A machine learning-based holistic approach for the diagnosis of patients suffering from different conditions within the Alzheimer's disease spectrum

Noemi Massetti [1,2], Alberto Granzotto [1,2,5], Manuela Bomba [1,2], Stefano Delli Pizzi[1,2], Alessandra Mosca[1,2], Reinhold Scherer[3], Marco Onofrj[1,2], and Stefano L. Sensi [1,2,4,*]

[1] Center for Advanced Studies and technology (CAST), University G. d'Annunzio of Chieti-Pescara, Italy
[2] Department of Neuroscience, Imaging, and Clinical Sciences, University G. d'Annunzio of Chieti-Pescara, Italy
[3] School of Computer Science and Electronic Engineering, University of Sussex, Brighton, United Kingdom
[4] Institute for Mind Impairments and Neurological Disorders – iMIND, University of California - Irvine, Irvine, USA
[5] Sue and Bill Gross Stem Cell Research Center, University of California – Irvine, Irvine, California, USA

* Corresponding author:
Prof. Stefano L. Sensi
Center of Excellence on Aging and Translational Medicine - CeSI-MeT,
University G. d'Annunzio of Chieti-Pescara,
Via Colle dell'Ara, Chieti 66100, Italy.
Tel.: +39 0871 541544;
fax: +39 0871 541542;
e-mail: ssensi@uci.edu

Abstract

Alzheimer's disease (AD) and related dementias are neurodegenerative conditions driven by a multifactorial etiology. Growing clinical and epidemiological evidence indicates that the convergence of many polygenic, epigenetic, environmental, vascular, and metabolic factors shapes the risk, onset, and progression of the disease. Capturing and framing the complexity of these factors represent an unmet clinical and etiological challenge. Machine learning (ML) is a powerful approach to infer the processes behind the complex interaction of converging pathogenic factors that work with different weights and mechanisms at different stages of the disease. We employed an ML-based Random Forest (RF) algorithm and the wealth of information offered by the ADNI database to investigate the relative contribution of clinically relevant factors for identifying subjects affected by Mild Cognitive Impairment (MCI), a transitional state between healthy aging and dementia. The results of our ML-based approach indicate that the tool cannot help to predict the clinical outcome and conversion to AD of MCI subjects. On the other hand, non converting (ncMCI) subjects were correctly classified and predicted. In particular, our findings indicate that two neuropsychological tests, the FAQ and ADAS13, are the most relevant features used by our ML for the classification and prediction of younger (under 70 y.o.) ncMCI subjects. In addition, structural MRI data combined with indices representing the status of energy and lipid metabolism, measurements of the patient cardiovascular status, and systemic indices are the most critical factors for the classification and prediction of older (over 70 y.o) ncMCI.

1. Introduction

Alzheimer's disease (AD) is a neurodegenerative condition, and one of the most prevalent forms of dementia[1]. Clinical and epidemiological evidence suggests that the initial disease-related changes occur decades before the onset of clinical symptom[2]. In that respect, Mild Cognitive Impairment (MCI) is a transitional state between healthy aging and dementia and a prodromal phase of AD[3]. To date, no disease-modifying strategies are available, and almost all the clinical trials have failed [4]. On the other hand, a growing body of evidence indicates that prevention strategies effectively delay dementia onset and progression[5]. In that regard, the development of cost-effective approaches set to identify MCI subjects at-risk of conversion to dementia is of critical clinical importance.

To date, the clinical identification of the MCI stage mainly occurs through the implementation of neuropsychological tests[6] and is primarily based on patient performance on episodic memory tests[7]. The diagnosis is eventually helped by the use of brain Magnetic Resonance Imaging (MRI) scans and the assays of AD-related proteins (like amyloid and tau) in the cerebrospinal fluid (CSF). Of note, evidence on the neuropathophysiology of AD indicates that subtle molecular and cellular changes precede by decades the appearance of cognitive symptoms[8,9,10] and the use of these early biomarkers is of capital therapeutic relevance.

Machine learning (ML) is a field of computer science that provides computational tools to help in automated data classification and event prediction. Applied to AD, ML is a powerful tool to capture the complex interactions between underlying molecular abnormalities and cognitive changes occurring in the early stages of the disease progression[11,12]. In that respect, recent studies have modeled the progression of AD by employing different ML architectures and helped to unravel subtle structural brain changes as assessed by mMRI[13–15] data. However, significant levels of accuracy in predicting the MCI transition to dementia have only been achieved through the combined implementation of multiple biomarkers like detailed MRI-based measurements of subtle changes in cortical thicknesses and brain volumes, alterations of pathogenic CSF proteins, the variation of cognition and behavior by neuropsychological data, evaluation of AD-related genetic mutations as well as the inference from sociodemographic data[16–19].

In general, the use of such an array of biomarkers has been limited to indices that reflect changes occurring within the Central Nervous System (CNS). However, recent paradigm shifts in the construct related to dementia indicate that the implementation of a more comprehensive systems medicine-based approach and the evaluation of peripheral and systemic changes may offer better opportunities for the comprehension of disease course within the AD spectrum. In line with this network- and systems medicine-based perspective, Loscalzo *et al.*[20] have proposed that chronic diseases, including dementia, need to be considered as the result of a converging perturbation of complex intra- and inter-cellular networks, with alterations that occur at many levels and are not limited to one organ or driven by a single molecular factor or pathogenic process.

Considering this view, we have therefore employed an ML-based approach to assessing the progression to AD of MCI subjects. To that aim, we have used ML to identify the subset of MCI individuals who are more prone to convert to dementia by taking advantage of a wealth of biological, functional, and MRI-based structural data as well as the clinical features that reflect the pathogenic events that occur within as well as outside of the CNS.

The object of our ML study was a cohort of 517 MCI patients and the related data obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. In this study cohort, we

employed an ML-based Random Forest (**RF**) algorithm to classify, in a supervised manner, converting to AD (**cMCI**) or stable, non-converting MCI (**ncMCI**) MCI subjects. Data related to the study population were analyzed with ML individually and in combination. Data were also assessed in terms of greater classification power. In the end, the ML generating the best classification of cMCI and ncMCI subjects was tested on a new cohort of patients (a test-dataset) to estimate the machine predictive ability. Finally, we performed a RF dividing the MCI study cohort into subjects who were older or younger than 70 years old. The decision was based on the assumption that disease processes occur with a different degree of relevance and pathogenic weight in "young" versus older patients as well as moving from the hypothesis that the AD-related underlying pathology may be different and only partially overlapping in these two groups. Thus, in a nutshell, we tested the hypothesis that younger MCI are mainly subjected to primary neurodegenerative processes while older MCI individuals suffer from a more combined variety of precipitating and age-dependent systemic factors like alterations of blood pressure, energy deficits, dysmetabolic conditions.

Overall, the study was also set to investigate the relative contribution in the identification of ncMCI and MCI subjects offered by the use of first-tier diagnostic tools like alterations of AD-related CSF proteins and variations of neuropsychological test scores or by the implementation of second-tier biomarkers like detailed structural MRI-related brain changes or alterations of peripheral parameters.

2.Materials

All the data used in this work are available on the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. ADNI is a public-private repository of clinical, imaging, genetic, and biochemical biomarker data obtained from North American subjects or patients (www.adni-info.org). ADNI is aimed at identifying the determinant processes leading to AD and diagnose pathological changes occurring at the earliest stage. The data downloaded and analyzed in this study were all acquired at the first appointment of the follow-up (baseline) required by the ADNI protocols. The management and selection of the data were carried out with custom made R-written codes[21].

2.1 Subjects

Subjects considered in this study were patients with a diagnosis of MCI extracted from the cohorts of ADNI-1, ADNI-GO, ADNI-2, and ADNI-3. The inclusion criteria for these patients were the ones provided by the ADNI protocols[22,23], and the diagnosis of MCI based on the assessment of global cognition as well as scores of neuropsychological evaluations related to episodic memory, the presence od subjective cognitive disorders, and the clinical evaluation through a semi-structured interview.

Two classes of MCI subjects were chosen: 1) patients with a baseline diagnosis of MCI who converted to AD within 24 months (cMCI) and patients with a diagnosis of MCI at baseline who remained stable for the following 36 months (ncMCI). The timeframe was chosen taking into consideration a report indicating that patients with a diagnosis of MCI have a high probability of converting to AD within 30 months[24]. Table.1 summarizes the demographics data of the patients involved in the study.

2.2 Neuropsychological assessments

For each patient, we took into account 16 neuropsychological tests employed to assess the status of different cognitive domains. This neuropsychological dataset included the Alzheimer's Disease Assessment Scale-Cognitive (ADAS-Cog) subscales to evaluate the severity of memory, learning, language (production and comprehension), praxis, and orientation deficits [25,26]; the Mini-Mental State Examination[27] to investigate global cognition; the 30-item Boston Naming Test[28] and the Animal Fluency[29] to assess semantic memory and language abilities; the Functional Activities Questionnaire **(FAQ)** for the assessment of daily living activities[30]; the Rey Auditory Verbal Learning Test and Logical Memory II, subscale of the Wechsler Memory Scale-Revised (WMS-R) to investigate recall and recognition[31,32] ; the Trail Making Test[33] , part A and B (time to completion) to assess attention/executive functions; the Clock Drawing Test to evaluate attention, working and visual memory, and auditory comprehension[34] ; the Clinical Dementia Rating Scale to quantify the patient's degree of dementia[35]. The Supplementary Table.1 shows the abbreviations used for each neuropsychological assessment and the domains and cognitive functions that each test is set to evaluate.

2.3 Biospecimen quantification

The Supplementary Table.2 shows the quantifications of the biospecimen considered in this work. The biospecimen selection, within the datasets available on the ADNI database (Biospecimen Inventory, http://adni.loni.usc.edu), was made by taking into account the number of samples and the consistency of the biospecimen measurements for the patients at different phases of the ADNI project (ADNI-1, ADNI-GO, ADNI-2, ADNI-3). To meet the second criterion and reduce the incidence of human error, we considered only data produced through automated techniques.

2.4 Additional indices and parameters

We extracted ADNI data related to years of education[36] and APOE-ε4 status[37] of MCI patients and also acquired data related to "vital signs" like systolic (VSBPSYS) and diastolic pressure (VSBPDIA), ventilation frequency (VSRESP), cardiac frequency (VSPULSE), and body temperature (VSTEMP).

2.5 Structural Magnetic Resonance data

The MRI data downloaded from ADNI (Image Collections, http://adni.loni.usc.edu) were acquired with a protocol of Magnetization Prepared-RApid Gradient Echo (MP-RAGE) and through a Philips 3-tesla MRI scanner (for details about MRI procedures and protocols see http://adni.loni.usc.edu/wp-content/uploads/2010/05/ADNI2_MRI_Training_Manual_FINAL.pdf) thereby limiting bias and technical issues associated with the use of different scanner types/brand. T1-weighted images were obtained using 3D Turbo Field-Echo sequences (slice thickness of 1.2 mm, repetition time/echo time of 6.8/3.1 ms). The structural MRI analysis was performed with Freesurfer (version 6.0). Automatic reconstruction and labeling of cortical and subcortical regions was achieved with the "recon-all-all" command line, according to Desikan-Killiany Atlas[38]. The volumes of the brain regions, computed with *asegstats2table*, were normalized by dividing to the total intracranial volume of each patient, while the thicknesses of the brain areas considered are

those calculated automatically by *aparcstats2table*. Table 4 shows the volumes and brain thicknesses analyzed in this study.

3. Methods

Our ML approach implemented a RF algorithm. The RF is a supervised non-linear classifier and its operation is based on the construction of binary decision trees obtained with the Bagging sampling method (an acronym for bootstrap aggregating)[39]. The RF classifier was chosen because the performance is robust and stable over an extensive range of parameters. Furthermore, it is independent of the distribution of data (i.e., neuropsychological scores, quantifications of lipoproteins, etc.) and has significant multi-class and advanced data-mining capabilities[40].

RF analyzed the data acquired at the ADNI follow-up baseline. In the training phase, the algorithm studied the non-linear interactions between the data (or features) of the subjects divided upstream into the two classes: cMCI and ncMCI. The goal at this stage was to identify the best subdivision/classification strategy. In the testing phase, this strategy was tested on a batch of new subjects.

In the training phase, the RF analyzed 85% of each dataset (randomly extracted) and performed 30 times the analysis; the number of trees was chosen manually and set to minimize the out-of-bag error (number of trees: 200, 250, 300, 350, 400, 450, 500, 550, 600).

The classic evaluation parameters of ML, ***Accuracy, Sensitivity,*** **and** ***Specificity***, refer to the overall performance of the machine, i.e., the ability to classify cMCI and ncMCI subjects. To evaluate the ability to predict whether a subject belongs to one of the two classes and, therefore, assess separately the classification strategy for cMCI and ncMCI we introduced the variable ***Predictivity***, which measures the percentage of subjects (test-dataset) correctly classified in each of the two classes.

In the study, we employed the Treebagger class function of the Statistics and Machine Learning Toolbox of MATLAB (The Mathworks Inc., Natick, MA, USA).

4. Results

Results shown in Table 2 indicate the best combinations of features that allow the most accurate classification of cMCI and ncMCI patients and also depicts the ML performance evaluation. These results are sorted by accuracy values. Predictivity values were considered significant if greater than or equal to 85%. The results of the whole ML analysis are shown in Supplementary Table 5 and Supplementary Table 6. In the case of cMCI patients, no combination of data and parameters allowed the ML to achieve a significant predictivity value. Disappointingly, we had to conclude that ML was not able to detect MCI subjects who are set to convert to AD if the machine is fed only with baseline data of the MCI individuals.

On the other hand, the ML performance in classifying ncMCI subjects was very effective. In this class of individuals, when employing data related to the neuropsychological tests, we reached an accuracy value of 0.95 and a predictivity rate of 84.9%. The result improved and reached a predictivity value of 86.4% when education levels were added. The neuropsychological test that had a decisive weight for the classification was the RAVLT immediate, a test employed to assess episodic memory. Surprisingly, data related to biospecimen quantifications allowed the best predictivity. Triglyceride and cholesterol (TRI_COL) data allowed the ML to correctly predict 87% of

ncMCI individuals with an accuracy of 0.92. The same accuracy value was reached when feeding the ML with data related to the energy substrates (ENER_SUB) and the systemic indexes (SYST_IND), even though only in the first two cases the predictivity approached the threshold for statistical significance and produced a predictivity value of 84.7%. The data that were determinant in the classification and, therefore, in the prediction were Triglycerides LDL (for the triglycerides and cholesterol dataset) and glucose (for the energy substrates dataset).

To investigate the effect of age, we divided the study cohort into cMCI subjects over and under 70 years of age. Table 3 shows the best results for the two cMCI groups sorted by taking into account the accuracy of the ML performance. No combination of data allowed the ML to classify and predict the fate of the cMCI patients.

The situation was different when investigating ncMCI patients, again divided into two groups, over and under 70 years of age. The predictivity values of the best ML increased when compared to the ML applied to the whole set of ncMCI, and the data combinations that helped ML differed completely. Table 4 shows the combination of datasets that allowed the ML to classify and predict ncMCI patients divided upon age. For older (**over 70)** ncMCI individuals, structural MRI data were the features that, in combination with other data, allowed the ML to achieve the highest values of accuracy and predictivity. The combination of MRI with data related to lipoproteins, glucogenesis-associated metabolites and amino acids (METABOL) allowed the best classification and prediction for older ncMCI. As for the MRI data, volumes of the left hippocampus was the feature that generated a prediction of 85.5% and an accuracy of 0.98. Feeding ML with MRI data associated with vital signs produced the second-best result, although the predictivity value dropped to 82.1%. Finally, combining MRI data with the systemic indexes like albumin and creatinine blood levels allowed the ML to classify ncMCI subjects with an accuracy value of 0.95 and a predictivity value of 86.3%.

For younger (**under 70)** ncMCI subjects, the scenario was completely different. Data related to neuropsychological scores alone allowed the ML to correctly classify ncMCI with a 100% accuracy and 90.2% predictivity. These values were achieved by the interaction of the employed 16 tests. Of note, the use of the FAQ (for the assessment of daily living activities) and the ADAS13 test (to assess the severity of impairment to memory, learning, language, practice, and orientation), as well as the RAVLT immediate test, were the major driving force for ML. Combining neuropsychological evaluations with levels of education decreased the predictivity to 88%, although the accuracy remained the same. In contrast, the combination of neuropsychological with MRI structural data was found to be less useful to help the ML in the classification and produced an accuracy value of 0.95 and a prediction rate of 86%.

In summary, the results of our RF-based approach indicate that the tool cannot help to predict the clinical outcome of cMCI. On the other hand, ncMCI subjects were correctly classified and predicted. In particular, our results indicate that 1) two neuropsychological tests (FAQ and ADAS13) are the most relevant for the classification and prediction of ncMCI under 70, and 2). structural MRI data combined with data representing the status of the energy and lipid metabolism and measurements of patient cardiovascular status and systemic indices were the most critical factors for the classification and prediction of ncMCI over 70.

5. Discussion

We aimed to identify which combination of ADNI-related data was the most effective to classify cMCI and ncMCI patients. We started with CNS features that are classically employed to assess the level of neurodegeneration and cognitive decline (i.e., altered levels of AD-related CSF proteins and neuropsychological evaluations). We then added, to refine the analyses, structural MRI data, and eventually peripheral biomarkers (i.e., blood/plasma levels of lipoproteins, gluconeogenesis-related metabolites, amino acids, bile acids, and purines). This large number of variables, measured at baseline, is still not enough for the ML to correctly classify and predict the fate of cMCI subjects. Even when we subdivided the group upon the +/- 70 years old divide, our RF failed to achieve significant accuracy and predictivity when dealing with conversion prediction. One possible explanation is that ML needs more than baseline data to work, and therefore, the availability of longitudinal data becomes mandatory.

A very different scenario emerged when assessing ncMCI subjects. Even with only baseline data feeding, our RF was able to classify and identify stable MCI patients accurately. The analysis of ML's variables employed for classifying ncMCI subjects indicates that neuropsychological tests, triglycerides, cholesterol, and energy substrates are major driving forces for the learning process. This result confirms the hypothesis that AD, like other chronic diseases, is not an organ-specific condition, and there is still a wealth of mostly unexplored pathological processes occurring outside of the CNS. Furthermore, when dividing ncMCI patients into old (over-70) and young (under-70), we found that the most effective diagnostic tools and biomarkers that the ML employed were very different in the two MCI subsets.

Elderly ncMCI subjects were best classified by the use of a combination of refined structural MRI and lipid- and glucid metabolism-related data (METABOL), along with cardiovascular measurements (VIT_SIGN) and with systemic indexes like the levels of albumin and creatinine (SYST_IND), thereby indicating the convergence of many systemic pathogenic processes occurring inside and outside of the brain. The analysis of the ML outputs shows that the most critical data employed in the classification were the one related to the left hippocampal volume, in line with the classic observations that this region is the most affected by the AD-related neurodegenerative processes.

On the other hand, when assessing young ncMCI individuals, the best performance in prediction and classification was achieved using data relative to cognitive status. The neuropsychological tests that emerged for importance in the classification were the FAQ and the ADAS-13. This finding is not surprising as the FAQ assesses the ability to deal with daily problems and personal care, a feature that is deeply affected by aging but is expected to be less deficient in younger patients and is therefore correctly interpreted as a significant warning sign that the ML is using.

The ADAS-13[41] test evaluates the overall cognitive status. In routine clinical work, the MMSE is preferred, but, surprisingly, ML found the test scores to be one of the least useful parameters for the classification of young patients. This result can be explained by taking into account the origin and score structure of the two tests.

The MMSE was created as an easy to use clinical tool while the ADAS-13[41] is more research-oriented. The score range is also different, higher [from 0 (no symptoms) to 70 (maximum symptoms)] in the ADAS-13 compared to the MMSE that spans from 0 (no symptoms) to 30

(maximum symptoms). Thus, the ADAS-13, more specific and with a more granular scale of values (0-70), is able to appreciate more nuanced cognitive abnormalities.

It is generally accepted that cognitive and memory deficits occurring at the early stages of the disease are related to the presence of synaptic failure[42–44]. These subtle changes are, however, behind the resolution power of clinically-employed MRI, and this might explain why the ML made little use of the MRI data for the classification of younger patients. On the other hand, in the case of older individuals, synaptic alterations disappear in the context of a broader degeneration that becomes more easily detectable by MRI. Therefore, these robust alterations allow the ML to perform well in the classification of ncMCI subjects over 70.

6. Conclusions

AD is a complex and multifactorial disease, and the characterization of patients in a prodromal stage of the disease (MCI) is a challenge for biomedical research and an unmet clinical and therapeutic needs. ML techniques and big data analysis are useful for understanding AD's etiopathology and identifying novel and unexpected disease features at different stages. For instance, a surprising finding of our study relates to the importance of considering peripheral biomarkers.

We were disappointed by the lack of the predicitivity power of our ML when dealing with cMCI subjects. The RF algorithm's robustness and effectiveness were not a problem, indicating that the lack of longitudinal inputs may be a severe limiting factor and a challenge.

Overall the study indicates that the analysis of variations that occurs inside and outside of the brain is a fruitful pathway to be pursued. These alterations deserve to be further investigated and are the gateway to precision medicine. Precision medicine and systems medicine and network-based approaches hold the promise to generate tailored diagnoses, predict disease risks, and produce customized treatments that maximize safety and efficacy. The use of Artificial Intelligence (AI) is mandatory for such an ambitious paradigm shift in health care. Thus, the use of AI will be essential for identifying, in an unbiased fashion, new pathogenic pathways and leveraging them for diagnostic and therapeutic use [45]. We hope that our study can make a humble contribution to motivate further the exploration of ML models for the study and the creation of better computational tools to understand the mystery of neurodegenerative processes and dementia in particular.

ACKNOWLEDGMENTS
The study was funded by the Alzheimer's Association Part the Cloud: Translational Research Funding for Alzheimer's Disease (PTC) PTC-19-602325 and the Alzheimer's Association - GAAIN Exploration to Evaluate Novel Alzheimer's Queries (GEENA-Q-19-596282) (SLS).

References

1. M., P. *et al.* The global prevalence of dementia: A systematic review and metaanalysis. *Alzheimer's and Dementia* (2013).

2. Vermunt, L. *et al.* SINGLE-SUBJECT GREY MATTER NETWORKS DETERIORATE 12 YEARS BEFORE DEMENTIA. *Alzheimer's Dement.* (2019). doi:10.1016/j.jalz.2019.06.4502

3. Morris, J. C. *et al.* Mild cognitive impairment represents early-stage Alzheimer disease. *Arch. Neurol.* (2001). doi:10.1001/archneur.58.3.397

4. Long, J. M. & Holtzman, D. M. Alzheimer Disease: An Update on Pathobiology and Treatment Strategies. *Cell* (2019). doi:10.1016/j.cell.2019.09.001

5. Rakesh, G., Szabo, S. T., Alexopoulos, G. S. & Zannas, A. S. Strategies for dementia prevention: latest evidence and implications. *Therapeutic Advances in Chronic Disease* (2017). doi:10.1177/2040622317712442

6. Petersen, R. C. *et al.* Mild cognitive impairment: Clinical characterization and outcome. *Arch. Neurol.* (1999). doi:10.1001/archneur.56.3.303

7. Jessen, F. *et al.* AD dementia risk in late MCI, in early MCI, and in subjective memory impairment. *Alzheimer's Dement.* (2014). doi:10.1016/j.jalz.2012.09.017

8. Dickerson, B. C. *et al.* The cortical signature of Alzheimer's disease: Regionally specific cortical thinning relates to symptom severity in very mild to mild AD dementia and is detectable in asymptomatic amyloid-positive individuals. *Cereb. Cortex* (2009). doi:10.1093/cercor/bhn113

9. Scahill, R. I., Schott, J. M., Stevens, J. M., Rossor, M. N. & Fox, N. C. Mapping the evolution of regional atrophy in Alzheimer's disease: Unbiased analysis of fluid-registered serial MRI. *Proc. Natl. Acad. Sci. U. S. A.* (2002). doi:10.1073/pnas.052587399

10. Thompson, P. M. *et al.* Cortical change in Alzheimer's disease detected with a disease-specific population-based brain atlas. *Cereb. Cortex* (2001). doi:10.1093/cercor/11.1.1

11. Kononenko, I. Machine learning for medical diagnosis: History, state of the art and perspective. *Artif. Intell. Med.* (2001). doi:10.1016/S0933-3657(01)00077-X

12. Beam, A. L. & Kohane, I. S. Big data and machine learning in health care. *JAMA - Journal of the American Medical Association* (2018). doi:10.1001/jama.2017.18391

13. Moradi, E., Pepe, A., Gaser, C., Huttunen, H. & Tohka, J. Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *Neuroimage* (2015). doi:10.1016/j.neuroimage.2014.10.002

14. Hua, X. *et al.* Mapping Alzheimer's disease progression in 1309 MRI scans: Power estimates for different inter-scan intervals. *Neuroimage* (2010). doi:10.1016/j.neuroimage.2010.01.104

15. Gaser, C., Franke, K., Klöppel, S., Koutsouleris, N. & Sauer, H. BrainAGE in Mild Cognitive Impaired Patients: Predicting the Conversion to Alzheimer's Disease. *PLoS One* (2013). doi:10.1371/journal.pone.0067346

16. Lee, G. *et al.* Predicting Alzheimer's disease progression using multi-modal deep learning approach. *Sci. Rep.* (2019). doi:10.1038/s41598-018-37769-z

17. Young, A. L. *et al.* Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with Subtype and Stage Inference. *Nat. Commun.* (2018). doi:10.1038/s41467-018-05892-0

18. Bhagwat, N., Viviano, J. D., Voineskos, A. N. & Chakravarty, M. M. Modeling and prediction of clinical symptom trajectories in Alzheimer's disease using longitudinal data. *PLoS Comput. Biol.* (2018). doi:10.1371/journal.pcbi.1006376

19. Casanova, R. *et al.* Alzheimer's disease risk assessment using large-scale machine learning methods. *PLoS One* (2013). doi:10.1371/journal.pone.0077949

20. Barabási, A. L., Gulbahce, N. & Loscalzo, J. Network medicine: A network-based approach to human disease. *Nature Reviews Genetics* (2011). doi:10.1038/nrg2918

21. Team, R. C. R: A language and environment for statistical computing. R Foundation for Statistical Computing. (2018).

22. Alzheimer's Disease Neuroimagine Initiative. ADNI2 Procedure Manual.

23. Initiative, A. D. N. ADNI3 Procedures Manual.

24. Fischer, P. *et al.* Conversion from subtypes of mild cognitive impairment to Alzheimer dementia. *Neurology* (2007). doi:10.1212/01.wnl.0000252358.03285.9d

25. Mohs, R. C. & Cohen, L. Alzheimer's Disease Assessment Scale (ADAS). *Psychopharmacol. Bull.* (1988).

26. Mohs, R. C. *et al.* Development of cognitive instruments for use in clinical trials of antidementia drugs: Additions to the Alzheimer's disease assessment scale that broaden its scope. *Alzheimer Dis. Assoc. Disord.* (1997). doi:10.1097/00002093-199700112-00003

27. Folstein, M. F., Folstein, S. E. & McHugh, P. R. 'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* (1975). doi:10.1016/0022-3956(75)90026-6

28. E, Goodglass H, W. S. The Boston Naming Test. *Philadelphia Lea Febiger.* (1983).

29. Morris, J. C. *et al.* The consortium to establish a registry for alzheimer's disease (CERAD). Part I. Clinical and neuropsychological assessment of alzheimer's disease. *Neurology* (1989). doi:10.1212/wnl.39.9.1159

30. Pfeffer, R. I., Kurosaki, T. T., Harrah, C. H., Chance, J. M. & Filos, S. Measurement of functional activities in older adults in the community. *Journals Gerontol.* (1982). doi:10.1093/geronj/37.3.323

31. Rey, A. *L'examen clinique en psychologie. [The clinical examination in psychology.]. L'examen clinique en psychologie.* (1964).

32. Wechsler, D. *WAIS--III: Wechsler Adult Intelligence Scale (3rd ed.) Administration and scoring manual. The Psychological Corporation, San Antonio, TX* (1997).

33. Spreen, O., y Strauss, E. A Compendium of Neuropsychological Tests. *Administration Norms And Commentary* (1998).

34. Nishiwaki, Y. *et al.* Validity of the Clock-Drawing Test as a screening tool for cognitive impairment in the elderly. *Am. J. Epidemiol.* (2004). doi:10.1093/aje/kwh288

35. Tractenberg, R. E., Schafer, K. & Morris, J. C. Interobserver disagreements on clinical dementia rating assessment: Interpretation and implications for training. *Alzheimer Dis. Assoc. Disord.* (2001). doi:10.1097/00002093-200107000-00007

36. Stern, Y. Cognitive reserve and Alzheimer disease. *Alzheimer Disease and Associated Disorders* (2006). doi:10.1097/01.wad.0000213815.20177.19

37. Blacker, D. *et al.* ApoE-4 and age at onset of Alzheimer's disease: The NIMH genetics initiative. *Neurology* (1997). doi:10.1212/WNL.48.1.139

38. Desikan, R. S. *et al.* An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* (2006). doi:10.1016/j.neuroimage.2006.01.021

39. Breiman, L. Bagging predictors. *Mach. Learn.* (1996). doi:10.1007/bf00058655

40. Steyrl, D., Scherer, R., Faller, J. & Müller-Putz, G. R. Random forests in non-invasive sensorimotor rhythm brain-computer interfaces: A practical and convenient non-linear classifier. *Biomed. Tech.* (2016). doi:10.1515/bmt-2014-0117

41. Kueper JK, Speechley M, M.-O. M. The Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog): Modifications and Responsiveness in Pre-Dementia Populations. A Narrative Review. *J Alzheimers Dis* **63(2)**, 423–444 (2018).

42. DeKosky, S. T. & Scheff, S. W. Synapse loss in frontal cortex biopsies in Alzheimer's disease: Correlation with cognitive severity. *Ann. Neurol.* (1990). doi:10.1002/ana.410270502

43. Selkoe, D. J. Alzheimer's disease is a synaptic failure. *Science* (2002). doi:10.1126/science.1074069

44. Sheng, M., Sabatini, B. L. & Südhof, T. C. Synapses and Alzheimer's disease. *Cold Spring Harb. Perspect. Biol.* (2012). doi:10.1101/cshperspect.a005777

45. Rajkomar, A. *et al.* Scalable and accurate deep learning with electronic health records. *npj Digit. Med.* (2018). doi:10.1038/s41746-018-0029-1

Figures and tables

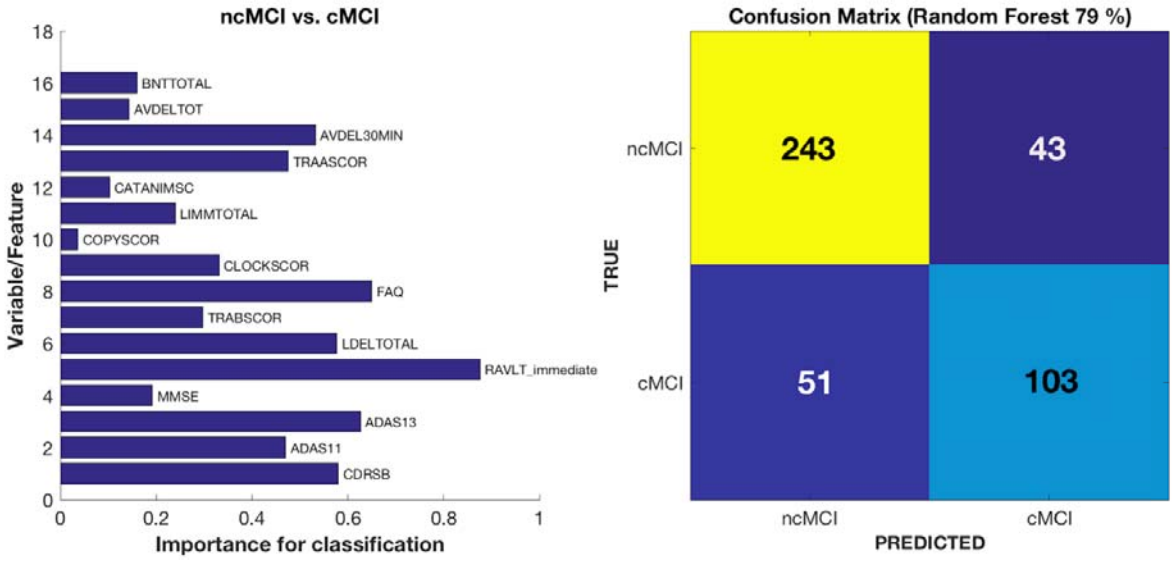

Fig. 1. ML training phase output. The figure shows the output shape of each of the 30 RFs using as feeding data the neuropsychological features. (Left) The histogram illustrates the classification weight, or importance, of each feature employed for predicting MCI subject conversion (cMCI) or non-conversion (ncMCI) to AD. Classification scores range from 0 to 1, with higher values indicating a better classification. (Right) The pictogram illustrates the supervised ML Confusion Matrix. The Matrix provides an overview of the ML ability to predict conversion (cMCI) or non-conversion (ncMCI) of MCI subjects to AD. The upper left box shows the number of true positive ncMCI subjects recognized by the ML; the upper right box, instead, shows the number of false positive ncMCI subjects identified by the algorithm. The bottom left square shows the number of cMCI individuals identified as ncMCI (false positive) by the ML; the bottom right box shows the number of true positive cMCI subjects recognized by the algorithm.

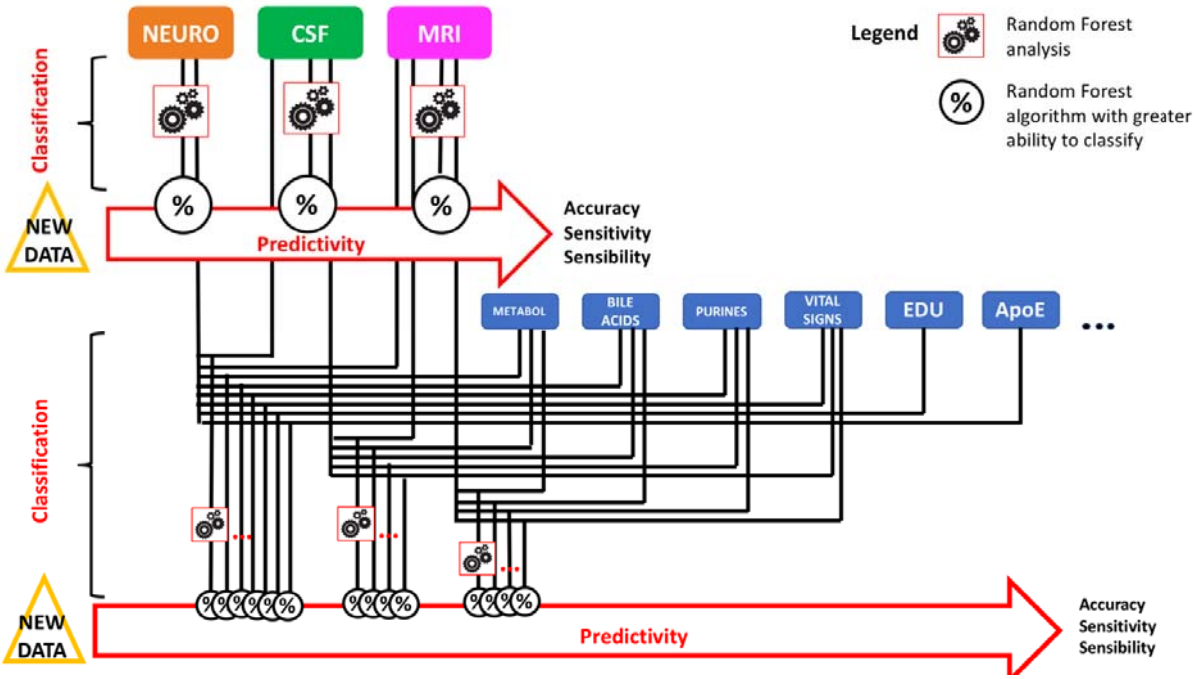

Fig. 2 Study workflow. The pictogram illustrates the workflow of the study. (Top) The pictogram illustrates the brain related data (features) that were considered in the first phase of the study. Each feature group (Neuro, CSF, and MRI) was studied individually and in combination. 85% of each dataset was used for selecting the ML approach leading to the highest classification scores. The remaining 15% ("new data") were employed for evaluation of the ML performance. (Bottom) The second phase of the study the features identified in the first stage were combined with peripheral biomarkers to evaluate the contribution of these parameters to ML predictivity.

Abbreviations: NEURO, neuropsychological tests; CSF, cerebrospinal fluid proteins; MRI, magnetic resonance imaging; METABOL, lipoproteins, glucogenesis-related metabolites and amino acids; EDU, years of education; ApoE, alipoprotein E ε4 allele.

Table 1. Study subject demographics. The table illustrates the demographics of the MCI cohort (t_0) as well as the demographics of the same subjects in the 24-month, for MCI converting to AD (cMCI, t_1), and in the 36-month follow-up, for MCI subjects non converting do AD (ncMCI, t_2).

	MCI (n=520) time: t_0	**cMCI** (n= 178) time: t_1	**ncMCI**(n= 342) time: t_2
Gender (female/male)	196/325	70/108	126/216
Age (yrs)	73,6 ± 7,6	75,2 ± 7,3	76,4 ± 7,8
Education (yrs)	15,8 ± 2,9	15,8 ± 2,7	15,8 ± 2,9
MMSE	27,3 ± 1,9	23,7 ± 2,9	27,1 ± 2,9
ADAS-13	17,7 ± 6,9	26,9 ± 6,8	16,6 ± 8,8
ApoE ε4 (0/1/2)	252/207/61	57/91/30	195/116/31

Abbreviations: MCI, mild cognitive impairment; cMCI, converter mild cognitive impairment; ncMCI, non-converter mild cognitive impairment; MMSE, Mini-Mental State Examination; ADAS-13, Alzheimer's Disease Assessment Scale-Cognitive subscales, 13 items; ApoE ε4 (0/1/2), alipoprotein E ε4 allele.

MCI-CONVERTER	Trees number	Predictivity	Accuracy	Sensitivity	Specificity
NEURO	500	66,8%	0,95	0,87	1,00
NEURO+EDU	300	66,2%	0,95	0,87	1,00
TRI_COL	300	12,70%	0,92	0,83	0,97
ENER_SUB	250	22,70%	0,92	0,79	1,00
SYST_IND	200	35,40%	0,92	0,83	0,97
AMMINO	150	31,20%	0,90	0,79	0,97
NEURO+METABOL	200	59,6%	0,86	0,86	0,87
CSF+ENER_SUB	300	44,70%	0,86	1,00	0,75
CSF+VIT_SIG	250	55,5%	0,82	0,64	0,83
NEURO+VIT_SIG	300	59,6%	0,81	0,54	0,93
NEURO+ApoEε4	300	64,9%	0,81	0,60	0,94
MRI+BIL_ACID	300	51,7%	0,80	0,75	0,84
MRI+CSF	250	58,9%	0,78	0,64	0,85
MRI+NEURO+CSF	250	69,4%	0,78	0,64	0,85
CSF+AMMINO	300	52%	0,78	0,68	0,81
NEURO+BIL_ACID	300	60,4%	0,77	0,60	0,86
NEURO+MRI	300	61,1%	0,75	0,56	0,87

MCI-NON CONVERTER	Trees number	Predictivity	Accuracy	Sensitivity	Specificity
NEURO	500	84,9%	0,95	1,00	0,87
NEURO+EDU	300	86,4%	0,95	1,00	0,87
TRI_COL	300	87%	0,92	0,97	0,83
ENER_SUB	250	84,70%	0,92	1,00	0,79
SYST_IND	200	74,10%	0,92	0,97	0,83
AMMINO	150	72,90%	0,90	0,97	0,79
NEURO+METABOL	200	85,7%	0,86	0,87	0,86
CSF+ENER_SUB	300	82,10%	0,86	0,75	1,00
CSF+VIT_SIG	250	77,3%	0,82	0,83	0,64
NEURO+VIT_SIG	300	84,2%	0,81	0,93	0,54
NEURO+ApoEε4	300	89,2%	0,81	0,94	0,60
MRI+BIL_ACID	300	89,9%	0,80	0,84	0,75
MRI+CSF	250	86,7%	0,78	0,85	0,64
MRI+NEURO+CSF	250	88,6%	0,78	0,85	0,64
CSF+AMMINO	300	79,20%	0,78	0,81	0,68
NEURO+BIL_ACID	300	87,0%	0,77	0,86	0,60
NEURO+MRI	300	89,0%	0,75	0,87	0,56

Table 2. ML prediction performance for MCI subjects converting (cMCI) and non-converting (ncMCI) to AD. The tables depict the ML ability to classify cMCI and ncMCI subjects in the test dataset. The ranking is based on the Accuracy value.

Abbreviations (see also supplementary tables Table_1, Table_2 and Table_3): NEURO, neuropsychological tests; EDU, years of education; TRI_COL, Triglycerides and cholesterol; ENER_SUB, energetic substrates; SYST_IND, systemic indices; AMMINO, amino acids; METABOL, lipoproteins, glucogenesis-related metabolites and amino acids; CSF, cerebrospinal fluid proteins; VIT_SIG, vital signs; ApoE ε4, alipoprotein E ε4 allele; BIL_ACID, bile acids; MRI, magnetic resonace imaging.

cMCI over 70	Trees number	Predictivity	Accuracy	Sensitivity	Specificity
MRI+METABOL	300	54,3%	0,98	0,95	1,00
MRI+ VIT_SIG	300	52,4%	0,98	0,95	1,00
MRI+SYST_IND	350	53%	0,95	0,90	1,00
TRI_COL	250	24,80%	0,93	0,85	1,00
ENER_SUB	250	25,50%	0,93	0,85	1,00
AMMINO	200	36,20%	0,93	0,85	1,00
NEURO	500	59,0%	0,89	0,84	0,91
NEURO+EDU	200	61,8%	0,89	0,84	0,91
SYST_IND	250	37,10%	0,88	0,85	0,91
NEURO+CSF	300	76,0%	0,88	0,70	0,96
NEURO+CSF+ApoEε4	200	73,3%	0,88	0,80	0,92
NEURO+ApoEε4	350	65,4%	0,87	0,79	0,91
CSF+ VIT_SIG	250	54,8%	0,83	0,75	0,87
NEURO+VIT_SIG	400	52,3%	0,81	0,65	0,91
NEURO+ENER_SUB	200	59%	0,78	0,55	0,94
NEURO+BIL_ACID	300	57,1%	0,77	0,50	0,89
CSF+METABOL	300	49,3%	0,76	0,56	0,83

cMCI under 70	Trees number	Predictivity	Accuracy	Sensitivity	Specificity
NEURO	400	67,4%	1,00	1,00	1,00
NEURO+EDU	200	71,4%	1,00	1,00	1,00
NEURO+MRI	350	63,8%	0,95	1,00	0,93
MRI+NEURO+CSF	300	65,4%	0,93	0,50	1,00
MRI	400	52,7%	0,90	0,67	1,00
CSF+ VIT_SIG	350	31,0%	0,88	1,00	0,86
CSF+SYST_IND	200	57,10%	0,88	0,00	1,00
CSF+AMMINO	200	53,60%	0,88	0,00	1,00
MRI+CSF	250	57,7%	0,86	0,00	1,00
NEURO+VIT_SIG	300	72,5%	0,78	0,43	0,94
NEURO+CSF	300	64,0%	0,76	0,63	0,89
NEURO+CSF+ApoEε4	200	68,0%	0,76	0,63	0,89
TRI_COL	250	15,60%	0,73	0,00	1,00
ENER_SUB	250	21,80%	0,73	0,00	1,00
SYST_IND	250	21,80%	0,73	0,00	1,00
AMMINO	200	25%	0,73	0,00	1,00
MRI+ENER_SUB	400	53,10%	0,73	0,50	0,82

Table 3. ML prediction performance for MCI subjects converting to AD (cMCI), following age stratification. The tables depict the ML ability to classify subjects following age stratification as indicated. Measurements of predictivity, accuracy, sensitivity and specificity refer to the ML performance obtained from the test-dataset. Please, note that, in both cases, the predictivity values does not reach the 85% threshold value. The ranking is is based on the accuracy score.

Abbreviations: NEURO, neuropsychological tests; EDU, years of education; TRI_COL, Triglycerides and cholesterol; ENER_SUB, energetic substrates; SYST_IND, systemic indices; AMMINO, amino acids; METABOL, lipoproteins, glucogenesis-related metabolites and amino acids; CSF, cerebrospinal fluid proteins; VIT_SIG, vital signs; ApoE ε4, alipoprotein E ε4 allele; BIL_ACID, bile acids; MRI, magnetic resonance imaging.

ncMCI over 70	Trees number	Predictivity	Accuracy	Sensitivity	Specificity
MRI+METABOL	300	85,5%	0,98	1,00	0,95
MRI+ VIT_SIG	300	82,1%	0,98	1,00	0,95
MRI+SYST_IND	350	86,30%	0,95	1,00	0,90
TRI_COL	250	81,40%	0,93	1,00	0,85
ENER_SUB	250	81,90%	0,93	1,00	0,85
AMMINO	200	73,90%	0,93	1,00	0,85
NEURO	500	84,6%	0,89	0,91	0,84
NEURO+EDU	200	85,1%	0,89	0,91	0,84
SYST_IND	250	77,10%	0,88	0,91	0,85
NEURO+CSF	300	81,8%	0,88	0,96	0,70
NEURO+CSF+ApoEε4	200	80,1%	0,88	0,92	0,80
NEURO+ApoEε4	350	86,6%	0,87	0,91	0,79
CSF+ VIT_SIG	250	74,4%	0,83	0,87	0,75
NEURO+VIT_SIG	400	85,0%	0,81	0,91	0,65
NEURO+BIL_ACID	300	87,0%	0,77	0,89	0,50
CSF+METABOL	300	76,9%	0,76	0,83	0,56
MRI+NEURO+CSF	300	87,8%	0,74	0,88	0,50

ncMCI under 70	Trees number	Predictivity	Accuracy	Sensitivity	Specificity
NEURO	400	90,2%	1,00	1,00	1,00
NEURO+EDU	200	88,0%	1,00	1,00	1,00
NEURO+MRI	350	86,0%	0,95	0,93	1,00
MRI+NEURO+CSF	300	89,5%	0,93	1,00	0,50
MRI	400	87,3%	0,90	1,00	0,67
CSF+ VIT_SIG	350	85,5%	0,88	0,86	1,00
CSF+SYST_IND	200	87,90%	0,88	1,00	0,00
CSF+AMMINO	200	89,40%	0,88	1,00	0,00
MRI+CSF	250	89,5%	0,86	1,00	0,00
NEURO+VIT_SIG	300	90,0%	0,78	0,94	0,43
NEURO+CSF	300	94,6%	0,76	0,89	0,63
NEURO+CSF+ApoEε4	200	91,9%	0,76	0,89	0,63
TRI_COL	250	91%	0,73	1,00	0,00
ENER_SUB	250	93,30%	0,73	1	0
SYST_IND	250	86,60%	0,73	1	0
AMMINO	200	86,60%	0,73	1	0
MRI+TRI_COL	300	12,80%	73,33	0,82	0,50

Table 4. ML prediction performance for MCI subjects non-converting to AD (ncMCI), following age stratification. The tables depict the ML ability to classify subjects following age stratification as indicated. Measurements of predictivity, accuracy, sensitivity and specificity refer to the ML performance obtained from the test-dataset. Please, note that, unlike cMCI (Table 3), the predictivity value exceed the 85% threshold of 85%, with ML performance for ncMCI subjects <70 years showing the highest predictivity scores. The ranking is based on the accuracy score.

Abbreviations: NEURO, neuropsychological tests; EDU, years of education; TRI_COL, Triglycerides and cholesterol; ENER_SUB, energetic substrates; SYST_IND, systemic indices; AMMINO, amino acids; METABOL, lipoproteins, glucogenesis-related metabolites and amino acids; CSF, cerebrospinal fluid proteins; VIT_SIG, vital signs; ApoE ε4, alipoprotein E ε4 allele; BIL_ACID, bile acids; MRI, magnetic resonance imaging.