

Detection of Alzheimer's Disease With Blood Plasma Proteins

Blessan. Y. R¹, P. Briskilla²

¹Dept of Computer Science and Engineering

²Assistant Professor, Dept of Computer Science and Engineering

^{1,2}Christian College of Engineering and Technology,

Oddanchatram, Dindigul, Tamil Nadu

Abstract- *The successful development of amyloid- based biomarkers and tests for Alzheimer's disease (AD) represents an important milestone in AD diagnosis. However, two major limitations remain. Amyloid-based diagnostic biomarkers and tests provide limited information about the disease process and they are unable to identify individuals with the disease before significant amyloid-beta accumulation in the brain develops. The objective in this study is to develop a method to identify potential blood-based non-amyloid biomarkers for early AD detection. The use of blood is attractive because it is accessible and relatively inexpensive. Our method is mainly based on machine learning (ML) techniques (support vector machines in particular) because of their ability to create multivariable models by learning patterns from complex data. Using novel feature selection and evaluation modalities, we identified 5 novel panels of non-amyloid proteins with the potential to serve as biomarkers of early AD. In particular, we found that the combination of A2M, ApoE, BNP, Eot3, RAGE and SGOT may be a key biomarker profile of early disease. Disease detection models based on the identified panels achieved sensitivity (SN) > 80%, specificity (SP) > 70%, and area under receiver operating curve (AUC) of at least 0.80 at prodromal stage (with higher performance at later stages) of the disease. Existing ML models performed poorly in comparison at this stage of the disease, suggesting that the underlying protein panels may not be suitable for early disease detection. Our results demonstrate the feasibility of early detection of AD using non-amyloid based biomarkers.*

Keywords- Alzheimer's disease, blood biomarker, dementia, machine learning, support vector machine.

I. INTRODUCTION

ALZHEIMER'S disease (AD) is the leading cause of dementia and poses a significant social and economic challenge. It is responsible for more than half of all cases of dementia [1]. Over 50 million individuals currently suffer from dementia worldwide with a projected increase to 152 million by 2050 [2]. No cure for AD has been discovered, but

there is intense effort to develop new clinical interventions that may slow or halt the disease. Such interventions are aimed at early (including preclinical and prodromal [3]) stages of the disease prior to extensive cell damage, when it is thought treatment is more likely to be effective.

To facilitate early diagnosis [4]–[6], the use of established biomarkers such as those based on amyloid-beta in cerebral spinal fluid (CSF) and molecular imaging of brain amyloid deposition using positron emission tomography (PET) is recommended [4]–[6].

However, despite progress with the development of amyloid- based biomarkers and tests for early AD diagnosis, they have two major constraints [7]–[9]. Amyloid-based biomarkers provide limited information about the disease pathological aetiology and pathways [10]–[12]. In addition, tests based on these biomarkers are unable to identify individuals at risk of AD prior to a significant amyloid-beta deposition in the brain. There is a need for biomarkers that have the potential to detect biological processes that precede brain amyloid-beta accumulation (amyloid pathology) during the disease development. Such biomarkers may advance understanding of the disease, aid identification of individuals at the early disease stages and the development of new interventions.

Studies suggest that AD is characterised by metabolic alterations [4] that may precede amyloid pathology [12]. Signatures of such metabolic abnormalities may therefore serve as biomarkers of earlier stages of the disease than amyloid biomarkers. Such biomarkers may be obtained from blood since blood has rich metabolic information content. The use of blood is also attractive because blood biomarker-based test is relatively non-invasive compared to CSF and may be more cost-effective than PET imaging. A number of studies have attempted to find non-amyloid biomarkers of disease by profiling a large array of non-amyloid proteins in blood and examining their association with the disease [13]–[15], but this approach is difficult to apply in practice.

A promising approach is the use of machine learning (ML) techniques to find appropriate combinations of non-amyloid proteins to detect AD as no single non-amyloid protein has been shown to reliably detect the disease. ML makes it possible to fit multivariable data to a model by learning complex patterns from data. Several studies [16]–[24] have applied ML to develop classifiers to differentiate between AD subjects and healthy controls. For example, O’Bryant et al. [19] developed a model with a panel of 30 serum proteins that classified Alzheimer’s disease dementia (ADD) subjects and HCs with sensitivity (SN), specificity (SP), and area under receiver operating curve (AUC) of 88%, 82%, and 0.91, respectively. Similarly, with 14 plasma proteins, a classifier model constructed by Llano et al. [22] classified ADD and HC subjects with 86.5% SN, 84.2% SP and AUC of 0.85. More recently, a panel of inflammatory markers in plasma was identified that classified ADD and HC with 84% SN, 70% SP, and AUC of 0.79 using a logistic regression model [25]. In another study, a 12-marker panel classified ADD and HC with 90% SN and 66.7% specificity, and higher performance in post-mortem confirmed AD cases [26]. Furthermore, a study [27] that explored the use of deep learning, random forest, and XGBoost algorithms for classification of ADD and HC achieved AUC of 0.88 with XGBoost algorithm and 0.85 with deep learning and random forest. Despite the promising results from these studies, most of the models were developed and evaluated using data from cognitively healthy controls and subjects at the later stages of the disease. The models were not evaluated in individuals at the early stages of the disease. Therefore, the panels underlying such models may not be suitable as biomarker signatures of early AD.

In this study, the main objective is to develop a ML-based method (support vector machines (SVM) in particular – see later) to identify blood biomarkers of early AD based on non-amyloid proteins with the potential to identify the disease prior to accumulation of amyloid-beta in the brains. We also assess the potential of existing ML-based methods to achieve early disease detection.

The rest of this paper is structured as follows. The materials and methods are described in Sections II and III. The results are presented in Section IV, and the discussion and conclusions are provided in Sections V and VI.

TABLE I
DEMOGRAPHIC INFORMATION OF SUBJECTS IN STUDY DATA

Clinical groups	Sample size	~Ave. age in years (SD)	~Ave. years of education (SD)	% Female
HC	58(54)	75(6)	16(2.8)	48(50)
MCI	136	75(7)	16(3.0)	45
ADD	108	75(8)	15(3.2)	46

SD: standard deviation; HC: healthy control; MCI: mild cognitive impairment; and ADD: Alzheimer’s dementia

II. MATERIALS

A. Blood Proteomic Data

Blood proteomic data used in this study were obtained from the Alzheimer’s disease neuroimaging initiative (ADNI) portal (<http://adni.loni.ucla.edu>). The quality-controlled data consist of 146 plasma proteins derived from 58 and 54 healthy controls (HCs) at baseline and 12 months later respectively, 136 individuals with mild cognitive impairment due to AD (MCI) at 12 months from baseline, and 108 Alzheimer’s dementia (ADD) patients at baseline. The MCI subjects were later diagnosed with AD dementia within about 10-year follow-up. A list of the 146 proteins are shown in the supplementary material. Mild dementia was diagnosed according to NINCDS-ADRDA criteria for probable ADD. A detailed description of the protocol may be found on the ADNI database. The demographic information of the subjects is shown in Table I. The subjects were age matched, over 70 years old and had about 16 years of education on average.

III. METHODS

A. Data Pre-processing

All study data were standardized as indicated in (1) to ensure that proteins with high numeric values relative to others would not cause bias in subsequent ML operations. Given a feature instance x , the standardised value z is given as,

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

Where μ and σ are the sample mean and standard deviation of the feature distribution, respectively.

To make optimal use of available data while minimizing susceptibility of our approach to overfitting problems, the pre-processed data were partitioned into two non-overlapping datasets; Datasets 1 and 2. Dataset 1 consists of baseline data from the ADDs and HCs. All existing methods evaluated in this study except [20] were originally developed based on Dataset 1. In our approach, Dataset 1 was

used to conduct a robust feature preselection (a key aspect in ML) and model development.

The resulting models were further evaluated with Dataset 2. Dataset 2 consists of month-12 data from MCIs and HCs. It was used to assess the performance of the developed models (trained on the entirety of Dataset 1) for MCI vs. HC classification. Models were trained with only Dataset 1 during model development using the entirety of it or its subsamples.

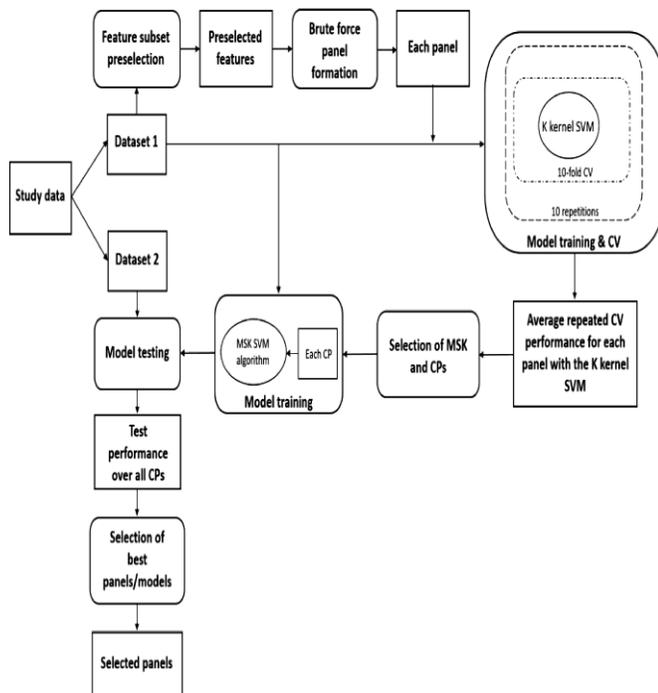


Fig. 1. Overall framework for identification of novel putative biomarker panels and model development for early AD detection. K: Different kernels of SVM including linear, 2 and 3 -degree polynomials, and radial basis function (RBF), respectively. MSK: Most stable kernel. A stable kernel is one that showed most moderate to high performance for most panels. CV: Cross-validation (CV). CP: Candidate panel. A candidate panel is one that meets our performance criteria (SN and SP of at least 70%) in the model training and CV step. Sensitivity and specificity have been described elsewhere [28].

B. Replication and Evaluation of Existing Methods

We replicated the ML models reported in previous studies for classification of ADD and HC subjects (Dataset 1) using 10-fold cross-validation with the average performance of the models taken after 10 repetitions. In 10-fold cross-validation, the dataset D is randomly split into 10 mutually exclusive subsets (the folds) D1, D2, ..., D10 of approximately equal size. The classifier is trained and tested 10 times; each time $t \in \{1, 2, \dots, 10\}$, it is trained on $D \setminus D_t$ and

tested on D_t [29]. The cross-validation estimate of the classifier performance is the overall performance over all the folds. Repeated cross-validation was implemented to ensure a robust estimation of performance [29]. The ability of the models to classify MCI and HC was then tested with Dataset 2 to assess their potential and hence the underlying protein panels to detect early AD.

C. Novel Panel Identification and Model Development

Fig. 1 shows the methodological framework that we used to identify novel blood protein panels and to develop the new ML models for early detection of AD. The framework is described in detail in the following subsections. Briefly, the framework consists of three major procedures which include feature subset preselection, protein panel formation, and ML-based model development and evaluation. A feature subset preselection process was performed to identify protein subsets that may have strong discriminatory power between disease subjects (ADD) and HCs. A brute force search was applied to the preselected feature subset to form several protein panels. Each of the panels was then used to develop and cross-validate SVM classifiers of different kernels (K) using Dataset 1. Data from ADD subjects were used in these initial procedures on the basis that dementia subjects are more likely to exhibit the metabolic alterations that are associated with the disease. The most stable kernel and candidate panels (promising models) trained on Dataset 1 were further evaluated for classification of individuals with MCI and HCs using Dataset 2. The promising models with best performance at this stage were selected as final. The protein panels that underlie the selected models are reported as potential blood-based non-amyloid biomarker signature of early disease.

Feature (Protein) Subset Preselection: 1) A feature subset preselection procedure was implemented with Dataset 1 using correlation-based feature subset selection (CFS) method [30]. The goal of this task was to make an initial selection of the most relevant and non-redundant features for classification of ADD and HC subjects and consequently reduce the dimension of the study data prior to model development. Reduction of the dimension of the study data was necessary because it would otherwise be computationally expensive to implement an exhaustive search to evaluate the classification performance of all possible feature subsets with ML algorithms. For N-dimensional data (where N is 146 in this case) there are 2^N possible feature subsets. Based on feature subset evaluation methods that attempt to remove irrelevant and redundant features from data by using correlation-based heuristic to determine the worth (merit) of a feature subset. This technique has been shown to compare favourably with wrapper-based approaches in selecting the best feature subsets that achieve

high classification accuracy while incurring far less computational cost [31]. It is based on a heuristic that evaluates the merit of feature subsets following the hypothesis that a good feature subset consists of features highly correlated with the class, yet uncorrelated with each other. Correlation in this sense refers to the predictability of one variable by another. Equation (2) shows the mathematical formulation of the CFS heuristics, a concept borrowed from test theory [32]

Merit is the heuristic merit of a feature subset consisting of k features, r_{fc} is the mean feature-class correlation and r_{ff} is the mean feature-feature inter-correlation. The parameters, r_{fc} and r_{ff} are measures of feature relevance and redundancy, respectively, based on the proposition that a feature is relevant if it is correlated with the class, otherwise it is irrelevant. Redundant features are correlated with one or more other features. To determine the correlations, continuous features were firstly discretized using the discretization method proposed in [33] to ensure that all features were uniformly handled. The correlations were calculated in terms of modified information gain known as symmetrical uncertainty (SU) [34] to cater for the bias of information gain in favour of features with more values. Values were normalised to the range [0, 1] to ensure that they were comparable and had similar effect.

Where gain is the information gain [35] for nominal features X and Y , $H(X)$ and $H(Y)$ are the entropy [36] of X and Y , respectively. The gain is formulated as

Novel Panel Formation and SVM-based Evaluation: Firstly, feature panels were formed from the CFS-preselected proteins based on a brute force approach. Each panel was then evaluated using a wrapper-based method to identify the ML algorithm and panels with best performance for classification of ADD and HC subjects. Using each panel, several SVM [37] classification models were constructed with different kernels including linear, 2nd and 3rd degree polynomials, and radial basis function (RBF) using Dataset 1. Average performance of each model to classify ADD and HC subjects was obtained using a 10-fold cross-validation [29] scheme repeated 10 times. Secondly, the performance of most stable models (SVM algorithm and feature panels) that met the performance criteria of average SN and SP $\geq 70\%$ for classification of ADD and HC subjects was tested with Dataset 2 for discrimination of MCI and HC groups. Finally, the models and underlying protein panels with best performance in classifying MCI and HC groups were selected as putative models and non-amyloid biomarker panels for early detection of AD.

Classification With Kernelized SVM: The choice of SVM for the model development task was informed by the fact that it is robust even with limited training data, and not prone to local

extremum [38], as well as our previous experience [24]. It is a very powerful tool widely applied in similar biomedical applications [39]. SVM classifies training instances belonging to either of two classes by fitting a separation boundary (hyperplane) between the classes such that the margin between the boundary and either class is maximized. The class of a new instance is decided depending on which side of the hyperplane it lies. Fig. 2 illustrates a 2-class SVM classifier. Given a 2-class problem with training data consisting of N examples $(x_1, y_1), (x_2, y_2), \dots, (x_{N-1}, y_{N-1}), (x_N, y_N)$, with input features $x_i \in \mathbb{R}^d$ and class $y_i \in \{-1, 1\}$, the goal of SVM is to define a hyperplane $h(x)$ that is given by subject to $y_i(x_i^T w + b) \geq 1$, where b is a constant, d is the dimension of the data, w is a vector of unknown length with d dimension pointing from the origin and normal to the margin, and m is shown to be equal. The resulting w from the optimization in (9) is of the form shown in (10), with α_i being nonzero for instances i (known as support vectors) where the constraint $y_i(w^T x_i + b) \geq 1$ is met.

With (10), b may be determined from (7), and following from (8), the decision rule for a new sample u of unknown class may be stated as,

Where α_i are Lagrangian multipliers resulting from the optimization of (9).

When the training data are not linearly separable by a hyperplane, SVM may transform the data to new space where they become linearly separable by using kernel functions. The kernel function simply computes dot products of features in the transformed space. One of such kernels is the polynomial kernel [40]. For example, given feature vectors v and z , a polynomial kernel K is formulated as,

Where r is the degree of the polynomial. Thus, for a SVM classifier with a polynomial kernel, the solution for the hyperplane (formally determined by substituting (10) in (7)) and decision rule for a new sample of unknown class are modified as shown in (13) and (14).

However, because a standard SVM seeks to fit a margin separating all positive and negative training instances without any error which is not often practicable, a concept known as soft margin [37] which permits minimum misclassification error is implemented in practical SVM algorithms with a slight modification of (9).

D. Implementation and Performance Evaluation Feature selection using CFS as discussed earlier was conducted with attribute selection toolbox in Weka software package [41]. All classification tasks were conducted with MATLAB and Weka

software packages. MATLAB codes are available on <https://github.com/chimastan/earlydetectionofAD>. In evaluating the models from previous studies, we used Weka where previous studies had used it for model development. Training of ML models and validation of performance for ADD vs. HC discrimination was based on 10-fold cross-validation scheme repeated 10 times. The data (Dataset 1) were randomly repartitioned after each run to ensure that data subsets used for training and validation varied from the ones used in the preceding run. This way, a more robust average performance is obtained. Classification performance metrics of primary consideration were measures of SN and SP in accordance with international recommendations for clinically usable AD biomarkers [42]. A performance threshold of 70% for SN and SP was adopted in the model development task. This is on the grounds that the diagnostic accuracy of human experts reaches 77% [43] with sensitivity and specificity reaching 81% and 70% [5], respectively. Moreover, sensitivity and specificity greater than 80% is the target performance for ideal AD biomarkers [42]. No class imbalance handling procedure was applied to the training dataset (Dataset 1) in model development as minority to majority class distribution was 35:65% which is acceptable in ML-based classification problems [44], [45].

IV. RESULTS

A. Replication and Evaluation of Existing Models

We successfully replicated 7 models for classification of ADD subjects and HCs. The model proposed by [20] could not be replicated because it was originally trained on a dataset not available to us. Nevertheless, we constructed a model with Dataset 1 based on the ML algorithm and blood protein panel proposed by the ([20]) study. Only existing models constructed with blood proteins available in our study dataset were investigated in this study. Table II shows the average cross-validated performance of the models repeated over 10 runs for classification of ADD and HC subjects. Nearly all the models achieved SN, SP, and AUC greater than 80%, 60%, and 0.70, respectively. However, when evaluated for possible detection of early AD by classifying MCI and HC with Dataset 2, the SN values of the models remained moderately high while their SP values drastically dropped (with only one model achieving up to 50%). This implies that the models may have undesirably high levels of false positives when applied for early disease detection. Consequently, the underlying protein panels may not serve as good biomarker signatures of early disease.

B. Feature Subset Preselection

Using our new methodological approach, sixteen proteins with a merit (Merit) of 0.36 were preselected with the CFS technique from the 146 proteins in the original study data. The 16 proteins are shown in Table III together with their statistical significance P as calculated with z-test. The z-test was used to estimate the statistical significance of the difference between the pair of clinical groups being considered together (AD vs. HC) and (MCI vs. HC) for the pre-selected features. All except a few features were statistically significant (p -value < 0.05) in the ADD vs. HC pair (Dataset 1). Most of the features were not statistically significant in the MCI vs. HC pair (Dataset 2). This may be due to the high imbalance between the samples sizes of MCI and the HC in the dataset.

C. Novel Panel Formation and SVM-based Evaluation

From the 16 CFS-preselected protein subset, 216 different protein panels were formed. Results from wrapper-based evaluation of all the panels for classification of ADD and HC groups using Dataset 1 showed that models constructed with 2-degree polynomial kernel had a better and more stable performance. Consequently, SVM with 2-degree polynomial kernel was selected as the algorithm of choice. Only (10,699) 2-degree polynomial kernelized SVM models that met our performance benchmark (SN and SP $\geq 70\%$) for ADD and HC classification were further evaluated for their potential to detect early disease with Dataset 2. Two models constructed with six and eight protein panels (A1M, A2M, ApoA2, CD5L, IL3, SGOT and A1M, A2M, ApoA2, BNP, BTC, CD5L, IL3, SGOT, respectively) achieved a remarkable cross-validated performance (SN of 92% and 93%, SP of 81% and 83%, AUC of 0.90 and 0.94 respectively) in classifying ADD and HC subjects. This perhaps highlights a possible performance benefit of the CFS-based feature preselection technique. Nevertheless, the two models performed poorly when evaluated for classification of MCI and HC subjects. The implication is that an excellent model at later stages of the disease does not necessarily imply a good disease detection model at the early disease stages. This may be attributed to subtle differences in the underlying patterns as well as noise in the data among other factors, thus highlighting the need for further evaluations. Five models constructed with panels shown in Table IV realized best performance for classification of MCI and HC groups. All but one of the models detected AD subjects with SN and SP above 80% and 70% respectively at dementia as well as MCI stage. A larger panel formed by combining all five panels in Table IV achieved a cross-validated SN, SP, and AUC of 85%, 70%, and 0.88, respectively in classifying ADD vs. HC. However, its specificity dropped drastically to 52% with 82% SN and 0.73 AUC when tested for MCI vs. HC classification. The

introduction of well-known risk factors of AD [46] such as age and level of education as covariates to the models did not improve performance significantly. APOE4 genotype was not used as a covariate to avoid bias since less than 9% of HC group have positive status.

V. DISCUSSION

In this study, we developed models and identified novel nonamyloid biomarker panels for early detection of AD following a new approach, and demonstrated that existing ML methods may not be suitable for early detection. The models and panels were selected based on their performance at both the prodromal and dementia stages of the disease, thus improving the chance that signals about the disease were captured rather than noise resulting from individual variations between study participants.

Ideally, the smaller the size of a panel, the better in terms of interpretability and cost of implementation in practical applications such as point of care technology. However, because our study was exploratory, it was important to flag all the panels that achieved reasonably good performance since it is unclear which panel or proteins are the most important. Gaining such clarification may require further investigation such as analysis of protein-protein interaction for the proposed panels (see later). We have also shown the performance of the larger panel derived by combining all five panels we identified, although it has a lower performance relative to the individual panels perhaps due to curse of dimensionality. Comparing our results (Table IV) with those of existing models we investigated (Table II); the best existing model identified AD subjects at MCI stage with high sensitivity and fairly good specificity (79% SN and 50% SP) while our model with the least panel size achieved a better performance with 80% SN and 70% SP. At dementia stage, our proposed models achieved a performance that is comparable to the best model from the investigated studies. Comparing our results with the three recent relevant studies (see Table V), we note that the panels identified in [25] and [26] classified ADD and HC with high performance, but the markers were reported by the authors to be poor at distinguishing between MCI and HC. Furthermore, while study [27] achieved high AUC of 0.88 with XGBoost model for classification of ADD and HC, the model’s performance has not been evaluated for disease detection at MCI stage. Due to unavailability of biomarkers used in the study in our study data, the performance of the models for MCI and HC classification was not investigated in this study. In contrast to the recent studies, our models achieved high performance for

disease detection at ADD stage (with one of the models shown in the table realising best AUC, with high sensitivity and specificity) as well as the MCI stage. Our proposed panels differ significantly from those of existing methods. This may be due to significant differences in the approaches including feature preselection and evaluation modalities which were deliberately applied in this study. We are not aware of the use of CFS for feature preselection in previous studies. We have provided details of the ML algorithm used including the kernel type and order as well as their selection process to ensure transparency of approach reproducibility

TABLE II
PERFORMANCE OF EXISTING BLOOD BIOMARKER PANELS FOR AD DETECTION

Study	Panel size	Panel	ML model	ADD vs. HC (Dataset 1)			MCI vs. HC (Dataset 2)		
				SN	SP	AUC	SN	SP	AUC
[20]	11	Adip, B2M, CRP, FABP, FVII, IL18, MCP1, PPP, TLSP, TNC, VCAM	Random forest	85.2	25.9	0.62	81.6	46.3	0.72
[21]	5	A1M, ApoE, BNP, IL16, SGOT	Logistic regression	85.2	74.1	0.90	79.0	50.0	0.70
	8	A1M, ApoA2, ApoE, BNP, Eot3, IGM, PLGF, SGOT		88.0	72.4	0.87	80.9	46.3	0.69
[22]	5	A1M, ApoA2, ApoE, BNP, SGOT		87.0	62.1	0.83	83.1	38.9	0.67
	13	ApoA2, ApoE, BNP, Eot3, HBEGF, IGM, IL16, PLGF, PYY, SGOT, TNC, TTR, Vit	Random forest	92.6	60.3	0.87	85.3	42.6	0.72
	14	A1M, A2M, ApoA2 ApoE, BNP, BTC, CRP, Eot3, IGM, IL16, MPO, PLGF, RAGE, SGOT		92.6	67.2	0.91	83.1	44.4	0.70
[23]	6	A1M, A2M, AAT, ApoE, CC3, PPP	Naive Bayes	86.1	63.8	0.82	78.3	37.0	0.62
[24]	5	A1M, A2M, CC3, IGM, TNC	SVM	81.1	60.5	0.77	75.7	35.2	0.65

TABLE III
CFS-BASED PRESELECTED PROTEINS

Protein	P	
	ADD vs. HC (Dataset 1)	MCI vs. HC (Dataset 2)
A1M	2.9E-6	3.3E-1
A2M	2.5E-3	3.2E-1
ApoA2	3.2E-8	1.1E-1
ApoE	1.1E-7	3.8E-4
BNP	7.7E-7	5.2E-2
BTC	4.4E-2	2.4E-1
CDSL	1.0 E-1	8.6E-1
Eot3	5.5E-5	6.2E-3
IGM	9.7E-7	3.9E-5
IL3	8.1E-3	6.9E-15
MCSF1	4.0E-1	8.4E-2
PAPPA	7.7E-4	1.6E-1
PLGF	1.3E-5	3.2E-1
PYY	2.7E-6	5.9E-1
RAGE	6.5E-3	6.3E-1
SGOT	9.2E-6	2.2E-6

TABLE IV
PERFORMANCE OF NOVEL CANDIDATE BLOOD BIOMARKER PANELS FOR EARLY DETECTION OF AD

Panel size	Panel	ADD vs. HC (Dataset 1)			MCI vs. HC (Dataset 2)		
		SN	SP	AUC	SN	SP	AUC
7	A2M, ApoE, BNP, Eot3, PLGF, RAGE, SGOT	88.5	70.4	0.87	80.1	70.4	0.80
7	A2M, ApoE, BNP, Eot3, PYY, RAGE, SGOT	88.9	73.8	0.89	77.9	74.1	0.80
8	A2M, ApoE, Eot3, IGM, MCSF1, PYY, RAGE, SGOT	85.3	71.6	0.86	83.8	70.4	0.83
9	A2M, ApoA2, ApoE, BNP, BTC, Eot3, PYY, RAGE, SGOT	85.0	75.0	0.89	80.1	72.2	0.80
10	A1M, A2M, ApoE, BNP, BTC, Eot3, IGM, MCSF1, PAPPA, SGOT	88.1	72.9	0.89	83.1	70.4	0.80

A1M: alpha-1 microglobulin ; A2M: alpha-2 macroglobulin; ApoA2: apolipoprotein A2; ApoE: apolipoprotein E; BNP: brain natriuretic peptide; BTC: betacellulin; CDSL: CDS; Eot3: Eotaxin 3; IGM: immunoglobulin M; IL3: interleukin 3; MCSF1: monocyte-colony stimulating factor 1; PAPPA: Pregnancy-Associated Plasma Protein A; PLGF: placenta growth factor; PPP: Pancreatic Polypeptide; PYY – peptide YY; RAGE: receptor for advanced glycosylation end; SGOT: serum glutamic oxaloacetic transaminase.

TABLE V
COMPARISON OF OUR RESULTS WITH RECENT RELEVANT STUDIES

Study	ML model	ADD vs. HC			MCI vs. HC		
		SN	SP	AUC	SN	SP	AUC
[25]	Logistic regression	84.0	70.0	0.79	poor		
[26]	Random forest	90.0	67.0	0.77	poor		
[27]	XGBoost	-	-	0.88	-	-	-
	Random forest	-	-	0.85	-	-	-
	Deep learning	-	-	0.85	-	-	-
Current study	SVM	85.0	75.0	0.89	80.1	72.2	0.80

In future, the study will be validated in independent cohorts and extended to preclinical stages of the disease. It is noteworthy that no existing AD model based on non-amyloid proteins has hitherto been evaluated for early disease detection using ADNI data. Regarding the proteins evaluated in this study, besides PAPP, which is rather highly associated with depressive symptoms in older adults [47] other proteins preselected by CFS have been previously identified in several studies [16]–[24] to have classification value in discriminating between ADD and HC groups. From the five selected panels shown in Table IV, six proteins (i.e., A2M, ApoE, BNP, Eot3, RAGE, and SGOT) appear as most prominent, featuring in nearly all the panels. A combination of the six proteins therefore seem to play a significant role in the discrimination of disease (prodromal and dementia) subjects and healthy controls. The panel classified both groups with sensitivity and specificity > 80% and 65%, respectively and AUC of at least 0.80. Several of these proteins are found in nearly all the previously reported models investigated in this study. Studies show that blood plasma levels of A2M are linked to mechanisms related to blood-brain barrier damage and neuronal injury as well as hippocampus metabolism in early AD [15], [48]. ApoE in blood is speculated as a dementia risk marker in preclinical AD [49]. BNP levels in plasma is associated with decline in cognitive function [50]. Plasma levels of RAGE are altered in AD [51]. RAGE has been reported to play a critical role in AD and considered as a potential therapeutic target [52]. SGOT is a biomarker of peripheral inflammation and an essential metabolic enzyme. It is often used as a clinical measure of liver function [53]. Interestingly, a recent finding has implicated liver function as a potential significant confounding factor in the onset of AD (https://www.alz.org/aaic/releases_2018/AAIC18-Tuesgut-liver-brain-axis.asp). However, this study has several limitations including the following:

Sample size and ML method: In this work, the sample size of study data was small. This is because of the

limited availability of relevant data due in part to the high cost of collection of such specialized data. As a result of the limited dataset, latest ML methods such as deep learning (DL) were not explored in this study owing to their requirement for large datasets. As more data become available, we shall explore DL methods such as convolutional and recurrent neural networks [54], [55]. Nevertheless, conventional machine learning methods are still attractive in this domain given their relative simplicity, cheaper cost, and usefulness for data modeling [56]. However, despite the high classification performance achieved by the traditional ML approach we applied, there are other methods such as ensemble learning [57] that have the potential to improve performance and therefore may be applied in future study. **Demographics:** Another limitation is that the study data only consist of older and educated subjects. Thus, our findings may not generalise well to other cohorts such as less educated individuals given that level of education is a well-known risk factor for AD. **Feature selection method:** Notwithstanding the usefulness of CFS feature preselection technique applied for dimensionality reduction and mitigation of model overfitting, some important markers with strong biological links to AD may have been eliminated as the process was blind to prior knowledge.

Protein-protein interaction analysis: In this study, aspects such as protein-protein interaction were not investigated as these were beyond the scope of the study. Potentially, analysis of the interactions between proteins in the identified panels may facilitate understanding of their joint role in AD process and clarify which panel(s) are more clinically relevant. In view of the limitations above, there is a need to conduct additional follow-up studies and validation of our findings in large and independent cohorts considering that validation of findings is an important step for clinical acceptance and translation into clinical practice. Besides proteomics-based biomarkers, there are also other nonamyloid-based blood biomarkers such as mRNA [58], [59] and autoantibodies [60] where progress is being made in AD detection and improving understanding of disease. For instance in [58], three mRNA biomarkers that suggest important dysregulated pathways in AD pathogenesis have been identified. Therefore, future studies should consider the exploitation of a range of blood-based biomarkers including proteomics and mRNA. This may lead to a more accurate panel of blood biomarkers to detect AD and improve the understanding of its aetiology. Overall, the results from this study suggest that it may be feasible to detect early AD using a profile of non-amyloid proteins in blood associated with the metabolic processes that accompany or precede the disease. Because the proteins are non-amyloid based, they have the potential to detect the disease even before amyloid pathology develops. It may be possible to develop new understanding of

the disease through further studies of these proteins and their protein-protein interactions in the disease pathogenesis. Such understanding may aid the development of new interventions in response to current failure of clinical trials targeting amyloid clearance. The main contributions of this study include the potential biomarker signatures identified and the methodological approach adopted in the search for these signatures in an effort to bridge an important study gap of early detection of AD with proteomic-based non-amyloid blood biomarkers.

VI. CONCLUSION

We have developed potential models and identified five novel candidate non-amyloid biomarker panels for early detection of AD utilizing a new approach. The developed models based on these panels classified prodromal AD as well as AD dementia and normal controls with sensitivity above 80%, specificity higher than 70%, and AUC of at least 0.80. A combination of A2M, ApoE, BNP, Eot3, RAGE and SGOT were identified as key protein profiles with significant contribution to the classifications performance. The results suggest that it may be feasible to detect early AD using a profile of non-amyloid proteins that identify the metabolic processes that accompany or precede the disease. It may be therefore possible to detect the disease with the proteins before amyloid pathology (the earliest signature current diagnostic biomarkers can detect) develops since they are not amyloid-based. This may aid identification of individuals at the earliest stages of AD who may benefit from early interventions. Furthermore, new insights about the disease may be gained from understanding the interactions between the proteins in disease subjects. Such enhanced understanding may contribute to the improvement of interventions in clinical trials.

VII. ACKNOWLEDGMENT

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (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. 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

- [1] A. Association, "2018 Alzheimer's disease facts and figures," *Alzheimer's Dementia*, vol. 14, no. 3, pp. 367–429, 2018.
- [2] M. Prince, A. Comas-Herrera, M. Knapp, M. Guerchet, and M. Karagian- nidou, *World Alzheimer Report 2016: Improving Healthcare for People Living with Dementia: Coverage, Quality and Costs now and in the Future*, London, U.K.: Alzheimer's Disease International, 2016.
- [3] B. Dubois et al., "Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria," *Alzheimer's Dementia*, vol. 12, no. 3, pp. 292–323, 2016.
- [4] M. S. Albert et al., "The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimer's Dementia*, vol. 7, no. 3, pp. 270–279, 2011.
- [5] G. M. McKhann et al., "The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimer's Dementia*, vol. 7, no. 3, pp. 263–269, 2011.
- [6] R. A. Sperling et al., "Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimer's Dementia*, vol. 7, no. 3, pp. 280–292, 2011.
- [7] G. P. Morris, I. A. Clark, and B. Vissel, "Questions concerning the role of amyloid- β in the definition, aetiology and diagnosis of Alzheimer's disease," *Acta Neuropathologica*, vol. 136, no. 5, pp. 663–689, 2018.
- [8] K. H. Tse and K. Herrup, "Re-imagining Alzheimer's disease—the diminishing importance of amyloid and a glimpse of what lies ahead," *J. Neurochemistry*, vol. 143, no. 4, pp. 432–444, 2017.
- [9] F. Zhang, J. Wei, X. Li, C. Ma, and Y. Gao, "Early candidate urine biomarkers for detecting Alzheimer's disease before amyloid- β plaque deposition in an APP (swe)/PSEN1 dE9 transgenic mouse model," *J. Alzheimer's Disease*, vol. 66, pp. 613–637, 2018.
- [10] F. Kametani and M. Hasegawa, "Reconsideration of amyloid hypothesis and tau hypothesis in Alzheimer's disease," *Frontiers Neuroscience*, vol. 12, 2018, Paper 25.
- [11] M. Gold, "Phase II clinical trials of anti-amyloid β antibodies: when is enough, enough?" *Alzheimer's Dementia: Translational Res. Clin. Interventions*, vol. 3, no. 3, pp. 402–409, 2017.

- [12] S. Makin, “The amyloid hypothesis on trial,” *Nature*, vol. 559, no. 7715, pp. S4–S4, 2018.
- [13] H. D. Soares, Y. Chen, M. Sabbagh, A. Rohrer, E. Schrijvers, and M. Breteler, “Identifying early markers of Alzheimer’s disease using quantitative multiplex proteomic immunoassay panels,” *Annals. New York Academy Sci.*, vol. 1180, no. 1, pp. 56–67, 2009.
- [14] H. D. Soares et al., “Plasma biomarkers associated with the apolipoprotein E genotype and Alzheimer disease,” *Archives Neurology*, vol. 69, no. 10, pp. 1310–1317, 2012.
- [15] M. Thambisetty et al., “Proteome-based identification of plasma proteins associated with hippocampal metabolism in early Alzheimer’s disease,” *J. Neurology*, vol. 255, no. 11, pp. 1712–1720, 2008.