Published in final edited form as:

Int Psychogeriatr. 2019 July ; 31(7): 937–945. doi:10.1017/S1041610218001618.

A clinically-translatable machine learning algorithm for the prediction of Alzheimer's disease conversion: further evidence of its accuracy via a transfer learning approach

Massimiliano Grassi¹, David A. Loewenstein^{2,3,4}, Daniela Caldirola¹, Koen Schruers⁵, Ranjan Duara^{3,6,7}, Giampaolo Perna^{1,2,5,8,9}

¹Department of Clinical Neurosciences, Hermanas Hospitalarias, Villa San Benedetto Menni Hospital, FoRiPsi, Albese con Cassano, Como, Italy ²Department of Psychiatry and Behavioral Sciences, Miller School of Medicine, University of Miami, Miami, Florida, USA ³Wien Center for Alzheimer's Disease and Memory Disorders, Mount Sinai Medical Center, Miami Beach, Florida, USA ⁴Center on Aging, Miller School of Medicine, University of Miami, Miami, Florida, USA ⁵Research Institute of Mental Health and Neuroscience and Department of Psychiatry and Neuropsychology, Faculty of Health, Medicine and Life Sciences, University of Maastricht, Maastricht, the Netherlands ⁶Courtesy Professor of Neurology, Department of Neurology, University of Florida College of Medicine, Gainesville, Florida, USA ⁷Herbert Wertheim College of Medicine, Florida International University, Miami, Florida, USA ⁸Department of Biomedical Sciences, Humanitas University, Rozzano, Milan, Italy ⁹Mantovani Foundation, Arconate, Italy

Abstract

Background: In a previous study, we developed a highly performant and clinically-translatable machine learning algorithm for a prediction of three-year conversion to Alzheimer's disease (AD) in subjects with Mild Cognitive Impairment (MCI) and Pre-mild Cognitive Impairment. Further tests are necessary to demonstrate its accuracy when applied to subjects not used in the original training process. In this study, we aimed to provide preliminary evidence of this via a transfer learning approach.

Methods: We initially employed the same baseline information (i.e. clinical and neuropsychological test scores, cardiovascular risk indexes, and a visual rating scale for brain atrophy) and the same machine learning technique (support vector machine with radial-basis function kernel) used in our previous study to retrain the algorithm to discriminate between

Correspondence should be addressed to: Massimiliano Grassi, Department of Clinical Neurosciences, Villa San Benedetto Menni, Hermanas Hospitalarias, FoRiPsi, via Roma 16, 22032 Albese con Cassano (Como), Italy. Phone: +39 031 4291511; Fax: +39 031427246. massi.gra@gmail.com.

Description of authors' roles

Massimiliano Grassi designed the study, planned and performed the analyses, interpreted the results, and wrote the paper. David A. Loewenstein designed and led the longitudinal study where the data were originally collected, advised on the current study design, and contributed to interpretation of the results and the manuscript writing. Daniela Caldirola and Koen Schruers contributed to the interpretation of the results and the manuscript writing. Ranjan Duara designed and led the longitudinal study where the data were originally collected and revised the final version of the manuscript. Giampaolo Perna advised on the current study design and contributed to interpretation of the results and the manuscript writing.

Conflict of interest

None.

participants with AD ($n = 75$) and normal cognition ($n = 197$). Then, the algorithm was applied to perform the original task of predicting the three-year conversion to AD in the sample of 61 MCI subjects that we used in the previous study.

Results: Even after the retraining, the algorithm demonstrated a significant predictive performance in the MCI sample (AUC = 0.821, 95% CI bootstrap = 0.705–0.912, best balanced accuracy = 0.779, sensitivity = 0.852, specificity = 0.706).

Conclusions: These results provide a first indirect evidence that our original algorithm can also perform relevant generalized predictions when applied to new MCI individuals. This motivates future efforts to bring the algorithm to sufficient levels of optimization and trustworthiness that will allow its application in both clinical and research settings.

Keywords

Alzheimer's disease; clinical prediction rule; machine learning; mild cognitive impairment; personalized medicine; precision medicine; transfer learning

Introduction

Alzheimer's disease (AD) is the most common neurodegenerative brain disorder, is the top cause for disabilities in later life, and is associated with huge global costs. Currently, no cure is available for AD, although, with new emerging treatment approaches, it is increasingly important to be able to identify subjects at a true risk of later developing AD. By identifying those persons at greatest risk for decline, there is a potential to make clinical trials of AD treatments more cost-effective and valid by better selecting subjects to recruit, as treatments will likely have the greatest impact when provided at the earliest possible stage of the disease process (Brooks and Loewenstein, 2010; Loewenstein *et al.*, 2017).

In a previous study, we presented a novel machine learning (ML) algorithm for the prediction of a three-year conversion to AD in subjects with Mild Cognitive Impairment (MCI) and preliminarily in subjects with Pre-mild Cognitive Impairment (PreMCI), which was developed with a sample of 123 MCI and PreMCI patients recruited in a collaborative longitudinal study by several centers located in the Miami (Florida, US) area (Grassi *et al.*, 2018). Differently from several other available ML algorithms, ours employed only non-invasive predictors that are either already routinely assessed or effectively introducible in current clinical practice, such as clinical and neuropsychological test scores, cardiovascular risk indexes, and clinician-rated levels of brain atrophy. Promisingly, the algorithm achieved a high predictive accuracy in our previous study, with a cross-validated balanced accuracy of 0.913 and an Area Under the Receiving Operating Curve (AUC) of 0.962 in the entire sample of MCI and PreMCI, and a cross-validated balanced accuracy of 0.874 and an AUC of 0.914 in the sole sample of MCI. Its level of accuracy is among the very best of the many algorithms available in literature, and the best achieved so far using only information easily collectable in clinical practice (Agarwal *et al.*, 2015; Apostolova *et al.*, 2014; Clark *et al.*, 2014; Dukart *et al.*, 2016; Hojjati *et al.*, 2017; Long *et al.*, 2017; Mathotaarachchi *et al.*, 2017; Minhas *et al.*, 2016; Moradi *et al.*, 2015; Plant *et al.*, 2010).

However, before an application can be safely proposed, a predictive algorithm needs to be tested in further independent samples of MCI and PreMCI subjects to demonstrate that its accuracy levels are also preserved when it is applied in generalized clinical and experimental contexts. To provide such evidence, a sample of distinct MCI and PreMCI subjects is currently under recruitment as part of a longitudinal study of over 450 persons at the University of Miami (DL). This sample will be used to test the algorithm we proposed in our previous study as soon as the three-year follow-up assessments are completed.

However, before such optimal testing strategy will become performable, another opportunity for a preliminary test of our algorithm can come from the application of the so-called transfer learning approach. In the ML field, this refers to the process of using knowledge from one problem to train part or an entire algorithm that will be later applied to another problem. Its application for the solution of many complex tasks has been growing in the last years (Weiss *et al.*, 2016), and such strategy has already been applied in developing several algorithms that predict the conversion of MCI subjects to AD (Cheng *et al.*, 2013; Collij *et al.*, 2016; Cui *et al.*, 2011; Dukart *et al.*, 2016; Hinrichs *et al.*, 2011; Nho *et al.*, 2010; Plant *et al.*, 2010; Retico *et al.*, 2015; Westman *et al.*, 2013; Young *et al.*, 2013). In these studies, the Authors initially trained the algorithms to discriminate between AD and Cognitively Normal individuals (CN), not using samples of MCI subjects but, instead, samples of solely AD and CN subjects. Then, they applied these ML algorithms in a different task, which is the prediction of future conversion to AD in MCI subjects. A prediction of conversion is made if the MCI subjects are classified as AD by the algorithm, while a prediction of non-conversion is made if the MCI subjects are classified as CN. Such strategy is motivated by the hypothesis that those MCI subjects, who will later convert to AD, already show AD-like unrecognized characteristics, which instead do not characterize the MCI subjects who will not convert.

Following the abovementioned approach, we employed the same predictors and ML technique (support vector machine with radial-basis function kernel) used in our previous study to retrain our algorithm to discriminate between AD and CN using a sample of subjects with either the former or the latter condition. Then, after retraining, we will use our ML algorithm to make a prediction of the three-years conversion to AD in the same sample of MCI subjects we used in our previous study. In the current study, although we will use the same predictors and ML technique, the MCI sample will be used only to test the algorithm and not during training. Thus, the results we will achieve will be able to provide a first indirect evidence of how our previously proposed ML algorithm can perform relevant predictions also when applied to a sample of subjects not used in the training process.

Compared to our previous study where both training and validation were performed in the MCI sample via a cross-validation procedure, we expect that the algorithm retrained in a separate sample of AD/CN individuals will achieve a reduced, but still relevant, predictive performance in the MCI sample, which will provide further complementary evidence in support of the results found in our previous study.

Materials and methods

Subjects

Data regarding 272 subjects with AD or CN, as well as the sample of 61 subjects with MCI used in our previous study (Grassi *et al.*, 2018), were included in the current one. Instead, considering that only three converters were available in the PreMCI sample, this was employed in the current study. All the included samples of subjects are part of a dataset that collects several patients recruited in a study investigating longitudinal changes associated with MCI and normal aging. This study involved community volunteers as well as subjects recruited from the Memory Disorders Clinic at the Wien Center for Alzheimer's disease, the Memory Disorders at Mount Sinai Medical Center, Miami Beach, Florida, and the Community and Memory Disorders Center at the University of South Florida. A common clinical and neuropsychological battery was administered to all subjects at all the sites. Considering the final aim of developing a predictive algorithm to be used in clinical practice, no other inclusion or exclusion criteria were applied beyond these diagnostic criteria and the occurrence of missing information in the variables used as predictors (see below).

Subjects were classified as having probable Alzheimer's disease (AD; $n = 75$, 27.07%) if at the time of the assessment they presented a Dementia syndrome by DSM-IV-TR criteria (American Psychiatric Association, 2000), and satisfied the National Institute of Neurological and Communicative Disorders and Stroke/Alzheimer's Disease and Related Disorders Association criteria for Alzheimer's disease (McKhann *et al.*, 1984). Subjects were classified as MCI if they presented subjective memory complaints by the participant and/or collateral informant, and evidence of decline from clinical history and evaluation, such as a global CDR score (Morris, 1993) of 0.5 and one or more memory measures (including the HVLTR, the SIT, Logical Memory Delay and Visual Reproduction of the WMS-IV, TMT-B, Category Fluency, Letter Fluency, and WAIS-III Block Design) of 1.5 standard deviation or greater below expected normative values. Finally, subjects were identified as CN ($n = 197$, 72.03%) if during assessment they had a global CDR of 0 and no neuropsychological deficits (1.5 standard deviation or greater above expected normative values).

The study was conducted with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. All subjects gave their written informed consent to the use of their clinical data for scientific research purposes.

Feature extraction

The same variables included as predictors in the best algorithm developed in our previous study were used to train the AD/CN algorithm, excluding the variable indicating the MCI/PreMCI sub-type that is not applicable to AD and CN subjects. These predictors were selected with a recursive feature elimination procedure starting from a larger set of 36 variables regarding sociodemographic characteristics, clinical and neuropsychological test scores, cardiovascular risk indexes, and a visual rating scale for brain atrophy.

- *Clinical scales:* The memory sum score of a modified informant-based version of CDR (ModCDR-M) (Duara *et al.*, 2010);
- *Visual Rating Scale for brain atrophy:* Left and right hemisphere HPC, ERC, and PRC atrophy levels were assessed with a 0–4 VRS (Duara *et al.*, 2008), an adaptation from the original Scheltens' VRS for the global assessment of medial temporal atrophy (Scheltens *et al.*, 1992). Ratings were performed on a Magnetic Resonance Imaging (MRI) image of a standardized coronal slice, perpendicular to the line joining the anterior and posterior commissures, intersecting the mammillary bodies, and on adjacent slices. Ratings are performed on a five-point scale (0 = no atrophy, 1 = minimal atrophy, 2 = mild atrophy, 3 = moderate atrophy, and 4 = severe atrophy) and excellent inter-rater (kappa, 0.75 to 0.94) and intra-rater (kappa, 0.84 to 0.94) agreements have been reported (Duara *et al.*, 2008; Urs *et al.*, 2009), thanks to a computer interface that provides a library of reference images. Only five of the six VRS scores were included in the algorithm, excluding the left PRC score as indicated by the feature selection procedure used in our previous study.
- *Neuropsychological tests:* The Hopkins Verbal Learning Test Revised—Total Recall (HVLTR-R) and Delayed Recall (HVLTR-D) scores (Benedict and Zgaljardic, 1998), the Semantic Interference Test—Total Retroactive (SIT-RT) and Total Recognition (SIT-RC) scores (Loewenstein *et al.*, 2004), the Trial Making the Logical Memory Test—Immediate Recall (LM-I) and Delayed Recall (LM-D) scores of the WMS-IV (Wechsler, 1997).
- *Cardiovascular risk indexes:* Heart rate and history of myocardial infarction.

Detailed descriptions of these variables can be found in (Grassi *et al.*, 2018). Continuous variables were standardized. In the end, 14 continuous and one dichotomous categorical predictors were used. The full list is available in Table 1.

Training with AD/CN participants

In this study, we used the same ML technique that generated the best performing algorithm in our previous study (Grassi *et al.*, 2018), which is Support Vector Machine (SVM) with radial basis function (Gaussian) kernel. It has two hyper-parameters (σ ; C) that allow a different tuning of the algorithm during the training process, and 200 random configurations of these hyper-parameters were attempted in order to identify the configuration that allows the algorithm to achieve the best predictive performance.

Specifically, we are interested in achieving the hyper-parameter configuration that results in the best possible performance when the algorithm is applied to discriminate new AD/CN cases that are not part of the training sample. We used cross-validation to provide an estimate of such generalized performance, but the sample size in this study was too large to apply the computationally expensive leave-pair-out cross-validation protocol, as we did in our previous study. Instead, a stratified cross-validation protocol was used. For each of the hyperparameter configurations, 75 folds were used, each including a single subject with AD. Training was performed excluding the cases in the fold from the training sample and

calculating the performance of the algorithm on them. Finally, the average performance metric is taken as estimate of the generalized performance of the algorithm created with that hyper-parameter configuration. As primary performance metric, the Area Under the Receiving Operating Curve (AUC) was used. At first, the algorithm outputs a continuous prediction score (range: 0–1; the closer to 1 the higher the predicted risk of conversion for that subject), then the class prediction is finally made setting a cut-off score (AD if above or equal to the cut-off score, CN if below).

A bootstrap procedure, (10,000 resampling with replacement) was used to calculate the confidence interval (CI) of the average cross-validated AUC. The distribution of the resampled 10,000 average AUCs was used to calculate 95% CI with the bias-corrected and accelerated (BCa) approach (Efron, 1987).

The hyper-parameter configuration for each technique that produced the best cross-validated AUC was retained, and a final algorithm with such configuration is finally trained on the whole dataset of AD and CN subjects.

Testing with MCI

Predictions of three-year conversion to AD for the MCI subjects was obtained using the algorithm trained with AD and CN subjects, considering a classification of AD as prediction of future conversion to AD, and CN as prediction of non-conversion. It is worth noting that, in this case, the MCI subjects were not used during the training of the model. The AUC in the MCI sample subsample were calculated and a stratified bootstrap procedure (10,000 resampling with replacement) was used to calculate the AUC confidence interval (CI). The distribution of the new 10,000 AUCs calculated was used to calculate 95% CI with the bias-corrected and accelerated (BCa) approach (Efron, 1987). The cut-off applied to the algorithm output scores was progressively increased starting from 0, and the thresholds providing the best balanced accuracy was identified, calculating also the sensitivity and specificity achieved. Moreover, the cross-validated levels of specificities and balanced accuracy values when sensitivity approached to 0.95, 0.9, 0.85, 0.8, 0.75 were calculated.

Results

Descriptive statistics of each feature in the AD and CN groups are reported in Table 1. Statistics of continuous features are reported before standardization was applied.

The final algorithm trained with the AD/CN sample shows very high cross-validated accuracy in discriminating between AD and CN individuals, with an AUC of 0.996 (C.I. 95% bootstrap = 0.983, 1).

When applied to the sample of MCI individuals to predict their risk of conversion to AD in the next three years, its predictive performance was relevant also in this task, with an AUC of 0.821 (C.I. 95% bootstrap = 0.705, 0.912) and a best balanced accuracy of 0.779 (sensitivity = 0.852, specificity = 0.706). The levels of specificities and balanced accuracy values when sensitivity approached 0.95, 0.9, 0.85, 0.8, 0.75 are reported in Table 2.

As expected, its predictive performance was smaller than the cross-validated one found in our previous study (AUC = 0.914, C.I. 95% bootstrap = 0.822, 0.975; best balanced accuracy = 0.874, sensitivity = 0.880, specificity = 0.867) but it demonstrated a predictive performance better than randomness (i.e. the AUC has a lower 95% bootstrap CI larger than 0.5).

Discussion

The aim of the current study was to provide a first indirect evidence in support of a clinically-translatable machine-learning algorithm for the identification of a three-year conversion to Alzheimer's disease in subjects with either MCI or PreMCI, which we presented in a previous paper (Grassi *et al.*, 2018). Such an algorithm showed a high cross-validated predictive performance, the highest among the currently available algorithms that are based only on information easily assessable in clinical practice (Grassi *et al.*, 2018).

A three-year follow-up assessment of a new sample of MCI subjects is currently ongoing and will allow a proper testing of our algorithm in a sample that is independent from the one employed in the training phase. Instead, in this study, we used the transfer learning approach to preliminary perform such testing. We employed the same feature and ML technique used in our previous study to retrain the algorithm to discriminate between AD and CN participants, and then we applied it to the sample of MCI subjects that we used in the previous study, considering a prediction of a three-year conversion to AD if the algorithm classifies a MCI subject as AD and a prediction of non-conversion if the algorithm classifies the subject as CN.

As hypothesized, after the retraining, the algorithm demonstrated a significant predictive performance in the MCI sample, although reduced in magnitude compared to the one achieved in our previous study (Grassi *et al.*, 2018). These results suggest that our algorithm can perform relevant predictions also when applied to new samples not used for training, further motivating future efforts to bring our algorithm at a clinical-ready level.

Previously, other investigators have applied a similar strategy to develop predictive algorithms for the conversion to AD in MCI subjects (Cheng *et al.*, 2013; Collij *et al.*, 2016; Cui *et al.*, 2011; Dukart *et al.*, 2016; Hinrichs *et al.*, 2011; Nho *et al.*, 2010; Plant *et al.*, 2010; Retico *et al.*, 2015; Westman *et al.*, 2013; Young *et al.*, 2013) based on the hypothesis that the MCI subjects, who will later convert to AD, already show characteristics of the AD, and that their MCI condition is caused by the same pathophysiological process that will later lead to a full-blown AD manifestation, which has already begun although not fully evident yet. Instead, the non-converters have MCI for other causes and their traits are distinct from those characterizing subjects with AD.

Results from previous studies that trained algorithms with an AD/CN sample and then applied them to predict the conversion to AD of MCI subjects are summarized in Table 3. Our algorithm achieved one of the best predictive performance available, with only the algorithms presented by Young and colleagues (Young *et al.*, 2013 #75) and Dukart and colleagues (Dukart *et al.*, 2016 #49) showing similar performances by the former and

higher performances by the latter, when compared to the ours. However, both these ML algorithms necessitate information that is, currently, not easily and routinely assessed in clinical practice, such as 18-fluorodeoxyglucose Positron Emission Tomography and the typing of the *APOE* gene. These results are consistent with the evidence from the previous study that our proposed algorithm is the best performing one among those based on only information easy to be clinically collected.

A reduced predictive performance compared to when the algorithm was trained directly on MCI and PreMCI subjects was expected. First, the training and the tuning of the model's hyperparameters were performed to accomplish a different classification task; that is, distinguishing AD and CN subjects. Even if it is hypothesized that MCI converters show AD-like characteristics while non-converters do not, the AD/CN and converters/non-converters classification tasks may share a common but not totally equal solution. Thus, the optimized hyper-parameter configuration of a ML algorithm identified to perform the former may be good but not the very best possible one to perform the latter, taking to a sub-optimal predictive accuracy. Moreover, training the algorithm with AD and CN subjects did not enable to include one of the predictors we have previously used, namely the MCI/PreMCI sub-type, which resulted in a particular relevance for the prediction. The lack of this piece of information may also have caused part of the fall in the predictive performance compared to what previously achieved. Despite these abovementioned issues, the retrained algorithm achieved a significant predictive capability in the MCI sample, which in this study was not directly employed in the training phase.

However, it is worth highlighting that our MCI sample cannot be viewed as perfectly independent from the AD/CN training sample, which is a potential limitation of the current study. As a matter of fact, both samples have been recruited in the same clinical centers as part of the same longitudinal study, all located in the area of Miami. The population referring to this study might have peculiar characteristics and the performance of the algorithm might give partially reduced results if used in different populations. Moreover, in the previous study, both the feature selection and the identification of the best ML technique was performed with a sample that included also the same MCI sample here applied as a test dataset. Because in this study we used the same features and ML technique that were selected in our previous study, some minor so-called information leakage may, indeed, have occurred. In the ML field, this indicates that some information may have passed from the training to the test process, which can cause a partial inflation of the estimate of the algorithm generalized performance obtained by its application to the test set. The inflation is expected to be more severe the greater the amount of information shared between training and testing, which, in our analyses, we expect to be limited and only related to the issues we have just discussed.

Albeit, taking into account these limitations, the results of this study further support that the baseline information we took into account together with the use of ML techniques can effectively allow a prediction of conversion to AD in MCI subjects and they offer motivation to proceed in the further development and testing of the algorithm to reach sufficient levels of optimization and trustworthiness for its application in both clinical and research settings.

Specifically, some main issues will be principally addressed in the next phase: first, a test of the generalizability of the algorithm will be performed by applying it to new MCI subjects, which are currently under recruitment in a longitudinal study of over 450 persons at the University of Miami (DL). In addition, we aim to test the algorithm's predictive performance in further subjects with PreMCI that convert to AD within three-years. This would allow the use of the algorithm to identify fast converters to AD at a very early stage of the degenerative process. Finally, a particular effort will be made to provide an explanation of which role each feature plays in the prediction. The current algorithm was chosen in our previous study because it proved to significantly outperform all the others we attempted. However, this algorithm results in a "black-box" at the moment, as it does not allow an easy interpretation of how the algorithm achieve to perform the predictions. A better interpretability of the algorithm will allow us to foster its application by its being better comprehended and accepted by all users, as well as it may allow researchers to reach further potential insights regarding the development process of AD.

Acknowledgments

This study was supported by the National Institute on Aging, United States Department of Health and Human Services (grant: 1P50AG025711-05, R01AG047649-01A1, P50 AG047726602).

References

- Agarwal S, Ghanty **PandPal** NR (2015). Identification of a small set of plasma signalling proteins using neural network for prediction of Alzheimer's disease. *Bioinformatics*, 31, 2505–2513. doi: 10.1093/bioinformatics/btv173. [PubMed: 25819077] **and**
- American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR* (Text Revision), 4 edn. Washington, DC: American Psychiatric Association.
- Apostolova LG, et al. (2014). ApoE4 effects on automated diagnostic classifiers for mild cognitive impairment and Alzheimer's disease. *Neuroimage Clinical*, 4, 461–472. doi: 10.1016/j.nicl.2013.12.012. [PubMed: 24634832]
- Benedict RH **and** Zgaljardic DJ (1998). Practice effects during repeated administrations of memory tests with and without alternate forms. *Journal of Clinical and Experimental Neuropsychology*, 20, 339–352. doi: 10.1076/jcen.20.3.339.822. [PubMed: 9845161] **and**
- Brooks LG **and** Loewenstein DA (2010). Assessing the progression of mild cognitive impairment to Alzheimer's disease: current trends and future directions. *Alzheimers Research & Therapy*, 2, 28. doi: 10.1186/alzrt52. **and**
- Cheng B, Zhang D, Chen S, Kaufer DI, Shen D **and** Alzheimer's Disease Neuroimaging Initiative. (2013). Semi-supervised multimodal relevance vector regression improves cognitive performance estimation from imaging and biological biomarkers. *Neuroinformatics*, 11, 339–353. doi: 10.1007/s12021-013-9180-7. [PubMed: 23504659] **and**
- Clark DG, et al. (2014). Latent information in fluency lists predicts functional decline in persons at risk for Alzheimer disease. *Cortex*, 55, 202–218. doi: 10.1016/j.cortex.2013.12.013. [PubMed: 24556551]
- Collij LE, et al. (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. [PubMed: 27383395] **.,**
- Cui Y, et al. (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. [PubMed: 21814561]

- Duara R, et al. (2008). Medial temporal lobe atrophy on MRI scans and the diagnosis of Alzheimer disease. *Neurology*, 71, 1986–1992. doi: 10.1212/01.wnl.0000336925.79704.9f. [PubMed: 19064880]
- Duara R, et al. (2010). Diagnosis and staging of mild cognitive impairment, using a modification of the clinical dementia rating scale: the mCDR. *International Journal of Geriatric Psychiatry*, 25, 282–289. doi: 10.1002/gps.2334. [PubMed: 19565573]
- Dukart J, Sambataro F and Bertolino A (2016). 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. and
- Efron B (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association*, 82, 171–185. doi: 10.1080/01621459.1987.10478410.
- Grassi M, et al. (2018). A clinically-translatable machine learning algorithm for the prediction of Alzheimer's disease conversion in individuals with mild and premild cognitive impairment. *Journal of Alzheimer's Disease*, 61, 1555–1573. doi: 10.3233/jad-170547.
- Hinrichs C, Singh V, Xu G, Johnson SC and Alzheimers Disease Neuroimaging Initiative. (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. [PubMed: 21146621] and
- Hojjati SH, Ebrahimzadeh A, Khazaei A, Babajani-Feremi A and Alzheimer's Disease Neuroimaging Initiative. (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. [PubMed: 28286064] and
- Loewenstein DA, Curiel RE, Duara R and Buschke H (2017). Novel cognitive paradigms for the detection of memory impairment in preclinical Alzheimer's disease. *Assessment*, 25, 348–359. doi: 10.1177/1073191117691608. [PubMed: 29214859] and
- Loewenstein DA, et al. (2004). Semantic interference deficits and the detection of mild Alzheimer's disease and mild cognitive impairment without dementia. *Journal of the International Neuropsychological Society*, 10, 91–100. doi: 10.1017/s1355617704101112. [PubMed: 14751011]
- Long X, Chen L, Jiang C, Zhang L and Alzheimer's Disease Neuroimaging Initiative. (2017). Prediction and classification of Alzheimer disease based on quantification of MRI deformation. *PLoS One*, 12, e0173372. doi: 10.1371/journal.pone.0173372. [PubMed: 28264071] and
- Mathotaarachchi S, et al. (2017). Identifying incipient dementia individuals using machine learning and amyloid imaging. *Neurobiology of Aging*, 59, 80–90. doi: 10.1016/j.neurobiolaging.2017.06.027. [PubMed: 28756942]
- McKhann G, et al. (1984). Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services task force on Alzheimer's disease. *Neurology*, 34, 939. doi: 10.1212/WNL.34.7.939. [PubMed: 6610841]
- Minhas S, Khanum A, Riaz F, Alvi A and Khan SA (2016). A non parametric 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. [PubMed: 28113683] and
- Moradi E, Pepe A, Gaser C, Huttunen H, Tohka J and Alzheimer's Disease Neuroimaging Initiative. (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. [PubMed: 25312773] and
- Morris JC (1993). The clinical dementia rating (CDR): current version and scoring rules. *Neurology*, 43, 2412–2414. doi: 10.1212/WNL.43.11.2412-a.
- Nho K, et al. (2010). Automatic prediction of conversion from mild cognitive impairment to probable Alzheimer's disease using structural magnetic resonance imaging. *AMIA Annual Symposium Proceedings*, 2010, 542–546. [PubMed: 21347037]
- Plant C, et al. (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. [PubMed: 19961938]

- Retico A, et al. (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. [PubMed: 25291354]
- Scheltens P, et al. (1992). Atrophy of medial temporal lobes on MRI in "probable" Alzheimer's disease and normal ageing: diagnostic value and neuropsychological correlates. *Journal of Neurology, Neurosurgery, and Psychiatry*, 55, 967–972. doi: 10.1136/jnnp.55.10.967.
- Urs Ret al. (2009). Visual rating system for assessing magnetic resonance images: a tool in the diagnosis of mild cognitive impairment and Alzheimer disease. *Journal of Computer Assisted Tomography*, 33, 73–78. doi: 10.1097/RCT.0b013e31816373d8. [PubMed: 19188789]
- Wechsler D (1997). WMS-III: Wechsler Memory Scale Administration and Scoring Manual. San Antonio: Psychological Corporation.
- Weiss K, Khoshgoftaar TM and Wang D (2016). A survey of transfer learning. *Journal of Big Data*, 3, 9. doi: 10.1186/s40537-016-0043-6. **and**
- Westman E, Aguilar C, Muehlboeck JS and 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. [PubMed: 22890700] **and**
- Young J, et al. (2013). Accurate multimodal probabilistic prediction of conversion to Alzheimer's disease in patients with mild cognitive impairment. *Neuroimage Clinical*, 2, 735–745. doi: 10.1016/j.nicl.2013.05.004. [PubMed: 24179825]

Table 1.

Descriptive statistics

CONTINUOUS PREDICTORS	COGNITIVELY NORMALS		ALZHEIMER'S DISEASE	
	MEAN	SD	MEAN	SD
Modified Clinical Dementia Rating Scale, Memory Sum Score	0.88	1.11	4.49	2.25
Right hippocampus atrophy (VRS)	0.56	0.68	1.79	1.23
Center hippocampus atrophy (VRS)	0.48	0.64	1.96	1.21
Right entorhinal cortex atrophy (VRS)	0.30	0.60	1.41	1.19
Center entorhinal cortex atrophy (VRS)	0.38	0.65	1.73	1.27
Right perirhinal cortex atrophy (VRS)	0.32	0.62	1.39	1.30
Hopkins Verbal Learning Test Revised, Total Recall	25.62	4.42	14.13	4.93
Hopkins Verbal Learning Test Revised, Delayed Recall	9.16	2.05	2.12	2.69
Semantic Interference Test, Retroactive Total Score	25.66	2.16	13.43	6.97
Semantic Interference Test, Recognition Total Score	28.40	1.74	18.55	6.74
Logical Memory, Immediate Recall Score (WMS-IV)	13.03	3.47	5.45	3.54
Logical Memory, Delayed Recall Score (WMS-IV)	11.38	3.42	2.96	3.22
Trial Making B, Errors	0.69	0.98	2.56	3.56
Heart Rate (in beats-per-minute)	69.72	8.40	70.37	9.44

Table 2.

CATEGORICAL PREDICTORS	COGNITIVELY NORMALS		ALZHEIMER'S DISEASE	
	N	%	N	%
History of myocardial infarction				
No	6	3.05%	7	9.33%
Yes	191	96.95%	68	90.67%

SD: Standard Deviation; N: number of subjects; WMS-IV: Wechsler Memory Scale – Fourth Edition; VRS: Visual Rating Scale.

Table 2.

Performance of the algorithm in the MCI sample

	SENSITIVITY (ACTUAL)	SPECIFICITY	BALANCED ACCURACY
MCI (AUC = 0.821)	Sensitivity of 0.95	0.471	0.717
	Sensitivity of 0.90	0.559	0.741
	Sensitivity of 0.85	0.706	0.779
	Sensitivity of 0.80	0.735	0.775
	Sensitivity of 0.75	0.735	0.756
	Best Balanced Accuracy	0.706	0.779

AUC: Area Under the Receiving Operating Curve, MCI: Mild Cognitive Impairment.

Table 3.

Comparison with previous algorithms

REFERENCE	FOLLOW-UP PERIOD	PREDICTORS	MACHINE LEARNING TECHNIQUE	AUC	SPECIFICITY	SENSITIVITY	BALANCED ACCURACY
Our study	3 years	Socio-demographic, clinical, neuropsychological, clinician-rated brain atrophy, cardiovascular risk scores	SVM with radial-basis function kernel	0.821	0.852	0.706	0.779
Collij <i>et al.</i> , 2016 #76	1–4 years	Arterial spin labeling perfusion MRI	Linear SVM	0.71	0.75	0.67	n.a.
Nho <i>et al.</i> , 2010 #78	3 years	Brain structural MRI	SVM with radial-basis function kernel	n.a.	0.688	0.753	n.a.
Retico <i>et al.</i> , 2015 #69	2 years	Brain structural MRI	Linear SVM	0.707	n.a.	n.a.	n.a.
Westman <i>et al.</i> , 2013 #72	1.5 years	Brain structural MRI	Partial least square to latent structures multivariate analysis	0.749	0.645	0.77	n.a.
Dukart <i>et al.</i> , 2016 #49	at least 2 years	Brain structural MRI, 18-fluorodeoxyglucose Positron Emission Tomography, and APOE typing	Naive Bayes	0.84	0.861	0.875	0.868
Plant <i>et al.</i> , 2010 #46	Approximately 2.5 years	Brain structural MRI	Voting Feature Intervals	n.a.	0.87	0.56	n.a.
Cheng <i>et al.</i> , 2013 #68	not specified	Brain structural MRI, 18-fluorodeoxyglucose Positron Emission Tomography, and cerebrospinal fluid markers	multimodal relevance vector regression	–	0.541	0.641	–
Cui <i>et al.</i> , 2011 #70	2 years	Brain structural MRI, cerebrospinal fluid markers, neuropsychological, and functional activity scores	SVM with radial-basis function kernel	0.796	0.482	0.964	–
Young <i>et al.</i> , 2013 #75	3 years	Brain structural MRI, 18-fluorodeoxyglucose Positron Emission Tomography, and APOE gene typing	GP	0.823	0.571	0.833	0.722
Hinrichs <i>et al.</i> , 2011 #67	2–3 years	Brain structural MRI, and 18-fluorodeoxyglucose Positron Emission Tomography	Multi-kernel SVM	0.738	–	–	–

AUC: Area Under the Receiving Operating Curve; SVM: Support Vector Machine. If several algorithms were presented in the paper, only the results of the best one is presented here.