HAL
open science

Predicting the Progression of Mild Cognitive Impairment Using Machine Learning: A Systematic, Quantitative and Critical Review

Manon Ansart, Stéphane Epelbaum, Giulia Bassignana, Alexandre Bône, Simona Bottani, Tiziana Cattai, Raphäel Couronné, Johann Faouzi, Igor Koval, Maxime Louis, et al.

To cite this version:

Manon Ansart, Stéphane Epelbaum, Giulia Bassignana, Alexandre Bône, Simona Bottani, et al.. Predicting the Progression of Mild Cognitive Impairment Using Machine Learning: A Systematic, Quantitative and Critical Review. Medical Image Analysis, 2021, 67, pp.101848. 10.1016/j.media.2020.101848 . hal-02337815v2

HAL Id: hal-02337815

https://hal.science/hal-02337815v2

Submitted on 1 Sep 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Predicting the Progression of Mild Cognitive Impairment Using Machine Learning: A Systematic, Quantitative and Critical Review

Manon Ansart[a,b,c,d,e,*], Stéphane Epelbaum[a,b,c,d,e,f], Giulia Bassignana[a,b,c,d,e], Alexandre Bône[a,b,c,d,e], Simona Bottani[a,b,c,d,e], Tiziana Cattai[a,b,c,d,e,1], Raphaël Couronné[a,b,c,d,e], Johann Faouzi[a,b,c,d,e], Igor Koval[a,b,c,d,e], Maxime Louis[a,b,c,d,e], Elina Thibeau-Sutre[a,b,c,d,e], Junhao Wen[a,b,c,d,e], Adam Wild[a,b,c,d,e], Ninon Burgos[a,b,c,d,e], Didier Dormont[a,b,c,d,e,g], Olivier Colliot[a,b,c,d,e,f,g], Stanley Durrleman[e,a,b,c,d]

[a] *Institut du Cerveau et de la Moelle épinière, ICM, F-75013, Paris, France*
[b] *Inserm, U 1127, F-75013, Paris, France*
[c] *CNRS, UMR 7225, F-75013, Paris, France*
[d] *Sorbonne Université, F-75013, Paris, France*
[e] *Inria, Aramis project-team, F-75013, Paris, France*
[f] *Institute of Memory and Alzheimer's Disease (IM2A), Centre of excellence of neurodegenerative disease (CoEN),National Reference Center for Rare or Early Dementias, Department of Neurology, Pitié-Salpêtrière Hospital, AP-HP, Boulevard de l'hôpital, F-75013, Paris, France*
[g] *AP-HP, Pitié-Salpêtrière hospital, Department of Neuroradiology, Paris, France*
[h] *Dept. of Information Engineering, Electronics and Telecommunication, Sapienza University of Rome, Italy*

Abstract

We performed a systematic review of studies focusing on the automatic prediction of the progression of mild cognitive impairment to Alzheimer's disease (AD) dementia, and a quantitative analysis of the methodological choices impacting performance. This review included 172 articles, from which 234 experiments were extracted. For each of them, we reported the used data set, the feature types, the algorithm type, performance and potential methodological issues. The impact of these characteristics on the performance was evaluated using a multivariate mixed effect linear regressions. We found that using cognitive, fluorodeoxyglucose-positron emission tomography or potentially electroencephalography and magnetoencephalography variables significantly improved predictive performance compared to not including them, whereas including other modalities, in particular T1 magnetic resonance imaging, did not show a significant effect. The good performance of cognitive assessments questions the wide use of imaging for predicting the progression to AD and advocates for exploring further fine domain-specific cognitive assessments. We also identified several methodological issues, including the absence of a test set, or its use for feature selection or parameter tuning in nearly a fourth of the papers. Other issues, found in 15% of the studies, cast doubts on the relevance of the method to clinical practice. We also highlight that short-term predictions are likely not to be better than predicting that subjects stay stable over

*Corresponding author
Email address: manon.ansart@inria.fr (Manon Ansart)

time. These issues highlight the importance of adhering to good practices for the use of machine learning as a decision support system for the clinical practice.

Keywords: quantitative review, Alzheimer's disease, Mild Cognitive Impairment, progression, automatic prediction, cognition

1. Introduction

The early diagnosis of Alzheimer's disease (AD) is crucial for patient care and treatment. Machine learning algorithms have been used to perform automatic diagnosis and predict the current clinical status at an individual level, mainly in research cohorts. Individuals suffering from mild cognitive impairment (MCI) are however likely to have a change of clinical status in the coming years, and to be diagnosed with AD or another form of dementia. Distinguishing between the MCI individuals that will remain MCI (MCI stable, or sMCI) from those who will progress to AD (pMCI) is an important task, that can allow for the early care and treatment of pMCI patients. In this article, we will review methods that have been proposed to automatically predict if an MCI patient will develop AD dementia in the future by performing a careful reading of published articles, and compare them through a quantitative analysis.

The application of machine learning to precision medicine is an emerging field, at the cross roads of different disciplines, such as computer science, radiology or neurology. Researchers working on the topic usually come from one field or the other, and therefore do not have all the skills that are necessary to design methods that would be efficient and following machine learning best practices, while being understandable and useful to clinicians.

Reviews of the automatic prediction of the patient's current diagnosis from clinical or imaging variables acquired at the same time in the context of AD have already been published, but none specifically target the prediction of progression from MCI to AD dementia. They focus on the use of magnetic resonance imaging (MRI) (Falahati et al., 2014; Leandrou et al., 2018), or of neuroimaging data more broadly (Rathore et al., 2017; Arbabshirani et al., 2017; Haller et al., 2011; Sarica et al., 2017). Several of them are systematic reviews such as Arbabshirani et al. (2017) with 112 studies on AD, Rathore et al. (2017) with 81 studies, Falahati et al. (2014) with 50 studies and Sarica et al. (2017) with 12 studies. They often gather the findings of each individual article and compare them, but no quantitative analysis of performance is proposed.

We propose here to perform a systematic and quantitative review of studies predicting the evolution of clinical diagnosis in individuals with MCI. We will report different quantitative and qualitative characteristics of the proposed method such as the sample size, type of algorithm, reported accuracy, identification of possible issues. We will then analyze this data to identify the characteristics which impact performance the most, and list several recommendations to ensure that the performance is well estimated, and that the algorithm would have the best chance to be useful in clinical practice.

2. Materials and Method

2.1. Selection process

The query used to find the relevant articles was composed of 4 parts:

1. As we study the progression from MCI to AD, the words MCI and AD should be present in the abstract ;
2. We removed the articles predicting only the patient's current diagnosis using variables acquired at the same time point by ensuring the words "prediction" and "progression" or associated terms are present in the abstract ;
3. A performance measure should be mentioned ;
4. A machine learning algorithm or classification related key-word should be in the abstract. This fourth part ensures the selected articles make individual predictions and reduces the presence of group analyses.

The full query can be found in Appendix A.1. Running it on Scopus on the 13th of December 2018 resulted in 330 articles. The abstracts were read to remove irrelevant articles, including studies of the progression of cognitively normal individuals to MCI, automatic diagnosis methods, review articles and group analyses. After this selection 206 articles were identified. As this first selection was quite conservative, 34 additional articles were removed from the selection for similar reasons during the reading process, leaving 172 studied articles which are listed in Appendix B. The selection process is described in Figure S1 in Appendix A.2.

2.2. Reading process

For each study, the number of individuals was first assessed and noted. Only studies including more than 30 sMCI and 30 pMCI (111 articles) were then fully read, as we considered that experience using fewer than 30 individuals cannot provide robust estimates of performance. Articles with fewer than 30 individuals in each category were still considered when studying the evolution of the number of articles with time, and of the number of individuals per article with time. The studies including enough individuals were then analyzed by one of the 19 readers participating in this review, and a final curation was performed by one of the authors (MA) to ensure homogeneity. 36 items, of which a list is available in Appendix A.3, were reported for each study, including the used features, the cohort, the method (time to prediction, algorithm, feature selection, feature processing), the evaluation framework and the performance measures, as well as identified biases in the method. When several experiments were available in an article, they were all reported in the table. A total of 234 experiments was thus studied.

A table containing the articles included in the review and the reported values can be found on `https://gitlab.com/icm-institute/aramislab/mci-progression-review`. The issues identified in each article were removed from this open-access table, to avoid negatively pointing at studies. They can be made available if requested to the corresponding author.

2.3. Quality check

Several methodological issues were identified during the reading process. This list of issues was not previously defined, it has been established as issues were encountered in the various studies. We identified the following list of issues:

- Lack of a test data set: use of the same data set for training and testing the algorithm, without splitting the data set or using any kind of cross-validation method. The performance computed this way is the training performance, whereas a test performance, computed on a different set of individuals, is necessary to measure the performance that could be obtained on another data set (i.e. generalizability of the method) .

- Automatic feature selection performed on the whole data set. When a large number of features is available, automatic feature selection can be performed in order to identify the most relevant features and use them as input. A variety of automatic algorithms exist to do this. Some studies performed this automatic feature selection on the whole data set, before splitting it into a training and a test set or performing cross-validation. An example of this issue is, first, using t-tests to identify features that best separate pMCI from sMCI, using the whole data set, then splitting the data set into a training and a test set, to respectively train the classification algorithm and evaluate its performance. In this example, the individuals from the test set have been used to perform the automatic feature selection and choose the most relevant features. This is an issue, as individuals in the test set should be used for performance evaluation only.

- Other data-leakage. More broadly, data leakage is the use of data from the test set outside of performance evaluation. Using the test data set for parameter tuning, or for choosing the best data set out of a large number of experiments, are two common examples of data leakage.

- Feature embedding performed on the whole data set. Feature embedding (for example principal components analysis) transforms the input features into a lower-dimension feature space. It is often used to reduce the input dimension when many features are available, but it does not use the individual labels (sMCI/pMCI) to do so, as feature selection often does. This issue is therefore similar to performing feature selection on the whole data set, except that only the features of the test individuals are used, and not their labels.

- Use of the date of AD diagnosis to select the input visit of pMCI individuals. An example of this issue is using the visit 3 years before progression to AD for pMCI subjects, and the first available visit for sMCI subjects, to predict the progression to AD at 3 years, even for testing the method. In this case, the date of progression to AD of the individuals of the test set was used to select the input visit, which is not possible in clinical practice, as the date of progression is not known. Such experimental designs are also likely to introduce biases between pMCI and sMCI subjects in age or in total observation periods for instance, which may lead to a better performance than what could be achieved in a real-life scenario.

Other methodological issues, not belonging to these categories, were also reported, such as incompatibility between different reported measures. The articles in which at least one of these issues was identified were not used when analyzing the performance of the methods. Only articles with no reported issues were used, however it is possible that some issues could not be detected from the elements given in the articles, and that some issues were not identified during reading.

2.4. Statistical analysis

2.4.1. General model

We studied the impact of various method characteristics (such as input feature and algorithm) on the performance of the classification task, separating sMCI from pMCI individuals. Several experiments were reported for each article, so we had to account for the dependency between experiments coming from the same article. In order to do so, we used linear mixed-effects models with a random intercept on the article.

For the model to have enough power, we grouped the characteristics in a hierarchical manner, creating broad categories that can be expanded into finer ones several times. The categories were created as such :

- linear models: linear regression, orthogonal partial least square (OPLS), linear discriminant analysis (LDA), manual threshold
- generalized linear models: linear support vector machine(SVM), logistic regression, survival analysis
- non-linear models: random forest, multi-kernel learning, non-linear SVM, bayesian methods, neural networks, others
- imaging features
  - T1 MRI
    - region-based features on selected regions of interest (T1-ROI)
    - region-based features on the whole brain
    - voxel-based features
  - positron emission tomography (PET)
    - fluorodeoxyglucose (FDG) PET
    - Amyloid PET
  - white matter hyper-intensities
  - electroencephalography (EEG) or magnetoencephalography (MEG)
  - diffusion tensor imaging (DTI)
  - fMRI
- cerebrospinal fluid (CSF) biomarkers
- cognitive features
  - general cognitive features

    - domain-targeted cognitive features
    - new, home-made cognitive features
- socio-demographic and genetic features
    - socio-demographic features
        * age
        * gender
    - Apolipoprotein E (APOE)
- other features
- longitudinal approach
- use of the Alzheimer's Disease Neuroimaging Initiative (ADNI) data set
- number of subjects

A first model was created with the broadest categories, and we used a two-sided t-test on the regression coefficients to identify the categories of characteristics which had a significant impact on performance. The next model was then created by expanding only the significant categories and keeping the non-significant one at a coarse level. The expansion and the creation of new models was repeated until we reached a model for which all significant coefficients belonged to categories that could not be expanded further. We report the results of the final model in 4. The intermediate models leading to the final one are reported in section Appendix A.4.1 of Supplementary Materials.

For each model, we only used the characteristics which were found in more than one article with an associated performance measure and with no identified issue. The performance measure used for these models was the area under the receiver operating characteristic (ROC) curve (AUC), experiments with no reported AUC were therefore not taken into account.

Only the experiments with no identified methodological issues were included in the model. This process was performed twice: once using all experiments without issues, and once using only the experiments performed on the ADNI database.

The p-values corrected for multiple comparisons were obtained by using the Benjamini-Hochberg procedure.

2.4.2. Individual feature models

We wanted to test whether T1 MRI, cognitive or FDG PET features are predictors of better performance if used alone or in combination with other features. To this purpose, for a given feature type F, we selected the experiments using this feature type and that had a reported AUC and no methodological issue. We then used a linear mixed-effect model, defined as:

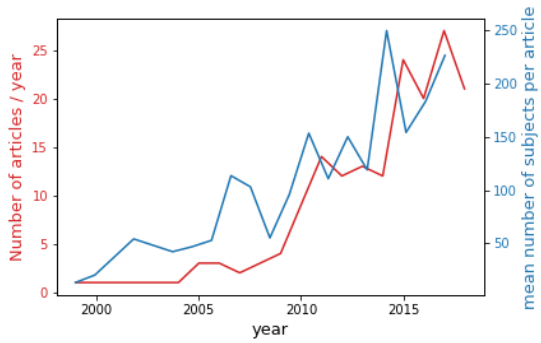
$$AUC_i = \alpha * \eta_i + \beta + \beta_{article_i}$$

where i is the experiment, $article_i$ is the article to which the experiment belongs (as several experiments can be reported in each article), and η_i is 0 when the experiment

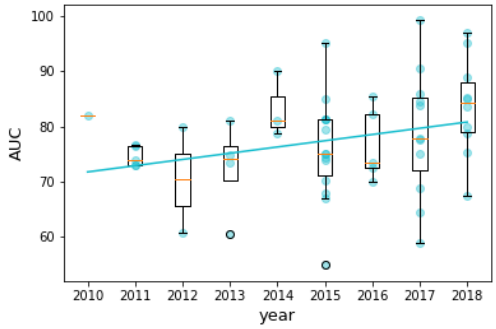
uses only the studied feature type $F$and 1 when it uses other feature types as well. We used a two-sided t-test on α to determine if including other feature types significantly changed the performance compared to using F alone.

This analysis was performed for F being: (a) T1 MRI features, (b) cognitive features, (c) FDG PET features. These were the features selected for the final general model as explained in 2.4.1, that have been used alone in at least 2 reported experiments, and that have been used in combination with other features in at least 2 experiments as well. Cognitive features were not divided into subsets so as to study the effect of cognitive assessments as a whole.

3. Descriptive analysis



(a) Evolution of number of article per year and of the number of individuals per article



(b) Evolution of the AUC with time

Figure 1: Recent trends. (a) Evolution of number of article per year (in red) and of the number of individuals per article with time (in blue). (b) Evolution of the area under the ROC (receiver operating characteristic) curve (AUC) with time. The AUC of each article is represented by a dot. The AUC of articles published the same year is represented as box-plots. The plain line corresponds to the regression of the AUC against time

3.1. A recent trend

Figure 1a shows that the number of articles published each year on the prediction of the progression of MCI to AD dementia has been steadily increasing since 2010.

Figure 1a also shows that the number of individuals used for the experiments is increasing over time (p= 10^{-5}, slope of 12.15 subjects per year, R^2 = 0.10). 84.6% of articles used data of the Alzheimer's Disease Neuroimaging Initiative (ADNI) study. Starting in 2004, this multicenter longitudinal study provides multiple modalities for the early detection of AD. As the recruitment of this largely used cohort is still ongoing, it is not surprising to see the number of included individuals increasing over the years. Studies often select individuals with a minimal follow-up time, of 3 years for example, and over the years more and more MCI individuals from the ADNI cohort fulfill these criteria, so more individuals can be included.

As shown in Figure 1b, the reported AUC are also increasing over time (p= 0.045, slope of 1.15 points of AUC per year), which can have multiple explanations. First, as new studies often compare their performance with those of previous methods, they tend to be published only when the obtained results seem competitive compared to previous ones. A more optimistic interpretation would be that algorithms tend to improve, and that newly available features might have a better predictive power. It has also been shown (Ansart et al., 2019; Domingos, 2012) that having a larger data set leads to a higher performance, so there may be a link between the increase in data set size and the increase in performance.

3.2. Features

T1 MRI, cognition and socio-demographic features are used in respectively 69.2%, 43.2% and 33.8% of experiments. On the other hand, FDG PET, APOE and CSF AD biomarkers are used in 15 to 20% of experiments, and the other studied features (white matter hyper-intensities, EEG, MEG, PET amyloid, amyloid binary status without considering the PET or CSF value, DTI and PET Tau) are used in fewer than 10% of experiments. No study using functional MRI has been identified.

Studies using T1 MRI mainly use selected regions of interest (46.8%), whereas 34.7% use the whole brain, separated into regions of interest, and 18.5% use voxel features. Studies using neuro-psychological tests mainly use aggregated tests evaluating multiple domains of cognition (51.2% of them), and 37.4% of them combine aggregated tests with domain-specific ones. Seven experiments use new or home-made cognitive tests. 35.7% of experiments use only T1 MRI (apart from socio-demographic features), and 15% use cognition only.

The prevalence of T1 MRI does not seem surprising, as researchers working on automatic diagnosis often come from the medical imaging community, and T1 MRI is the most widely available modality. The prevalence of the imaging community can also explain the choice of cognitive features, and why more detailed and targeted cognitive tests are not used as much as more general and more well-known ones.

3.3. Algorithm

Support vector machines (SVM) and logistic regressions are the most used algorithms, being used in respectively 32.6% and 15.0% of experiments. Among the experiments using an SVM, 63.2% use a non-linear kernel, 30.3% use a linear kernel and

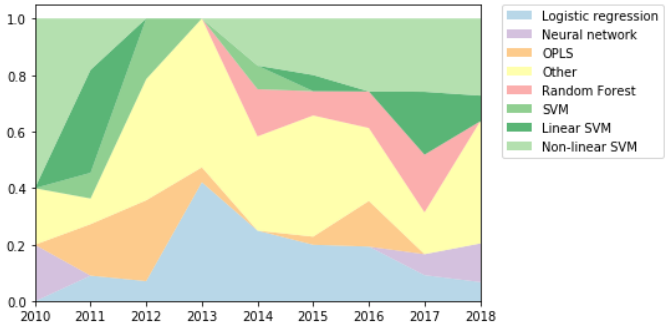


Figure 2: Evolution of the use of various algorithms with time. SVM with unknown kernel are simply noted as "SVM". OPLS: orthogonal partial least square; SVM: support vector machine

6.6% do not mention the used kernel. Other algorithms are used in fewer than 10% of cases. Figure 2 shows the evolution of the algorithm use over time.

The high proportion of methods using an SVM has already been shown for the prediction of the current diagnosis in Falahati et al. (2014) and Rathore et al. (2017), it is therefore not surprising that this algorithm is also commonly used for the prediction of future diagnosis. We see that random forests started being used around 2014, but the proportion of methods using this algorithm, even recently, stays low compared to the proportion of methods using an SVM. Neural networks started being used during the last two years, as it can be seen in Figure 2, and we can assume the phenomenon has been too recent to be visible just yet in the field. Overall, even if the proportion of SVM has been decreasing until 2013, the field has not been so prompt to use new algorithms as one could have expected. A possible explanation is that the choice of algorithm does not significantly impact performance.

3.4. Validation method

For evaluating their performance, 29.1 % of experiments use a 10-fold, and 12.8% use a k-fold with k different from 10. Leave-one individual out is also quite popular, being used in 17.5% of cases. We noted that 7.3% of experiments were trained and tested on the same individuals, and 7.3% train the method on a first cohort and test it on a different one.

It should be kept in mind when comparing the performance of different studies that the cross-validation methods can impact the performance. Using a larger training set and smaller test set is more favorable, hence the same method might result in a better performance when evaluated using a leave-one out validation than using a 10-fold validation, as shown in Lin et al. (2018). Bias and variance also vary across validation methods (Efron, 1983). Varoquaux et al. (2017) also studied the impact of the cross-validation strategy on a range of classification tasks performed on fMRI and MEG data sets, and showed that differences in performance tend to be smaller than the variance of the estimated performance using a cross-validation method, mitigating the importance of the choice of cross-validation strategy. This study still warns against the use of leave-one-out validation, leading to less stable estimates.

Reporting variance or confidence intervals is an important best practice to compare results from different studies and experiments. We did not collect this information, and further work regarding the adoption of this practice could complete this study.

4. Performance analyses

Characteristic	**coeff.**	**p-value**	**corrected p-value**	**number of exp.**
intercept	78	**0**	**0**	NA
linear model	-0.47	0.79	0.94	23
generalized linear model	0.19	0.89	0.96	28
non linear model	0.83	0.55	0.79	50
T1 features	0.92	0.26	0.76	77
amyloid PET	1.3	0.35	0.79	5
FDG PET	2.6	**0.023**	0.13	24
white matter hyper-intensities	-0.58	0.49	0.79	3
EEG/MEG	3.4	$\mathbf{2.9{*}10^{-03}}$	**0.029**	5
general cognitive features	-0.14	0.91	0.96	49
domain targeted cognitive features	2.6	**0.026**	0.13	25
new or specific cognitive features	0.89	0.52	0.79	2
socio-demographic features	1.2	0.43	0.79	43
APOE	2.27	**0.049**	0.19	26
biomarkers	0.75	0.39	0.79	19
other features	0.53	0.54	0.79	12
longitudinal	0.25	0.80	0.95	13
ADNI	0.011	0.99	0.99	106
number of subjects	-0.39	0.76	0.94	NA
individual intercept	NA	0.072	0.24	NA

Table 1: Impact of method characteristics. This table shows the coefficients obtained using the linear mixed-effect model described in section 2.4.1 on all experiments, the associated p-values and corrected p-values. The last columns shows the number of experiments using the given characteristic, out of the 120 experiments included in the model. Benjamini-Hochberg procedure was applied to get corrected p-values. coeff.:coefficient of the characteristics in the mixed effect model; FDG: fluorodeoxyglucose; PET: positron emission tomography; EEG: electroencephalography; MEG: magnetoencephalography; APOE: Apolipoprotein E; ADNI: Alzheimer's Disease Neuroimaging Initiative; NA: not applicable

The results of the linear mixed-effect model used to model the AUC based on the method characteristics are shown in Table 1, and the details of the intermediate models can be found in section Appendix A.4.1 of Supplementary Materials. The performance is significantly better when using EEG and MEG (coefficient=3.4, p=$3{*}10^{-3}$), domain-targeted cognitive features (coefficient=2.6, p = 0.026), FDG PET (coefficient=2.2, p=0.023) or APOE (coefficient=2.3, p=0.049). The use of the ADNI cohort and of longitudinal data are not shown to be significant. The impact of the algorithm type and of the number of subjects are not shown to be significant either.

We also run the performance analysis using only the experiments performed on the ADNI cohort. The only characteristics with a significant impact on the AUC are the use of T1-ROI features (coefficient = 1.7, p=0.014) (and not the other T1-based features which are regions based features on the whole brain and voxel-based features), FDG PET features (coefficient = 4.4, $p < 1 * 10^{-7}$) and domain-targeted cognitive features (coefficient=2.4, $p = 9*10^{-3}$). The complete results can be found in section Appendix A.4.2 of Supplementary Materials.

We considered the impact of using each feature alone compared to a combination of them. It is significantly better to combine T1 MRI with other features than to use it solely (coefficient = 5.5, $p = 9*10^{-3}$). The effect is not significant for cognition (coefficient=3.0, p=0.19) and FDG PET (coefficient = -6.1, p=0.38).

4.1. Cognition

Cognitive variables can be easily collected in clinical routine, at a low cost, and they are proven to increase the performance of the methods, so their use should be encouraged. This finding is consistent with comparisons performed in several studies. Minhas et al. (2018); Kauppi et al. (2018); Ardekani et al. (2017); Tong et al. (2017); Gavidia-Bovadilla et al. (2017); Moradi et al. (2015); Hall et al. (2015); Fleisher et al. (2008) showed that using cognition and T1 MRI performed better than using T1 MRI only. Dukart et al. (2015); Cui et al. (2011); Thung et al. (2018); Li et al. (2018) showed that adding cognition to other modalities also improved the results.

More surprisingly, we showed that using other modalities does not significantly improve the results compared to using cognition only. Although Fleisher et al. (2008) shows that using T1 MRI in addition to cognition does not improve the performance compared to using cognition only, several studies show the opposite on various modalities (Samper-Gonzalez et al., 2019; Moradi et al., 2015; Ardekani et al., 2017; Li et al., 2018; Kauppi et al., 2018). However, when taking all studies into account, it appears that the improvement one gains by including other modalities along with cognitive variables is not significant. As the cost of collecting cognitive variables compared to performing an MRI or a FDG PET is quite low, the non-significant improvement in performance might not be worth the cost, logistics and patient inconvenience arising from the collection of other modalities. Methods focusing on cognition only, such as proposed by Johnson et al. (2014), should therefore be further explored. Such methods should include domain-specific cognitive scores, which have shown to increase the performance.

4.2. Medical imaging and biomarkers

Imaging modalities are not as widely available as cognitive features, but they can represent a good opportunity to better understand the disease process by showing the changes that appear before the individuals progress to AD dementia.

Among the used imaging modalities, we showed that using FDG PET leads to a better performance. Using T1-ROI features also leads to a better performance on the ADNI experiments, but this effect is not significant when considering all experiments. All the experiments using T1 MRI are performed on the ADNI database, so one can assume the effect on performance is small and is diluted when considering all experiments instead of the experiments performed on ADNI. Even considering the ADNI

experiments, the effect of using T1-ROI features is 2.6 times smaller than the effect of using FDG PET, and 1.45 times smaller than the effect of using cognitive features. We also showed that using T1 MRI features alone performs significantly worse than using other features as well. Over all, T1 MRI features should not be used alone and PET images could represent a better alternative for the imaging community. Similar observations have been made by Samper-Gonzalez et al. (2018). FDG PET was included as a supportive feature in AD diagnosis criteria in 2007 (Dubois et al., 2007), and although it was removed - along with structural MRI - from IWG-2 diagnostic criteria in 2014, Dubois et al. (2014) stressed that FDG PET can be useful to differentiate between AD and other types of dementia and to measure disease progression. According to the model hypothesized in (Jack et al., 2010a) changes in FDG PET appear earlier in the AD process than changes in structural MRI, which has been corroborated by different quantitative studies (Chetelat et al., 2007; Reiman et al., 1998; Jagust et al., 2006). These changes might already be visible in MCI individuals several years before their progression to AD, which can explain why FDG PET is more predictive of this progression.

Only one method using Tau PET has been identified in this review so we could not evaluate the impact on performance. This new modality should also be affected early in the disease process, and could therefore represent great hopes for the imaging community. However, surprisingly, Amyloid PET or CSF value, which is also one of the earliest markers, did not have a significant impact on the prediction performance. Although amyloid load saturates several years before symptom onset (Jack et al., 2010b; Yau et al., 2015), several studies show that MCI individuals who are amyloid positive are more likely to convert to dementia in the next 2 to 4 years than those who are amyloid negative (Landau et al., 2012; Jack et al., 2010b; Okello et al., 2009).

The use of EEG or MEG has a significant impact on the performance. However, only 5 experiments using these features were included in the model, it is therefore difficult to conclude if this effect is real, and if it is not due to methodological issues that have not been identified during the quality check.

4.3. Longitudinal data

Longitudinal data could give a better view of the evolution of the patient, and hence be more predictive of the progression to AD than cross-sectional data. Nonetheless, we did not find the use of longitudinal data to have a significant effect on the performance. Similar findings are reported in Aksman (2017) for the classification of AD and in Schuster et al. (2015) for progressive diseases in general. Longitudinal analyses are more difficult to design in age-related diseases since there is no temporal marker of disease progression especially before diagnosis. Patients are also seen at different time-points and not all features are acquired at each visit, leading to many missing values. Methodologies for such designs are more exploratory than for cross-sectional approaches (Schiratti et al., 2015; Venkatraghavan et al., 2019)

4.4. Algorithm

Table 1 shows that the choice of algorithm has no significant impact on performance. Even if non-linear models seem to be associated to a higher coefficient (0.83)

Characteristic	**coeff.**	**p-value**	**corrected p-value**	**number of exp.**
intercept	78	**0**	**0**	NA
linear model	-8.1	$\mathbf{1*10^{-03}}$	$\mathbf{5.5*10^{-03}}$	23
generalized linear model	-3.7	0.12	0.25	28
non linear model	-0.13	0.96	0.96	50
imaging features	-1.6	0.42	0.52	94
cognitive features	2	**0.028**	0.073	53
socio-demographic features and APOE	2.4	**0.012**	**0.04**	49
biomarkers	0.92	0.28	0.45	19
other features	0.87	0.31	0.45	12
longitudinal	0.35	0.72	0.82	13
ADNI	-1.42	0.26	0.45	106
number of subjects	-0.066	0.96	0.96	NA
interaction: linear model and imaging features	7.85	$\mathbf{5.8*10^{-04}}$	$\mathbf{4.6*10^{-03}}$	19
interaction: generalized linear model and imaging features	4.41	**0.036**	0.083	21
interaction: non linear model and imaging features	2.13	0.41	0.52	38
individual intercept	2.27	$\mathbf{8.6*10^{-03}}$	**0.034**	NA

Table 2: Impact of method characteristics, taking into account the interaction between the model type and the use of imaging features. This table shows the coefficients obtained using the linear mixed-effect model described in section 2.4.1 on all experiments, the associated p-values and corrected p-values. The last columns shows the number of experiments using the given characteristic, out of the 120 experiments included in the model. Benjamini-Hochberg procedure was applied to get corrected p-values. coeff.:coefficient of the characteristics in the mixed effect model; APOE: Apolipoprotein E; ADNI: Alzheimer's Disease Neuroimaging Initiative; NA: not applicable

than linear and generalized linear models (-0.47 and 0.19 respectively), these coefficients are far from significant.

The model displayed in Table 2 takes into account the interaction between the model choice and the usage of imaging features. These results show that linear models perform significantly worse than other models (coefficient=-8.11, p=0.001), however the interaction between linear models and imaging features is significantly positive (coefficient = 7.85, p=0.0006); using imaging features therefore leads to a significant increase in performance when using a linear model. Similar conclusions can be drawn from the interaction between generalized linear model and imaging features (coefficient = 4.41, p=0.04), whereas this effect is not significant for non-linear models (coefficient=2.13, p=0.4). By combining the different coefficients, one can see that the best results are obtained using non-linear models. In this case, the use of imaging feature does not significantly impact performance. A possible explanation is that non-linear models are more powerful and better leverage the information contained in non-imaging data, whereas linear and generalized-linear models have a lower performance on non-imaging data. They therefore benefit from the addition of imaging data, leading to a performance similar to the one obtained using non-linear models.

4.5. Other methodological characteristics

One could expect the performance to increase when the data set size increases, however we find that the effect of the number of subjects is not significant (coefficient=-0.39, p=0.76). The impact of data set size is further investigated in 5.1.2.

The impact of using the ADNI data set is not significant (coefficient=0.011, p=0.995). This finding is mitigated by the fact that our results slightly vary when using all experiments or only the ADNI experiment. As only 14 included experiments do not use the ADNI database it is difficult to estimate the impact of its usage independently from the other characteristics.

Although we used a hierarchical grouping of the variables in order to have more statistical power, few p-values and fewer corrected p-values are significant. This small number of significant effects means that the variance of the reported performance measures is high compared to the effect sizes.

5. Design of the decision support system and methodological issues

5.1. Identified issues

5.1.1. Lack or misuse of test data

The lack of a test data set is observed in 7.3% of experiments. In 16% of articles using feature selection, it is performed on the whole data set, and 8% of articles do not describe this step well enough to draw conclusions. Other data leakage (use of the test set for decision making) is identified in 8% of experiments, and is unclear for 4%.

Overall, 26.5% of articles use the test set in the training process, to train the algorithm, choose the features or tune the parameters. This issue, and in particular performing feature selection on the whole data set, has also been pointed out by Arbabshirani et al. (2017) in the context of brain disorder prediction.

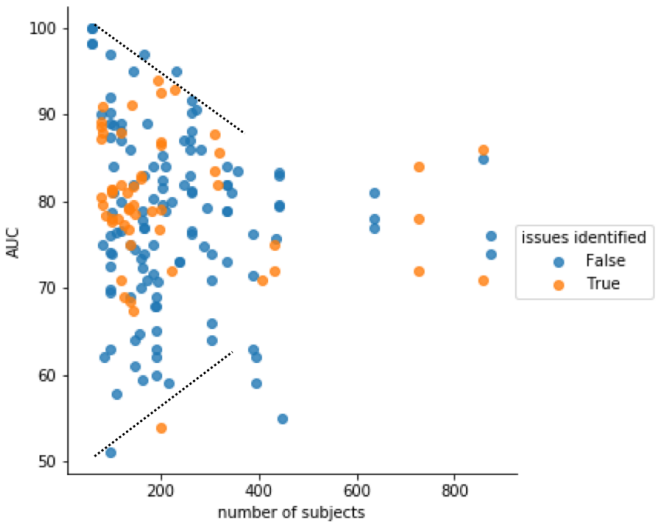


Figure 3: Relationship between the AUC (area under the ROC curve) and the number of individuals. The black dotted lines represent the upper and lower limits.

5.1.2. Performance as a function of data set size

We plotted the AUC against the number of individuals for each experiment in Figure 3, with the colored dots representing experiments with identified issues. The colored dots show that there is a higher prevalence of studies with identified issues among high-performance studies: a methodological issue has been identified in 18.5% of experiments with an AUC below 75%, whereas this proportion rises to 36.4% for experiments with an AUC of 75% or higher (significant difference, with $p = 0.006$).

We can observe an upper-limit (shown in dashed line) decreasing when the number of individuals increases, suggesting that high-performance achieved with a small number of subjects might be due to overfitting. This phenomenon has already been identified by Arbabshirani et al. (2017) and Varoquaux (2018) regarding the use of neuroimaging for brain disorders.

A lower limit is also visible, with the AUC increasing with the number of individuals. This may reflect the fact that, on average, methods generalize better when correctly trained on larger data sets. But it might also suggest that it is harder to publish a method with a relatively low performance if it has been trained on a large number of subjects, such a paper being then considered as reporting a negative result. Within papers also, authors tend to focus on their best performing method, and rarely explain what they learned to achieve this. The machine learning field has the chance to have simple metrics, such as AUC or accuracy, to compare different methods on an objective basis. However, we believe that one should use such metrics wisely not to discourage the publication of innovative methodological works even if it does not yield immediately better prediction performance, and not to overshadow the need to better understand why some methods work better than others.

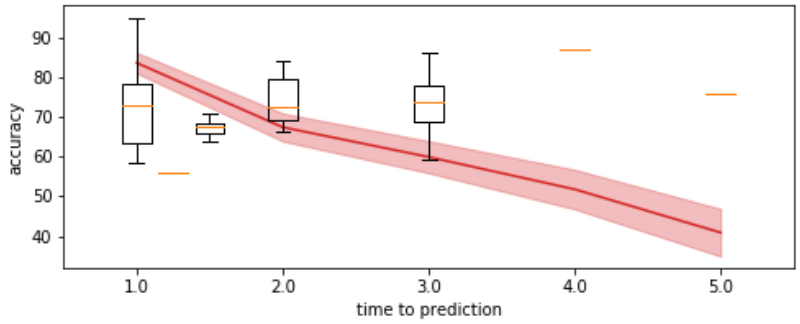


Figure 4: Evolution of the performance with respect to the time to prediction. Box plots represent the accuracy reported in the articles using ADNI included in this review. The straight line represents the accuracy that we computed by predicting that all MCI subjects remain MCI, that is the proportion of MCI subjects in ADNI who remain MCI at the follow-up visit. The shaded area corresponds to the 90% confidence interval. Although some papers in the literature use a sub-set of ADNI and not all ADNI, this plot still shows that results reported in the literature do not out-perform the naive constant prediction for time-to-predictions smaller than 3 years. This comparison is rarely done in the articles.

As the number of subjects increases, the two lines seem to converge to an AUC of about 75%, which might represent the true performance for current state-of-the-art methods.

5.1.3. Use of features of test subjects

Feature embedding is performed on the whole data set in 6.8% of experiments, meaning that the features of the test individuals are used for feature embedding during the training phase. As the diagnosis of the test individuals is often not used for feature embedding, as it is for feature selection, performing it on test individual can be considered a less serious issue than for feature selection. It however requires to re-train the algorithm each time the prediction has to be made on a new individual, which is not suited for a use in clinical practice.

5.1.4. Use of the diagnosis date

In 5.6% of the experiments, the date of AD diagnosis is used to select the input visit of pMCI individuals, for training and testing. As explained in section 2.3, this practice can prevent the generalization of the method to the clinical practice, as the progression date of test individuals is by definition unknown.

This type of experiments answers the question "may one detect some characteristics in the data of a MCI patient 3 years before the diagnosis which, at the same time, is rarely present in stable MCI subjects?". Which should not be confused with: "can such characteristics predict that a MCI patient will progress to AD within the next 3 years". What misses to conclude about the predictive ability is to consider the MCI subjects who have the found characteristics and count the proportion of them who will not develop AD within 3 years.

This confusion typically occurred after the publication of Ding et al. (2018). The paper attracted a great attention from general media, including a post on Fox News (Wooller,

2018), stating "Artificial intelligence can predict Alzheimer's 6 years earlier than medics". However, the authors state in the paper that "final clinical diagnosis after all follow-up examinations was used as the ground truth label", thus without any control of the follow-up periods that vary across subjects. Therefore, a patient may be considered as a true negative in this study, namely as a true stable MCI subject, whereas this subject may have been followed for less than 6 years. There is no guarantee that this subject is not in fact a false negative for the prediction of diagnosis at 6 years.

5.1.5. Choice of time-to-prediction

We found that 22.6% of experiments work on separating pMCI from sMCI, regardless of their time to progression to dementia. We advise against this practice, as the temporal horizon at which the individuals are likely to progress is an important information in clinical practice. Methods predicting the exact progression dates, such as what is asked in the Tadpole challenge (Marinescu et al., 2018), should be favored over methods predicting the diagnosis at a given date.

The other experiments have set a specific time to prediction, often between 1 and 3 years, meaning that they intend to predict the diagnosis of the individual at the end of this time interval. Figure 4 shows the evolution of the accuracy of these methods tested on ADNI with respect to the time to prediction. The time to prediction did not have a significant effect on AUC, accuracy, balanced accuracy, specificity nor sensitivity. Figure 4 also shows the accuracy that one would get on ADNI when using a constant prediction, that is predicting that all individuals stay MCI at future time points. The accuracy of this constant prediction has been computed using the proportion of MCI remaining stable at each visit. We show that most methods predicting the progression to AD within a short-term period smaller than 3 years do not perform better than this constant prediction. This finding is consistent with results from the Tadpole challenge (Marinescu et al., 2020), in which no participants significantly outperformed the constant prediction, to which a random noise was added, on prediction of cognitive scores within a 18 month period. We therefore advise to use a time to prediction of at least 3 years. For shorter time intervals the proportion of MCI individuals progressing to AD is so small that predicting that all individuals remain stable gives a better accuracy than most proposed methods.

This fact also shows that the accuracy may be arbitrarily increased by using a cohort with a large proportion of stable subjects. The algorithm may then yield high accuracy by mimicking a constant predictor. This effect may be alleviated by optimizing the balanced accuracy instead of the accuracy.

5.1.6. Problem formulation and data set choice

A common theme that arises from the previous issues is that the methods are not always designed to be the most useful in clinical practice. It is for example true of methods that do not use a specific time-to-prediction, or that use the date of AD diagnosis to select the included visits.

More generally, we think the most useful decision support system should not only focus on Alzheimer's disease but perform differential diagnosis. Clinicians do not usually need to distinguish between individuals who will develop AD and individuals

who will not develop any neurological disorder. They most likely need help to determine which disorder an MCI individual is likely to develop. Unfortunately, no widely available data set allows the development methods for differential diagnosis to date. Methods focusing on AD should therefore target individuals who have already been identified as at risk of developing AD, by providing insight on the date at which this conversion is likely to happen. Such methods could be trained on MCI subjects that are at risk to develop Alzheimer's disease, defined for instance as the ones who have a MMSE of 27 or smaller and are amyloid positive. In addition to being closer to what can be expected in clinical practice, such data sets of at risk subjects should include a larger proportion of pMCI, leading to a better performance compared to the constant prediction. For example in ADNI, 71.6% of MCI subjects stay stable 2 years after inclusion, whereas this proportion drops to 53.7% for MCI subjects who are amyloid positive and have a MMSE of 27 or lower. Similarly, one should think carefully of the possible other biases introduced by the selection of sMCI or pMCI sub-sets, for instance bias in age, gender or cognitive state. One choice is to match the two sub-sets for these factors. This choice is justified for the detection of the features that are specific to the progressers and to the stable MCI. However, to analyze the performance of a decision support system, one should better reproduce the biases within the population that will be tested by the system in a real case scenario.

The diagnosis of Alzheimer's disease highly depends on the clinical practice, and varies greatly across sites and countries (Beach et al., 2012). Therefore, the short-term prediction of progression to Alzheimer's disease is unlikely to generalize well outside the well controlled environment of a research study. Studies on clinical data sets, such as performed in Archetti et al. (2019) regarding the prediction of current diagnosis, could assess how these methods would perform in clinical settings. An interesting alternative may be to predict the changes in the imaging or clinical biomarkers in time rather than the change in diagnosis, such as in Koval et al. (2020), Iddi et al. (2019) and Marinescu et al. (2020).

5.2. Need to adhere to best practice guidelines

Given the number of methodological issues that we found in the preparation of this review and that we have discussed above, we feel the need to list here several best practices recommendations.

We first list general guidelines to ensure best generalization of the method and limit the risk for overfitting, following Hastie et al. (2009); Bishop (2006); Géron (2019); Poldrack et al. (2019):

- Separate train and test data sets by using independent cohorts or, if not available, cross-validation. Following Hastie et al. (2009); Varoquaux et al. (2017); Borra and Di Ciaccio (2010); Davison and Hinkley (1997); Kohavi (1995), guidelines for best practices recommend to favor k-fold, repeated k-fold and repeated hold-out over leave-one-out method.

- No element of the test data set, both labels and features, should be used except for performance evaluation. In particular, parameter tuning should not be performed on test data, therefore nested cross-validation or train, validation and test splits should be used to tune the algorithm parameters.

- Use a large data set or pool different cohorts to obtain a large data set. Figure 3 shows that overfitting is reduced for more than 300 subjects, at which point the maximum AUC seems to stabilize. This is concordant with results from Arbabshirani et al. (2017), showing a similar point around 200 subjects. Similarly, Poldrack et al. (2019) recommends using data sets of at least several hundred subjects.

We also compile a list of guidelines to carefully design the experiments so that they could support the conclusion about the predictive performance of the method which, in this particular context, includes:

- pre-registration of the time window within which one aims to predict conversion to AD, as we show that performance may greatly vary depending on the time-window and that no conclusion could be drawn regarding the ability to predict the future without it,
- definition of data sets that best reflect the use of the method in the clinical practice, for instance by selecting subjects that would be considered at risk of developing the disease rather than all possible subjects in ADNI, or by using sex ratio, distribution of age, cognitive state and other similar factors that best mimic the population characteristics that will be tested by the system.
- systematic benchmark of the method against the prediction that the subjects will remain stable over time, as we show that this naive method often outperforms proposed method with a time-to-prediction smaller than 3 years.

6. Conclusion

We conducted a systematic and quantitative review on the automatic prediction of the evolution of clinical status of MCI individuals. We reported results from 234 experiments coming from 111 articles. We showed that studies using cognitive variables or FDG PET reported significantly better results than studies that did not, and that including other feature types does not significantly improve performance compared to using cognition or FDG PET alone. These modalities should be further explored, cognition because it can be easily collected in clinical routine, and FDG PET for the interest it might represent for the imaging community and for increasing our understanding of the disease. On the other hand, we showed that using solely T1 MRI yields a significantly lower performance, despite the great number of methods developed for this imaging modality. These findings call into question the role of imaging, and more particularly of MRI, for the prediction of the progression of MCI individuals to dementia. In light of this review, we believe that one should give higher priority to other modalities. More specific cognitive tests could be created, and the impact of using digitized tests, that can be frequently used at home by the patients themselves, should be studied. The creation of digitized tests for clinical routine, such as proposed by Souillard-Mandar et al. (2016); Müller et al. (2017); Schinle et al. (2018) is a first step in this direction.

We identified several key points that should be checked when creating a method which aims at being used as a clinical decision support. When possible, an independent test set should be used to evaluate the performance of the method, otherwise a test

set can be separated by carefully splitting the cohort. In any case, the test individuals should not be used to make decisions regarding the method, such as the selection of the features or parameter tuning. The time window in which one aims at predicting the progression to AD should be pre-registered, as the temporal horizon at which an individual is likely to progress to AD is a useful information for clinicians. Alzheimer's disease being a very slowly progressive disease, algorithm performance should be systematically compared with the prediction that no change will occur in the future. We have shown indeed that the constant prediction may yield very high performance depending on the time frame of the prediction and the composition of the cohort. Finally, the cohort on which the method is tested should be carefully chosen and defined, so as to reflect the future use in clinical practice as best as possible. We noticed that there is often a confusion between two different objectives : understanding the specificities of subjects who will or will not convert to dementia on the one hand, and predicting the progression to dementia on the other hand. Experiments are often designed to address the first objective, but results are then misinterpreted in relation with the second objective. Addressing each objective requires indeed a rather different experimental design.

Following the guidelines will help to design better systems that would eventually lead to similar results in real life. In any case, the final evaluation of such systems will be done in a prospective manner, either in the framework of a challenge like the TADPOLE challenge (Marinescu et al., 2018, 2019, 2020), or even better in a prospective clinical trial (Bruun et al., 2019).

This review focused on the prediction of progression to dementia, as this problem has, by far, attracted most attention from the scientific community. Nevertheless, predicting the future values of the biomarkers or the images may be of greater interest for such clinical decision support systems to be adopted in practice (Marinescu et al., 2020; Koval et al., 2020; Ansart, 2019).

Acknowledgements

Federica Cacciamani, Baptiste Couvy-Duchesne, Pascal Lu and Wen Wei participated in reading articles to conduct this review.

We thank the reviewers for their insightful comments that helped us to improve the manuscript, including Gaël Varoquaux who purposely disclosed his name.

The research leading to these results has received funding from the program "Investissements d'avenir" ANR-10-IAIHU-06 (Agence Nationale de la Recherche-10-IA Institut Hospitalo-Universitaire-6) from the European Union H2020 program (project EuroPOND, grant number 666992, project HBP SGA1 grant number 720270), from the ICM Big Brain Theory Program (project DYNAMO, project PredictICD), from the Inria Project Lab Program (project Neuromarkers), from the European Research Council (to Dr Durrleman project LEASP, grant number 678304), from the Abeona Foundation (project Brain@Scale). OC is supported by a "contrat d'interface local" from AP-HP. China Scholarship Council supports J.W's work on this topic.

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI

and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

References

Aksman, L.M., 2017. Longitudinal neuroimaging features for discriminating early neurodegeneration. Ph.D. thesis. King's College London.

Ansart, M., 2019. Design of data driven decision support systems for the early detection of subjects at risk to develop Alzheimer's disease. phdthesis. Sorbonne Université. URL: https://tel.archives-ouvertes.fr/tel-02425735.

Ansart, M., Epelbaum, S., Gagliardi, G., Colliot, O., Dormont, D., Dubois, B., Hampel, H., Durrleman, S., for the Alzheimer's Disease Neuroimaging Initiative* and the INSIGHT-preAD study, 2019. Reduction of recruitment costs in preclinical AD trials: validation of automatic pre-screening algorithm for brain amyloidosis. Statistical Methods in Medical Research , 0962280218823036doi:10.1177/0962280218823036.

Arbabshirani, M.R., Plis, S., Sui, J., Calhoun, V.D., 2017. Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. NeuroImage 145, 137–165. doi:10.1016/j.neuroimage.2016.02.079.

Archetti, D., Ingala, S., Venkatraghavan, V., Wottschel, V., Young, A.L., Bellio, M., Bron, E.E., Klein, S., Barkhof, F., Alexander, D.C., Oxtoby, N.P., Frisoni, G.B., Redolfi, A., 2019. Multi-study validation of data-driven disease progression models to characterize evolution of biomarkers in Alzheimer's disease. NeuroImage: Clinical 24, 101954. URL: http://www.sciencedirect.com/science/article/pii/S2213158219303043, doi:10.1016/j.nicl.2019.101954.

Ardekani, B.A., Bermudez, E., Mubeen, A.M., Bachman, A.H., 2017. Prediction of Incipient Alzheimer's Disease Dementia in Patients with Mild Cognitive Impairment. Journal of Alzheimer's Disease 55, 269–281. doi:10.3233/JAD-160594.

Beach, T.G., Monsell, S.E., Phillips, L.E., Kukull, W., 2012. Accuracy of the Clinical Diagnosis of Alzheimer Disease at National Institute on Aging Alzheimer's Disease Centers, 2005–2010. Journal of Neuropathology and Experimental Neurology 71, 266–273. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3331862/, doi:10.1097/NEN.0b013e31824b211b.

Bishop, C.M., 2006. Pattern recognition and machine learning. Information science and statistics, Springer, New York.

Borra, S., Di Ciaccio, A., 2010. Measuring the prediction error. A comparison of cross-validation, bootstrap and covariance penalty methods. Computational Statistics & Data Analysis 54, 2976–2989. URL: http://www.sciencedirect.com/science/article/pii/S0167947310001064, doi:10.1016/j.csda.2010.03.004.

Bruun, M., Frederiksen, K.S., Rhodius-Meester, H.F.M., Baroni, M., Gjerum, L., Koikkalainen, J., Urhemaa, T., Tolonen, A., van Gils, M., Rueckert, D., Dyremose, N., Andersen, B.B., Lemstra, A.W., Hallikainen, M., Kurl, S., Herukka, S.K., Remes, A.M., Waldemar, G., Soininen, H., Mecocci, P., van der Flier, W.M., Lötjönen, J., Hasselbalch, S.G., 2019. Impact of a clinical decision support tool on prediction of progression in early-stage dementia: a prospective validation study. Alzheimer's Research & Therapy 11, 25. URL: https://alzres.biomedcentral.com/articles/10.1186/s13195-019-0482-3, doi:10.1186/s13195-019-0482-3.

Chetelat, G., Desgranges, B., Landeau, B., Mezenge, F., Poline, J.B., de la Sayette, V., Viader, F., Eustache, F., Baron, J.C., 2007. Direct voxel-based comparison between grey matter hypometabolism and atrophy in Alzheimer's disease. Brain 131, 60–71. URL: https://academic.oup.com/brain/article-lookup/doi/10.1093/brain/awm288, doi:10.1093/brain/awm288.

Cui, Y., Liu, B., Luo, S., Zhen, X., Fan, M., Liu, T., Zhu, W., Park, M., Jiang, T., Jin, J.S., Initiative, t.A.D.N., 2011. Identification of Conversion from Mild Cognitive Impairment to Alzheimer's Disease Using Multivariate Predictors. PLOS ONE 6, e21896. doi:10.1371/journal.pone.0021896.

Davison, A.C., Hinkley, D.V., 1997. Bootstrap methods and their application. volume 1. Cambridge university press.

Ding, Y., Sohn, J.H., Kawczynski, M.G., Trivedi, H., Harnish, R., Jenkins, N.W., Lituiev, D., Copeland, T.P., Aboian, M.S., Mari Aparici, C., Behr, S.C., Flavell, R.R., Huang, S.Y., Zalocusky, K.A., Nardo, L., Seo, Y., Hawkins, R.A., Hernandez Pampaloni, M., Hadley, D., Franc, B.L., 2018. A Deep Learning Model to Predict a Diagnosis of Alzheimer Disease by Using 18f-FDG PET of the Brain. Radiology 290, 456–464. URL: https://pubs.rsna.org/doi/10.1148/radiol.2018180958, doi:10.1148/radiol.2018180958.

Domingos, P., 2012. A few useful things to know about machine learning. Communications of the ACM 55, 78. doi:10.1145/2347736.2347755.

Dubois, B., Feldman, H.H., Jacova, C., Dekosky, S.T., Barberger-Gateau, P., Cummings, J., Delacourte, A., Galasko, D., Gauthier, S., Jicha, G., Meguro, K., O'brien, J., Pasquier, F., Robert, P., Rossor, M., Salloway, S., Stern, Y., Visser, P.J., Scheltens, P., 2007. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. The Lancet. Neurology 6, 734–746. doi:10.1016/S1474-4422(07)70178-3.

Dubois, B., Feldman, H.H., Jacova, C., Hampel, H., Molinuevo, J.L., Blennow, K., DeKosky, S.T., Gauthier, S., Selkoe, D., Bateman, R., Cappa, S., Crutch, S., Engelborghs, S., Frisoni, G.B., Fox, N.C., Galasko, D., Habert, M.O., Jicha, G.A., Nordberg, A., Pasquier, F., Rabinovici, G., Robert, P., Rowe, C., Salloway, S., Sarazin, M., Epelbaum, S., Souza, L.C.d., Vellas, B., Visser, P.J., Schneider, L., Stern, Y., Scheltens, P., Cummings, J.L., 2014. Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. The Lancet Neurology 13, 614–629.

URL: http://www.thelancet.com/journals/laneur/article/PIIS1474-4422(14)70090-0/abstract, doi:10.1016/S1474-4422(14)70090-0.

Dukart, J., Sambataro, F., Bertolino, A., 2015. Accurate prediction of conversion to Alzheimer's disease using imaging, genetic, and neuropsychological biomarkers. Journal of Alzheimer's Disease 49, 1143–1159. doi:10.3233/JAD-150570. 00022.

Efron, B., 1983. Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation. Journal of the American Statistical Association 78, 316–331. doi:10.1080/01621459.1983.10477973.

Falahati, F., Westman, E., Simmons, A., 2014. Multivariate Data Analysis and Machine Learning in Alzheimer's Disease with a Focus on Structural Magnetic Resonance Imaging. Journal of Alzheimer's Disease 41, 685–708. doi:10.3233/JAD-131928.

Fleisher, A., Sun, S., Taylor, C., Ward, C., Gamst, A., Petersen, R., Jack, C., Aisen, P., Thal, L., 2008. Volumetric MRI vs clinical predictors of Alzheimer disease in mild cognitive impairment. Neurology 70, 191–199. doi:10.1212/01.wnl.0000287091.57376.65. 00178.

Gavidia-Bovadilla, G., Kanaan-Izquierdo, S., Mataroa-Serrat, M., Perera-Lluna, A., 2017. Early prediction of Alzheimer's disease using null longitudinal model-based classifiers. PLoS ONE 12. doi:10.1371/journal.pone.0168011.

Géron, A., 2019. Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'reilly media ed.

Hall, A., Mattila, J., Koikkalainen, J., Lötjonen, J., Wolz, R., Scheltens, P., Frisoni, G., Tsolaki, M., Nobili, F., Freund-Levi, Y., Minthon, L., Frölich, L., Hampel, H., Visser, P., Soininen, H., 2015. Predicting progression from cognitive impairment to alzheimer's disease with the disease state index. Current Alzheimer Research 12, 69–79. doi:10.2174/1567205012666141218123829.

Haller, S., Lovblad, K.O., Giannakopoulos, P., 2011. Principles of Classification Analyses in Mild Cognitive Impairment (MCI) and Alzheimer Disease. Journal of Alzheimer's Disease 26, 389–394. doi:10.3233/JAD-2011-0014.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. Springer Series in Statistics. 2 ed., Springer-Verlag, New York. URL: https://www.springer.com/gp/book/9780387848570, doi:10.1007/978-0-387-84858-7.

Iddi, S., Li, D., Aisen, P.S., Rafii, M.S., Thompson, W.K., Donohue, M.C., for the Alzheimer's Disease Neuroimaging Initiative, 2019. Predicting the course of Alzheimer's progression. Brain Informatics 6, 6. URL: https://doi.org/10.1186/s40708-019-0099-0, doi:10.1186/s40708-019-0099-0.

Jack, C.R., Knopman, D.S., Jagust, W.J., Shaw, L.M., Aisen, P.S., Weiner, M.W., Petersen, R.C., Trojanowski, J.Q., 2010a. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet neurology 9, 119. doi:10.1016/S1474-4422(09)70299-6.

Jack, C.R., Wiste, H.J., Vemuri, P., Weigand, S.D., Senjem, M.L., Zeng, G., Bernstein, M.A., Gunter, J.L., Pankratz, V.S., Aisen, P.S., Weiner, M.W., Petersen, R.C., Shaw, L.M., Trojanowski, J.Q., Knopman, D.S., Alzheimer's Disease Neuroimaging Initiative, 2010b. Brain beta-amyloid measures and magnetic resonance imaging atrophy both predict time-to-progression from mild cognitive impairment to Alzheimer's disease. Brain: A Journal of Neurology 133, 3336–3348. doi:10.1093/brain/awq277.

Jagust, W., Gitcho, A., Sun, F., Kuczynski, B., Mungas, D., Haan, M., 2006. Brain imaging evidence of preclinical Alzheimer's disease in normal aging. Annals of Neurology 59, 673–681. URL: http://doi.wiley.com/10.1002/ana.20799, doi:10.1002/ana.20799.

Johnson, P., Vandewater, L., Wilson, W., Maruff, P., Savage, G., Graham, P., Macaulay, L., Ellis, K., Szoeke, C., Martins, R., Rowe, C., Masters, C., Ames, D., Zhang, P., 2014. Genetic algorithm with logistic regression for prediction of progression to Alzheimer's disease. BMC Bioinformatics 15. doi:10.1186/1471-2105-15-S16-S11. 00027.

Kauppi, K., Fan, C., McEvoy, L., Holland, D., Tan, C., Chen, C.H., Andreassen, O., Desikan, R., Dale, A., 2018. Combining polygenic hazard score with volumetric MRI and cognitive measures improves prediction of progression from mild cognitive impairment to Alzheimer's disease. Frontiers in Neuroscience 12. doi:10.3389/fnins.2018.00260.

Kohavi, R., 1995. A Study of Cross-validation and Bootstrap for Accuracy Estimation and Model Selection, in: Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA. pp. 1137–1143. URL: http://dl.acm.org/citation.cfm?id=1643031.1643047.

Koval, I., Bône, A., Louis, M., Bottani, S., Marcoux, A., Samper-Gonzalez, J., Burgos, N., CHARLIER, B., Bertrand, A., Epelbaum, S., Colliot, O., Allassonnière, S., Durrleman, S., 2020. Simulating Alzheimer's disease progression with personalised digital brain models. URL: https://hal.inria.fr/hal-01964821.

Landau, S.M., Mintun, M.A., Joshi, A.D., Koeppe, R.A., Petersen, R.C., Aisen, P.S., Weiner, M.W., Jagust, W.J., 2012. Amyloid deposition, hypometabolism, and longitudinal cognitive decline. Annals of Neurology 72, 578–586. URL: http://onlinelibrary.wiley.com/doi/abs/10.1002/ana.23650, doi:10.1002/ana.23650.

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

Li, Y., Yao, Z., Zhang, H., Hu, B., for, t.A.D.N.I., 2018. Indirect relation based individual metabolic network for identification of mild cognitive impairment. Journal of Neuroscience Methods 309, 188–198. doi:10.1016/j.jneumeth.2018.09.007.

Lin, W., Tong, T., Gao, Q., Guo, D., Du, X., Yang, Y., Guo, G., Xiao, M., Du, M., Qu, X., 2018. Convolutional neural networks-based MRI image analysis for the Alzheimer's disease prediction from mild cognitive impairment. Frontiers in Neuroscience 12. doi:10.3389/fnins.2018.00777.

Marinescu, R.V., Oxtoby, N.P., Young, A.L., Bron, E.E., Toga, A.W., Weiner, M.W., Barkhof, F., Fox, N.C., Eshaghi, A., Toni, T., Salaterski, M., Lunina, V., Ansart, M., Durrleman, S., Lu, P., Iddi, S., Li, D., Thompson, W.K., Donohue, M.C., Nahon, A., Levy, Y., Halbersberg, D., Cohen, M., Liao, H., Li, T., Yu, K., Zhu, H., Tamez-Pena, J.G., Ismail, A., Wood, T., Bravo, H.C., Nguyen, M., Sun, N., Feng, J., Yeo, B.T.T., Chen, G., Qi, K., Chen, S., Qiu, D., Buciuman, I., Kelner, A., Pop, R., Rimocea, D., Ghazi, M.M., Nielsen, M., Ourselin, S., Sorensen, L., Venkatraghavan, V., Liu, K., Rabe, C., Manser, P., Hill, S.M., Howlett, J., Huang, Z., Kiddle, S., Mukherjee, S., Rouanet, A., Taschler, B., Tom, B.D.M., White, S.R., Faux, N., Sedai, S., Oriol, J.d.V., Clemente, E.E.V., Estrada, K., Aksman, L., Altmann, A., Stonnington, C.M., Wang, Y., Wu, J., Devadas, V., Fourrier, C., Raket, L.L., Sotiras, A., Erus, G., Doshi, J., Davatzikos, C., Vogel, J., Doyle, A., Tam, A., Diaz-Papkovich, A., Jammeh, E., Koval, I., Moore, P., Lyons, T.J., Gallacher, J., Tohka, J., Ciszek, R., Jedynak, B., Pandya, K., Bilgel, M., Engels, W., Cole, J., Golland, P., Klein, S., Alexander, D.C., 2020. The Alzheimer's Disease Prediction Of Longitudinal Evolution (TADPOLE) Challenge: Results after 1 Year Follow-up. arXiv:2002.03419 [q-bio, stat] URL: http://arxiv.org/abs/2002.03419. arXiv: 2002.03419.

Marinescu, R.V., Oxtoby, N.P., Young, A.L., Bron, E.E., Toga, A.W., Weiner, M.W., Barkhof, F., Fox, N.C., Golland, P., Klein, S., Alexander, D.C., 2019. TADPOLE Challenge: Accurate Alzheimer's Disease Prediction Through Crowdsourced Forecasting of Future Data, in: Rekik, I., Adeli, E., Park, S.H. (Eds.), Predictive Intelligence in Medicine, Springer International Publishing, Cham. pp. 1–10. doi:10.1007/978-3-030-32281-6_1.

Marinescu, R.V., Oxtoby, N.P., Young, A.L., Bron, E.E., Toga, A.W., Weiner, M.W., Barkhof, F., Fox, N.C., Klein, S., Alexander, D.C., Consortium, t.E., Initiative, f.t.A.D.N., 2018. TADPOLE Challenge: Prediction of Longitudinal Evolution in Alzheimer's Disease. arXiv preprint arXiv:1805.03909 .

Minhas, S., Khanum, A., Riaz, F., Khan, S., Alvi, A., 2018. Predicting progression from mild cognitive impairment to Alzheimer's disease using autoregressive modelling of longitudinal and multimodal biomarkers. IEEE Journal of Biomedical and Health Informatics 22, 818–825. doi:10.1109/JBHI.2017.2703918.

Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J., 2015. Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. NeuroImage 104, 398–412. doi:10.1016/j.neuroimage.2014.10.002.

Müller, S., Preische, O., Heymann, P., Elbing, U., Laske, C., 2017. Increased Diagnostic Accuracy of Digital vs. Conventional Clock Drawing Test for Discrimination of Patients in the Early Course of Alzheimer's Disease from Cognitively Healthy Individuals. Frontiers in Aging Neuroscience

9. URL: https://www.frontiersin.org/articles/10.3389/fnagi.2017.00101/full, doi:10.3389/fnagi.2017.00101.

Okello, A., Koivunen, J., Edison, P., Archer, H.A., Turkheimer, F.E., Någren, K., Bullock, R., Walker, Z., Kennedy, A., Fox, N.C., Rossor, M.N., Rinne, J.O., Brooks, D.J., 2009. Conversion of amyloid positive and negative MCI to AD over 3 years: an 11C-PIB PET study. Neurology 73, 754–760. doi:10.1212/WNL.0b013e3181b23564.

Poldrack, R.A., Huckins, G., Varoquaux, G., 2019. Establishment of Best Practices for Evidence for Prediction: A Review. JAMA psychiatry doi:10.1001/jamapsychiatry.2019.3671.

Rathore, S., Habes, M., Iftikhar, M.A., Shacklett, A., Davatzikos, C., 2017. A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages. NeuroImage 155, 530–548. doi:10.1016/j.neuroimage.2017.03.057.

Reiman, E.M., Uecker, A., Caselli, R.J., Lewis, S., Bandy, D., de Leon, M.J., De Santi, S., Convit, A., Osborne, D., Weaver, A., Thibodeau, S.N., 1998. Hippocampal volumes in cognitively normal persons at genetic risk for Alzheimer's disease. Annals of Neurology 44, 288–291. doi:10.1002/ana.410440226.

Samper-Gonzalez, J., Burgos, N., Bottani, S., Fontanella, S., Lu, P., Marcoux, A., Routier, A., Guillon, J., Bacci, M., Wen, J., et al., 2018. Reproducible evaluation of classification methods in alzheimer's disease: Framework and application to mri and pet data. NeuroImage 183, 504–521.

Samper-Gonzalez, J., Burgos, N., Bottani, S., Habert, M.O., Evgeniou, T., Epelbaum, S., Colliot, O., 2019. Reproducible evaluation of methods for predicting progression to Alzheimer's disease from clinical and neuroimaging data, in: Angelini, E.D., Landman, B.A. (Eds.), Medical Imaging 2019: Image Processing, SPIE, San Diego, United States. p. 30. doi:10.1117/12.2512430.

Sarica, A., Cerasa, A., Quattrone, A., 2017. Random Forest Algorithm for the Classification of Neuroimaging Data in Alzheimer's Disease: A Systematic Review. Frontiers in Aging Neuroscience 9, 329. doi:10.3389/fnagi.2017.00329.

Schinle, M., Wyszka, D., Schwarzler, F., Volz, K., Ruby, M.A., Sejdinovic, E., Stork, W., 2018. An Approach to digitalize Psychological Tests to support Diagnosis of Alzheimer's Disease in Ambulatory Care, in: 2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA), IEEE, Rome. pp. 1–6. URL: https://ieeexplore.ieee.org/document/8438761/, doi:10.1109/MeMeA.2018.8438761.

Schiratti, J.B., Allassonniere, S., Colliot, O., Durrleman, S., 2015. Learning spatiotemporal trajectories from manifold-valued longitudinal data, in: Cortes, C., Lawrence, N.D., Lee, D.D., Sugiyama, M., Garnett, R. (Eds.), Advances in Neural Information Processing Systems 28, Curran Associates, Inc.. pp. 2404–2412.

URL: http://papers.nips.cc/paper/5663-learning-spatiotemporal-trajectories-from-manifold-valued-longitudinal-data.pdf.

Schuster, C., Elamin, M., Hardiman, O., Bede, P., 2015. Presymptomatic and longitudinal neuroimaging in neurodegeneration—from snapshots to motion picture: a systematic review. Journal of Neurology, Neurosurgery & Psychiatry 86, 1089–1096. doi:10.1136/jnnp-2014-309888.

Souillard-Mandar, W., Davis, R., Rudin, C., Au, R., Libon, D.J., Swenson, R., Price, C.C., Lamar, M., Penney, D.L., 2016. Learning classification models of cognitive conditions from subtle behaviors in the digital Clock Drawing Test. Machine Learning 102, 393–441. URL: https://doi.org/10.1007/s10994-015-5529-5, doi:10.1007/s10994-015-5529-5.

Thung, K.H., Yap, P.T., Adeli, E., Lee, S.W., Shen, D., 2018. Conversion and time-to-conversion predictions of mild cognitive impairment using low-rank affinity pursuit denoising and matrix completion. Medical Image Analysis 45, 68–82. doi:10.1016/j.media.2018.01.002.

Tong, T., Gao, Q., Guerrero, R., Ledig, C., Chen, L., Rueckert, D., 2017. A novel grading biomarker for the prediction of conversion from mild cognitive impairment to Alzheimer's disease. IEEE Transactions on Biomedical Engineering 64, 155–165. doi:10.1109/TBME.2016.2549363.

Varoquaux, G., 2018. Cross-validation failure: Small sample sizes lead to large error bars. NeuroImage 180, 68–77. URL: http://www.sciencedirect.com/science/article/pii/S1053811917305311, doi:10.1016/j.neuroimage.2017.06.061.

Varoquaux, G., Raamana, P.R., Engemann, D.A., Hoyos-Idrobo, A., Schwartz, Y., Thirion, B., 2017. Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. NeuroImage 145, 166–179. URL: https://linkinghub.elsevier.com/retrieve/pii/S105381191630595X, doi:10.1016/j.neuroimage.2016.10.038.

Venkatraghavan, V., Bron, E.E., Niessen, W.J., Klein, S., 2019. Disease progression timeline estimation for Alzheimer's disease using discriminative event based modeling. NeuroImage 186, 518–532. URL: https://linkinghub.elsevier.com/retrieve/pii/S1053811918321025, doi:10.1016/j.neuroimage.2018.11.024.

Wooller, S., 2018. Artificial intelligence can predict Alzheimer's 6 years earlier than medics, study finds. URL: https://www.foxnews.com/health/artificial-intelligence-can-predict-alzheimers-6-years-earlier-than-medics-study-finds.

Yau, W.Y.W., Tudorascu, D.L., McDade, E.M., Ikonomovic, S., James, J.A., Minhas, D., Mowrey, W., Sheu, L.K., Snitz, B.E., Weissfeld, L., Gianaros, P.J., Aizenstein, H.J., Price, J.C., Mathis, C.A., Lopez, O.L., Klunk, W.E., 2015. Lon-

gitudinal assessment of neuroimaging and clinical markers in autosomal dominant Alzheimer's disease: a prospective cohort study. The Lancet Neurology 14, 804–813. URL: http://www.sciencedirect.com/science/article/pii/S1474442215001350, doi:10.1016/S1474-4422(15)00135-0.

Appendix A. Supplementary Materials

Appendix A.1. Query

The full query was:

```
TITLE-ABS-KEY ("alzheimer's" OR alzheimer OR ad) AND TITLE-ABS-
    KEY ("Mild Cognitive Impairment" OR "MCI") AND TITLE-ABS-
    KEY ((predicting OR prediction OR predictive) AND (
    conversion OR decline OR progression OR onset) OR prognosis
    ) AND TITLE-ABS-KEY (accuracy OR roc OR auc OR specificity
    OR sensitivity) AND (TITLE-ABS-KEY ("Deep learning" OR "
    neural network" OR "neural networks" OR "convolutional
    network" OR "convolutional networks" OR "bayesian network"
    OR "bayesian networks") OR TITLE-ABS-KEY ("Matrix
    completion" OR "Support vector machine" OR "linear mixed-
    effect" OR "logistic regression" OR "Random Forest" OR "
    kernel classifier" OR "kernel" OR "decision tree" OR "
    decision trees" OR "least-squares") OR TITLE-ABS-KEY ("
    Machine learning" OR "pattern recognition" OR "pattern
    classification" OR "classifier" OR "algorithm" OR "
    classification"))
```

Appendix A.2. Selection process diagram

The process used to select the articles included in the review is shown in Figure S1.

Appendix A.3. Reported items

For each article, the following elements were reported:

- number of MCI subjects progressing to AD;
- number of stable MCI subjects;
- time to prediction;
- used cohorts;
- use of socio-demographic features (yes/no);
- use of APOE (yes/no);
- use of general cognitive features (yes/no);
- use of domain-targeted cognitive features (yes/no);
- use of new, home-made cognitive features (yes/no);
- use of voxel based features from T1 MRI (yes/no);

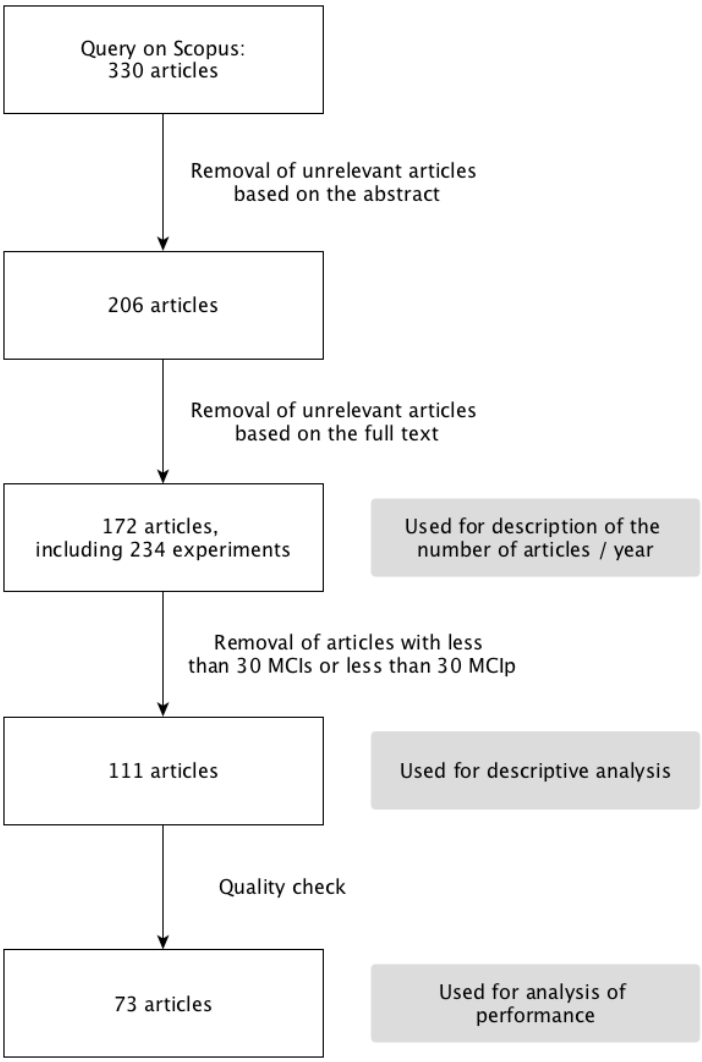


Figure S1: Diagram representing who the articles were selected

- use of regions of interest on the whole brain, from T1 MRI (yes/no);
- use of selected regions of interest from T1 MRI (yes/no);
- use of white matter hyper-intensities (yes/no);
- use of PET FDG features (yes/no);
- use of PET amyloid features (yes/no);
- use of PET tau features (yes/no);
- use of CSF features (yes/no);
- use of amyloid status (yes/no);
- use of DTI features (yes/no);
- use of functional MRI features (yes/no);
- use of EEG or MEG features (yes/no);
- use of other features (yes/no, precision given as a free note);
- use of longitudinal features (yes/no);
- is feature selection performed (yes/no);
- used algorithm (categories defined below);
- validation method (categories defined bellow);
- feature selection performed on the whole data set (yes/no/unclear);
- feature embedding performed on the whole data set (yes/no/unclear);
- selection of the input visit of the test subjects using their date of progression to AD (yes/no);
- other data leakage (use of the test set to make decisions) (yes/no/unclear);
- other issue (yes/no)
- AUC value;
- accuracy value;
- balanced accuracy value;
- sensitivity value;
- specificity value;

Free notes describing the issues, or important points that did not fit in the previous list, were added.

The possible algorithm categories were added by the readers and aggregated. The final list was: bayesian algorithms, classification by clinicians, gaussian process, linear discriminant analysis (LDA), low rank matrix completion (LRMC), linear regression, logistic regression, manifold learning, multiple kernel learning, neural network, orthogonal partial least square (OPLS), random forest, regularized logistic regression, linear support vector machine (linear SVM), non-linear SVM, SVM with unknown kernel (simply noted as SVM), survival analysis, use of a threshold and others (including home-made algorithms).

The same process was used to create the cross-validation category list, composed of: 10-fold, k-fold, repeated k-fold, leave one out, out of the bag, single split, repeated single split, validation on an independent cohort, validation on different groups (when the algorithm is trained on separating AD and CN subjects, and tested on predicting the progression of MCI subjects), none, not described (when the use of cross-validation is mentioned but the used validation method is not described) and not needed (for thresholding with a manually chosen threshold for example).

Appendix A.4. Additional performance analysis

Appendix A.4.1. On all experiments

As explained in 2.4.1, we first built a coarse model, grouping the characteristics into broad categories. The results are shown in Table S1. As several of the broad categories had a significant effect on performance (imaging and cognitive features), these categories were expanded into finer ones, building the model shown in Table S2. The significant categories were expanded once again (PET and socio-demographic features and APOE, EEG/MEG was already expanded at maximum level), leading to the model shown in table 1. As all the significant characteristics were expanded at the maximum level, no further model was created.

Appendix A.4.2. On the ADNI experiments

We also analysed the impact of method characteristic on performance using only the experiments performed on ADNI. Table S3 shows the model obtained by using the broadest characteristic categories. Table S4 shows the results obtained by expanding the categories significant in Table S3 (imaging and cognitive features). Table S5 shows the results obtained by expanding the categories significant in Table S4 (PET and T1 features, domain targeted cognitive features being expanded at maximum level).

Characteristic	**coeff.**	**p-value**	**corrected p-value**	**number of exp.**
intercept	78	**0**	**0**	NA
linear model	-1.4	0.32	0.46	23
generalized linear model	0.34	0.81	0.90	28
non linear model	1.87	0.17	0.37	50
imaging features	2.2	**0.012**	0.074	94
cognitive features	2	**0.033**	0.12	53
socio-demographic features and APOE	1.7	0.093	0.24	49
biomarkers	1	0.24	0.41	19
other features	0.46	0.58	0.75	12
longitudinal	0.22	0.83	0.90	13
ADNI	-1.4	0.25	0.41	106
number of subjects	-0.018	0.99	0.99	NA
individual intercept	1.2	**0.017**	0.074	NA

Table S1: Impact of method characteristics on all experiments, using the broadest categories. This table shows the coefficients obtained using the linear mixed-effect model described in section 2.4.1 on all experiments, the associated p-values and corrected p-values. The last columns shows the number of experiments using the given characteristic, out of the 120 experiments included in the model. Benjamini-Hochberg procedure was applied to get corrected p-values. coeff.:coefficient of the characteristics in the mixed effect model; APOE: Apolipoprotein E; ADNI: Alzheimer's Disease Neuroimaging Initiative; NA: not applicable

Characteristic	**coeff.**	**p-value**	**corrected p-value**	**number of exp.**
intercept	78	**0**	**0**	NA
linear model	-0.79	0.67	0.86	23
generalized linear model	0.2	0.89	0.92	28
non linear model	0.95	0.49	0.73	50
T1 features	0.99	0.21	0.53	77
PET	2.82	$\mathbf{6.8*10^{-03}}$	**0.03**	25
white matter hyper-intensities	-0.64	0.44	0.72	3
EEG/MEG	3.5	$\mathbf{1.1*10^{-03}}$	**0.01**	5
general cognitive features	0.5	0.67	0.86	49
domain targeted cognitive features	2.2	0.051	0.15	25
new or specific cognitive features	1.1	0.4	0.72	2
socio-demographic features and APOE	2.8	$\mathbf{4.7*10^{-03}}$	0.028	49
biomarkers	0.84	0.32	0.72	19
other features	0.75	0.37	0.72	12
longitudinal	0.34	0.74	0.88	13
ADNI	0.2	0.9	0.92	106
number of subjects	-0.13	0.92	0.92	NA
individual intercept	2.7	**0.03**	0.11	NA

Table S2: Impact of method characteristics on all experiments, after refining the categories that were significant in table S1. This table shows the coefficients obtained using the linear mixed-effect model described in section 2.4.1 on all experiments, the associated p-values and corrected p-values. The last columns shows the number of experiments using the given characteristic, out of the 120 experiments included in the model. Benjamini-Hochberg procedure was applied to get corrected p-values. coeff.:coefficient of the characteristics in the mixed effect model; PET: positron emission tomography; EEG: electroencephalography; MEG: magnetoencephalography; APOE: Apolipoprotein E; ADNI: Alzheimer's Disease Neuroimaging Initiative; NA: not applicable

Characteristic	**coeff.**	**p-value**	**corrected p-value**	**number of exp.**
intercept	78	**0**	**0**	NA
linear model	0.5	0.72	0.86	16
generalized linear mode	0.91	0.51	0.68	27
non linear mode	1.7	0.19	0.45	48
imaging features	2.3	$\mathbf{3.5{*}10^{-04}}$	$\mathbf{2.1{*}10^{-03}}$	87
cognitive features	2.4	$\mathbf{3{*}10^{-03}}$	**0.012**	48
socio-demographic features and APOE	0.099	0.92	0.96	42
biomarkers	0.6	0.46	0.68	16
other features	0.79	0.26	0.52	9
longitudinal	0.86	0.34	0.59	13
number of subjects	0.074	0.96	0.96	NA
individual intercept	4.1	$\mathbf{8.9{*}10^{-03}}$	**0.027**	NA

Table S3: Impact of method characteristics on ADNI experiments, using the broadest categories. This table shows the coefficients obtained using the linear mixed-effect model described in section 2.4.1 on the ADNI experiments, the associated p-values and corrected p-values. The last columns shows the number of experiments using the given characteristic, out of the 106 experiments included in the model. Benjamini-Hochberg procedure was applied to get corrected p-values. coeff.:coefficient of the characteristics in the mixed effect model; APOE: Apolipoprotein E; NA: not applicable; ADNI: Alzheimer's Disease Neuroimaging Initiative

Characteristic	**coeff.**	**p-value**	**corrected p-value**	**number of exp.**
intercept	77	**0**	**0**	NA
linear model	1.9	0.11	0.25	16
generalized linear model	1.4	0.26	0.4	27
non linear model	1.3	0.23	0.4	48
T1 features	1.4	$\mathbf{6.3*10^{-03}}$	**0.022**	77
PET	4.6	$\mathbf{3.8*10^{-09}}$	$\mathbf{2.7*10^{-08}}$	25
general cognitive features	0.64	0.48	0.56	46
domain targeted cognitive features	2.3	$\mathbf{9.1*10^{-03}}$	**0.026**	23
socio-demographic features and APOE	0.69	0.43	0.53	42
biomarkers	0.18	0.8	0.8	16
other features	0.55	0.37	0.51	9
longitudinal	0.94	0.22	0.4	13
number of subjects	0.73	0.62	0.67	NA
individual intercept	9.9	$\mathbf{4.7*10^{-03}}$	**0.022**	NA

Table S4: Impact of method characteristics on ADNI experiments, after refining the categories that were significant in table S3. This table shows the coefficients obtained using the linear mixed-effect model described in section 2.4.1 on the ADNI experiments, the associated p-values and corrected p-values. The last columns shows the number of experiments using the given characteristic, out of the 106 experiments included in the model. Benjamini-Hochberg procedure was applied to get corrected p-values. coeff.:coefficient of the characteristics in the mixed effect model; PET: positron emission tomography; APOE: Apolipoprotein E; NA: not applicable; ADNI: Alzheimer's Disease Neuroimaging Initiative

Characteristic	**coeff.**	**p-value**	**corrected p-value**	**number of exp.**
intercept	77	**0**	**0**	NA
linear model	2	0.095	0.26	16
generalized linear model	1.4	0.27	0.42	27
non linear model	1.3	0.24	0.41	48
T1 region-based features on selected regions of interest	1.1	0.11	0.26	17
T1 region-based features on the whole brain	0.71	0.31	0.44	20
T1 voxel-based features	1.7	**0.014**	**0.046**	42
amyloid PET	1.5	0.38	0.5	5
FDG PET	4.38	$\mathbf{1.5{*}10^{-08}}$	$\mathbf{1.3{*}10^{-07}}$	24
general cognitive features	0.31	0.74	0.83	46
domain targeted cognitive features	2.4	$\mathbf{9.1{*}10^{-03}}$	**0.039**	23
socio-demographic features and APOE	1.1	0.18	0.37	23
biomarkers	0.15	0.83	0.83	16
other features	0.39	0.53	0.64	9
longitudinal	0.96	0.2	0.37	13
number of subjects	0.39	0.8	0.83	NA
individual intercept	11	$\mathbf{5{*}10^{-03}}$	**0.029**	NA

Table S5: Impact of method characteristics on ADNI experiments, after refining the categories that were significant in table S4. This table shows the coefficients obtained using the linear mixed-effect model described in section 2.4.1 on the ADNI experiments, the associated p-values and corrected p-values. The last columns shows the number of experiments using the given characteristic, out of the 106 experiments included in the model. Benjamini-Hochberg procedure was applied to get corrected p-values. coeff.:coefficient of the characteristics in the mixed effect model; PET: positron emission tomography; FDG: fluorodeoxyglucose; APOE: Apolipoprotein E; NA: not applicable; ADNI: Alzheimer's Disease Neuroimaging Initiative

Appendix A.5. Journals and conference proceedings

Table S6 shows the journals and conference proceedings in which more than one included article has been published, and the associated number of articles.

Journal or conference proceedings	**Number of included articles**
Journal of Alzheimer's Disease	12
NeuroImage	11
Lecture Notes in Computer Science	7
PLoS ONE	9
Neurobiology of Aging	6
Neurology	3
Brain Topography	3
Current Alzheimer Research	3
Medical Image Analysis	3
Frontiers in Aging Neuroscience	3
Scientific Reports	2
Frontiers in Neuroscience	2
IEEE Journal of Biomedical and Health Informatics	2
IEEE Transactions on Biomedical Engineering	2
NeuroImage: Clinical	2
Journal of Neuroscience Methods	2

Table S6: Number of included articles published in each journal or conference proceedings. Only the journals with more than one included article are shown here. The articles taken into account are the one considered for analysis, and that use a large enough data set.

Appendix B. Articles included in the review

Aguilar, C., Muehlboeck, J.S., Mecocci, P., Vellas, B., Tsolaki, M., Kloszewska, I., Soininen, H., Lovestone, S., Wahlund, L.O., Simmons, A., Westman, E., 2014. Application of a MRI based index to longitudinal atrophy change in Alzheimer disease, mild cognitive impairment and healthy older individuals in the AddNeuroMed cohort. Frontiers in Aging Neuroscience 6. doi:10.3389/fnagi.2014.00145. 00000.

Aguilar, C., Westman, E., Muehlboeck, J.S., Mecocci, P., Vellas, B., Tsolaki, M., Kloszewska, I., Soininen, H., Lovestone, S., Spenger, C., Simmons, A., Wahlund, L.O., 2013. Different multivariate techniques for automated classification of MRI data in Alzheimer's disease and mild cognitive impairment. Psychiatry Research: Neuroimaging 212, 89–98. doi:10.1016/j.pscychresns.2012.11.005. 00082.

Arbizu, J., Prieto, E., Martínez-Lage, P., Martí-Climent, J., García-Granero, M., Lamet, I., Pastor, P., Riverol, M., Gómez-Isla, M., Peñuelas, I., Richter, J., Weiner, M., 2013. Automated analysis of FDG PET as a tool for single-subject probabilistic prediction and detection of Alzheimer's disease dementia. European Journal of Nuclear Medicine and Molecular Imaging 40, 1394–1405. doi:10.1007/s00259-013-2458-z.

Arco, J., Ramírez, J., Górriz, J., Puntonet, C., Ruz, M., 2016. Short-term prediction of MCI to AD conversion based on longitudinal MRI analysis and neuropsychological tests. Smart Innovation, Systems and Technologies 45, 385–394. doi:10.1007/978-3-319-23024-5_35.

Ardekani, B.A., Bermudez, E., Mubeen, A.M., Bachman, A.H., 2017. Prediction of Incipient Alzheimer's Disease Dementia in Patients with Mild Cognitive Impairment. Journal of Alzheimer's Disease 55, 269–281. doi:10.3233/JAD-160594.

Ayub, A., Farhan, S., Fahiem, M., Tauseef, H., 2017. A novel approach for the prediction of conversion from mild cognitive impairment to Alzheimer's disease using MRI images. Advances in Electrical and Computer Engineering 17, 113–122. doi:10.4316/AECE.2017.02015.

Beer, J., Snitz, B., Chang, C.C., Loewenstein, D., Ganguli, M., 2018. Does a cognitive stress test predict progression from mild cognitive impairment to dementia equally well in clinical versus population-based settings? International Psychogeriatrics 30, 1435–1445. doi:10.1017/S1041610217002666.

Beheshti, I., Demirel, H., Matsuda, H., 2017a. Classification of Alzheimer's disease and prediction of mild cognitive impairment-to-Alzheimer's conversion from structural magnetic resource imaging using feature ranking and a genetic algorithm. Computers in Biology and Medicine 83, 109–119. doi:10.1016/j.compbiomed.2017.02.011.

Beheshti, I., Maikusa, N., Daneshmand, M., Matsuda, H., Demirel, H., Anbarjafari, G., 2017b. Classification of Alzheimer's disease and prediction of mild cognitive

impairment conversion using histogram-based analysis of patient-specific anatomical brain connectivity networks. Journal of Alzheimer's Disease 60, 295–304. doi:10.3233/JAD-161080.

Belleville, S., Gauthier, S., Lepage, E., Kergoat, M.J., Gilbert, B., 2014. Predicting decline in mild cognitive impairment: A prospective cognitive study. Neuropsychology 28, 643–652. doi:10.1037/neu0000063. 00037.

Borroni, B., Anchisi, D., Paghera, B., Vicini, B., Kerrouche, N., Garibotto, V., Terzi, A., Vignolo, L., Di Luca, M., Giubbini, R., Padovani, A., Perani, D., 2006. Combined 99mtc-ECD SPECT and neuropsychological studies in MCI for the assessment of conversion to AD. Neurobiology of Aging 27, 24–31. doi:10.1016/j.neurobiolaging.2004.12.010. 00160.

Buratti, L., Balestrini, S., Altamura, C., Viticchi, G., Falsetti, L., Luzzi, S., Provinciali, L., Vernieri, F., Silvestrini, M., 2015. Markers for the Risk of Progression from Mild Cognitive Impairment to Alzheimer's Disease. Journal of Alzheimer's Disease 45, 883–890. doi:10.3233/JAD-143135.

Buscema, M., Grossi, E., Capriotti, M., Babiloni, C., Rossini, P., 2010. The I.F.A.S.T. model allows the prediction of conversion to alzheimer disease in patients with mild cognitive impairment with high degree of accuracy. Current Alzheimer Research 7, 173–187. doi:10.2174/156720510790691137. 00031.

Cabranes, J., De Juan, R., Encinas, M., Marcos, A., Gil, P., Fernández, C., De Ugarte, C., Barabash, A., 2004. Relevance of functional neuroimaging in the progression of mild cognitive impairment. Neurological Research 26, 496–501. doi:10.1179/016164104225016155.

Callahan, B., Ramirez, J., Berezuk, C., Duchesne, S., Black, S., 2015. Predicting Alzheimer's disease development: A comparison of cognitive criteria and associated neuroimaging biomarkers. Alzheimer's Research and Therapy 7. doi:10.1186/s13195-015-0152-z. 00021.

Caminiti, S., Ballarini, T., Sala, A., Cerami, C., Presotto, L., Santangelo, R., Fallanca, F., Vanoli, E., Gianolli, L., Iannaccone, S., Magnani, G., Perani, D., Parnetti, L., Eusebi, P., Frisoni, G., Nobili, F., Picco, A., Scarpini, E., BIOMARKAPD, P., 2018. FDG-PET and CSF biomarker accuracy in prediction of conversion to different dementias in a large multicentre MCI cohort. NeuroImage: Clinical 18, 167–177. doi:10.1016/j.nicl.2018.01.019.

Chen, R., Young, K., Chao, L., Miller, B., Yaffe, K., Weiner, M., Herskovits, E., 2012. Prediction of conversion from mild cognitive impairment to Alzheimer disease based on Bayesian data mining with ensemble learning. Neuroradiology Journal 25, 5–16. doi:10.1177/197140091202500101.

Chen, X., Zhou, Y., Wang, R., Cao, H., Reid, S., Gao, R., Han, D., 2016. Potential clinical value of multiparametric pet in the prediction of Alzheimer's disease progression. PLoS ONE 11. doi:10.1371/journal.pone.0154406.

Cheng, B., Liu, M., Suk, H.I., Shen, D., Zhang, D., 2015a. Multimodal manifold-regularized transfer learning for MCI conversion prediction. Brain Imaging and Behavior 9, 913–926. doi:10.1007/s11682-015-9356-x.

Cheng, B., Liu, M., Zhang, D., Munsell, B., Shen, D., 2015b. Domain Transfer Learning for MCI Conversion Prediction. IEEE Transactions on Biomedical Engineering 62, 1805–1817. doi:10.1109/TBME.2015.2404809. 00036.

Chincarini, A., Bosco, P., Calvini, P., Gemme, G., Esposito, M., Olivieri, C., Rei, L., Squarcia, S., Rodriguez, G., Bellotti, R., Cerello, P., De Mitri, I., Retico, A., Nobili, F., 2011. Local MRI analysis approach in the diagnosis of early and prodromal Alzheimer's disease. NeuroImage 58, 469–480. doi:10.1016/j.neuroimage.2011.05.083. 00122.

Cho, Y., Seong, J.K., Jeong, Y., Shin, S., 2012. Individual subject classification for Alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data. NeuroImage 59, 2217–2230. doi:10.1016/j.neuroimage.2011.09.085.

Choi, H., Jin, K., 2018. Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging. Behavioural Brain Research 344, 103–109. doi:10.1016/j.bbr.2018.02.017.

Choo, I., Ni, R., Schöll, M., Wall, A., Almkvist, O., Nordberg, A., 2013. Combination of 18f-FDG PET and cerebrospinal fluid biomarkers as a better predictor of the progression to Alzheimer's disease in mild cognitive impairment patients. Journal of Alzheimer's Disease 33, 929–939. doi:10.3233/JAD-2012-121489. 00041.

Clark, D., 2011. Residual vectors for Alzheimer disease diagnosis and prognostication. Brain and Behavior 1, 142–152. doi:10.1002/brb3.19. 00007.

Clark, D., Kapur, P., Geldmacher, D., Brockington, J., Harrell, L., DeRamus, T., Blanton, P., Lokken, K., Nicholas, A., Marson, D., 2014. Latent Information in Fluency Lists Predicts Functional Decline in Persons at Risk for Alzheimer Disease. Cortex; a journal devoted to the study of the nervous system and behavior 55, 202–218. doi:10.1016/j.cortex.2013.12.013. 00017.

Clark, D., McLaughlin, P., Woo, E., Hwang, K., Hurtz, S., Ramirez, L., Eastman, J., Dukes, R.M., Kapur, P., DeRamus, T., Apostolova, L., 2016. Novel verbal fluency scores and structural brain imaging for prediction of cognitive outcome in mild cognitive impairment. Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring 2, 113–122. doi:10.1016/j.dadm.2016.02.001.

Collij, L., Heeman, F., Kuijer, J., Ossenkoppele, R., Benedictus, M., Möller, C., Verfaillie, S., Sanz-Arigita, E., Van Berckel, B., Van Der Flier, W., Scheltens, P., Barkhof, F., Wink, A., 2016. Application of machine learning to arterial spin labeling in mild cognitive impairment and Alzheimer disease. Radiology 281, 865–875. doi:10.1148/radiol.2016152703.

Convit, A., De Asis, J., De Leon, M., Tarshish, C., De Santi, S., Rusinek, H., 2000. Atrophy of the medial occipitotemporal, inferior, and middle temporal gyri in non-demented elderly predict decline to Alzheimer's disease. Neurobiology of Aging 21, 19–26. doi:10.1016/S0197-4580(99)00107-4. 00306.

Costafreda, S., Dinov, I., Tu, Z., Shi, Y., Liu, C.Y., Kloszewska, I., Mecocci, P., Soininen, H., Tsolaki, M., Vellas, B., Wahlund, L.O., Spenger, C., Toga, A., Lovestone, S., Simmons, A., 2011. Automated hippocampal shape analysis predicts the onset of dementia in mild cognitive impairment. NeuroImage 56, 212–219. doi:10.1016/j.neuroimage.2011.01.050. 00149.

Cui, Y., Liu, B., Luo, S., Zhen, X., Fan, M., Liu, T., Zhu, W., Park, M., Jiang, T., Jin, J.S., Initiative, t.A.D.N., 2011. Identification of Conversion from Mild Cognitive Impairment to Alzheimer's Disease Using Multivariate Predictors. PLOS ONE 6, e21896. doi:10.1371/journal.pone.0021896.

Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehéricy, S., Habert, M.O., Chupin, M., Benali, H., Colliot, O., 2011. Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database. NeuroImage 56, 766–781. doi:10.1016/j.neuroimage.2010.06.013.

Davatzikos, C., Bhatt, P., Shaw, L., Batmanghelich, K., Trojanowski, J., 2011. Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. Neurobiology of Aging 32, 2322.e19–2322.e27. doi:10.1016/j.neurobiolaging.2010.05.023. 00377.

Devanand, D., Liu, X., Tabert, M., Pradhaban, G., Cuasay, K., Bell, K., de Leon, M., Doty, R., Stern, Y., Pelton, G., 2008. Combining Early Markers Strongly Predicts Conversion from Mild Cognitive Impairment to Alzheimer's Disease. Biological Psychiatry 64, 871–879. doi:10.1016/j.biopsych.2008.06.020. 00261.

Devanand, D., Pradhaban, G., Liu, X., Khandji, A., De Santi, S., Segal, S., Rusinek, H., Pelton, G., Honig, L., Mayeux, R., Stern, Y., Tabert, M., De Leon, M., 2007. Hippocampal and entorhinal atrophy in mild cognitive impairment: Prediction of Alzheimer disease. Neurology 68, 828–836. doi:10.1212/01.wnl.0000256697.20968.d7. 00599.

Devanand, D., Van Heertum, R., Kegeles, L., Liu, X., Jin, Z., Pradhaban, G., Rusinek, H., Pratap, M., Pelton, G., Prohovnik, I., Stern, Y., Mann, J., Parsey, R., 2010. 99mtc hexamethyl-propylene-aminoxime single-photon emission computed tomography prediction of conversion from mild cognitive impairment to Alzheimer disease. American Journal of Geriatric Psychiatry 18, 959–972. doi:10.1097/JGP.0b013e3181ec8696.

Dierckx, E., Engelborghs, S., De Raedt, R., Van Buggenhout, M., De Deyn, P., Verté, D., Ponjaert-Kristoffersen, I., 2009. Verbal cued recall as a predictor of conversion to Alzheimer's disease in Mild Cognitive Impairment. International Journal of Geriatric Psychiatry 24, 1094–1100. doi:10.1002/gps.2228. 00055.

Doyle, O., Westman, E., Marquand, A., Mecocci, P., Vellas, B., Tsolaki, M., Kłoszewska, I., Soininen, H., Lovestone, S., Williams, S., Simmons, A., 2014. Predicting progression of Alzheimer's disease using ordinal regression. PLoS ONE 9. doi:10.1371/journal.pone.0105542. 00028.

Drzezga, A., Grimmer, T., Riemenschneider, M., Lautenschlager, N., Siebner, H., Alexopoulus, P., Minoshima, S., Schwaiger, M., Kurz, A., 2005. Prediction of individual clinical outcome in MCI by means of genetic assessment and 18f-FDG PET. Journal of Nuclear Medicine 46, 1625–1632.

Duchesne, S., Bocti, C., De Sousa, K., Frisoni, G., Chertkow, H., Collins, D., 2010. Amnestic MCI future clinical status prediction using baseline MRI features. Neurobiology of Aging 31, 1606–1617. doi:10.1016/j.neurobiolaging.2008.09.003.

Dukart, J., Sambataro, F., Bertolino, A., 2015. Accurate prediction of conversion to Alzheimer's disease using imaging, genetic, and neuropsychological biomarkers. Journal of Alzheimer's Disease 49, 1143–1159. doi:10.3233/JAD-150570. 00022.

Çitak Er, F., Goularas, D., Ormeci, B., The, A.D.N.I., 2017. A novel convolutional neural network model based on voxel-based morphometry of imaging data in predicting the prognosis of patients with mild cognitive impairment. Journal of Neurological Sciences 34, 52–69.

Eskildsen, S., Coupé, P., Fonov, V., Pruessner, J., Collins, D., 2015. Structural imaging biomarkers of Alzheimer's disease: Predicting disease progression. Neurobiology of Aging 36, S23–S31. doi:10.1016/j.neurobiolaging.2014.04.034. 00058.

Eskildsen, S., Coupé, P., García-Lorenzo, D., Fonov, V., Pruessner, J., Collins, D., 2013. Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning. NeuroImage 65, 511–521. doi:10.1016/j.neuroimage.2012.09.058. 00137.

Ewers, M., Buerger, K., Teipel, S., Scheltens, P., Schröder, J., Zinkowski, R., Bouwman, F., Schönknecht, P., Schoonenboom, N., Andreasen, N., Wallin, A., DeBernardis, J., Kerkman, D., Heindl, B., Blennow, K., Hampel, H., 2007. Multicenter assessment of CSF-phosphorylated tau for the prediction of conversion of MCI. Neurology 69, 2205–2212. doi:10.1212/01.wnl.0000286944.22262.ff. 00161.

Ewers, M., Walsh, C., Trojanowski, J., Shaw, L., Petersen, R., Jack, C., Feldman, H., Bokde, A., Alexander, G., Scheltens, P., Vellas, B., Dubois, B., Weiner, M., Hampel, H., 2012. Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance. Neurobiology of Aging 33, 1203–1214.e2. doi:10.1016/j.neurobiolaging.2010.10.019.

Falahati, F., Ferreira, D., Soininen, H., Mecocci, P., Vellas, B., Tsolaki, M., Kłoszewska, I., Lovestone, S., Eriksdotter, M., Wahlund, L.O., Simmons, A.,

Westman, E., 2016. The Effect of Age Correction on Multivariate Classification in Alzheimer's Disease, with a Focus on the Characteristics of Incorrectly and Correctly Classified Subjects. Brain Topography 29, 296–307. doi:10.1007/s10548-015-0455-1.

Fayed, N., Modrego, P., García-Martí, G., Sanz-Requena, R., Marti-Bonmatí, L., 2017. Magnetic resonance spectroscopy and brain volumetry in mild cognitive impairment. A prospective study. Magnetic Resonance Imaging 38, 27–32. doi:10.1016/j.mri.2016.12.010.

Ferreira, F., Cardoso, S., Silva, D., de Mendonça, A., Madeira, S., 2017. Improving prognostic prediction from mild cognitive impairment to Alzheimer's disease using genetic algorithms. Advances in Intelligent Systems and Computing 616, 180–188. doi:10.1007/978-3-319-60816-7_22.

Fleisher, A., Sun, S., Taylor, C., Ward, C., Gamst, A., Petersen, R., Jack, C., Aisen, P., Thal, L., 2008. Volumetric MRI vs clinical predictors of Alzheimer disease in mild cognitive impairment. Neurology 70, 191–199. doi:10.1212/01.wnl.0000287091.57376.65. 00178.

Fritzsche, K., Stieltjes, B., Schlindwein, S., Van Bruggen, T., Essig, M., Meinzer, H.P., 2010. Automated MR morphometry to predict Alzheimer's disease in mild cognitive impairment. International Journal of Computer Assisted Radiology and Surgery 5, 623–632. doi:10.1007/s11548-010-0412-0.

Frölich, L., Peters, O., Lewczuk, P., Gruber, O., Teipel, S., Gertz, H., Jahn, H., Jessen, F., Kurz, A., Luckhaus, C., Hüll, M., Pantel, J., Reischies, F., Schröder, J., Wagner, M., Rienhoff, O., Wolf, S., Bauer, C., Schuchhardt, J., Heuser, I., Rüther, E., Henn, F., Maier, W., Wiltfang, J., Kornhuber, J., 2017. Incremental value of biomarker combinations to predict progression of mild cognitive impairment to Alzheimer's dementia. Alzheimer's Research and Therapy 9. doi:10.1186/s13195-017-0301-7.

Gallagher, D., Mhaolain, A., Coen, R., Walsh, C., Kilroy, D., Belinski, K., Bruce, I., Coakley, D., Walsh, J., Cunningham, C., Lawlor, B., 2010. Detecting prodromal Alzheimer's disease in mild cognitive impairment: Utility of the CAMCOG and other neuropsychological predictors. International Journal of Geriatric Psychiatry 25, 1280–1287. doi:10.1002/gps.2480.

García-Herranz, S., Díaz-Mardomingo, M., Peraita, H., 2016. Neuropsychological predictors of conversion to probable Alzheimer disease in elderly with mild cognitive impairment. Journal of Neuropsychology 10, 239–255. doi:10.1111/jnp.12067.

Gaser, C., Franke, K., Klöppel, S., Koutsouleris, N., Sauer, H., 2013. BrainAGE in Mild Cognitive Impaired Patients: Predicting the Conversion to Alzheimer's Disease. PLoS ONE 8. doi:10.1371/journal.pone.0067346. 00097.

Gavidia-Bovadilla, G., Kanaan-Izquierdo, S., Mataroa-Serrat, M., Perera-Lluna, A., 2017. Early prediction of Alzheimer's disease using null longitudinal model-based classifiers. PLoS ONE 12. doi:10.1371/journal.pone.0168011.

Geslani, D., Tierney, M., Herrmann, N., Szalai, J., 2005. Mild cognitive impairment: An operational definition and its conversion rate to Alzheimer's disease. Dementia and Geriatric Cognitive Disorders 19, 383–389. doi:10.1159/000084709. 00178.

Griffith, H., Netson, K., Harrell, L., Zamrini, E., Brockington, J., Marson, D., 2006. Amnestic mild cognitive impairment: Diagnostic outcomes and clinical prediction over a two-year time period. Journal of the International Neuropsychological Society 12, 166–175. doi:10.1017/S1355617706060267. 00118.

Guo, S., Lai, C., Wu, C., Cen, G., 2017. Conversion discriminative analysis on mild cognitive impairment using multiple cortical features from MR images. Frontiers in Aging Neuroscience 9. doi:10.3389/fnagi.2017.00146.

Hall, A., Mattila, J., Koikkalainen, J., Lötjonen, J., Wolz, R., Scheltens, P., Frisoni, G., Tsolaki, M., Nobili, F., Freund-Levi, Y., Minthon, L., Frölich, L., Hampel, H., Visser, P., Soininen, H., 2015. Predicting progression from cognitive impairment to alzheimer's disease with the disease state index. Current Alzheimer Research 12, 69–79. doi:10.2174/1567205012666141218123829.

Hinrichs, C., Singh, V., Xu, G., Johnson, S.C., 2011. Predictive markers for AD in a multi-modality framework: An analysis of MCI progression in the ADNI population. NeuroImage 55, 574–589. doi:10.1016/j.neuroimage.2010.10.081.

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

Hojjati, S., Ebrahimzadeh, A., Khazaee, A., Babajani-Feremi, A., Alzheimer's, D.N.I., 2018. Predicting conversion from MCI to AD by integrating rs-fMRI and structural MRI. Computers in Biology and Medicine 102, 30–39. doi:10.1016/j.compbiomed.2018.09.004.

Huang, C., Wahlund, L.O., Svensson, L., Winblad, B., Julin, P., 2002. Cingulate cortex hypoperfusion predicts Alzheimer's disease in mild cognitive impairment. BMC Neurology 2. doi:10.1186/1471-2377-2-9. 00222.

Huang, M., Yang, W., Feng, Q., Chen, W., Weiner, M.W., Aisen, P., Petersen, R., Jack, C.R., Jagust, W., Trojanowki, J.Q., Toga, A.W., Beckett, L., Green, R.C., Saykin, A.J., Morris, J., Shaw, L.M., Kaye, J., Quinn, J., Silbert, L., Lind, B., Carter, R., Dolen, S., Schneider, L.S., Pawluczyk, S., Beccera, M., Teodoro, L., Spann, B.M., Brewer, J., Vanderswag, H., Fleisher, A., Heidebrink, J.L., Lord, J.L., Mason, S.S., Albers, C.S., Knopman, D., Johnson, K., Doody, R.S., Villanueva-Meyer, J., Chowdhury, M., Rountree, S., Dang, M., Stern, Y., Honig, L.S., Bell, K.L., Ances, B., Morris, J.C., Carroll, M., Creech, M.L., Franklin, E., Mintun, M.A., Schneider, S., Oliver, A., Marson, D., Griffith, R., Clark, D., Geldmacher, D., Brockington, J., Roberson, E., Love, M.N., Grossman, H., Mitsis, E., Shah, R.C., deToledo Morrell, L., Duara, R., Varon, D., Greig, M.T., Roberts, P., Albert, M., Onyike, C., D'Agostino, D., Kielb, S., Galvin, J.E., Cerbone, B., Michel, C.A., Pogorelec, D.M.,

Rusinek, H., de Leon, M.J., Glodzik, L., De Santi, S., Doraiswamy, P.M., Petrella, J.R., Borges-Neto, S., Wong, T.Z., Coleman, E., Smith, C.D., Jicha, G., Hardy, P., Sinha, P., Oates, E., Conrad, G., Porsteinsson, A.P., Goldstein, B.S., Martin, K., Makino, K.M., Ismail, M.S., Brand, C., Mulnard, R.A., Thai, G., Mc-Adams-Ortiz, C., Womack, K., Mathews, D., Quiceno, M., Levey, A.I., Lah, J.J., Cellar, J.S., Burns, J.M., Swerdlow, R.H., Brooks, W.M., Apostolova, L., Tingus, K., Woo, E., Silverman, D.H.S., Lu, P.H., Bartzokis, G., Graff-Radford, N.R., Parfitt, F., Kendall, T., Johnson, H., Farlow, M.R., Hake, A.M., Matthews, B.R., Brosch, J.R., Herring, S., Hunt, C., van Dyck, C.H., Carson, R.E., MacAvoy, M.G., Varma, P., Chertkow, H., Bergman, H., Hosein, C., Black, S., Stefanovic, B., Caldwell, C., Hsiung, G.Y.R., Feldman, H., Mudge, B., Assaly, M., Finger, E., Pasternack, S., Rachisky, I., Trost, D., Kertesz, A., Bernick, C., Munic, D., Mesulam, M.M., Lipowski, K., Weintraub, S., Bonakdarpour, B., Kerwin, D., Wu, C.K., Johnson, N., Sadowsky, C., Villena, T., Turner, R.S., Johnson, K., Reynolds, B., Sperling, R.A., Johnson, K.A., Marshall, G., Yesavage, J., Taylor, J.L., Lane, B., Rosen, A., Tinklenberg, J., Sabbagh, M.N., Belden, C.M., Jacobson, S.A., Sirrel, S.A., Kowall, N., Killiany, R., Budson, A.E., Norbash, A., Johnson, P.L., Obisesan, T.O., Wolday, S., Allard, J., Lerner, A., Ogrocki, P., Tatsuoka, C., Fatica, P., Fletcher, E., Maillard, P., Olichney, J., DeCarli, C., Carmichael, O., Kittur, S., Borrie, M., Lee, T.Y., Bartha, R., Johnson, S., Asthana, S., Carlsson, C.M., Potkin, S.G., Preda, A., Nguyen, D., Tariot, P., Burke, A., Trncic, N., Fleisher, A., Reeder, S., Bates, V., Capote, H., Rainka, M., Scharre, D.W., Kataki, M., Adeli, A., Zimmerman, E.A., Celmins, D., Brown, A.D., Pearlson, G.D., Blank, K., Anderson, K., Flashman, L.A., Seltzer, M., Hynes, M.L., Santulli, R.B., Sink, K.M., Gordineer, L., Williamson, J.D., Garg, P., Watkins, F., Ott, B.R., Querfurth, H., Tremont, G., Salloway, S., Malloy, P., Correia, S., Rosen, H.J., Miller, B.L., Perry, D., Mintzer, J., Spicer, K., Bachman, D., Pomara, N., Hernando, R., Sarrael, A., Relkin, N., Chaing, G., Lin, M., Ravdin, L., Smith, A., Raj, B.A., Fargher, K., 2017. Longitudinal measurement and hierarchical classification framework for the prediction of Alzheimer's disease. Scientific Reports 7, 39880. doi:10.1038/srep39880.

Johnson, P., Vandewater, L., Wilson, W., Maruff, P., Savage, G., Graham, P., Macaulay, L., Ellis, K., Szoeke, C., Martins, R., Rowe, C., Masters, C., Ames, D., Zhang, P., 2014. Genetic algorithm with logistic regression for prediction of progression to Alzheimer's disease. BMC Bioinformatics 15. doi:10.1186/1471-2105-15-S16-S11. 00027.

Junwei Ding, Qiu Huang, 2017. Prediction of MCI to AD conversion using Laplace Eigenmaps learned from FDG and MRI images of AD patients and healthy controls, in: 2017 2nd International Conference on Image, Vision and Computing (ICIVC), IEEE, Chengdu, China. pp. 660–664. doi:10.1109/ICIVC.2017.7984638. 00000.

Kauppi, K., Fan, C., McEvoy, L., Holland, D., Tan, C., Chen, C.H., Andreassen, O., Desikan, R., Dale, A., 2018. Combining polygenic hazard score with volumetric MRI and cognitive measures improves prediction of progression from mild cognitive impairment to Alzheimer's disease. Frontiers in Neuroscience 12. doi:10.3389/fnins.2018.00260.

Kebets, V., Richiardi, J., Assche, M., Goldstein, R., Meulen, M., Vuilleumier, P., Ville, D., Assal, F., 2015. Predicting Pure Amnestic Mild Cognitive Impairment Conversion to Alzheimer's Disease Using Joint Modeling of Imaging and Clinical Data, in: Proceedings - 2015 International Workshop on Pattern Recognition in NeuroImaging, PRNI 2015, pp. 85–88. doi:10.1109/PRNI.2015.23. 00001.

Khan, W., Aguilar, C., Kiddle, S., Doyle, O., Thambisetty, M., Muehlboeck, S., Sattlecker, M., Newhouse, S., Lovestone, S., Dobson, R., Giampietro, V., Westman, E., Simmons, A., 2015a. A subset of cerebrospinal fluid proteins from a multianalyte panel associated with brain atrophy, disease classification and prediction in Alzheimer's disease. PLoS ONE 10. doi:10.1371/journal.pone.0134368.

Khan, W., Westman, E., Jones, N., Wahlund, L.O., Mecocci, P., Vellas, B., Tsolaki, M., Kłoszewska, I., Soininen, H., Spenger, C., Lovestone, S., Muehlboeck, J.S., Simmons, A., 2015b. Automated Hippocampal Subfield Measures as Predictors of Conversion from Mild Cognitive Impairment to Alzheimer's Disease in Two Independent Cohorts. Brain Topography 28, 746–759. doi:10.1007/s10548-014-0415-1.

Kim, J.W., Byun, M.S., Sohn, B.K., Yi, D., Seo, E.H., Choe, Y.M., Kim, S.G., Choi, H.J., Lee, J.H., Chee, I.S., Woo, J.I., Lee, D.Y., 2017. Clinical Dementia Rating Orientation Score as an Excellent Predictor of the Progression to Alzheimer's Disease in Mild Cognitive Impairment. Psychiatry Investigation 14, 420. doi:10.4306/pi.2017.14.4.420. 00006.

Klöppel, S., Peter, J., Ludl, A., Pilatus, A., Maier, S., Mader, I., Heimbach, B., Frings, L., Egger, K., Dukart, J., Schroeter, M., Perneczky, R., Häussermann, P., Vach, W., Urbach, H., Teipel, S., Hüll, M., Abdulkadir, A., 2015. Applying automated MR-based diagnostic methods to the memory clinic: A prospective study. Journal of Alzheimer's Disease 47, 939–954. doi:10.3233/JAD-150334.

Koppara, A., Wolfsgruber, S., Kleineidam, L., Schmidtke, K., Frölich, L., Kurz, A., Schulz, S., Hampel, H., Heuser, I., Peters, O., Reischies, F., Jahn, H., Luckhaus, C., Hüll, M., Gertz, H.J., Schröder, J., Pantel, J., Rienhoff, O., Rüther, E., Henn, F., Wiltfang, J., Maier, W., Jessen, F., Kornhuber, J., Wagner, M., 2015. The latent dementia phenotype δ is associated with cerebrospinal fluid biomarkers of Alzheimer's disease and predicts conversion to dementia in subjects with mild cognitive impairment. Journal of Alzheimer's Disease 49, 547–560. doi:10.3233/JAD-150257.

Korolev, I.O., Symonds, L.L., Bozoki, A.C., Initiative, A.D.N., 2016. Predicting Progression from Mild Cognitive Impairment to Alzheimer's Dementia Using Clinical, MRI, and Plasma Biomarkers via Probabilistic Pattern Classification. PLOS ONE 11, e0138866. doi:10.1371/journal.pone.0138866.

Kruczyk, M., Zetterberg, H., Hansson, O., Rolstad, S., Minthon, L., Wallin, A., Blennow, K., Komorowski, J., Andersson, M., 2012. Monte Carlo feature selection and rule-based models to predict Alzheimer's disease in mild cognitive impairment. Journal of Neural Transmission 119, 821–831. doi:10.1007/s00702-012-0812-0.

Lehallier, B., Essioux, L., Gayan, J., Alexandridis, R., Nikolcheva, T., Wyss-Coray, T., Britschgi, M., 2016. Combined plasma and cerebrospinal fluid signature for the prediction of midterm progression from mild cognitive impairment to Alzheimer disease. JAMA Neurology 73, 203–212. doi:10.1001/jamaneurol.2015.3135.

Lei, B., Chen, S., Ni, D., Wang, T., 2016. Discriminative learning for Alzheimer's disease diagnosis via canonical correlation analysis and multimodal fusion. Frontiers in Aging Neuroscience 8. doi:10.3389/fnagi.2016.00077.

Lei, B., Yang, P., Wang, T., Chen, S., Ni, D., 2017. Relational-Regularized Discriminative Sparse Learning for Alzheimer's Disease Diagnosis. IEEE Transactions on Cybernetics 47, 1102–1113. doi:10.1109/TCYB.2016.2644718. 00018.

Li, H., Liu, Y., Gong, P., Zhang, C., Ye, J., 2014. Hierarchical interactions model for predicting Mild Cognitive Impairment (MCI) to Alzheimer's Disease (AD) conversion. PLoS ONE 9. doi:10.1371/journal.pone.0082450. 00016.

Li, Y., Jiang, J., Shen, T., Wu, P., Zuo, C., 2018a. Radiomics features as predictors to distinguish fast and slow progression of Mild Cognitive Impairment to Alzheimer's disease, in: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, pp. 127–130. doi:10.1109/EMBC.2018.8512273.

Li, Y., Yao, Z., Zhang, H., Hu, B., for, t.A.D.N.I., 2018b. Indirect relation based individual metabolic network for identification of mild cognitive impairment. Journal of Neuroscience Methods 309, 188–198. doi:10.1016/j.jneumeth.2018.09.007.

Lin, W., Tong, T., Gao, Q., Guo, D., Du, X., Yang, Y., Guo, G., Xiao, M., Du, M., Qu, X., 2018. Convolutional neural networks-based MRI image analysis for the Alzheimer's disease prediction from mild cognitive impairment. Frontiers in Neuroscience 12. doi:10.3389/fnins.2018.00777.

Liu, H., Zhou, X., Jiang, H., He, H., Liu, X., 2016. A semi-mechanism approach based on MRI and proteomics for prediction of conversion from mild cognitive impairment to Alzheimer's disease. Scientific Reports 6. doi:10.1038/srep26712.

Liu, X., Tosun, D., Weiner, M., Schuff, N., 2013. Locally linear embedding (LLE) for MRI based Alzheimer's disease classification. NeuroImage 83, 148–157. doi:10.1016/j.neuroimage.2013.06.033. 00069.

Long, X., Chen, L., Jiang, C., Zhang, L., 2017. Prediction and classification of Alzheimer disease based on quantification of MRI deformation. PLoS ONE 12. doi:10.1371/journal.pone.0173372.

Lonie, J., Parra-Rodriguez, M., Tierney, K., Herrmann, L., Donaghey, C., O'Carroll, R., Ebmeier, K., 2010. Predicting outcome in mild cognitive impairment: 4-Year follow-up study. British Journal of Psychiatry 197, 135–140. doi:10.1192/bjp.bp.110.077958.

López, M., Turrero, A., Cuesta, P., López-Sanz, D., Bruña, R., Marcos, A., Gil, P., Yus, M., Barabash, A., Cabranes, J., Maestú, F., Fernández, A., 2016. Searching for Primary Predictors of Conversion from Mild Cognitive Impairment to Alzheimer's Disease: A Multivariate Follow-Up Study. Journal of Alzheimer's Disease 52, 133–143. doi:10.3233/JAD-151034.

Mangialasche, F., Westman, E., Kivipelto, M., Muehlboeck, J.S., Cecchetti, R., Baglioni, M., Tarducci, R., Gobbi, G., Floridi, P., Soininen, H., Kłoszewska, I., Tsolaki, M., Vellas, B., Spenger, C., Lovestone, S., Wahlund, L.O., Simmons, A., Mecocci, P., 2013. Classification and prediction of clinical diagnosis of Alzheimer's disease based on MRI and plasma measures of α-/γ-tocotrienols and γ-tocopherol. Journal of Internal Medicine 273, 602–621. doi:10.1111/joim.12037. 00055.

Marcos, A., Gil, P., Barabash, A., Rodriguez, R., Encinas, M., Fernández, C., Cabranes, J., 2006. Neuropsychological markers of progression from mild cognitive impairment to Alzheimer's disease. American Journal of Alzheimer's Disease and other Dementias 21, 189–196. doi:10.1177/1533317506289348. 00041.

Marra, C., Gainotti, G., Fadda, L., Perri, R., Lacidogna, G., Scaricamazza, E., Piccininni, C., Quaranta, D., 2015. Usefulness of an Integrated Analysis of Different Memory Tasks to Predict the Progression from Mild Cognitive Impairment to Alzheimer's Disease: The Episodic Memory Score (EMS). Journal of Alzheimer's Disease 50, 61–70. doi:10.3233/JAD-150613.

Martinez-Murcia, F., Górriz, J., Ramírez, J., Segovia, F., Salas-Gonzalez, D., Castillo-Barnes, D., Ortiz, A., 2018. Assessing mild cognitive impairment progression using a spherical brain mapping of magnetic resonance imaging. Journal of Alzheimer's Disease 65, 713–729. doi:10.3233/JAD-170403.

Martínez-Torteya, A., Rodríguez-Rojas, J., Celaya-Padilla, J., Galván-Tejada, J., Treviño, V., Tamez-Peña, J., 2014. MRI signal and texture features for the prediction of MCI to Alzheimer's disease progression, in: Progress in Biomedical Optics and Imaging - Proceedings of SPIE, p. 903526. doi:10.1117/12.2043903. 00002.

Mathotaarachchi, S., Pascoal, T.A., Shin, M., Benedet, A.L., Kang, M.S., Beaudry, T., Fonov, V.S., Gauthier, S., Rosa-Neto, P., 2017. Identifying incipient dementia individuals using machine learning and amyloid imaging. Neurobiology of Aging 0. doi:10.1016/j.neurobiolaging.2017.06.027.

Mazzeo, S., Santangelo, R., Bernasconi, M., Cecchetti, G., Fiorino, A., Pinto, P., Passerini, G., Falautano, M., Comi, G., Magnani, G., 2016. Combining Cerebrospinal Fluid Biomarkers and Neuropsychological Assessment: A Simple and Cost-Effective Algorithm to Predict the Progression from Mild Cognitive Impairment to Alzheimer's Disease Dementia. Journal of Alzheimer's Disease 54, 1495–1508. doi:10.3233/JAD-160360.

Minhas, S., Khanum, A., Riaz, F., Alvi, A., Khan, S., 2015. Early alzheimer's disease prediction in machine learning setup: Empirical analysis with missing value computation. Lecture Notes in Computer Science (including subseries Lecture Notes in

Artificial Intelligence and Lecture Notes in Bioinformatics) 9375 LNCS, 424–432. doi:10.1007/978-3-319-24834-9_49.

Minhas, S., Khanum, A., Riaz, F., Alvi, A., Khan, S., 2017a. A Nonparametric Approach for Mild Cognitive Impairment to AD Conversion Prediction: Results on Longitudinal Data. IEEE Journal of Biomedical and Health Informatics 21, 1403–1410. doi:10.1109/JBHI.2016.2608998.

Minhas, S., Khanum, A., Riaz, F., Khan, S., Alvi, A., 2017b. Trajectory based predictive modeling of conversion from mild cognitive impairment to Alzheimer's disease, in: 2017 IEEE EMBS International Conference on Biomedical and Health Informatics, BHI 2017, pp. 385–388. doi:10.1109/BHI.2017.7897286.

Minhas, S., Khanum, A., Riaz, F., Khan, S., Alvi, A., 2018. Predicting progression from mild cognitive impairment to Alzheimer's disease using autoregressive modelling of longitudinal and multimodal biomarkers. IEEE Journal of Biomedical and Health Informatics 22, 818–825. doi:10.1109/JBHI.2017.2703918.

Misra, C., Fan, Y., Davatzikos, C., 2009. Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: results from ADNI. NeuroImage 44, 1415–1422. doi:10.1016/j.neuroimage.2008.10.031. 00427.

Mitchell, J., Arnold, R., Dawson, K., Nestor, P., Hodges, J., 2009. Outcome in subgroups of mild cognitive impairment (MCI) is highly predictable using a simple algorithm. Journal of Neurology 256, 1500–1509. doi:10.1007/s00415-009-5152-0. 00089.

Modrego, P., Fayed, N., Pina, M., 2005. Conversion from mild cognitive impairment to probable Alzheimer's disease predicted by brain magnetic resonance spectroscopy. American Journal of Psychiatry 162, 667–675. doi:10.1176/appi.ajp.162.4.667.

Modrego, P.J., Fayed, N., Sarasa, M., 2011. Magnetic resonance spectroscopy in the prediction of early conversion from amnestic mild cognitive impairment to dementia: a prospective cohort study. BMJ Open 1, e000007–e000007. doi:10.1136/bmjopen-2010-000007. 00055.

Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J., 2015. Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. NeuroImage 104, 398–412. doi:10.1016/j.neuroimage.2014.10.002.

Moretti, D., Frisoni, G., Fracassi, C., Pievani, M., Geroldi, C., Binetti, G., Rossini, P., Zanetti, O., 2011. MCI patients' EEGs show group differences between those who progress and those who do not progress to AD. Neurobiology of Aging 32, 563–571. doi:10.1016/j.neurobiolaging.2009.04.003.

Mubeen, A., Asaei, A., Bachman, A., Sidtis, J., Ardekani, B., 2017. A six-month longitudinal evaluation significantly improves accuracy of predicting incipient Alzheimer's disease in mild cognitive impairment. Journal of Neuroradiology 44, 381–387. doi:10.1016/j.neurad.2017.05.008.

Nesteruk, M., Nesteruk, T., Styczynska, M., Mandecka, M., Barczak, A., Barcikowska, M., 2016. Combined use of biochemical and volumetric biomarkers to assess the risk of conversion of mild cognitive impairment to Alzheimer's disease. Folia Neuropathologica 54, 369–374. doi:10.5114/fn.2016.64815.

Nho, K., Shen, L., Kim, S., Risacher, S., West, J., Foroud, T., Jack, C., Weiner, M., Saykin, A., 2010. Automatic Prediction of Conversion from Mild Cognitive Impairment to Probable Alzheimer's Disease using Structural Magnetic Resonance Imaging. AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium 2010, 542–546.

Ning, K., Chen, B., Sun, F., Hobel, Z., Zhao, L., Matloff, W., Toga, A., Alzheimer's, D.N.I., 2018. Classifying Alzheimer's disease with brain imaging and genetic data using a neural network framework. Neurobiology of Aging 68, 151–158. doi:10.1016/j.neurobiolaging.2018.04.009.

Ota, K., Oishi, N., Ito, K., Fukuyama, H., 2015. Effects of imaging modalities, brain atlases and feature selection on prediction of Alzheimer's disease. Journal of Neuroscience Methods 256, 168–183. doi:10.1016/j.jneumeth.2015.08.020.

Ota, K., Oishi, N., Ito, K., Fukuyama, H., 2016. Prediction of Alzheimer's Disease in Amnestic Mild Cognitive Impairment Subtypes: Stratification Based on Imaging Biomarkers. Journal of Alzheimer's Disease 52, 1385–1401. doi:10.3233/JAD-160145.

Papaliagkas, V., Kimiskidis, V., Tsolaki, M., Anogianakis, G., 2008. Usefulness of event-related potentials in the assessment of mild cognitive impairment. BMC Neuroscience 9. doi:10.1186/1471-2202-9-107. 00106.

Park, H., Yang, J.J., Seo, J., Lee, J.M., 2013. Dimensionality reduced cortical features and their use in predicting longitudinal changes in Alzheimer's disease. Neuroscience Letters 550, 17–22. doi:10.1016/j.neulet.2013.06.042. 00014.

Parnetti, L., Chiasserini, D., Eusebi, P., Giannandrea, D., Bellomo, G., De Carlo, C., Padiglioni, C., Mastrocola, S., Lisetti, V., Calabresi, P., 2012. Performance of Aβ1-40, Aβ1-42, total tau, and phosphorylated tau as predictors of dementia in a cohort of patients with mild cognitive impairment. Journal of Alzheimer's Disease 29, 229–238. doi:10.3233/JAD-2011-111349. 00000.

Rocha de Paula, M., Gómez Ravetti, M., Berretta, R., Moscato, P., 2011. Differences in abundances of cell-signalling proteins in blood reveal novel biomarkers for early detection of clinical Alzheimer's disease. PLoS ONE 6. doi:10.1371/journal.pone.0017481. 00032.

Peters, F., Villeneuve, S., Belleville, S., 2014. Predicting progression to dementia in elderly subjects with mild cognitive impairment using both cognitive and neuroimaging predictors. Journal of Alzheimer's Disease 38, 307–318. doi:10.3233/JAD-130842. 00044.

Plant, C., Teipel, S., Oswald, A., Böhm, C., Meindl, T., Mourao-Miranda, J., Bokde, A., Hampel, H., Ewers, M., 2010. Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease. NeuroImage 50, 162–174. doi:10.1016/j.neuroimage.2009.11.046.

Plocharski, M., Østergaard, L.R., 2019. Prediction of Alzheimer's Disease in Mild Cognitive Impairment Using Sulcal Morphology and Cortical Thickness, in: World Congress on Medical Physics and Biomedical Engineering 2018, Springer, Singapore. pp. 69–74. doi:10.1007/978-981-10-9035-6_13.

Poil, S.S., de Haan, W., van der Flier, W., Mansvelder, H., Scheltens, P., Linkenkaer-Hansen, K., 2013. Integrative EEG biomarkers predict progression to Alzheimer's disease at the MCI stage. Frontiers in Aging Neuroscience 5. doi:10.3389/fnagi.2013.00058. 00063.

Popuri, K., Balachandar, R., Alpert, K., Lu, D., Bhalla, M., Mackenzie, I., Hsiung, R.Y., Wang, L., Beg, M., the, A.D.N.I., 2018. Development and validation of a novel dementia of Alzheimer's type (DAT) score based on metabolism FDG-PET imaging. NeuroImage: Clinical 18, 802–813. doi:10.1016/j.nicl.2018.03.007.

Quaranta, D., Gainotti, G., Vita, M., Lacidogna, G., Scaricamazza, E., Piccininni, C., Marra, C., 2016. Are raw scores on memory tests better than age- and education-adjusted scores for predicting progression from amnesic mild cognitive impairment to Alzheimer disease? Current Alzheimer Research 13, 1414–1420. doi:10.2174/1567205013666160314145522.

Raamana, P., Weiner, M., Wang, L., Beg, M., 2015. Thickness network features for prognostic applications in dementia. Neurobiology of Aging 36, S91–S102. doi:10.1016/j.neurobiolaging.2014.05.040.

Rabin, L., Paré, N., Saykin, A., Brown, M., Wishart, H., Flashman, L., Santulli, R., 2009. Differential memory test sensitivity for diagnosing amnestic mild cognitive impairment and predicting conversion to Alzheimer's disease. Aging, Neuropsychology, and Cognition 16, 357–376. doi:10.1080/13825580902825220. 00174.

Retico, A., Bosco, P., Cerello, P., Fiorina, E., Chincarini, A., Fantacci, M., 2015. Predictive Models Based on Support Vector Machines: Whole-Brain versus Regional Analysis of Structural MRI in the Alzheimer's Disease. Journal of Neuroimaging 25, 552–563. doi:10.1111/jon.12163.

Rhodius-Meester, H., Koikkalainen, J., Mattila, J., Teunissen, C., Barkhof, F., Lemstra, A., Scheltens, P., L'otj'onen, J., Van Der Flier, W., 2015. Integrating Biomarkers for Underlying Alzheimer's Disease in Mild Cognitive Impairment in Daily Practice: Comparison of a Clinical Decision Support System with Individual Biomarkers. Journal of Alzheimer's Disease 50, 261–270. doi:10.3233/JAD-150548.

Richard, E., Schmand, B., Eikelenboom, P., Van Gool, W., 2013. MRI and cerebrospinal fluid biomarkers for predicting progression to Alzheimer's disease in patients with mild cognitive impairment: A diagnostic accuracy study. BMJ Open 3. doi:10.1136/bmjopen-2012-002541. 00061.

Ritter, K., Schumacher, J., Weygandt, M., Buchert, R., Allefeld, C., Haynes, J.D., 2015. Multimodal prediction of conversion to Alzheimer's disease based on incomplete biomarkers. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring 1, 206–215. doi:10.1016/j.dadm.2015.01.006.

Rodriguez, J., Ramirez, J., Gorriz, J., Padilla, P., Ortiz, A., 2016. Short-term MCI-to-AD prediction using MRI, neuropsychological scores and ensemble tree learning techniques, in: 2015 IEEE Nuclear Science Symposium and Medical Imaging Conference, NSS/MIC 2015, pp. 1–3. doi:10.1109/NSSMIC.2015.7582226.

Sabuncu, M., 2013. A Bayesian algorithm for image-based time-to-event prediction. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 8184 LNCS, 74–81. doi:10.1007/978-3-319-02267-3_10. 00006.

Sanroma, G., Andrea, V., Benkarim, O., Manjón, J., Coupé, P., Camara, O., Piella, G., González Ballester, M., 2017. Early prediction of alzheimer's disease with non-local patch-based longitudinal descriptors. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 10530 LNCS, 74–81. doi:10.1007/978-3-319-67434-6_9.

Shaffer, J., Petrella, J., Sheldon, F., Choudhury, K., Calhoun, V., Edward Coleman, R., Murali Doraiswamy, P., 2013. Predicting cognitive decline in subjects at risk for Alzheimer disease by using combined cerebrospinal fluid, MR imaging, and PET biomarkers. Radiology 266, 583–591. doi:10.1148/radiol.12120010. 00158.

Shen, T., Jiang, J., Li, Y., Wu, P., Zuo, C., Yan, Z., 2018. Decision Supporting Model for One-year Conversion Probability from MCI to AD using CNN and SVM, in: Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference, pp. 738–741. doi:10.1109/EMBC.2018.8512398.

Shmulev, Y., Belyaev, M., The, A.D.N.I., 2018. Predicting conversion of mild cognitive impairments to alzheimer's disease and exploring impact of neuroimaging. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 11044 LNCS, 83–91. doi:10.1007/978-3-030-00689-1_9.

Si, B., Yakushev, I., Li, J., 2017. A sequential tree-based classifier for personalized biomarker testing of Alzheimer's disease risk. IISE Transactions on Healthcare Systems Engineering 7, 248–260. doi:10.1080/24725579.2017.1367979.

Snaedal, J., Johannesson, G., Gudmundsson, T., Gudmundsson, S., Pajdak, T., Johnsen, K., 2010. The use of EEG in Alzheimer's disease, with and without scopolamine - A pilot study. Clinical Neurophysiology 121, 836–841. doi:10.1016/j.clinph.2010.01.008. 00027.

Sohn, B., Yi, D., Seo, E., Choe, Y., Kim, J., Kim, S., Choi, H., Byun, M., Jhoo, J., Woo, J., Lee, D., 2015. Comparison of regional gray matter atrophy, white matter

alteration, and glucose metabolism as a predictor of the conversion to alzheimer's disease in mild cognitive impairment. Journal of Korean Medical Science 30, 779–787. doi:10.3346/jkms.2015.30.6.779.

Suk, H.I., Lee, S.W., Shen, D., 2017. Deep ensemble learning of sparse regression models for brain disease diagnosis. Medical Image Analysis 37, 101–113. doi:10.1016/j.media.2017.01.008.

Summers, M., Saunders, N., 2012. Neuropsychological measures predict decline to alzheimer's dementia from mild cognitive impairment. Neuropsychology 26, 498–508. doi:10.1037/a0028576. 00119.

Tanpitukpongse, T., Mazurowski, M., Ikhena, J., Petrella, J., 2017. Predictive utility of marketed volumetric software tools in subjects at risk for Alzheimer disease: Do regions outside the hippocampus matter? American Journal of Neuroradiology 38, 546–552. doi:10.3174/ajnr.A5061.

Teipel, S., Kurth, J., Krause, B., Grothe, M., 2015. The relative importance of imaging markers for the prediction of Alzheimer's disease dementia in mild cognitive impairment - Beyond classical regression. NeuroImage: Clinical 8, 583–593. doi:10.1016/j.nicl.2015.05.006.

Thung, K.H., Yap, P.T., Adeli, E., Lee, S.W., Shen, D., 2018. Conversion and time-to-conversion predictions of mild cognitive impairment using low-rank affinity pursuit denoising and matrix completion. Medical Image Analysis 45, 68–82. doi:10.1016/j.media.2018.01.002.

Tondelli, M., Bedin, R., Chiari, A., Molinari, M., Bonifacio, G., Lelli, N., Trenti, T., Nichelli, P., 2015. Role of cerebrospinal fluid biomarkers to predict conversion to dementia in patients with mild cognitive impairment: A clinical cohort study. Clinical Chemistry and Laboratory Medicine 53, 453–460. doi:10.1515/cclm-2014-0414.

Tong, T., Gao, Q., 2015. Extraction of features from patch based graphs for the prediction of disease progression in AD. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 9226, 500–509. doi:10.1007/978-3-319-22186-1_50.

Tong, T., Gao, Q., Guerrero, R., Ledig, C., Chen, L., Rueckert, D., 2017. A novel grading biomarker for the prediction of conversion from mild cognitive impairment to Alzheimer's disease. IEEE Transactions on Biomedical Engineering 64, 155–165. doi:10.1109/TBME.2016.2549363.

Trzepacz, P., Yu, P., Sun, J., Schuh, K., Case, M., Witte, M., Hochstetler, H., Hake, A., 2014. Comparison of neuroimaging modalities for the prediction of conversion from mild cognitive impairment to alzheimer's dementia. Neurobiology of Aging 35, 143–151. doi:10.1016/j.neurobiolaging.2013.06.018. 00084.

Valencia, N., Lehrner, J., 2018. Screening for dementia with the vienna visuo-constructional test 3.0 screening (vvt 3.0 screening). neuropsychiatrie 32, 196–203.

Vecchio, D., Miraglia, D., Iberite, D., Lacidogna, D., Guglielmi, D., Marra, D., Pasqualetti, D., Tiziano, D., Rossini, P., 2018. Sustainable method for Alzheimer dementia prediction in mild cognitive impairment: Electroencephalographic connectivity and graph theory combined with apolipoprotein E. Annals of Neurology 84, 302–314. doi:10.1002/ana.25289.

Visser, P., Scheltens, P., Verhey, F., Schmand, B., Launer, L., Jolles, J., Jonker, C., 1999. Medial temporal lobe atrophy and memory dysfunction as predictors for dementia in subjects with mild cognitive impairment. Journal of Neurology 246, 477–485. doi:10.1007/s004150050387. 00340.

Vivar, G., Zwergal, A., Navab, N., Ahmadi, S.A., 2018. Multi-modal Disease Classification in Incomplete Datasets Using Geometric Matrix Completion. arXiv:1803.11550 [cs] ArXiv: 1803.11550.

Wang, P., Chen, K., Yao, L., Hu, B., Wu, X., Zhang, J., Ye, Q., Guo, X., 2016. Multimodal classification of mild cognitive impairment based on partial least squares. Journal of Alzheimer's Disease 54, 359–371. doi:10.3233/JAD-160102.

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

Westman, E., Aguilar, C., Muehlboeck, J.S., Simmons, A., 2013. Regional magnetic resonance imaging measures for multivariate analysis in Alzheimer's disease and mild cognitive impairment. Brain Topography 26, 9–23. doi:10.1007/s10548-012-0246-x. 00134.

Westman, E., Cavallin, L., Muehlboeck, J.S., Zhang, Y., Mecocci, P., Vellas, B., Tsolaki, M., Kłoszewska, I., Soininen, H., Spenger, C., Lovestone, S., Simmons, A., Wahlund, L.O., 2011a. Sensitivity and specificity of medial temporal lobe visual ratings and multivariate regional MRI classification in Alzheimer's disease. PLoS ONE 6. doi:10.1371/journal.pone.0022506. 00102.

Westman, E., Muehlboeck, J.S., Simmons, A., 2012. Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion. NeuroImage 62, 229–238. doi:10.1016/j.neuroimage.2012.04.056. 00216.

Westman, E., Simmons, A., Muehlboeck, J.S., Mecocci, P., Vellas, B., Tsolaki, M., Kloszewska, I., Soininen, H., Weiner, M., Lovestone, S., Spenger, C., Wahlund, L.O., 2011b. AddNeuroMed and ADNI: Similar patterns of Alzheimer's atrophy and automated MRI classification accuracy in Europe and North America. NeuroImage 58, 818–828. doi:10.1016/j.neuroimage.2011.06.065. 00120.

Westman, E., Simmons, A., Zhang, Y., Muehlboeck, J.S., Tunnard, C., Liu, Y., Collins, L., Evans, A., Mecocci, P., Vellas, B., Tsolaki, M., Kloszewska, I., Soininen, H., Lovestone, S., Spenger, C., Wahlund, L.O., 2011c. Multivariate analysis of MRI data for Alzheimer's disease, mild cognitive impairment and healthy controls. NeuroImage 54, 1178–1187. doi:10.1016/j.neuroimage.2010.08.044. 00130.

Xu, L., Wu, X., Li, R., Chen, K., Long, Z., Zhang, J., Guo, X., Yao, L., 2016. Prediction of progressive mild cognitive impairment by multi-modal neuroimaging biomarkers. Journal of Alzheimer's Disease 51, 1045–1056. doi:10.3233/JAD-151010.

Yang, X., Tan, M., Qiu, A., 2012. CSF and Brain Structural Imaging Markers of the Alzheimer's Pathological Cascade. PLoS ONE 7. doi:10.1371/journal.pone.0047406. 00024.

Ye, D., Pohl, K., Davatzikos, C., 2011. Semi-supervised pattern classification: Application to structural MRI of Alzheimer's disease, in: Proceedings - International Workshop on Pattern Recognition in NeuroImaging, PRNI 2011, pp. 1–4. doi:10.1109/PRNI.2011.12.

Ye, J., Farnum, M., Yang, E., Verbeeck, R., Lobanov, V., Raghavan, N., Novak, G., DiBernardo, A., Narayan, V., 2012. Sparse learning and stability selection for predicting MCI to AD conversion using baseline ADNI data. BMC Neurology 12. doi:10.1186/1471-2377-12-46. 00093.

Young, A., Oxtoby, N., Daga, P., Cash, D., Fox, N., Ourselin, S., Schott, J., Alexander, D., 2014. A data-driven model of biomarker changes in sporadic Alzheimer's disease. Brain 137, 2564–2577. doi:10.1093/brain/awu176.

Young, J., Modat, M., Cardoso, M., Ashburner, J., Ourselin, S., 2016. An oblique approach to prediction of conversion to alzheimer's disease with multikernel gaussian processes. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 9444 LNAI, 122–128. doi:10.1007/978-3-319-45174-9_13.

Yu, P., Dean, R., Hall, S., Qi, Y., Sethuraman, G., Willis, B., Siemers, E., Martenyi, F., Tauscher, J., Schwarz, A., 2012. Enriching amnestic mild cognitive impairment populations for clinical trials: Optimal combination of biomarkers to predict conversion to dementia. Journal of Alzheimer's Disease 32, 373–385. doi:10.3233/JAD-2012-120832. 00029.

Zhang, D., Shen, D., 2012. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. NeuroImage 59, 895–907. doi:10.1016/j.neuroimage.2011.09.069. 00337.

Zhang, D., Shen, D., Initiative, A.D.N., 2012. Predicting Future Clinical Changes of MCI Patients Using Longitudinal and Multimodal Biomarkers. PLOS ONE 7, e33182. doi:10.1371/journal.pone.0033182. 00144.

Zhang, H., Therriault, J., Kang, M., Ng, K., Pascoal, T., Rosa-Neto, P., Gauthier, S., 2018. Cerebrospinal fluid synaptosomal-associated protein 25 is a key player in synaptic degeneration in mild cognitive impairment and Alzheimer's disease. Alzheimer's Research and Therapy 10. doi:10.1186/s13195-018-0407-6.

Zhao, Y., Yao, Z., Zheng, W., Yang, J., Ding, Z., Li, M., Lu, S., 2017. Predicting MCI progression with individual metabolic network based on longitudinal FDG-PET, in: Proceedings - 2017 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2017, pp. 1894–1899. doi:10.1109/BIBM.2017.8217948.

Zheng, W., Yao, Z., Xie, Y., Fan, J., Hu, B., 2018. Identification of Alzheimer's Disease and Mild Cognitive Impairment Using Networks Constructed Based on Multiple Morphological Brain Features. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging 3, 887–897. doi:10.1016/j.bpsc.2018.06.004. 00001.

Zhu, X., Suk, H.I., Shen, D., 2014a. A novel matrix-similarity based loss function for joint regression and classification in AD diagnosis. NeuroImage 100, 91–105. doi:10.1016/j.neuroimage.2014.05.078. 00120.

Zhu, X., Suk, H.I., Shen, D., 2014b. A novel multi-relation regularization method for regression and classification in AD diagnosis. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 8675 LNCS, 401–408. doi:10.1007/978-3-319-10443-0_51. 00021.

Zhu, X., Suk, H.I., Wang, L., Lee, S.W., Shen, D., 2017. A novel relational regularization feature selection method for joint regression and classification in AD diagnosis. Medical Image Analysis 38, 205–214. doi:10.1016/j.media.2015.10.008.