2016-08

Identification of blood biomarkers for use in point of care diagnosis tool for Alzheimer's disease

Jammeh, Emmanuel

10.1109/embc.2016.7591217
2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)
IEEE

Identification of blood biomarkers for use in point of care diagnosis tool for Alzheimer's disease

E. Jammeh, P. Zhao, C Carroll, S Pearson and E. Ifeachor

Abstract— **Early diagnosis of Alzheimer's Disease (AD) is widely regarded as necessary to allow treatment to be started before irreversible damage to the brain occur and for patients to benefit from new therapies as they become available. Low-cost point-of-care (PoC) diagnostic tools that can be used to routinely diagnose AD in its early stage would facilitate this, but such tools require reliable and accurate biomarkers. However, traditional biomarkers for AD use invasive cerebrospinal fluid (CSF) analysis and/or expensive neuroimaging techniques together with neuropsychological assessments. Blood-based PoC diagnostics tools may provide a more cost and time efficient way to assess AD to complement CSF and neuroimaging techniques. However, evidence to date suggests that only a panel of biomarkers would provide the diagnostic accuracy needed in clinical practice and that the number of biomarkers in such panels can be large. In addition, the biomarkers in a panel vary from study to study. These issues make it difficult to realise a PoC device for diagnosis of AD. An objective of this paper is to find an optimum number of blood biomarkers (in terms of number of biomarkers and sensitivity/specificity) that can be used in a handheld PoC device for AD diagnosis. We used the Alzheimer's disease Neuroimaging Initiative (ADNI) database to identify a small number of blood biomarkers for AD. We identified a 6-biomarker panel (which includes A1Micro, A2Macro, AAT, ApoE, complement C3 and PPP), which when used with age as covariate, was able to discriminate between AD patients and normal subjects with a sensitivity of 85.4% and specificity of 78.6%.**

Index Terms— AD, Machine Learning, Classification, ADNI.

I. Introduction

AD is an age related neurodegenerative disease [1] which leads to old age disability, institutionalisation and death within 10 years of clinical symptoms [2]. In the UK, it is estimated that 2 million people will suffer from dementia by 2050, with a total annual cost of care of £78 million [3]. There is currently no cure for AD but simple and inexpensive PoC diagnostic tools that can routinely diagnose AD in its early stage are widely regarded as necessary [4]. This would allow treatment to be started before irreversible damage to the brain occurs and enable patients to get maximum benefit from new therapies when they become available. Although traditional biomarkers for AD have reached clinical application, they may be difficult to use in handheld PoC diagnostic tools because they are based on invasive cerebrospinal fluid (CSF) analysis and/or expensive neuroimaging techniques together with neuropsychological assessments that take a long time [5]. Blood-based AD biomarkers may provide a more cost and time efficient way to assess AD to complement traditional biomarkers because obtaining blood samples is far less complex. Several studies have identified promising blood biomarkers for AD [6]–[12]. A recent review identified 163 candidate blood biomarkers for AD from 21 studies. These biomarkers had different concentration level in the blood of AD patients compared to normal elderly subjects [13]. However, there has been a failure to replicate many blood biomarker discovery studies due to many factors, including the use of different proteomic technologies, the use of different research cohorts, data over-fitting, and differences in sample collection methods [14]. Some studies have tried to identify blood biomarkers that are specific to a particular aspect of AD, e.g. cognitive decline [15] brain atrophy [12] and neocortical amyloid burden [6]. Yet despite much progress none of the identified blood biomarkers have been used in PoC tools for AD [16].

Despite the limitations, there is evidence from various studies to suggest that a panel of biomarkers would provide the diagnostic accuracy needed in clinical practice [17]-[21]. Each study identified different panels of different biomarkers with differing panel sizes. Furthermore, some of the identified panels are quite large making their realisation in PoC devices difficult. As yet there has not been any panel identified that can be used in a handheld PoC diagnostic tool for AD. However, technological advances in machine learning and computing power may enable the identification of an optimum number of biomarkers (in terms of the number of biomarkers and sensitivity/specificity) from proteomic datasets.

An objective of the research project on which this paper is based is to develop a novel handheld graphene-based PoC diagnostic tool for diagnosis of AD. In this study, the focus is to investigate validated blood-based biomarkers with a view to finding an optimal panel of blood biomarkers that could be used in such a tool. A literature review of blood biomarker studies was firstly conducted to identify the most validated blood biomarkers for AD. The concentrations of identified biomarkers were characterised for AD and controls using ADNI proteomic database. Univariate analysis of individual biomarkers was carried out to determine candidate biomarkers that may be included in biomarker panels. Machine learning was then used to different panels from candidate biomarkers with good diagnostic value.

The panel that had acceptable clinical performance with the least number of biomarkers was selected as the optimal panel. The study identified an optimal panel of 6 biomarkers panel consisting of α-1 Macroglobulin, α-2-macroglobulin, α-1 Antitrypsin, Apolipoprotein E, complement C3 and Pancreatic Polypeptide. The panel of biomarkers was able to discriminate between AD patients and normal elderly subjects with sensitivity and specificity values of 85.4% and 78.6%, respectively.

II. METHODS

The ADNI proteomic database (http://adni.loni.ucla.edu) was used in this study. More information about the ADNI proteomic dataset can be obtained from http://www.adniinfo.org. A literature review was carried out to identify the most cross-validated biomarkers for AD. The biomarkers identified from the literature review were extracted from the ADNI proteomic database for AD and controls subjects. The extracted data was then pre-processed to remove records in which no plasma concentrations were recorded. Any in which the MMSE score of any subject did not correlate with their dementia diagnosis status was also removed. The data contained concentrations of biomarkers identified from the literature for 106 AD patients and 51 controls. The mean age of AD patients is 74.88 and 74.56 for the normal elderly subjects. The average MMSE scores were 22.75 for AD patients and 28.36 for controls..

The concentration of each biomarker for all AD and controls and the age of subjects were extracted from the dataset. This was used to train supervised machine learning algorithms to learn to discriminate between AD and controls based only on age and concentrations of individual biomarker in blood of the subjects. The diagnostic value of each candidate biomarker to discriminate between AD patients and control subjects was measured in terms of the area under the receiver operating curve (AUC) [22][23]..

Candidate biomarkers that obtained an AUC of not less than 0.6 were considered as candidate biomarkers for inclusion in panels. Several machine learning based feature selection algorithms were used to search through the space of candidate biomarkers with an AUC of over 0.6. This process allowed the identification of feature subsets, or panels of biomarkers, that may have good diagnostic performance.

Each identified panel was used to train different machine learning classifiers to learn to discriminate AD from controls. We trained Naïve Bayes, Logistic, Multilayer Perceptron, Random Forest and Support Vector supervised learning algorithms. Naïve Bayes classifier had the best results. Naïve Bayes classifiers assume that the presence or absence of a feature associated with a class is not related to the absence or presence of other features within the class. It is suited for complex real world applications and has been effectively used in clinical decision support systems to identify undiagnosed dementia in primary care [24]. In a Naïve Bayes classifier,

$$p(c_j \mid d) = \frac{p(d \mid c_j)p(c_j)}{p(d)} \qquad (1)$$

where $p(c_j \mid d)$, is the probability that instance *d* belongs to class c_j, $p(c_j)$ is the probability of occurrence of class c_j or the relative prevalence of the class in the dataset, and $p(d)$ is the probability of instance d in the dataset.

All data analysis, feature selection and classification tasks were performed using Weka open source machine learning toolbox [22]. The performance of each panel to classify AD versus control subjects was evaluated using 10-fold cross validation training and testing methodology. The dataset is randomly divided into 10 sub-datasets. One is considered as new and unseen test data and the remaining sub datasets are considered as the pool of example data cases used to train a supervised machine learning algorithm. This process is carried out 10 times each time leaving out a different set that is used for testing. The performance is accessed using the sensitivity, specificity and AUC for the task of classifying AD and control subjects. The acceptable diagnostic accuracy that is required in clinical practice is a sensitivity and specificity greater than or equal to 80% [25]. However, even with moderate sensitivity and specificity, a PoC diagnostic tool for dementia may still be useful for routine screening of AD. Therefore, a threshold of 75% was set for sensitivity and specificity in determining acceptable panel of biomarkers performance. The smallest panel that achieved sensitivity and specificity values of over 75% was considered the optimal biomarker panel for use in the proposed handheld PoC diagnostic device for AD.

III. RESULTS

A review of the literature identified a total of 163 candidate blood biomarkers from 22 studies that have been validated at least twice. The concentration of these biomarkers for AD and control subjects were extracted from the ADNI database were analysed. Table 2 shows the concentration levels of some of these biomarkers in AD and control subjects that were identified from the review and which have been validated in at least three different studies. The data shows that the biomarkers are differently expressed in AD and controls. These biomarkers have been validated at least twice by different biomarker discovery studies. Three of these biomarkers, A2Macro (α2-Macroglobulin), Fibrinogen and CFH (complement factor H) have been shown to be associated with brain amyloid burden [6]. C3 (complement C3), fibrinogen and A1Micro (α1 microglobulin) have been shown by [26] to be strong predictors of AD pathology.

A total of 63 biomarkers gave an AUC of greater than 0.6 when used with age as covariate for the task of classifying AD and controls. Table 3 shows the results of the univariate analysis biomarkers that have the most diagnostic value in discriminating between AD and controls are shown. Eotaxin_3 (Eotaxin 3), IgM (Immunoglobulin), ApoE (Apolipoprotein E), A1Micro (Alpha-1 Microglobulin), PLGF (Placenta Growth Factor), PYY (Peptide YY) , PPP (Pancreatic Polypeptide), A2Macro, CRP (C-Reactive Protein) and EGF (Epidermal Growth Factor), with age as covariate, achieved the highest performance with an AUC of greater than 0.648. Table 1 shows the description of some of the biomarkers. Table 1 shows the description of some of the biomarkers.

Table 1: Description of some of the biomarkers

Biomarker	Name	Biomarker	Name
CLU	Clusterin	ANG2	Angiopoietin 2
AAT	Alpha-1 Antitrypsin	CD40	CD 40 Antigen
IL_3	Interleukin-3	Cortisol	Cortisol
IGFBP	Insulin-like Growth Factor-Binding Protein	C Peptide	C Peptide

Table 2: Characterisation of candidate biomarkers

Biomarker	Min (max)	Avg. (Std.)	Min (Max)	Avg. (Std.)
Eotaxin_3	1.544 (3.09)	2.618 (0.21)	1.544 (3.31)	2.383 (0.41)
IgM	0.509 (0.90)	0.187 (0.24)	-0.06 (0.90)	0.377 (0.23)
ApoE	1.23 (2.23)	1.712 (0.19)	1.505 (2.19)	1.855 (0.14)
A1Micro	0.875 (1.38)	1.115 (0.10)	0.839 (1.32)	1.036 (0.10)
PLGF	4.65 (77)	27.99 (15.0)	16 (61)	36.53 (10.4)
PYY	1.398 (3.076)	2.075 (0.32)	1.398 (2.42)	1.848 (0.30)
PPP	-0.004 (3.15)	2.149 (0.41)	1.23 (2.62)	1.969 (0.32)
A2Macro	-0.119 (0.45)	0.075 (0.09)	-0.137 (0.34)	0.025 (0.10)
CRP	-0.824 (1.71)	0.148 (0.63)	-0.77 (1.34)	0.315 (0.47)
EGF	0.176 (2.90)	1.532 (0.59)	0.602 (2.70)	1.688 (0.491
IGFBP	1.477 (2.19)	1.91 (0.11)	1.491 (2.40)	1.883 (0.19
IL_3	-2.495 (-1.06)	-1.723 (0.28)	-2.046 (-1.15)	-1.618(0.21)
B2M	0.114 (0.924)	0.346 (0.17)	0.079 (0.69)	0.29 (0.11)
Fibrinogen	0.007 (14)	6.769 (2.06)	3 (10)	6.41 (1.31)
ANG2	0.114 (0.954)	0.614 (0.12)	0.322 (1.07)	0.599 (0.14)
CD40	-0.367 (0.255)	-0.094 (0.12)	0.337 (0.14)	-0.111 (0.11)
Cortisol	1.982 (2.635)	2.207 (0.11)	1.869 (2.58)	2.174 (0.14)
C Peptide	-0.167 (0.74)	0.36 (0.18)	-0.046 (0.99)	0.374 (0.21)
C3	0.709 (0.46)	0.285 (0.08)	0.079 (0.46)	0.259 (0.07)
CFH	829 (7990)	4055 (1325)	1080 (6520)	3718 (1175)
CLU	2.35 (2.81)	2.516 (0.09)	2.356 (2.643)	2.502 (0.079
EGRF	2.5 (6.5)	4.515 (0.76)	2.5 (6)	4.517 (0.656)
AAT	0.079 (0.65)	0.47 (0.096)	0.204 (0.69)	0.443 (0.85)
Age	55 (89)	74.87 (8.052)	62 (90)	75.15 (5.78)
Education	4 (20)	15.116 (3.207)	8 (20)	15.67 (2.78)

Table 3: Identified panels of biomarkers

Biomarker	AUC	Biomarker	AUC
Eotaxin_3	0.731	B2M	0.644
IgM	0.727	Fibrinogen	0.639
ApoE	0.723	ANG2	0.638
A1Micro	0.717	CD40	0.638
PLGF	0.706	Cortisol	0.638
PYY	0.698	C Peptide	0.624
PPP	0.68	C3	0.621
A2Macro	0.678	CFH	0.605
CRP	0.675	CLU	0.602
EGF	0.661	EGRF	0.602
IGFBP	0.648	AAT	0.602
IL_3	0.648		

Table 4 shows four panels that were identified and their performance investigated. The panels have different sizes ranging from 4 to 16 and used different combinations of candidate biomarkers. ApoE was used in all panels and C3, CRP, IGM and PYY were used in three out of the four panels. A2Macro, AAT, CD40, CLU, Cortisol and IL_3 were used in two panels.

Table 4: Identified biomarker panels

Panel 1	Panel 2	Panel 3	Panel 4
ApoE	ApoE	ApoE	A1Micro
A2Macro	CRP	C3	A2Macro
AAT	IGM	CD40	AAT
ANG2	PYY	CLU	ApoE
B2M		Cortisol	C3
C3		CRP	PPP
CD40		IGM	
CFH		IL_3	
CLU		PYY	
Cortisol			
C_Peptide			
CRP			
IGM			
IL_3			
PYY			
VCAM			

Each panel of biomarkers was used to train a machine learning classifier. The performance of the biomarkers was evaluated in terms of the sensitivity, specificity, AUC and accuracy in classifying AD and controls. The results of evaluating the performance of the panels are presented in Table 5. The minimum sensitivity for all the panels was 0.68 and maximum was 0.854 and the minimum and maximum of the specificity was 0.75 and 0.796, respectively. The average sensitivity and specificity value was 0.77. Only two of the panels achieved a sensitivity and specificity of more than 75%. Panel 4 obtained the optimal performance of sensitivity value of 0.85, specificity of 0.78, an AUC value of 0.85 and an accuracy of classifying AD and controls of 83.6%.

Table 5: Biomarker panel performance

Panel	1	2	3	4
Size	16	4	9	6
Sensitivity	0.7	0.68	0.78	0.85
Specificity	0.7	0.75	0.79	0.78
AUC	0.8	0.82	0.83	0.85
Accuracy	76.3	70.3	78.9	83.6

The receiver operating curve for Panel 4 is shown in Figure 1. This panel has the smallest number of biomarkers and meets the specified performance threshold for use in the proposed handheld PoC tool for routine clinical use.

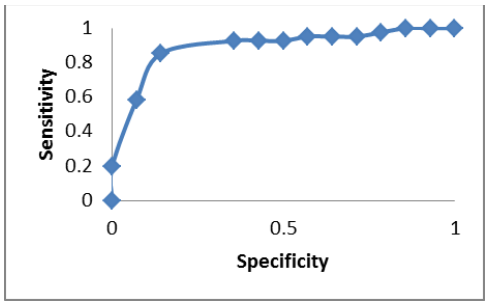

Figure 1: The receiver operating curve for 6-panel biomarker panel

IV. CONCLUSIONS AND DISCUSSIONS

This study investigated the determination of an optimum number of blood biomarkers that can be used in a diagnostic PoC tool for routine diagnosis of dementia from blood

samples in minutes. ADNI proteomic database was used to characterise the concentration of the identified biomarkers in AD and control patients. A univariate data analysis was performed to determine which of the identified biomarkers have significant diagnostic value for them to be potentially used in diagnoses of AD. These were then used in a multivariate data analysis to identify a combination of biomarkers which can be used as a panel of biomarkers that could be used to diagnose AD with clinically acceptable performance.

A panel of 6 biomarkers which consists of A1Micro, A2Macro, AAT, ApoE, C3 and PPP, together with subjects' age as covariate, was able to classify AD and control subjects with a sensitivity of 85.4% and specificity of 78.6%. This panel has a high enough performance to be used in a handheld PoC tool for dementia diagnosis. Results of this study are promising and suggest that, a panel of only 6 blood biomarkers may be used in a PoC tool for the diagnosis of AD based on blood samples. As future work, the results will be validated using a different proteomic dataset.

ACKNOWLEDGMENT

This paper presents independent research funded by the EPSRC (Grant Reference Number: EP/M006301/1. The views expressed in this article are those of the authors and not necessarily those of the EPSRC. The authors are grateful for the ADNI for the data used in this study.

REFERENCES

[1] M. Veitinger et al, "A platelet protein biochip rapidly detects an Alzheimer's disease-specific phenotype," *Acta Neuropathol.*, vol. 128, no. 5, pp. 665–677, 2014.

[2] E. Dodd et al, "An evaluation of primary care led dementia diagnostic services in Bristol.," *BMC Health Serv. Res.*, vol. 14, no. 1, p. 592, 2014.

[3] M. Prince et al , *Dementia UK: Update*. 2014.

[4] C. Laske et al, "Innovative diagnostic tools for early detection of Alzheimer's disease," *Alzheimer's Dement.*, vol. 11, no. 5, pp. 561–578, 2015.

[5] K. Henriksen et al, "The future of blood- based biomarkers for Alzheimer ' s disease Karsdal MA , 13 Weiner MW . For the Blood-Based Biomarker Interest Group (BBBIG). Abstract Treatment of Alzheimer ' s disease (AD) is significantly hampered by the lack of easily accessib."

[6] N. J. Ashton et al, "Blood protein predictors of brain amyloid for enrichment in clinical trials?," *Alzheimer's Dement. Diagnosis, Assess. Dis. Monit.*, vol. 1, no. 1, pp. 48–60, 2015.

[7] P.-C. Liao et al, "Proteomics analysis of plasma for potential biomarkers in the diagnosis of Alzheimer's disease.," *Proteomics. Clin. Appl.*, vol. 1, no. 5, pp. 506–512, 2007.

[8] M. G. Ravetti and P. Moscato, "Identification of a 5-protein biomarker molecular signature for predicting Alzheimer's disease," *PLoS One*, vol. 3, no. 9, 2008.

[9] H. D. Soares et al, "Identifying early markers of alzheimer's disease using quantitative multiplex proteomic immunoassay panels," *Ann. N. Y. Acad. Sci.*, vol. 1180, pp. 56–67, 2009.

[10] J. Marksteiner et al, "Five out of 16 plasma signaling proteins are enhanced in plasma of patients with mild cognitive impairment and Alzheimer ' s disease ,," vol. 32, no. 3, pp. 539–540, 2011.

[11] S. E. O'Bryant et al, "A blood-based screening tool for Alzheimer's disease that spans serum and plasma: Findings from TARC and ADNI," *PLoS One*, vol. 6, no. 12, 2011.

[12] V. Devanarayan and A. Simon, "Evaluation of the plasma proteomics data from the ADNI database for Alzheimer's disease-state classification and prediction of 12-month progression from MCI to AD," *Alzheimer's Dement.*, vol. 7, no. 4, p. S545, 2011.

[13] S. J. Kiddle et al, "Candidate blood proteome markers of Alzheimer's disease onset and progression: A systematic review and replication study," *J. Alzheimer's Dis.*, vol. 38, no. 3, pp. 515–531, 2014.

[14] J. Marksteine et al, "Five out of 16 plasma signaling proteins are enhanced in plasma of patients with mild cognitive impairment and Alzheimer's disease," *Neurobiol. Aging*, vol. 32, no. 3, pp. 539–540, 2011.

[15] A. Güntert et al, "Plasma gelsolin is decreased and correlates with rate of decline in Alzheimer's disease," *J. Alzheimer's Dis.*, vol. 21, no. 2, pp. 585–596, 2010.

[16] D. Galasko and T. E. Golde, "Biomarkers for Alzheimer's disease in plasma, serum and blood - conceptual and practical problems.," *Alzheimers. Res. Ther.*, vol. 5, no. 2, p. 10, 2013.

[17] S. J. Kiddle et al, "Plasma Based Markers of [11C] PiB-PET Brain Amyloid Burden," *PLoS One*, vol. 7, no. 9, 2012.

[18] S. M. Laws, "Blood-Based Protein Biomarkers for Diagnosis of Alzheimer Disease<alt-title>Blood-Based Protein Biomarkers for Diagnosis of AD</alt-title>," *Arch. Neurol.*, p. 1, 2012.

[19] W. T. Hu et al, "Plasma multianalyte profiling in mild cognitive impairment and Alzheimer disease," *Neurology*, vol. 79, no. 9, pp. 897–905, 2012.

[20] D. Johnstone et al, "Multivariate protein signatures of pre-clinical alzheimer's disease in the alzheimer's disease neuroimaging initiative (adni) plasma proteome dataset," *PLoS One*, vol. 7, no. 4, 2012.

[21] D. A Llano, V. Devanarayan, and A. J. Simon, "Evaluation of plasma proteomic data for Alzheimer disease state classification and for the prediction of progression from mild cognitive impairment to Alzheimer disease.," *Alzheimer Dis. Assoc. Disord.*, vol. 27, no. 3, pp. 233–43, 2013.

[22] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed. Elsevier, 2011.

[23] K. Hajian-Tilaki, "Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation," *Casp. J Intern Med*, vol. 4, no. 2, pp. 627–635, 2013.

[24] E. Jammeh et al, "Using NHS primary care data to identify undiagnosed dementia," *J Neurol Neurosurg Psychiatry*, vol. 86, no. 4, pp. 28–30, 2015.

[25] J. Escudero et al, "Machine Learning-Based Method for Personalized and Cost-Effective Detection of Alzheimer ' s Disease," 2012.

[26] M. Thambisetty et al, "Plasma biomarkers of brain atrophy in Alzheimer's disease," *PLoS One*, vol. 6, no. 12, pp. 1–7, 2011.