A Machine Learning-Based Holistic Approach to Predict the Clinical Course of Patients within the Alzheimer’s Disease Spectrum

Noemi Massetta\textsuperscript{a,b,1}, Mirella Russo\textsuperscript{a,b,1}, Raffaella Franciotti\textsuperscript{b,1}, Davide Nardin\textsuperscript{c}, Giorgio Mandolini\textsuperscript{c}, Alberto Granzotto\textsuperscript{a,b,d}, Manuela Bomba\textsuperscript{a,b}, Stefano Delli Pizzi\textsuperscript{a,b}, Alessandra Mosca\textsuperscript{a,b}, Reinhold Scherer\textsuperscript{c}, Marco Onofri\textsuperscript{a,b} and Stefano L. Sensi\textsuperscript{a,b,f,}\textsuperscript{*} for the Alzheimer’s Disease Neuroimaging Initiative (ADNI)\textsuperscript{2} and the Alzheimer’s Disease Metabolomics Consortium (ADMC)\textsuperscript{3}

\textsuperscript{1}Center for Advanced Studies and Technology – CAST, University G. d’Annunzio of Chieti-Pescara, Italy
\textsuperscript{2}Department of Neuroscience, Imaging, and Clinical Sciences, University G. d’Annunzio of Chieti-Pescara, Italy
\textsuperscript{3}Biomedical Unit, ASC 27 s.r.l., Rome, Italy
\textsuperscript{4}Sue and Bill Gross Stem Cell Research Center, University of California - Irvine, Irvine, CA, USA
\textsuperscript{5}Brain-Computer Interfaces and Neural Engineering Laboratory, School of Computer Science and Electronic Engineering, University of Essex, Colchester, United Kingdom
\textsuperscript{6}Institute for Mind Impairments and Neurological Disorders – iMIND, University of California - Irvine, Irvine, CA, USA

Accepted 24 November 2021

Abstract.

Background: Alzheimer’s disease (AD) is a neurodegenerative condition driven by multifactorial etiology. Mild cognitive impairment (MCI) is a transitional condition between healthy aging and dementia. No reliable biomarkers are available to predict the conversion from MCI to AD.

Objective: To evaluate the use of machine learning (ML) on a wealth of data offered by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) and Alzheimer’s Disease Metabolomics Consortium (ADMC) database in the prediction of the MCI to AD conversion.

Methods: We implemented an ML-based Random Forest (RF) algorithm to predict conversion from MCI to AD. Data related to the study population (587 MCI subjects) were analyzed by RF as separate or combined features and assessed for classification power. Four classes of variables were considered: neuropsychological test scores, AD-related cerebrospinal fluid (CSF) biomarkers, peripheral biomarkers, and structural magnetic resonance imaging (MRI) variables.

\cite{1These authors contributed equally to this work.}

\cite{2Data used in preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (http://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}

\cite{3Data used in preparation of this article were generated by the Alzheimer’s Disease Metabolomics Consortium (ADMC). As such, the investigators within the ADMC provided data but did not participate in analysis or writing of this report. A complete listing of ADMC investigators can be found at: https://sites.duke.edu/adnimetab/team/}

\cite{Correspondence to: Prof. Stefano L. Sensi, Center for Advanced Studies and Technology – CAST, 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.}
Results: The ML-based algorithm exhibited 86% accuracy in predicting the AD conversion of MCI subjects. When assessing the features that helped the most, neuropsychological test scores, MRI data, and CSF biomarkers were the most relevant in the MCI to AD prediction. Peripheral parameters were effective when employed in association with neuropsychological test scores. Age and sex differences modulated the prediction accuracy. AD conversion was more effectively predicted in females and younger subjects.

Conclusion: Our findings support the notion that AD-related neurodegenerative processes result from the concerted activity of multiple pathological mechanisms and factors that act inside and outside the brain and are dynamically affected by age and sex.

Keywords: Alzheimer’s disease, conversion, dementia, machine learning, mild cognitive impairment, random forest

INTRODUCTION

Alzheimer’s disease (AD) is one of the most prevalent causes of early-onset dementia [1]. Clinical and epidemiological evidence indicate that AD-related pathological changes occur decades before the onset of clinical symptoms [2–4]. Mild cognitive impairment (MCI) is a critical prodromal phase of AD that offers a window of opportunity for therapeutic intervention [5, 6]. A few highly debated disease-modifying options are becoming available [7–12]. On the other hand, a growing body of evidence shows that prevention strategies may delay AD onset and progression [13–21]. Therefore, the development of cost-effective approaches to identify MCI subjects at risk of conversion to dementia and who will benefit from early therapeutic intervention is paramount.

To date, the clinical identification of the MCI stage has been achieved through the combined implementation of neuropsychological tests, the use of brain magnetic resonance imaging (MRI) scans, and the evaluation of altered levels of AD-related proteins [(i.e., amyloid-β and tau in the cerebrospinal fluid (CSF) or brain parenchyma)] [5, 6, 22].

Machine learning (ML) is a computer science field that provides computational tools to perform automated data classification and generate event predictions. ML is finding a variety of applications in medicine and neurology [23, 24]. Applied to dementia, the approach can help capture the complex molecular interactions of pathogenic events that occur in the early AD stages and/or facilitate disease progression [24, 25]. For instance, ML, fed with MRI data relative to subtle structural brain changes, has successfully helped unravel the disease continuum that spans from brain aging to AD via MCI [26–30]. Accuracy higher than 80% has also been achieved by employing multimodal approaches that combine the computation of detailed MRI-based measurements, the analysis of brain or CSF alterations of amyloid-β and tau levels, neuropsychological and behavioral tests, and dementia-related omics [31–36].

The use of such a wide array of biomarkers has been mainly limited to changes occurring within the central nervous system (CNS). However, promising alternative diagnostic venues are now offered by using systems medicine and network-based approaches and evaluating peripheral and systemic changes [37–40]. The implementation of this holistic strategy relies on the notion that chronic diseases, including dementia, are likely the result of converging perturbations of complex intra- and intercellular networks as well as alterations that occur at many levels and are not limited to one organ or driven by a single molecular factor or pathogenic mechanism [41–46].

Moving from this conceptual framework, we have employed an ML-based approach to identify, in a cohort of 587 MCI subjects, individuals more prone to convert to dementia. To that aim by taking advantage of the wealth of data that reflect pathogenic events occurring inside as well as outside of the CNS. The study evaluated data obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database and implemented an ML-based Random Forest (RF) algorithm [47].

METHODS

Data used in the preparation of this article were obtained from the ADNI database (http://adni.loni.usc.edu). ADNI is a public-private repository of clinical, imaging, genetic, and biochemical biomarker data obtained from North American subjects or patients (http://www.adni-info.org). ADNI aims to identify the determinant processes leading to AD and diagnose pathological changes occurring at
the earliest stage. All ADNI data collected at baseline were downloaded and managed with custom-made R-written codes.

Subjects

Subjects considered in this study were patients diagnosed with MCI extracted from the cohorts of ADNI-1, ADNI-GO, ADNI-2, and ADNI-3. The inclusion criteria were the ones provided by the ADNI protocol. Thus, all subjects were classified as MCI based on memory deficits but the relative preservation of other cognitive domains and maintained autonomy in the activities of daily living (http://adni.loni.usc.edu/study-design). To be included in the analysis, the subjects need to have completed all the baseline neuropsychological assessments. Subjects were followed for at least 36 months. The timeframe was chosen considering that MCI subjects have a high probability of converting to AD within 30 months [48].

All the variables included in the database were grouped into four classes: psychometric features, MRI-related data, AD-related biomarkers, and peripheral biomarkers.

Psychometric variables

Psychometric variables included neuropsychological test scores. For each subject, sixteen neuropsychological tests were employed to assess the status of different cognitive domains. The neuropsychological dataset included the Alzheimer’s Disease Assessment Scale-Cognitive (ADAS-Cog), subscales used to evaluate the severity of memory, learning, language (production and comprehension), praxis, and orientation deficits [49, 50]; the Mini-Mental State Examination [51] used to assess global cognition; the 30-item Boston Naming Test [52] and the Animal Fluency [53] to evaluate semantic memory and language abilities; the Functional Activities Questionnaire (FAQ) for the assessment of daily living activities [54]; the Rey Auditory Verbal Learning Test and Logical Memory II, subscales of the Wechsler Memory Scale-Revised (WMS-R) to investigate recall and recognition [55, 56]; the Trail Making Test [57], 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 [58]; the Clinical Dementia Rating Scale to quantify the patients’ severity of cognitive impairment related to the autonomy in daily living activities [59].

Table 1 summarizes the domains and cognitive functions investigated by each test.

AD-related biomarkers

AD-related biomarkers included CSF levels of amyloid-β peptide 1–42 (Aβ42), total-Tau (t-Tau), phosphorylated-Tau (p-Tau), and p-Tau/Aβ42 ratio. The APOE ε4 genotype [60] was included. The procedures of acquisition, stock, processing, and analysis of the biospecimens are available online (see http://adni.loni.usc.edu/methods/documents/).

Peripheral biomarkers

Peripheral biomarkers were obtained from the human plasma and serum. Supplementary Table 2 shows all the biospecimens 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 considering the number of samples and the consistency of measurements within the 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.

MRI variables

MRI variables included cortical thickness values and normalized volumes of relevant deep structures, as shown in Supplementary Table 3. Specifically, the MRI data downloaded from the ADNI database (Image Collections, http://adni.loni.usc.edu) were acquired with a Philips 3T scanner (see details at http://adni.loni.usc.edu/wp-content/uploads/2010/05/ADNI2_MRI_Training_Manual_FINAL.pdf), thereby limiting bias and technical issues related to the use of different scanner types or brands. T1-weighted images were acquired using 3D Turbo Field-Echo sequences (slice thickness = 1.2 mm; repetition time/echo time = 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 [61]. The volumes of the brain regions, computed with asegestats2table, 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.
ML analysis

Our ML approach used an RF algorithm as implemented by the scikit-learn library [62] written in Python. The RF is a supervised non-linear classifier. Its operation is based on the construction of binary decision trees obtained with the Bagging sampling method (an acronym for bootstrap aggregating) [63]. This model was chosen due to its robust performance and stability over an extensive range of parameters. Furthermore, the model is independent of the distribution of data and exhibits significant multi-class and advanced data-mining capabilities [64].

During the training phase, the algorithm explored the non-linear interactions between ADNI variables (or features) of the study subjects divided into two classes: individuals who converted to AD during the follow-up (cMCI) or not (ncMCI). The goal at this stage was to identify the best subdivision/classification strategy.

In the training phase, the RF analyzed 85% of the dataset’s subjects (who were randomly extracted). We used grid search and random search as hyperparameters optimization techniques [65]. Specifically, we focused on the number of trees, the depth of each tree, the number of samples for leaf, and the number of variables. Once the training phase was completed, we assessed feature importance to understand the role of each variable in the production of the classification and decision process. After the training, we entered the testing phase, and the RF strategy was applied to the remaining 15% of the dataset.

After a global analysis of the entire sample of MCI patients, the cohort was divided into four groups according to age quartiles (age brackets: 55–68, 69–74, 75–78, 79–88 years old). The RF was then repeated on the four groups separately. Differences due to sex were evaluated by analyzing separately male and female subjects.

RF performance in classifying cMCI and ncMCI subjects was assessed by taking into account accuracy values (ACC), positive predictive values (PPV), negative predictive values (NPV), sensitivity, and specificity.

RESULTS

Demographics and baseline data

Of the overall sample of 587 MCI patients, 236 (40%) converted to AD (cMCI) within the 36-month follow-up. Of these, 42% were males, and the mean age was 74.0 ± 7.1 years. The remaining 351 (39% males, mean age 72.2 ± 7.4 years) remained clinically stable (ncMCI). The demographics and baseline data of the study cohort are summarized in Table 1.

Global analysis

The use of RF allows the analysis of the features that offer the best predictive power. In our study, the RF-related features that had the higher impact in helping to identify cMCI subjects were psychometric data in combination with AD-related biomarkers (ACC = 0.86, sensitivity = 0.73 and specificity = 0.93) or MRI parameters (ACC = 0.83, sensitivity = 0.70 and specificity = 0.93) (Table 2). The combined use of AD biomarkers and MRI data also generated good accuracy (ACC = 0.81, sensitivity = 0.69 and specificity = 0.89).

Furthermore, on a ranking scale, psychometric variables at baseline were the most accurate classifiers (ACC = 0.80, sensitivity = 0.81 and specificity = 0.79), followed by MRI-related data (ACC = 0.75, sensitivity = 0.64 and specificity = 0.85) and AD-related biomarkers (ACC = 0.70, sensitivity = 0.77).
Sex stratification

Finally, we investigated sex differences in the predictive performance of the algorithm. As shown in Table 4, the accuracy was higher in female subjects. Differences in RF accuracy were modest for some classes (i.e., MRI data, AD-related biomarkers, psychometric scores). They became more robust in the case of peripheral biomarkers (ACC = 0.73 for females versus 0.57 for males). When considering the order of importance (Fig. 3), higher anatomical and functional relevance were observed for frontal lobe-related data (i.e., MRI and TRAIL-B scores) of male patients. RF also showed differences in peripheral biomarker relevance (Fig. 3). In that respect, glutamine was the most significant variable in both groups. Sex-related differences emerged. HDL cholesterol and butyrate were more helpful in predicting the conversion process of females, while pyruvate was most helpful in male subjects. BA levels were highly relevant in both groups.

### DISCUSSION

This study investigated which combination of ADNI-related data was the most effective for predicting the MCI conversion to dementia. To that aim, we took into account neuropsychological test scores, CSF levels of AD-related proteins, detailed structural MRI features, and peripheral biomarkers (Table 2). The ADNI database has been used by many authors to classify patients using ML algorithms [66–71]. In line
Fig. 1. (Continued)
Fig. 1. Global analysis. Features importance obtained in the Random Forest performed on the train dataset (85% of MCI subjects). The figure contains the classes which showed an accuracy value greater or equal to 0.80 in the test dataset (i.e., psychometric tests, AD-related biomarkers, structural MRI and peripheral biomarker, see Table 2). For each class, the histograms depict the weight, or importance, of each feature in the training phase of the machine learning. The importance scores range from 0 to 1, with higher values indicating greater weight in the classification process. AD biomarkers, Alzheimer’s disease-related biomarkers including cerebrospinal fluid biomarkers of neurodegeneration + APOE e4; MCI, mild cognitive impairment; MRI, magnetic resonance imaging; Peripheral biomarkers, amino acids + bile acids + energetic substrates + purines + systemic indices + triglycerides and cholesterol; Psychometric, neuropsychological tests. See Supplementary Tables for detailed variables enclosed in each category.

Table 3
Random Forest (RF) prediction performance for MCI conversion to AD within 36 months, after age stratification. The table depicts the RF ability to correctly classify converter and non-converter MCI subjects (cMCI and ncMCI, respectively) in the test dataset (15% of the total sample size), after the division of the whole cohort in four age quartiles. The ranking is based on accuracy values.

<table>
<thead>
<tr>
<th>Age</th>
<th>Accuracy</th>
<th>PPV</th>
<th>NPV</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Total sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychometric</td>
<td>55–68</td>
<td>0.86</td>
<td>0.86</td>
<td>86.7</td>
<td>0.75</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>69–74</td>
<td>0.81</td>
<td>0.63</td>
<td>88.9</td>
<td>0.71</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>75–78</td>
<td>0.72</td>
<td>0.40</td>
<td>84.6</td>
<td>0.50</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>79–88</td>
<td>0.70</td>
<td>0.60</td>
<td>76.9</td>
<td>0.67</td>
<td>0.71</td>
</tr>
<tr>
<td>MRI</td>
<td>55–68</td>
<td>0.85</td>
<td>1.00</td>
<td>0.83</td>
<td>0.33</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>69–74</td>
<td>0.77</td>
<td>0.78</td>
<td>0.75</td>
<td>0.88</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>75–78</td>
<td>0.80</td>
<td>1.00</td>
<td>0.71</td>
<td>0.60</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>79–88</td>
<td>0.77</td>
<td>1.00</td>
<td>0.57</td>
<td>0.67</td>
<td>1.00</td>
</tr>
<tr>
<td>Peripheral biomarkers</td>
<td>55–68</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.75</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>69–74</td>
<td>0.75</td>
<td>0.57</td>
<td>1.00</td>
<td>0.62</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>75–78</td>
<td>0.62</td>
<td>0.67</td>
<td>0.50</td>
<td>0.80</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>79–88</td>
<td>0.53</td>
<td>0.50</td>
<td>0.67</td>
<td>0.86</td>
<td>0.25</td>
</tr>
<tr>
<td>AD-related biomarkers</td>
<td>55–68</td>
<td>0.84</td>
<td>1.00</td>
<td>0.83</td>
<td>0.25</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>69–74</td>
<td>0.72</td>
<td>0.43</td>
<td>0.91</td>
<td>0.75</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>75–78</td>
<td>0.71</td>
<td>0.80</td>
<td>0.67</td>
<td>0.57</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>79–88</td>
<td>0.71</td>
<td>0.67</td>
<td>0.80</td>
<td>0.86</td>
<td>0.57</td>
</tr>
<tr>
<td>Psychometric + AD-related biomarkers</td>
<td>55–68</td>
<td>0.94</td>
<td>1.00</td>
<td>0.94</td>
<td>0.75</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>69–74</td>
<td>0.89</td>
<td>0.57</td>
<td>1.00</td>
<td>0.79</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>75–78</td>
<td>0.85</td>
<td>1.00</td>
<td>0.78</td>
<td>0.71</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>79–88</td>
<td>0.85</td>
<td>1.00</td>
<td>0.78</td>
<td>0.71</td>
<td>1.00</td>
</tr>
<tr>
<td>Psychometric + MRI</td>
<td>55–68</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.75</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>69–74</td>
<td>0.84</td>
<td>0.88</td>
<td>0.80</td>
<td>0.88</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>75–78</td>
<td>0.90</td>
<td>1.00</td>
<td>0.83</td>
<td>0.80</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>79–88</td>
<td>0.77</td>
<td>0.88</td>
<td>0.60</td>
<td>0.78</td>
<td>0.75</td>
</tr>
<tr>
<td>Psychometric + peripheral biomarkers</td>
<td>55–68</td>
<td>0.86</td>
<td>0.75</td>
<td>1.00</td>
<td>0.75</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>69–74</td>
<td>0.50</td>
<td>0.38</td>
<td>0.75</td>
<td>0.75</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>75–78</td>
<td>0.87</td>
<td>0.83</td>
<td>1.00</td>
<td>0.67</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>79–88</td>
<td>0.80</td>
<td>0.70</td>
<td>1.00</td>
<td>1.00</td>
<td>0.62</td>
</tr>
</tbody>
</table>

Measurements of accuracy, predictive values, sensitivity, and specificity refer to performances obtained from the test dataset. AD, Alzheimer’s disease; AD-related biomarkers, CSF biomarkers of neurodegeneration + APOE e4; MCI, mild cognitive impairment; MRI, magnetic resonance imaging biomarkers; NPV, Negative Predictive Value; Peripheral biomarkers, amino acids + bile acids + energetic substrates + purines + systemic indices + triglycerides and cholesterol; PPV, Positive Predictive Value; Psychometric, neuropsychological tests. See Supplementary Tables for detailed variables enclosed in each category.

with our study, some studies had used an RF-based classification strategy on structural MRI features [67, 68]. However, contrary to our study, these single-modality reports had used, in the training phase, mixed cohorts of healthy controls, ncMCI/cMCI and AD subjects [67, 68]. Conversely, we employed a multimodal approach and embraced a holistic viewpoint of the disease. Our prediction model supports the notion of neurodegenerative processes as the converging point of pathological processes occurring inside and outside the brain, factors also affected by age and sex-related factors.

ML is a powerful tool that significantly helps the diagnostic and therapeutic process, but care should be applied to maximize its heuristic power [24, 26–29, 31–35]. Applied to AD, evidence indicates that ML performances are greatly influenced by the time extent of the conversion process. Indeed a recent systematic review [72] assessing ML approaches employed to predict the conversion to AD of MCI...
Fig. 2. Age stratification. Features importance for psychometric tests obtained in the Random Forest performed on the train dataset (85% of MCI subjects). The figure shows the results of the whole cohort stratification according to four age quartiles. For each age bracket, the histograms depict the weight or importance, of the psychometric tests’ features in the training phase of the machine learning. The importance scores range from 0 to 1, with higher values indicating greater weight in the classification process. See Supplementary Table 2 for detailed variables enclosed in the Psychometric category.

Table 4
Random Forest (RF) prediction performance for MCI conversion to AD within 36 months, after sex stratification. The table depicts the RF ability to correctly classify converter and non-converter MCI subjects (cMCI and ncMCI, respectively) in the test dataset (15% of the total sample size), after the division of the whole cohort in two groups (male and female subjects). The ranking is based on accuracy values.

<table>
<thead>
<tr>
<th>Sex</th>
<th>Accuracy</th>
<th>PPV</th>
<th>NPV</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Total sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychometric</td>
<td>Female</td>
<td>0.86</td>
<td>0.86</td>
<td>0.87</td>
<td>0.75</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>0.81</td>
<td>0.63</td>
<td>0.89</td>
<td>0.71</td>
<td>0.84</td>
</tr>
<tr>
<td>MRI</td>
<td>Female</td>
<td>0.79</td>
<td>0.57</td>
<td>0.92</td>
<td>0.80</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>0.73</td>
<td>0.70</td>
<td>0.75</td>
<td>0.58</td>
<td>0.83</td>
</tr>
<tr>
<td>Peripheral biomarkers</td>
<td>Female</td>
<td>0.73</td>
<td>0.71</td>
<td>1.00</td>
<td>1.00</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>0.57</td>
<td>0.47</td>
<td>0.78</td>
<td>0.80</td>
<td>0.44</td>
</tr>
<tr>
<td>AD-related biomarkers</td>
<td>Female</td>
<td>0.81</td>
<td>0.64</td>
<td>1.00</td>
<td>1.00</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>0.79</td>
<td>0.83</td>
<td>0.77</td>
<td>0.62</td>
<td>0.91</td>
</tr>
<tr>
<td>Psychometric + AD-related biomarkers</td>
<td>Female</td>
<td>0.89</td>
<td>0.80</td>
<td>0.94</td>
<td>0.89</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>0.87</td>
<td>0.92</td>
<td>0.84</td>
<td>0.75</td>
<td>0.95</td>
</tr>
<tr>
<td>Psychometric + MRI</td>
<td>Female</td>
<td>0.95</td>
<td>1.00</td>
<td>0.93</td>
<td>0.80</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>0.80</td>
<td>0.73</td>
<td>0.87</td>
<td>0.85</td>
<td>0.76</td>
</tr>
<tr>
<td>Psychometric + Peripheral biomarkers</td>
<td>Female</td>
<td>0.87</td>
<td>0.83</td>
<td>1.00</td>
<td>1.00</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>0.58</td>
<td>0.47</td>
<td>0.78</td>
<td>0.8</td>
<td>0.44</td>
</tr>
</tbody>
</table>

Measurements of accuracy, predictive values, sensitivity, and specificity refer to performances obtained from the test dataset. AD, Alzheimer’s disease; AD-related biomarkers, CSF biomarkers of neurodegeneration + APOE e4; MCI, mild cognitive impairment; MRI, magnetic resonance imaging biomarkers; NPV, Negative Predictive Value; Peripheral biomarkers, amino acids + bile acids + energetic substrates + purines + systemic indices + triglycerides and cholesterol; PPV, Positive Predictive Value; Psychometric, neuropsychological tests; See Supplementary Tables for detailed variables enclosed in each category.
Fig. 3. Sex stratification. Features importance obtained in the Random Forest performed on the train dataset (85% of MCI subjects). The figure contains some classes shown in Table 4 (i.e., psychometric tests, structural MRI and peripheral biomarkers) which showed differences following sex stratification. For each class, the histograms depict the weight, or importance, of each feature in the training phase of the machine learning. The importance scores range from 0 to 1, with higher values indicating greater weight in the classification process. MCI, mild cognitive impairment; MRI, magnetic resonance imaging; Peripheral biomarkers, amino acids + bile acids + energetic substrates + purines + systemic indices + triglycerides and cholesterol; Psychometric, neuropsychological tests. See Supplementary Tables for detailed variables enclosed in each category.

Subjects indicates that optimal results can be produced with the implementation of a 3-year follow-up. The same review [72] suggested that the composition of the cohort should be carefully chosen accordingly to the ML-based approach that one is implementing.

In the final analysis, we employed longitudinal data to test the RF accuracy to predict AD progression, taking advantage of a dataset of MCI patients not previously used in the ML training phase. The analysis did not consider possible confounders like baseline comorbidities, ethnicity, lifestyle, living environment (i.e., urban versus rural areas), generating accuracy bias.

Combining baseline psychometric variables and AD-related biomarkers produced significant (>0.85) accuracy (Table 2). Overall, “classic” AD biomarkers
(i.e., psychometric test scores, CSF levels of AD-related biomarkers + APOE status, and brain MRI data) were the most accurate predictors for conversion.

Our RF-based approach indicated that, among psychometric data, verbal memory test scores, ADAS scales, and FAQ parameters were the most significant classifiers. It should be stressed that ADAS scales evaluate in great detail the overall cognitive status [73]. However, in routine clinical settings, the MMSE is preferred to the ADAS13 or 11 tests. Surprisingly, our RF found that MMSE scores were the least valuable classifiers. MMSE became relevant only after the age stratification of the cohort (as shown by Fig. 2). The different predictive weights of the two tests can be explained by their distinct score structure and overall purpose. The MMSE was created as an easy-to-use clinical tool, while the ADAS is more research-oriented [73]. The score range is also different, more granular (0–70 points) in the ADAS than the limited MMSE 30 points. Thus, the ADAS is more sensitive and specific and offers a more detailed scale of values to assess subtle cognitive abnormalities [74].

Our RF fed with CSF biomarker values and MRI data confirmed the higher relevance of the p-Tau/AB ratio and levels of temporal lobe atrophy (Fig. 1). These results are in line with a large body of evidence supporting the temporal lobe’s strategic role for memory-related tasks [75–78].

Sex-related analysis revealed that data relative to the atrophy of the medial orbital cortex were helping the predictive process only for the male group, thereby suggesting the presence of sex-related differences in the regional trajectories of the neurodegenerative processes [79, 80].

The combination of peripheral biomarkers and psychometric measures showed the same predictive power of psychometric test scores alone but exhibited greater sensitivity and predictive values (both positive and negative). Thus, one can speculate that, in the future, a matrix of peripheral biomarkers and neuropsychological tests may be employed as a first-line practical and cost-efficient way to facilitate the diagnostic process of the early stages of the disease. Among all peripheral biomarkers, variations of levels of glutamine, purine, lipids, and BA were the most significant feature to help the RF-based decision process (Fig. 1). The results are in accordance with findings based on graph modeling that suggest that glutamine is a central hub of metabolic imbalance in the context of dementia [81, 82]. Normal glutamate-glutamine cycling (GGC) plays a pivotal role in cognitive processes, as indicated by the presence of severely disrupted memory processes in hepatic encephalopathy (where high ammonium levels interfere with astrocytic GGC) [83]. Altered levels of glutamine have been frequently found in AD patients’ serum and CSF [84, 85]. The reduced activity of glutamine-synthase in AD patients has also been reported, a phenomenon deemed to impair the glutamate conversion to glutamine [81, 82].

On a speculative note, processes affecting glutamate accumulation in astrocytes [85] can concur to induce AD-related excitotoxicity [86–89]. At the same time, the imbalance of the glutamate-glutamine cycle may impinge on other AD-related alterations like the impaired γ-aminobutyric acid (GABA) synthesis or changes in anaplerotic reactions that generate mitochondrial bioenergetic dysfunctions [82].

Lipid and energy-related dysmetabolism have also been previously reported in AD patients [36, 90–92]. Altered blood [93] and brain levels of BA [94] have been described. Interestingly, these metabolites were found to be highly relevant to drive our RF-based predictive process. This intriguing finding is in line with a growing body of evidence supporting the presence of a gut-brain connection in neurodegeneration [95–100] and the role played by the liver in AD-related processes [96, 97]. The notion is also supported by a recent study indicating the association between altered BA profiles with higher degrees of brain atrophy, brain hypometabolism (as assessed by FDG-PET), and alterations of CSF AD-related biomarkers in AD patients [93]. These findings also agree with a study in which AD patients exhibited significantly low plasma levels of several medium-chain acylcarnitines [101]. These changes indicate underlying hepatic dysfunctions as most of the fatty acid oxidation, the mechanism that regulates acylcarnitine production [102] is controlled by the liver. Defective hepatic fatty acid oxidation impairs ketogenesis and produces lower levels of plasma ketones [103]. As ketones are the brain’s energy substrates alternative to glucose, the impairment of hepatic ketogenesis found in AD patients may exacerbate energetic brain deficits and be a critical aggravating factor in the disease progression. Interestingly, in preclinical AD models as well as in MCI or AD patients, ketogenic diets and/or pharmacologic manipulations set to favor ketogenesis have been shown to improve cognitive performances [104–108]. Given the high concentration of lipids within the CNS and the role played by these molecules in...
several neurodegenerative disorders, including AD [109–114], lipidomic-based approaches are becoming diagnostic tools of great potential. In that regard, further research on the interplay between lipid dysmetabolism and dementia should carefully consider sex differences, an emerging and promising area of investigation [80].

Little is known about the imbalance of the purine metabolic pathway in AD. A study indicated that compared to healthy subjects, AD patients exhibit increased serum levels of xanthosine. The study also found a significant correlation between high CSF levels of purine and t-tau [115]. Reduced levels of xanthosine have also been found in the entorhinal cortex of deceased AD patients [116].

To better understand the role of different disease modulators along with aging, we stratified the cohort into four age brackets and performed an ex-novo RF analysis. We found that the accuracy of all the classifiers was better in younger patients (Table 3).

These results support the notion that cognitive impairment in older patients results from the pathological convergence of multiple intermingled factors [117, 118].

Also, it should be emphasized that lipids acting as energy substrates may differently affect the fuel economy of the brain accordingly with pre-existing comorbidity (diabetes, metabolic syndrome, etc.). Thus, a current limitation of our study is the lack of information on such comorbidity in the investigated study subjects. Nevertheless, our results align with the general view that energetic changes are critical early biomarkers of the MCI-AD continuum even before the deposition of Aβ and expression of the cognitive decline [119, 120].

Finally, intriguing findings were generated in an RF analysis applied after dividing the cohort according to sex. Predictive performances were better in female patients (Table 4), and the most striking differences concerned the implementation of peripheral biomarkers (ACC = 0.73 for females versus 0.57 for males). In that respect, differences related to HDL cholesterol levels were more relevant to help the prediction process in women.

A potential limitation concerns differences in RF performances in the female sub-cohort. The better output in this group could be partially justified by the difference, when compared to males, in sample size and conversion rates per age bracket. These results nevertheless support the research endeavor on sex-related neurobiology of neurodegeneration [79, 80].

CONCLUSIONS

AD is a complex and multifactorial condition. The characterization of patients in a prodromal stage of the disease like MCI represents a challenge for biomedical research and unmet clinical and therapeutic needs.

A monumental effort in financial and human resources has been employed to reduce these aggregated proteins in the past thirty years. The rationale behind this strategy is that protein deposits are “toxic” and their physical disaggregation halts the neurodegenerative progression [121]. Except for a few highly debated clinical trials, the strategy has failed, thereby casting some fundamental doubts on the construct’s validity [122–126].

Our study, based on a multimodal approach, provides support for a holistic viewpoint of the disease. The valuable performance of our prediction model supports the notion of neurodegenerative processes as the converging point of pathological processes occurring inside and outside the brain that are also affected by age and sex-related factors.

ML techniques and big-data analysis can help identify novel and unexpected disease features and escape the dogmatic loop we are currently entrapped. For instance, a surprising finding of our study concerns the importance of peripheral biomarkers.

This set of combined systemic alterations is the gateway to precision medicine and offers fertile ground for innovative research. Precision medicine, systems medicine, and network-based approaches are in a position to generate tailored diagnoses, predict disease risks, and produce customized treatments that maximize safety and efficacy [43, 46, 79, 117, 118, 127].

Finally, a word of caution is needed when resting many diagnostic hopes in implementing AI-based approaches. A bottleneck in using many clinical parameters to be fed into ML is that most are phenotypic features with no precise alignment with underlying biology. Indeed, as recently suggested [128, 129], clinical phenotypes are considered the phenotypical mirror of distinct, specific, and unique underlying biological features. We believe that a reverse order of development and a switch from phenotypes to biotypes is required in precision medicine-based approaches to neurodegenerative conditions [129]. Indeed, AI-driven strategies may greatly help shift the attention from phenotypes to the importance of individualized biotypes.
In that vein, we hope our study helps further explore ML-based models set to unravel the complexity of neurodegenerative processes and dementia.

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 - GAANN Exploration to Evaluate Novel Alzheimer’s Queries (GEENA-Q-19-596282) (SLS). AG is supported by the European Union’s Horizon 2020 Research and Innovation Program under the Marie Sklodowska-Curie grant agreement iMIND—No. 841665.

Data collection and sharing for this project was funded by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbvie, Alzheimer’s Association; Alzheimer’s Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Luminosity; Lundbeck; Merek & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (http://www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer’s Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. Data collection and sharing for this project was also funded by the Alzheimer’s Disease Metabolomics Consortium (National Institute on Aging R01AG046171, RF1AG051550 and 3U01AG024904-09S4).

Authors’ disclosures available online (https://www.j-alz.com/manuscript-disclosures/21-0573r2).

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: https://dx.doi.org/10.3233/JAD-210573.

REFERENCES


