Edinburgh Research Explorer

Machine Learning models for detection and assessment of progression in Alzheimer's disease based on blood and cerebrospinal fluid biomarkers

Citation for published version:

Digital Object Identifier (DOI):
10.1109/EMBC40787.2023.10341203

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Proceedings of the 45th Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Machine Learning models for detection and assessment of progression in Alzheimer’s disease based on blood and cerebrospinal fluid biomarkers

Saturnino Luz1, Fasih Haider1,2,5 and Paul De Sousa3,4
for the Alzheimer’s Disease Neuroimaging Initiative*

Abstract—Machine-learning techniques were applied to human blood plasma and cerebrospinal fluid (CSF) biomarker data related to cognitive decline in Alzheimer’s Disease (AD) patients available via Alzheimer Disease Neuroimaging Initiative (ADNI) study. We observed the accuracy of AD diagnosis is greatest when protein biomarkers from cerebrospinal fluid are combined with plasma proteins using Support Vector Machines (SVM); this is not improved by adding age and sex. The area under the receiver operator characteristic (ROC) curve for our model of AD diagnosis based on a full (unbiased) set of plasma proteins was 0.94 in cross-validation and 0.82 on an external validation (test) set. Taking plasma in combination with CSF, the model reaches 0.98 area under the ROC curve on the test set. Accuracy of prediction of risk of mild cognitive impairment progressing to AD is the same for blood plasma biomarkers as for CSF and is not improved by combining them or adding age and sex as covariates.

Clinical relevance—The identification of accurate and cost-effective biomarkers to screen for risk of developing AD and monitoring its progression is crucial for improved understanding of its causes and stratification of patients for treatments under development. This paper demonstrates the feasibility of AD detection and prognosis based on blood plasma biomarkers.

I. INTRODUCTION

Alzheimer’s Disease (AD) is a prevalent neurological disease affecting ageing populations whose causes (i.e. aetiology) are complex and variable. The overwhelming majority incidence is sporadic and late onset. Pathological hallmarks include extra- and intra-neuronal proteinopathies in the form of amyloid-beta (Aβ) plaques and neurofibrillary tangles (NFT), neuronal dystrophy and loss, neuroinflammation involving astrogliosis, microglial migration and phagocytosis, and vascular alterations including compromised blood brain barrier permeability. The latter exposes the brain to peripheral immune challenges likely to exacerbate disease and confound treatment options. The aetiology and pathogenesis of AD continue to be subject of ongoing research and debate.

The amyloid cascade hypothesis, which posits that accumulation of aggregated forms of Aβ in the brain triggers or drives disease progression is now questioned as the exclusive cause or intervening link between pathophysiological phenotype and clinical manifestation of cognitive decline and dementia.

The identification of biomarkers to screen for risk of developing AD and monitoring its progression is a vital part of improved understanding of the causes of AD and stratification of patients for treatments under development [1]. A broad spectrum have been developed. Cerebrospinal fluid (CSF) and amyloid positron emission tomography (PET) imaging markers of Aβ and tau are highly accurate detecting AD associated pathophysiological and neuropathological changes. However, high cost, insufficient accessibility and invasiveness limit their use as first line tools for detecting patterns of pathophysiology. A multitstage, tiered approach is needed prioritizing development of an initial screen to exclude from these tests the high numbers of people with cognitive decline who do not demonstrate evidence of underlying AD pathophysiology.

Given the complex phenomenology of AD and other neurodegenerative diseases, clinical, neurological and neuropsychological exams are still an integral component of accurate late stage detection of clinically symptomatic individuals. However waiting times for these can be substantial and critical for patients. Memory clinics or general neurology clinics receive a broad range of referrals. Streamlining referrals to memory clinics could substantially benefit health care utilisation and costs. Cognitive exams are frequently administered, scored and interpreted incorrectly in primary care owing to a lack of training and expertise. A process that aids primary care practitioners in deciding which patients should receive a referral to a memory clinic would be of substantial benefit to specialists and GPs by decreasing numbers of unnecessary referrals and diagnostic procedures. To this end biomarker based diagnostics can aid multi-stage selection of patients for appropriate centres.

In this paper, we assess the predictive accuracy of machine learning methods to diagnose and predict risk of developing AD using human blood plasma and CSF biomarkers. In particular, we examine the potential of a blood biomarker panel containing 190 blood plasma proteins, in comparison and combination with CSF biomarkers. An accurate biomarker panel based solely on blood plasma could potentially contribute to a better understanding of AD pathophysiology and provide more cost-effective and accessible biomarkers for drug trials and clinical practice. Our results show that a
support vector machine (SVM) model based on a selection of 146 blood plasma biomarkers can reach diagnostic accuracy comparable to that of CSF biomarkers, and surpass CSF accuracy in a 4-year prognosis task. In diagnosis, combining CSF and blood plasma biomarkers in an SVM model results in an increase in the accuracy in relation to each set of biomarkers taken separately, reaching 0.985 area under the ROC curve (AUC).

A. Related research

Recent studies have demonstrated significant potential for blood plasma proteins to be used as cost effective biomarkers for AD in research and clinical practice [2], [3]. Plasma $A\beta$ biomarkers derived from an assay that combines immunoprecipitation techniques with mass spectrometry have been proposed which demonstrated high precision in predicting brain $A\beta$ burden in relation to established AD biomarkers such as CSF $A\beta_{1-42}$ and $A\beta$ PET [4].

Phosphorylated tau (pTau) forms can also be detected in blood plasma (as well as in CSF) at different sites through high-accuracy mass spectrometry and immunoassays [5], [6]. Neurofilament light chain protein (NFL) and glial fibrillary acid protein (GFAP) blood plasma based biomarkers have also been investigated [3].

The study involving machine learning models of blood plasma proteins that most closely resembles ours was conducted by Das et. al [7]. They used the same blood plasma assay we employed in our study, but selected 14 blood analytes found to be associated with AD pathology in the literature. This selected feature set was used to produce an interpretable grid of these protein levels that reached diagnostic accuracy of 76% [7].

II. METHODS

A. Data preparation

The data used in the preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database. Ethical approval was given by the local ethical committees of all involved sites of ADNI, and the research was conducted in accordance with the Helsinki Declaration. The ADNI was launched in 2003 as a public-private partnership, with the primary goal of testing whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer’s disease (AD). Additional information is found at www.adni-info.org.

We built the dataset used in the study presented in this paper by linking (merging) data from a number of ADNI studies. The data were divided into two main groups according to the machine learning task performed, namely: a) AD diagnosis, and b) prognosis, which we defined as a change in from the baseline examination to month 12.

The distribution of patients per diagnostic category is shown in Table I. The Blood dataset contains the Biomarkers Consortium Plasma Proteomics Project Rules-Based Medicine multiplex data (RBMPM), created by the ADNI project [8], [9]. We created models for the full set of plasma biomarkers (Blood), for a subset containing 14 plasma proteins reported in the literature as being linked to AD pathology [7], and for combined plasma and CSF features – in this case for a subset of the ADNI participants for whom both RBMPM and CSF data were available (Blood+CSF). The 14-protein data used in [7] is a subset of our Blood dataset. The number of patients per diagnostic (change in diagnostic in relation to the baseline) is show in the table at the bottom.

<table>
<thead>
<tr>
<th></th>
<th>AD</th>
<th>NC</th>
<th>MCI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood</td>
<td>158</td>
<td>61</td>
<td>275</td>
</tr>
<tr>
<td>Blood+CSF</td>
<td>106</td>
<td>113</td>
<td></td>
</tr>
</tbody>
</table>

These datasets were created using the ADNImerge package and R for data generation and missing value imputation. Variables that contained only missing features for the plasma were removed, leaving a total 146 blood protein features for classifier training. For the remaining data, missing values were imputed as the mean values for the corresponding features. In the full dataset there were only 8 such features that contained missing values. The resulting data were split into training, validation and test sets.

B. Classification models

The following classifiers were trained on the feature sets of plasma protein concentrations for those individuals for whom all 146 protein concentration levels were available or could be readily imputed, as described in the previous section. The data were split into training (70%) and test (30%) sets, with no overlap between these sets.

We employed the following classification methods: a) SVM, with an auto-scaled linear kernel, with cost parameter (“box constraint”) set to 0.5 [10], b) Decision Tree (DT) with minimum leaf size set to 3 [11], c) AdaBoost, using the M1 algorithm, with an ensemble of 100 weak tree learners [12], and d) K-nearest-neighbour (KNN), with k set to 5. This choice of classifiers aimed to compare typical machine learning approaches (linear, kernel, symbolic and instance-based) rather than necessarily obtain the best performance.

We performed cross-validation (CV) on the training set by partitioning it randomly into 10 subsets , and then selected the model that performed best in CV for testing on the test set. To assess the performance of different plasma biomarkers in relation to CSF biomarkers commonly used in research and clinical practice and to the set of 14 analytes previously employed in machine learning modelling [7], we built classifiers for each set of biomarkers separately. We also
created models for a combination of CSF and the full set of 146 plasma biomarkers.

III. RESULTS

A summary of results for the AD classification (i.e. diagnosis) experiments is shown in Table II, expressed in terms of AUC. Ten-fold CV results are reported for reference. Prediction accuracy is greatest for models that combine protein biomarkers from CSF and all 146 available in blood plasma concentration using SVM. Performance is not improved by adding age and sex as model parameters. The 14-protein model had high overall sensitivity (90.1%) but low specificity (59.5%) in CV, whereas these measures dropped to 79.2% and 57.9% on the test set. Our full plasma protein model improved considerably on these results, achieving 95.5% sensitivity and 83.3% specificity in CV, and 85.4% sensitivity and 73.7% specificity on the test set. Combining blood plasma with CSF parameter yielded only a small improvement over the plasma-only feature set, namely 95.9% sensitivity in CV, and 90.6 sensitivity on the held-out test set.

TABLE II
MODEL PERFORMANCE ON THE AD DIAGNOSIS TASK (AUC VALUES).

<table>
<thead>
<tr>
<th>Biomarkers</th>
<th>Method</th>
<th>Biomarkers only CV</th>
<th>Test</th>
<th>Biomarkers, age, sex CV</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>14 plasma</td>
<td>Adaboost</td>
<td>0.854</td>
<td>0.714</td>
<td>0.889</td>
<td>0.764</td>
</tr>
<tr>
<td>146 plasma</td>
<td>SVM</td>
<td>0.938</td>
<td>0.820</td>
<td>0.939</td>
<td>0.823</td>
</tr>
<tr>
<td>CSF</td>
<td>KNN</td>
<td>0.984</td>
<td>0.947</td>
<td>0.983</td>
<td>0.947</td>
</tr>
<tr>
<td>CSF+Plasma</td>
<td>SVM</td>
<td>0.986</td>
<td>0.985</td>
<td>0.985</td>
<td>0.941</td>
</tr>
</tbody>
</table>

The results of the AD prognosis task are summarised in Table III, for models containing biomarkers only, and for models that included age and sex as covariates. Contrary to diagnosis, the accuracy of prediction of the risk of mild cognitive decline progressing to AD was the same for blood plasma biomarkers as for CSF, and it was not improved by combining them or by adding age and sex as covariates. The prognosis models perform considerably more poorly than the diagnosis models. This is in part due to the scarcity of data on progression, and the class imbalance that characterise these data. For imbalanced classes, it is usual to report performance in terms of unweighted average recall (UAR), that is the arithmetic mean of sensitivity scores across all classes. In the case of prognosis we have 6 different categories of disease progression (including stability), as shown in Table I, bottom. Sensitivity scores were computed for each of these classes and summarised as UAR scores in Table III. Unsurprisingly, the highest values for sensitivity were among the stable categories, with MCI-MCI reaching 80% sensitivity for the 14-protein panel in CV and 63.9% in testing. For the full 146-blood plasma protein panel MCI stability prediction had 96.9% sensitivity in CV and 97.3% in testing. This performance was considerably better than CSF alone, which recorded 79.7% sensitivity in CV and 70.6% sensitivity on the test set for the MCI-MCI category. For this category, combining CSF and blood plasma proteins did not improve prediction; sensitivity was 96.2% in CV and 88.2% in testing. Performance on the prediction of MCI to AD conversion was remarkably worse than prediction of stability. The best blood plasma protein sensitivity was 18.2% in CV and only 15.8% on the test set. CSF underperformed blood plasma in CV, with sensitivity of 13% in CV, but outperformed in testing (40% sensitivity).

TABLE III
MODEL PERFORMANCE ON THE AD PROGNOSIS TASK (UAR VALUES).

<table>
<thead>
<tr>
<th>Biomarkers</th>
<th>Method</th>
<th>Biomarkers only CV</th>
<th>Test</th>
<th>Biomarkers, age, sex CV</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>14 plasma</td>
<td>DT</td>
<td>0.445</td>
<td>0.383</td>
<td>0.424</td>
<td>0.344</td>
</tr>
<tr>
<td>146 plasma</td>
<td>SVM</td>
<td>0.620</td>
<td>0.619</td>
<td>0.627</td>
<td>0.623</td>
</tr>
<tr>
<td>CSF</td>
<td>DT</td>
<td>0.622</td>
<td>0.607</td>
<td>0.607</td>
<td>0.607</td>
</tr>
<tr>
<td>CSF+Plasma</td>
<td>Adaboost</td>
<td>0.654</td>
<td>0.613</td>
<td>0.648</td>
<td>0.613</td>
</tr>
</tbody>
</table>

IV. DISCUSSION

Our diagnosis predictions employing machine learning models of all available RBMPM multiplex data (excluding protein concentrations with values missing for all participants) outperformed models built using a set 14 of proteins handpicked from the AD literature, and nearly matched the performance of the CSF protein panel, which is regarded by many as a gold standard in AD diagnosis.

Using all blood proteins yielded a 26% improvement in accuracy (88.9% for the full feature set) in comparison to the 14 feature set derived from the AD literature and used by Das et al [7] (70.1%). Notice that our results differ somewhat from the ones reported in [7] as we used a different, larger dataset. Our results suggest that there are blood plasma analytes that could contribute to AD diagnosis but have not yet been identified in biomedical research studies. If these improvements can be sustained in an interpretable model, the newly identified features may be useful in the formulation of research hypotheses for further experimental work. Age and sex do not seem to contribute much to diagnostic accuracy, possibly due to the fact that the distributions of these features over the datasets was fairly well balanced.

Nakamura et al. assessed the potential of the ratios of Aβ precursor protein (APP)_{60-71} to Aβ_{1-42} and Aβ_{1-40} to Aβ_{1-42}, as well as Aβ_{1-42} alone, and a composite biomarker set to predict individual brain Aβ status in two independent cross-sectional datasets, where Aβ PIB-PET was regarded as the gold standard for Aβ-positive or negative status [4]. They achieved high performance predicting Aβ status using the composite biomarker, ranging between 0.88 and 0.94 AUC for the validation set, to 0.96 in a “discovery” (cross-validation) set. However, this study did not attempt to differentiate between AD and non-AD individuals, as we do in the present study. In fact, Teunissen et Al. claim that the predictive potential of plasma Aβ has not been investigated in any of the studies featured in their recent review [3].

Tissot et Al. [13] examined how plasma pTau181 and pTau231 correlated to a PET imaging gold standard for Aβ positivity and found moderate statistically significant correlations between each biomarker and 18F-MK-6240 tau PET. They obtained AUROC of approximately 0.84 in Aβ status prediction for both pTau181 and pTau231, and
AUROC of 0.94 and 0.97 in distinguishing AD patients from unimpaired healthy young individuals for pTau181 and pTau231 respectively. Despite good overall concordance with tau PET, discrepancies were observed between individual pTau231 and pTau181 predictions, leading the authors to suggest that these biomarkers might be reflecting different stages of progression. However, they did not attempt to model progression, as we do in this study. Simrén et al. [14] investigated the predictive power of pTau181, comparing it to other blood based biomarkers and finding that it significantly outperformed them in AD diagnosis, with AUC of 0.91. Another study, assessed plasma pTau181 for AD diagnosis obtaining AUC between 0.90 and 0.98 against cognitively normal older adults, and good accuracy in distinguishing AD from other neurodegenerative disorders [5].

Meta-analyses suggest that low sensitivity low sensitivity has hindered the development of blood based NFL biomarkers [2]. However, recent high-sensitivity techniques have shown blood NFL to be a promising biomarker for AD diagnosis and for the assessment of disease severity [3]. Palmqvist et al. [6] investigated plasma pTau217, Aβ342/A340, NFL in isolation and in combination with imaging, APOE staus and cognitive testing. Using standard statistical modelling (logistic regression) and the Akaike information criterion for model selection they found that pTau217 exhibited an AUC of 0.83 on its own and 0.91 when combined with cognitive testing in predicting progression to AD within 4 years. Tested on an independent cohort (ADNI) the model yielded AUC of 0.90 using pTau181.

Prediction of progression proved a much more challenging task. In this task, our results ranged from high sensitivity for prediction of stability, but low sensitivity for prediction of conversion from MCI to AD, which is arguably the most relevant prediction. This is likely due to the small dataset available for model training, compounded by the relatively short interval between baseline and subsequent diagnoses.

V. CONCLUSION

Applying machine learning methods to model multi-parametric and modal biomarkers in a open, longitudinal cohort dataset, as we have done in this paper, provides a route to AD diagnosis which is as good or better than current and recent approaches, and trends to focus on single biomarkers. This also provides a route to unbiased discovery of interactomes augmenting understanding of the disease. In future work we intend to explore such avenues, and gather further data to improve our prognosis models towards clinical use standards.

ACKNOWLEDGMENT

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; Lumosity; Lundbeck; Merck & 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. 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.

REFERENCES