to more diverse samples; and, on the other, by finding the trade-off between the higher precision achieved when including complex information and a sufficient performance using only the clinical data commonly available for the clinicians.

## **Abbreviations**

AUC: Area Under the Curve

AD: Alzheimer's Disease

ADNI: Alzheimer's Disease Neuroimage Initiative

HC: Healthy Controls

MCI: Mild Cognitive Impairment

ML: Machine Learning

MRI: Magnetic Resonance Image

PET: Positron Emission Tomography

pMCI: Progressive Mild Cognitive Impairment

ROC: Receiver Operating Characteristic

sMCI: Stable Mild Cognitive Impairment

SVM: Support Vector Machine

## **Declarations**

### **Ethics approval and consent to participate:**

Not applicable

### **Consent for publication:**

Not applicable

### **Availability of data and materials:**

Not applicable

### **Competing interests:**

The authors declare that they have no competing interests

### **Funding:**

Not applicable

### **Author's contributions:**

SG conceived the original idea for the review, designed the search strategy, performed the search, article selection, and data extraction, also wrote the first draft of the manuscript and contributed to the subsequent reviews and final version. RVS supervised the review process, made an independent article screening, revised the manuscript, and contributed to writing of the final version.

## **Acknowledgements:**

Not applicable

## **References**

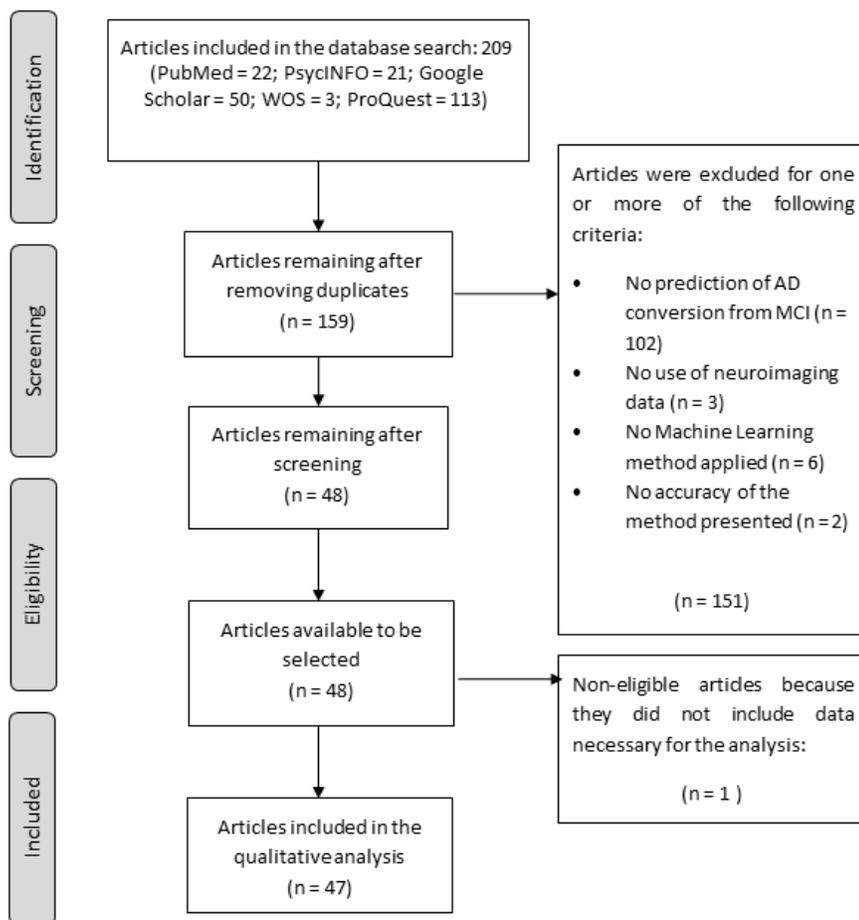
1. Menéndez G. La revolución de la longevidad : cambio tecnológico, envejecimiento poblacional y transformación cultural. Rev Ciencias Soc. 2017;
2. Prince M, Wimo A, Guerchet M, Gemma-Claire A, Wu Y-T, Prina M. World Alzheimer Report 2015: The Global Impact of Dementia - An analysis of prevalence, incidence, cost and trends. *Alzheimer's Dis Int*. 2015;
3. Dementia [Internet]. [cited 2021 Feb 9]. Available from: [https://www.who.int/health-topics/dementia#tab=tab\\_1](https://www.who.int/health-topics/dementia#tab=tab_1)
4. APA. American Psychiatric Association, 2013. Diagnostic and statistical manual of mental disorders (5th ed.). American Journal of Psychiatry. 2013.
5. Robinson L, Tang E, Taylor J-P. Dementia: timely diagnosis and early intervention. *BMJ* [Internet]. 2015 Jun 16 [cited 2019 Dec 21];350(jun15 14):h3029–h3029. Available from: <http://www.bmj.com/cgi/doi/10.1136/bmj.h3029>
6. Gauthier S, Reisberg B, Zaudig M, Petersen RC, Ritchie K, Broich K, et al. International Psychogeriatric Association Expert Conference on mild cognitive impairment. *Lancet* (London, England). 2006;
7. Braak H, Braak E. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathologica*. 1991.
8. Hinrichs C, Singh V, Xu G, Johnson SC. Predictive markers for AD in a multi-modality framework: An analysis of MCI progression in the ADNI population. *Neuroimage*. 2011 Mar 15;55(2):574–89.
9. Zhang D, Shen D. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. *Neuroimage*. 2012 Jan 16;59(2):895–907.
10. Pellegrini E, Ballerini L, Hernandez M del CV, Chappell FM, González-Castro V, Anblagan D, et al. Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: A systematic review. *Alzheimer's Dement Diagnosis, Assess Dis Monit*. 2018;
11. Samper-González J, Burgos N, Fontanella S, Bertin H, Habert MO, Durrleman S, et al. Yet another ADNI machine learning paper? Paving the way towards fully-reproducible research on classification of Alzheimer's disease. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2017.
12. Davatzikos C, Fan Y, Wu X, Shen D, Resnick SM. Detection of prodromal Alzheimer's disease via pattern classification of magnetic resonance imaging. *Neurobiol Aging*. 2008;
13. Klöppel S, Chu C, Tan GC, Draganski B, Johnson H, Paulsen JS, et al. Automatic detection of preclinical neurodegeneration: Presymptomatic Huntington disease. *Neurology*. 2009;
14. Lao Z, Shen D, Xue Z, Karacali B, Resnick SM, Davatzikos C. Morphological classification of brains via high-dimensional shape transformations and machine learning methods. *Neuroimage*. 2004;
15. Arbabshirani MR, Plis S, Sui J, Calhoun VD. Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *Neuroimage* [Internet]. 2017 Jan 15 [cited 2021 Feb 9];145(Pt B):137–65. Available from: </pmc/articles/PMC5031516/>
16. Jo T, Nho K, Saykin AJ. Deep Learning in Alzheimer's Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data. *Front Aging Neurosci* [Internet]. 2019 Aug 20 [cited 2021 Jan 25];11. Available from: <https://pubmed.ncbi.nlm.nih.gov/31481890/>
17. Sarica A, Cerasa A, Quattrone A. Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: A systematic review. *Front Aging Neurosci*. 2017;9(OCT):1–12.
18. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol*. 2009;
19. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. In: *Journal of clinical epidemiology*. 2009.
20. Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0*. The Cochrane Collaboration. 2011.
21. Kohoutová L, Heo J, Cha S, Lee S, Moon T, Wager TD, et al. Toward a unified framework for interpreting machine-learning models in neuroimaging. *Nat Protoc*. 2020;

22. Capecchi E, Doborjeh ZG, Mammone N, La Foresta F, Morabito FC, Kasabov N. Longitudinal study of Alzheimer's disease degeneration through EEG data analysis with a NeuCube spiking neural network model. In: Proceedings of the International Joint Conference on Neural Networks. 2016.
23. Plant C, Teipel SJ, Oswald A, Böhm C, Meindl T, Mourao-Miranda J, et al. Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease. *Neuroimage*. 2010 Mar;50(1):162–74.
24. Costafreda SG, Dinov ID, Tu Z, Shi Y, Liu CY, Kloszewska I, et al. Automated hippocampal shape analysis predicts the onset of dementia in mild cognitive impairment. *Neuroimage*. 2011 May 1;56(1):212–9.
25. Chincarini A, Bosco P, Calvini P, Gemme G, Esposito M, Olivieri C, et al. Local MRI analysis approach in the diagnosis of early and prodromal Alzheimer's disease. *Neuroimage*. 2011 Sep 15;58(2):469–80.
26. Filipovych R, Davatzikos C. Semi-supervised pattern classification of medical images: Application to mild cognitive impairment (MCI). *Neuroimage*. 2011 Apr 1;55(3):1109–19.
27. Westman E, Simmons A, Zhang Y, Muehlboeck JS, Tunnard C, Liu Y, et al. Multivariate analysis of MRI data for Alzheimer's disease, mild cognitive impairment and healthy controls. *Neuroimage*. 2011 Jan 15;54(2):1178–87.
28. Zhang D, Wang Y, Zhou L, Yuan H, Shen D. Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage*. 2011;
29. Cho Y, Seong JK, Jeong Y, Shin SY. Individual subject classification for Alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data. *Neuroimage*. 2012;
30. Gray KR, Wolz R, Heckemann RA, Aljabar P, Hammers A, Rueckert D. Multi-region analysis of longitudinal FDG-PET for the classification of Alzheimer's disease. *Neuroimage*. 2012 Mar;60(1):221–9.
31. Toussaint PJ, Perlberg V, Bellec P, Desarnaud S, Lacomblez L, Doyon J, 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 Nov 1;63(2):936–46.
32. Casanova R, Hsu FC, Sink KM, Rapp SR, Williamson JD, Resnick SM, et al. Alzheimer's disease risk assessment using large-scale machine learning methods. *PLoS One*. 2013 Nov 8;8(11).
33. Liu X, Tosun D, Weiner MW, Schuff N. Locally linear embedding (LLE) for MRI based Alzheimer's disease classification. *Neuroimage* [Internet]. 2013 Dec [cited 2019 Nov 14];83:148–57. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1053811913006708>
34. Wee CY, Yap PT, Shen D. Prediction of Alzheimer's disease and mild cognitive impairment using cortical morphological patterns. *Hum Brain Mapp*. 2013 Dec;34(12):3411–25.
35. Young J, Modat M, Cardoso MJ, Mendelson A, Cash D, Ourselin S. Accurate multimodal probabilistic prediction of conversion to Alzheimer's disease in patients with mild cognitive impairment. *NeuroImage Clin*. 2013;2(1):735–45.
36. Guerrero R, Wolz R, Rao AW, Rueckert D. Manifold population modeling as a neuro-imaging biomarker: Application to ADNI and ADNI-GO. *Neuroimage*. 2014 Jul 1;94:275–86.
37. Lebedev A V, Westman E, Van Westen GJP, Kramberger MG, Lundervold A, Aarsland D, et al. Random Forest ensembles for detection and prediction of Alzheimer's disease with a good between-cohort robustness. *NeuroImage Clin*. 2014;6:115–25.
38. Liu M, Zhang D, Shen D. Identifying informative imaging biomarkers via tree structured sparse learning for AD diagnosis. *Neuroinformatics*. 2014;12(3):381–94.
39. Liu F, Wee CY, 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 Jan 1;84:466–75.
40. Min R, Wu G, Cheng J, Wang Q, Shen D. Multi-atlas based representations for Alzheimer's disease diagnosis. 2014 [cited 2019 Nov 11];35(10):5052–70. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.22531>
41. Suk H II, Lee SW, Shen D. Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. *Neuroimage*. 2014;
42. Cheng B, Liu M, Suk H II, Shen D, Zhang D. Multimodal manifold-regularized transfer learning for MCI conversion prediction. *Brain Imaging Behav*. 2015 Dec 1;9(4):913–26.
43. Moradi E, Pepe A, Gaser C, Huttunen H, Tohka J. Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *Neuroimage*. 2015 Jan 1;104:398–412.
44. Ritter K, Schumacher J, Weygandt M, Buchert R, Allefeld C, Haynes JD. Multimodal prediction of conversion to Alzheimer's disease based on incomplete biomarkers. *Alzheimer's Dement Diagnosis, Assess Dis Monit*. 2015;1(2):206–15.
45. Salvatore C, Cerasa A, Battista P, Gilardi MC, Quattrone A, Castiglioni I. Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: A machine learning approach. *Front Neurosci*. 2015;9(SEP).

46. Collij LE, Heeman F, Kuijjer JPA, Ossenkoppele R, Benedictus MR, Möller C, et al. Application of machine learning to arterial spin labeling in mild cognitive impairment and Alzheimer disease. *Radiology*. 2016 Dec 1;281(3):865–75.
47. Li Y, Wang X, Li Y, Sun Y, Sheng C, Li H, et al. Abnormal Resting-State Functional Connectivity Strength in Mild Cognitive Impairment and Its Conversion to Alzheimer's Disease. 2016 [cited 2019 Nov 11]; Available from: <http://dx.doi.org/10.1155/2016/4680972>
48. López ME, Turrero A, Cuesta P, López-Sanz D, Bruña R, Marcos A, et al. Searching for Primary Predictors of Conversion from Mild Cognitive Impairment to Alzheimer's Disease: A Multivariate Follow-Up Study / Searching for Primary Predictors of Conversion from MCI to AD. *J Alzheimer's Dis*. 2016;52:133–43.
49. Thung KH, Wee CY, Yap PT, Shen D. Identification of progressive mild cognitive impairment patients using incomplete longitudinal MRI scans. *Brain Struct Funct*. 2016 Nov 1;221(8):3979–95.
50. Long X, Chen L, Jiang C, Zhang L. Prediction and classification of Alzheimer disease based on quantification of MRI deformation. *PLoS One*. 2017 Mar 1;12(3).
51. Donnelly-Kehoe PA, Pascariello GO, Gómez JC. Looking for Alzheimer's Disease morphometric signatures using machine learning techniques. *J Neurosci Methods*. 2018 May 15;302:24–34.
52. Gao N, Tao LX, Huang J, Zhang F, Li X, O'Sullivan F, et al. Contourlet-based hippocampal magnetic resonance imaging texture features for multivariant classification and prediction of Alzheimer's disease. *Metab Brain Dis*. 2018 Dec 1;33(6):1899–909.
53. Khanna S, Domingo-Fernández D, Iyappan A, Emon MA, Hofmann-Apitius M, Fröhlich H. Using Multi-Scale Genetic, Neuroimaging and Clinical Data for Predicting Alzheimer's Disease and Reconstruction of Relevant Biological Mechanisms. *Sci Rep*. 2018 Dec 1;8(1).
54. Popuri K, Balachandar R, Alpert K, Lu D, Bhalla M, Mackenzie IR, et al. Development and validation of a novel dementia of Alzheimer's type (DAT) score based on metabolism FDG-PET imaging. *NeuroImage Clin*. 2018 Jan 1;18:802–13.
55. Gupta Y, Lama RK, Kwon G-R. Prediction and Classification of Alzheimer's Disease Based on Combined Features From Apolipoprotein-E Genotype, Cerebrospinal Fluid, MR, and FDG-PET Imaging Biomarkers. *Front Comput Neurosci [Internet]*. 2019 Oct 16 [cited 2019 Nov 11];13. Available from: <https://www.frontiersin.org/article/10.3389/fncom.2019.00072/full>
56. Lee E, Choi J-S, Kim M, Suk H-I. Toward an interpretable Alzheimer's disease diagnostic model with regional abnormality representation via deep learning. *Neuroimage*. 2019 Nov;202:116113.
57. Pusil S, Dimitriadis SI, López ME, Pereda E, Maestú F. Aberrant MEG multi-frequency phase temporal synchronization predicts conversion from mild cognitive impairment-to-Alzheimer's disease. *NeuroImage Clin*. 2019;
58. Spasov S, Passamonti L, Duggento A, Liò P, Toschi N. A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease. *Neuroimage*. 2019 Apr 1;189:276–87.
59. Wee CY, Liu C, Lee A, Poh JS, Ji H, Qiu A. Cortical graph neural network for AD and MCI diagnosis and transfer learning across populations. *NeuroImage Clin*. 2019 Jan 1;23.
60. Abrol A, Bhattarai M, Fedorov A, Du Y, Plis S, Calhoun V. Deep residual learning for neuroimaging: An application to predict progression to Alzheimer's disease. *J Neurosci Methods*. 2020;
61. Gao F, Yoon H, Xu Y, Goradia D, Luo J, Wu T, et al. AD-NET: Age-adjust neural network for improved MCI to AD conversion prediction. *NeuroImage Clin*. 2020;
62. Giorgio J, Landau SM, Jagust WJ, Tino P, Kourtzi Z. Modelling prognostic trajectories of cognitive decline due to Alzheimer's disease. *NeuroImage Clin*. 2020;
63. Lin W, Gao Q, Yuan J, Chen Z, Feng C, Chen W, et al. Predicting Alzheimer's Disease Conversion From Mild Cognitive Impairment Using an Extreme Learning Machine-Based Grading Method With Multimodal Data. *Front Aging Neurosci*. 2020;
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;
65. Ramon-Julvez U, Hernandez M, Mayordomo E, Adni. Analysis of the Influence of Diffeomorphic Normalization in the Prediction of Stable VS Progressive MCI Conversion with Convolutional Neural Networks. In: *Proceedings - International Symposium on Biomedical Imaging*. 2020.
66. Xiao R, Cui X, Qiao H, Zheng X, Zhang Y. Early diagnosis model of Alzheimer's Disease based on sparse logistic regression. *Multimed Tools Appl [Internet]*. 2020 Sep 25 [cited 2020 Oct 16];1–12. Available from: <https://doi.org/10.1007/s11042-020-09738-0>
67. Xu M, Sanz DL, Garces P, Maestu F, Li Q, Pantazis D. A graph Gaussian embedding method for predicting Alzheimer's disease progression with MEG brain networks. *arXiv*. 2020.
68. Steardo L, Carbone EA, Filippis R de, Pisanu C, Segura-Garcia C, Squassina A, et al. Application of support vector machine on fmri data as biomarkers in schizophrenia diagnosis: A systematic review. *Front Psychiatry*. 2020;

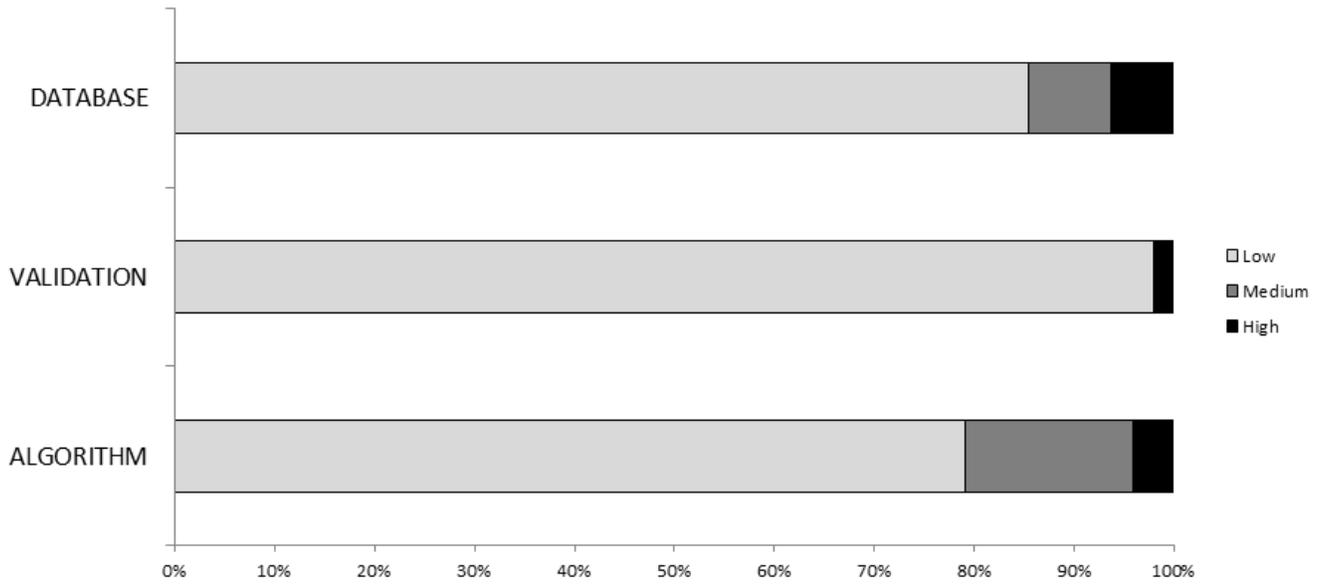
69. Metz CE. Receiver Operating Characteristic Analysis: A Tool for the Quantitative Evaluation of Observer Performance and Imaging Systems. *J Am Coll Radiol.* 2006;
70. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology.* 1982;
71. Ebrahimighahnaveh MA, Luo S, Chiong R. Deep learning to detect Alzheimer's disease from neuroimaging: A systematic literature review. *Comput Methods Programs Biomed.* 2020;
72. Chan D, Fox NC, Scahill RI, Crum WR, Whitwell JL, Leschziner G, et al. Patterns of temporal lobe atrophy in semantic dementia and Alzheimer's disease. *Ann Neurol.* 2001;
73. Fan Y, Batmanghelich N, Clark CM, Davatzikos C. Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. *Neuroimage.* 2008;
74. Jack CR, Petersen RC, Xu Y, O'Brien PC, Smith GE, Ivnik RJ, et al. Rate of medial temporal lobe atrophy in typical aging and Alzheimer's disease. *Neurology.* 1998;
75. Lerch JP, Pruessner J, Zijdenbos AP, Collins DL, Teipel SJ, Hampel H, et al. Automated cortical thickness measurements from MRI can accurately separate Alzheimer's patients from normal elderly controls. *Neurobiol Aging.* 2008;
76. Singh V, Chertkow H, Lerch JP, Evans AC, Dorr AE, Kabani NJ. Spatial patterns of cortical thinning in mild cognitive impairment and Alzheimer's disease. *Brain.* 2006;

## Figures



**Figure 1**

PRISMA workflow for study selection. Note. This diagram shows the PRISMA workflow followed for the sample selection adapted from Moher et al. (18). WOS: Web Of Science; AD: Alzheimer's disease;



**Figure 2**

Risk of bias analysis