Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 248 (2021) 119188

Contents lists available at ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

journal homepage: www.elsevier.com/locate/saa

Raman spectroscopy and machine learning for biomedical applications: Alzheimer's disease diagnosis based on the analysis of cerebrospinal fluid

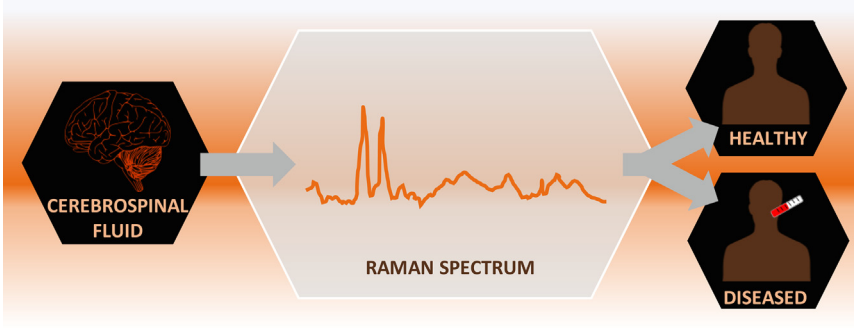
Elena Ryzhikova [a], Nicole M. Ralbovsky [a], Vitali Sikirzhytski [a], Oleksandr Kazakov [b], Lenka Halamkova [a], Joseph Quinn [c], Earl A. Zimmerman [d], Igor K. Lednev [a,*]

[a] Department of Chemistry, University at Albany, SUNY, 1400 Washington Avenue, Albany, NY 12222, USA
[b] Department of Physics, University at Albany, SUNY, 1400 Washington Avenue, Albany, NY 12222, USA
[c] Layton Aging and Alzheimer's Center, Oregon Health & Science University, Portland, OR 97239, USA
[d] Alzheimer's Center, Department of Neurology of Albany Medical Center, Albany, NY 12222, USA

HIGHLIGHTS

- Raman spectroscopy and machine learning is a universal tool for medical diagnosis.
- Alzheimer's disease is the most common form of dementia worldwide.
- Raman spectroscopy is used for detecting Alzheimer's disease in cerebrospinal fluid.
- Statistical analysis improves the capability of the method for accurate diagnosis.
- A novel method for detecting and diagnosing Alzheimer's disease is proposed.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 7 August 2020
Received in revised form 29 October 2020
Accepted 1 November 2020
Available online 13 November 2020

Keywords:
Alzheimer's disease
Early diagnosis
Cerebrospinal fluid
Raman spectroscopy
Machine learning

ABSTRACT

Current Alzheimer's disease (AD) diagnostics is based on clinical assessments, imaging and neuropsychological tests that are efficient only at advanced stages of the disease. Early diagnosis of AD will provide decisive opportunities for preventive treatment and development of disease-modifying drugs. Cerebrospinal fluid (CSF) is in direct contact with the human brain, where the deadly pathological process of the disease occurs. As such, the CSF biochemical composition reflects specific changes associated with the disease and is therefore the most promising body fluid for AD diagnostic test development. Here, we describe a new method to diagnose AD based on CSF via near infrared (NIR) Raman spectroscopy in combination with machine learning analysis. Raman spectroscopy is capable of probing the entire biochemical composition of a biological fluid at once. It has great potential to detect small changes specific to AD, even at the earliest stages of pathogenesis. NIR Raman spectra were measured of CSF samples acquired from 21 patients diagnosed with AD and 16 healthy control (HC) subjects. Artificial neural networks (ANN) and support vector machine discriminant analysis (SVM-DA) statistical methods were used for differentiation purposes, with the most successful results allowing for the differentiation of AD and HC subjects with 84% sensitivity and specificity. Our classification models show high discriminative power, suggesting the method has a great potential for AD diagnostics. The reported Raman spectroscopic examination of CSF can complement current clinical tests, making early AD detection fast, accurate, and

* Corresponding author.
E-mail address: ilednev@albany.edu (I.K. Lednev).

https://doi.org/10.1016/j.saa.2020.119188

inexpensive. While this study shows promise using a small sample set, further method validation on a larger scale is required to indicate the true strength of the approach.

1. Introduction

Raman spectroscopy has been shown to be a suitable diagnostic tool which can provide very specific information regarding the structure, conformation, and composition of such biomacromolecules as nucleic acids, proteins, and lipids for biomedical applications [1,2]. Recently, Raman spectroscopy has been utilized to diagnose a variety of diseases including different types of cancer, diabetes, atherosclerosis, and Parkinson's disease [3–6]. The advantage of Raman spectroscopy resides in its ability to probe the entire biochemical composition of a heterogeneous sample simultaneously [7]. The result is an accurate and statistically significant characterization of the molecular makeup of a sample, allowing for detection of multiple biomarkers of a disease at the same time. Through concurrent detection of these important biomolecules, the specificity of the method for investigating diseases is greatly improved. Even further, combining Raman spectroscopic analysis with machine learning methods creates the opportunity to build statistical models which can be used for accurate and objective medical diagnostic applications. In fact, the combination of Raman spectroscopy with machine learning has already been applied for investigating Duchenne muscular dystrophy in a mouse model through analysis of serum, with the results showing the method as being more successful than those currently used for diagnosis [8]. In human studies, single red blood cells were probed as a novel approach for detecting Celiac disease in a minimally invasive manner; machine learning models here achieved 100% diagnostic accuracy [9]. Relevant to the current study, diagnosis of Alzheimer's disease (AD) based on analysis of saliva was achieved with greater than 99% accuracy at the donor level [10]. Similarly, the same method was applied for investigating serum for diagnosing AD using both spontaneous Raman spectroscopy and surface enhanced Raman spectroscopy, with each study reaching very high levels of accuracy [11,12]. It was the goal of this study to further expand upon the method of Raman spectroscopy with machine learning to identify changes which are specific to AD that occur in cerebrospinal fluid (CSF) for improving the diagnostic sensitivity of the approach.

As the most common form of dementia afflicting the elderly population worldwide, AD is one of the most difficult health problems facing all nations. This is especially true when considering the United States, where 5.7 million people suffer from the disease [13]. Such a large number of individuals with the disease results in devastating consequences to the healthcare level and socioeconomic wellbeing of millions of people and their families [14]. AD is a progressive neurodegenerative disease, and the outcome is lethal, as there is no known cure and available diagnostic methods are limited to identifying the disease in its later stages [15,16].

In the span of only a few years, AD clinical symptoms progress from mild memory loss to profound cognitive failure [17]. Neuropathology of disease progression is linked to severe alterations of the brain tissue including significant synaptic and neuronal loss and degeneration, accumulation of extracellular plaques containing amyloid-β peptide core, and formation of intracellular neurofibrillary tangles (NFT) which consist of abnormally phosphorylated microtubule-associated tau proteins [18]. A clinical diagnosis of probable AD can only be confirmed based on these histopathological features identified after post-mortem examination [19].

Because the destructive pathophysiological processes last for many years and initially only shows general symptoms such as perturbations in episodic memory, AD diagnostics is often a challenging and complicated process that heavily depends on clinical assessment criteria. However, when a diagnosis of AD can be established, the brain tissue has already been irreversibly damaged [20]. As such, the "preclinical" phase of the disease is thought to be the most critical time for therapeutic intervention. If intervention occurs early on, the progression of AD can be slowed for as long as possible with the possibility to even prevent development completely [16]. As such, the early and accurate diagnosis of AD has become of enormous importance, and the need to advance current diagnostic tools has lead researchers to investigate disease-specific biomarkers [21].

CSF chemistry has been widely studied in the search for AD biomarkers because of its direct contact with the extracellular space of the brain [22]. Previous studies resulted in the development of a laboratory test that is now widely used in clinics. The test includes measurements of three different CSF biomarkers: the concentration levels of the total amount of tau protein (T-tau) which indicates the level of neuroaxonal degeneration; the level of abnormally hyperphosphorylated tau protein (P-tau) which is indicative of the tangle pathology; and the 42 amino acid isoform of the Aβ peptide ($A\beta_{42}$) which is inversely correlated with plaque pathology [19,23]. By interpreting the biomarkers together, the CSF test allows for the detection of AD pathogenesis and can sometimes even provide a pre-dementia diagnosis of AD [19,24,25]. Although this CSF test has been implemented in clinical trials and practice, it has become evident that some major problems remain to be solved. It was found that biomarker concentrations vary considerably between AD studies due to the lack of standardization of methods [26,27]. There are many factors that can contribute to these variations, including differences between laboratories in their pre-analytical or analytical protocols related to lumbar puncture and CSF sample handling, as well as the choice of assay type and analytical kit applied for measurements [26,28]. Many other CSF components have been found to be indicative of AD development, such as inflammatory factors, apolipoprotein E, fatty acids, and phospholipids [29–32]. Therefore, the development of a unified multi-biomarker profile signature in CSF based on total composition could be highly beneficial for disease detection strategies. We envision that a simple, efficient and universal way to simultaneously detect various CSF biomarkers would be tremendously useful for diagnosing AD. Thus, we explored the potential of Raman spectroscopy in combination with machine learning to differentiate CSF samples obtained from AD patients and healthy controls. Several multivariate statistical methods were employed, including principal component analysis, genetic algorithm, SVM-DA, and ANN to characterize and classify Raman spectroscopic data with great levels of success. Further studies are needed to verify the approach with a larger number of samples.

2. Materials and methods

2.1. Clinical subjects and protocols

Evaluated subjects were recruited at the Layton Aging and Alzheimer's Center, Oregon Health & Science University, Portland, OR. There were two groups of subjects in the study. The first group

consisted of unrelated AD patients who had both their medical history reviewed and a clinical evaluation regarding level of dementia performed by a trained neurologist. The assessment of AD subjects' cognitive functions was done using Clinical Dementia Rating (CDR). The other group of donors included age-matched control subjects who were not diagnosed with any neurological or psychiatric disorders. All recruited subjects signed a written informed consent prior to their participation. Institutional Review Boards of both the University at Albany and the Oregon Health & Science University reviewed and approved the research protocol for human studies. The demographic information for study subjects is presented in Table 1.

Although the number of individuals studied in this project is small, it is comparable to the number of samples used in other proof-of-concept studies which use both Raman spectroscopy and surface enhanced Raman spectroscopy for identifying Alzheimer's disease in body fluids [10–12,33,34].

2.2. Sample preparation

CSF samples were collected at the time of lumbar puncture and transported on ice to be immediately centrifuged at 100 rpm for 10 min and aliquoted. All samples were stored at −80 °C until use. Each CSF sample was subjected to a maximum of one freeze–thaw cycle before Raman measurements. A 40-μL drop of each CSF sample was transferred to a standard microscopic glass slide covered with aluminum foil and dried completely under gentle air flow for 5 min.

2.3. Raman spectroscopic measurements

All Raman spectra were collected using a Renishaw inVia confocal Raman spectrometer equipped with a research-grade Leica microscope and a 50x long-range objective with 0.50 numerical aperture. All Raman spectroscopic measurements were performed in the mapping mode using a Renishaw PRIOR stage system. WiRE 3.2 software was employed for spectral measurements in the selected spectral range of 400–1800 cm^{-1} under a 785 nm diode laser excitation. The laser power was reduced to 55 mW to prevent photodegradation of the samples. The mapping procedure was preset for scanning the sample within an area of 2 × 2 mm and 121 spectra were measured for each sample, and each spectrum was measured with two 10 s accumulations. This allowed for a thorough representation of the sample to be obtained as well as to account for its intrinsic heterogeneity.

2.4. Data treatment and analysis

For data treatment and advanced statistical analysis, MATLAB R2013b software was used (8.2.0.701). Receiver operating characteristic (ROC) curve analysis was carried out using the "pROC" package, implemented in R project [35]. The original Raman spectra were baseline corrected, normalized by total area, and mean centered [36]. The adaptive iteratively reweighted penalized least squares (airPLS) algorithm[37] was used for baseline correction. Raman spectra with low signal-to-noise ratios were excluded from the dataset. After preprocessing, further analysis was performed using PLS_Toolbox 7.5.2 (Eigenvector Research, Inc.). Genetic Algorithm (GA) was used to select variables for subsequent multivariate classification models. The criteria used by GA is the predictive ability of the PLS models based on cross-validation. The fitness function was determined according to the root mean square error of cross-validation (RMSECV) of the partial least squares. Principal component analysis (PCA), support vector machine discriminative analysis (SVM-DA) and artificial neural networks (ANN) were all employed for classification purposes.

Table 1
Demographic information on the healthy control (HC) and Alzheimer's disease (AD) subjects.

Parameters	Diagnostic category	
	HC (n = 17)	AD (n = 21)
Age in years ± SD	71 ± 12	72 ± 5.3
Male (%)	41	62
Female (%)	59	38

3. Results and discussion

3.1. Raman spectroscopy of CSF

Individual Raman spectra collected from the CSF of healthy and diseased donors showed some variations in shape and intensity, which reflect changes in the biochemical composition of the samples. To illustrate these variations, the average Raman spectra and corresponding standard deviations for each of the two classes of CSF (AD and HC) are shown in Fig. 1A and 1B. The most prominent differences between AD and HC Raman data can be visually identified within the 1035–1050 cm^{-1} spectral region (Fig. 1A).

The Raman peak at 1045 cm^{-1} is present in ~20% of AD Raman spectra only and could be considered characteristic for the AD class. Interestingly, the majority of AD Raman spectra do not have this Raman band at all. About ~10% of the HC Raman spectra have a very low intensity peak at 1045 cm^{-1}. Less prominent differences between the two average spectra are observed at 1065 cm^{-1} and other spectral regions. As shown in Fig. 1B, the calculated difference between the two classes lays within ±2 standard deviations for both classes. This indicates that further analysis through machine learning methods is required to determine whether the aforementioned differences are useful for successful discrimination between the two classes. A summary of the machine learning methods used is shown in Table 2.

3.2. Genetic algorithm (GA)

To reveal statistically significant spectral differences between the AD and HC donors, the data was analyzed via GA. GA is a mathematical method derived from the biological principle of natural selection [38]. It is used to select variables with the lowest root-mean-square prediction errors. During GA, the most promising subsets of variables (chromosomes) are selected to form a population which will be continually replaced by new populations until the best performance, in terms of prediction error, is achieved and the chromosomes contributing the most influence are identified. A detailed explanation of GA for variable selection and its application was published by Niazi and Leardi [39]. Briefly, as the first step for GA, a large number of random selections of variables must be generated and the RMSECV is calculated for each of the given subsets. Each subset of variables is called an individual and the pool of all tested individuals is called the population. An individual is characterized by a set of parameters (variables) known as genes, where binary values are used to encode the genes. The RMSECV values, described as the fitness of the individual, determines how fit an individual's selection of variables is for the Y block. In the selection phase of GA, the fittest individuals are selected and pass their genes to the next generation. Individuals with greater fitness have a higher chance of being selected for reproduction, and the individuals with low fitness are discarded. In our study, the breeding was done by double cross-over, where the genes from two random individuals are split at two random points and

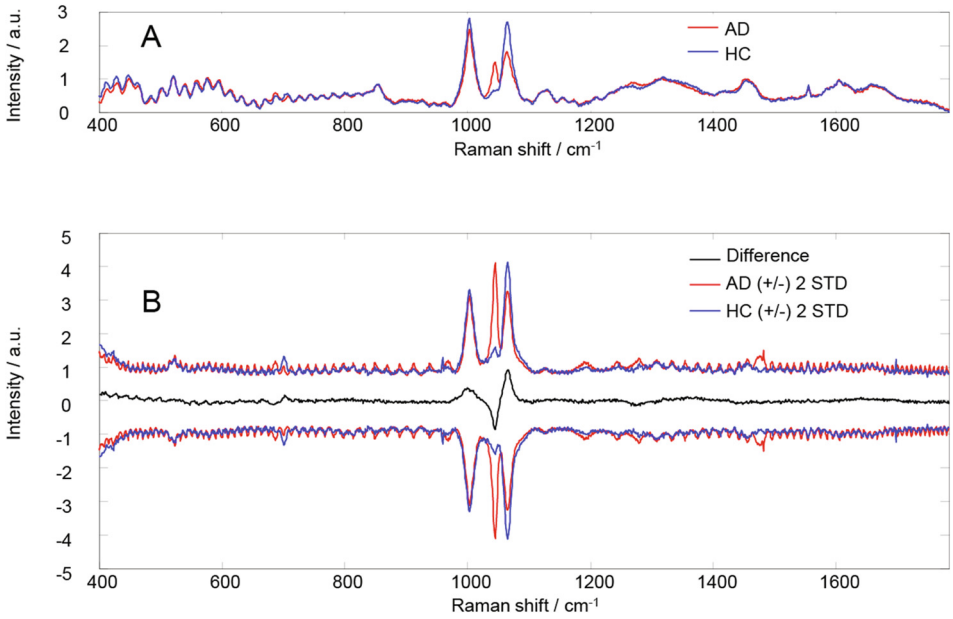


Fig. 1. Mean and difference Raman spectra of cerebrospinal fluid. (A) Mean Raman spectra of cerebrospinal fluid from AD (red line) and HC (blue line) cohorts. (B) Difference spectrum (black line) and spectral variations around the mean (±2 standard deviations).

Table 2
Summary of machine learning methods employed in the manuscript with main results obtained.

Step (#)	Method Used	Main Result of Method
1	Principal Component Analysis	Remove outliers within spectral dataset
2	Genetic Algorithm	Spectral variable selection
3	Support Vector Machine Discriminant Analysis	Supervised learning for classification
4	Receiver Operating Characteristic Curve Analysis	Evaluate performance of SVM-DA model
5	Artificial Neural Networks	Alternative supervised learning for classification

portions of the two individuals are swapped. The newly added individuals also have the possibility for mutation. When all individuals have been paired and bred, the whole population goes back to the original size and the process can continue again.

GA can be used for the identification of Raman bands and regions that are the most relevant and useful for differentiation between groups of spectral data. Therefore, the identified regions can be considered spectral biomarkers. This spectral information can be clinically relevant for biomarker development since the identified spectral regions can be attributed to particular biomolecules. This approach of relating signals to metabolites/chemicals has been applied in several studies [40–43].

After preprocessing of the data was accomplished, two additional steps were conducted. First, multivariate outlier removal was carried out using principal component analysis (PCA). The PCA outlier removal and screening algorithm was used to remove spectra of low quality and yielded a final number of 1037 spectra from the original 1663. Second, the quality of the spectra was evaluated after baseline correction by a single screening algorithm, which was developed in R. The total integration of the spectral dataset was log transformed and the mean spectra and standard deviation was calculated (Fig. 1A and B). To ensure at least 70% of the total spectra was kept, a "threshold" was used to adjust filtering according to the noise present in the dataset, i.e. mean plus/minus multiples of standard deviation was used to filter out bad spectra. Low quality spectra were removed from further processing, since it has been shown that such quality assessment prior to statistical analysis is beneficial [44]. Further details regarding this algorithm can be found elsewhere [45].

After PCA and the screening algorithm were completed, the evolutionary feature selection technique, GA, was applied to eliminate the non-informative and redundant variables from the dataset. GA considers all variables within the Raman spectral dataset and their significance toward accomplishing the discrimination between classes. GA reduces the original Raman spectral dataset to a subset of wavenumbers with the best discriminatory power to achieve better prediction performance. In order to drive the optimization process in GA, PLS was used as the fitness function with a special binary 'dummy' y-variable. The RMSECV was used to evaluate model performance. The variables that frequently appeared in better-performance models were considered useful for class differentiation after many generations of iteration. To simulate a natural evolutionary process, cross-over and mutation were allowed to occur after each generation.

The GA algorithm was implemented, as included in PLS Toolbox (version 7.5.2) using function "genalg", for variable selection operating with the PLS-based variable selection procedure. The population size was set to 70, the maximum number of generations was set to 100, the breeding crossover rule was set to double crossover, and the default mutation rate was used (0.005). Finally, a total of 100 runs were performed. The best solution provided by the GA identified 360 informative variables in the spectra compared to the original 1492.

GA is especially helpful in cases when the spectral dataset consists of hundreds or thousands variables. Briefly, using GA, seven fingerprint regions that are relevant to the spectroscopic signature for AD were identified (Fig. 2). The bands located in the regions selected by GA and the presumptive contributions of biomolecules responsible for those bands are given in Table S1 (Supplementary Material). Interestingly, the Raman bands which exhibited the greatest differences between the two classes of CSF samples, although still small and within two standard deviations of the mean, correspond to various amino acids which are present in known AD biomarkers such as tau and $A\beta_{42}$ proteins. Specifically,

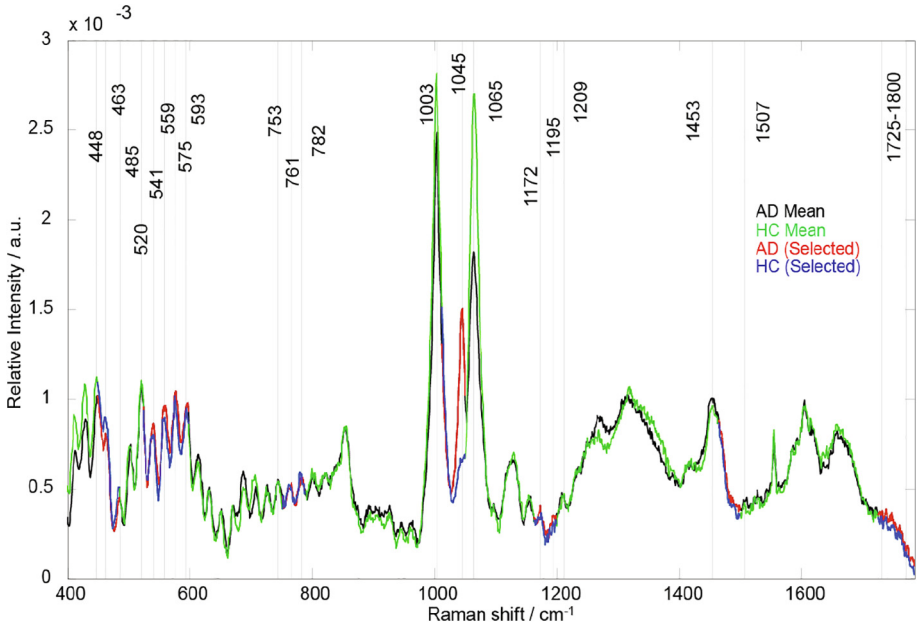


Fig. 2. Raman spectral regions identified by GA as the most useful for differentiation between CSF samples. GA-identified regions are highlighted on the Raman spectral pattern of the average class spectra: AD in black with red highlights and HC in green with blue highlights. Vertical lines indicate band positions.

the 1045 cm^{-1} band intensity increases for AD samples and correlates with an increase in glycine and proline contribution due to the tau protein; on the other hand, the 1065 cm^{-1} band intensity decreases in AD samples which is consistent with a decrease in (i) arginine contribution due to the tau protein and (ii) histidine and valine contributions due to $A\beta_{42}$. The reported amino acid residue composition of these two notable AD biomarkers are noted in comparison to the composition of an "average" protein [46]; differences in amino acid composition exist, indicating that changes in Raman spectra may occur due to these biomarkers. Further interpretation of these bands requires additional investigation to establish their clinical relevance with respect to current CSF biomarkers. Those areas not selected by GA are not addressed, as these spectral regions were not deemed useful for differentiation between the classes of data.

3.3. Classification of CSF Raman data using support vector machine discriminate analysis

It has been demonstrated that higher classification accuracy can be obtained using statistical machine learning methods which are able to model nonlinear relationships between variables, such as SVM-DA and ANN. The following machine learning methods were applied for AD diagnostics in this study: a multilayer perceptron (MLP) ANN and SVM-DA.

SVM-DA was applied to distinguish between AD subjects and HC subjects using the features which were selected by GA from the original Raman spectral dataset as an input. SVM-DA is a supervised machine learning method, which means it can be used for classification tasks. First described by Vapnik in 1995 [47], it has since found many applications in different scientific areas including bioinformatics [43] and has already been applied to solve various diagnostic problems [48,49]. SVM-DA can classify data by determining a set of support vectors. During the training process, a set of support vectors that outline a hyperplane in the feature space are selected. Then, a generic mechanism that fits the hyperplane surface to the training inputs by means of a kernel function is determined, where the specific kernel function is chosen by the user [43]. In this way, the behavior of SVM-DA is modified to determine the hyperplane that separates classes by maximizing the margin, or the distance between each class. For this study, the radial basis function (RBF) kernel was selected and was optimized by a combined approach of five-fold cross validation (leaving five samples out) and a systematic grid search of the parameters.

To evaluate the subject-independent accuracy performance of the SVM-DA model, a re-sampling scheme using subject-wise cross-validation was used. The model was built repeatedly (37 times), each time leaving out all spectral data from one donor. The data from the subject that was left out was then used to test the calculated SVM-DA model. For each left-out subject, classification results were calculated in terms of probabilities, where the probability of each spectrum belonging to a possible class was calculated using the distance between the spectral data and the hyperplane which separates the two classes. For each of the 37 subsets of spectral data, the predicted probability for each spec-

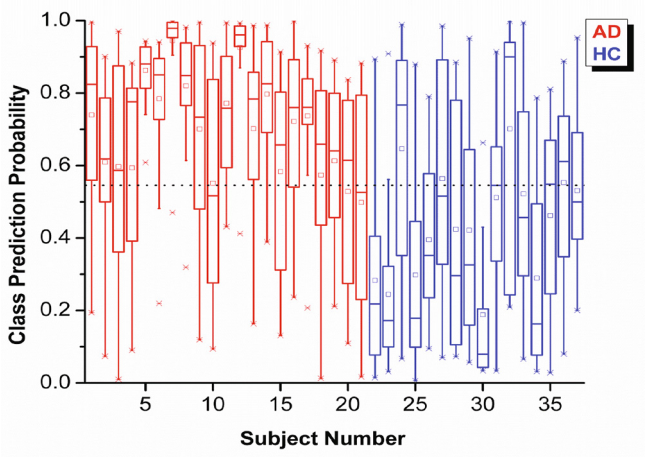


Fig. 3. Cross-validated SVM-DA classification of CSF samples' Raman spectra. Calculated probability for each single Raman spectrum to be classified as resulting from CSF of AD (red) or HC (blue) subjects. A threshold of 55.7% is established as the cut-off.

trum and for each subject as belonging to a particular class was determined. Fig. 3 shows SVM-DA results for the AD versus HC binary SVM-DA classification model. Less than 30 total spectra were misclassified. The result of classification depends on threshold values which can be arbitrarily set to a particular value. The best achieved cross-validated sensitivity and specificity parameters were 75% and 64% for individual spectra and 91% and 81% for overall subjects, respectively, using a threshold of 55.7% (Fig. 3).

ROC curve analysis was conducted following SVM-DA. ROC analysis and the calculated area under curve (AUC) are commonly used for the purpose of establishing discrimination thresholds. The sensitivity (true positive rate) is plotted against the specificity (true negative rate) to analyze various probability thresholds for class prediction purposes. The AUC is a single value which evaluates the probability of assigning a subject to the correct class. The AUCs of ROC curves were estimated by the trapezoidal method of integration with corresponding 95% confidence intervals (CI) using the method described by De Long et al. [50]. The ROC curves built herein (Fig. 4) were generated from calibration probability estimations to determine the best threshold for class predictions, and the threshold was used to classify all samples as either AD or HC during cross-validation.

AD patients were discriminated from HC using SVM-DA with an AUC of 0.75 (95% CI: 0.72–0.78) based on single spectra and an AUC of 0.90 (95% CI: 0.81–1.00) based on individual subjects (Fig. 4). The discriminatory power of the SVM-DA model is reduced for single spectra as compared to the subject-wise model. This can be explained by the presence of poor quality spectra in the dataset and, as such, a lower reliability for their correct prediction. It also can be noted that not all spectra collected from CSF of AD donors will reflect a noticeable contribution from AD biomarkers.

3.4. Classification of CSF Raman data using Artificial neural networks

In an effort to improve classification results, a second statistical method was employed. A multilayer perceptron (MLP) network is a type of ANN where computations are performed using a set of simple units with weighted connections between them. Training the learning algorithms can be thought of as a search for optimal weight values that result in a neural network which gives an output with the most correct answer. In this work, various three-layer MLPs, which consist of an input layer, a hidden layer, and output layer, were considered.

The MLPs were built using the MATLAB Neural Network Toolbox. Initially, using significant factor analysis, it was determined that at least 5 significant factors were required for optimal characterization of the spectroscopic data [51]. PCA was then employed to calculate the number of principle components (PCs) required and to reduce the number of input variables for better classification. The model architecture considered between 5 and 20 neurons in the input layer. The number of neurons in the hidden layer were varied between 10 and 800 until an optimal solution was found. A log sigmoidal activation function was chosen because of its ability to model complicated decision boundaries, and five different back-propagating algorithms (Levenberg-Marquardt, Gradient descent, Powell-Beale conjugate gradient, Bayesian regularization, and Fletcher–Powell conjugate gradient) were tested.

Various network structures were tested by randomly splitting the AD and HC datasets into training (70%), testing (15%), and validation (15%) subsets. To test the performance of each network structure, the sensitivity (true positive rate) and specificity (true negative rate) were calculated. The desired optimal performance was set so that the difference between sensitivity and specificity of each class for the testing and validation subsets did not exceed 10%. If the difference exceeded the 10% limit, it was assumed the classification model was the result of convergence to a local minima, and thus displayed weak generalization properties. In this case, the network structure was disregarded. To avoid possible over-fitting, each network architecture was tested 10 times. Classification models that demonstrated the required performance were considered for further analysis.

Table 3 presents the best results achieved for selected MLP networks for HC vs. AD classification. In the column labeled "Network Structure", the descriptors for the network are presented which identify the number of neurons in the input, hidden, and output layers. The sensitivity of correct classification of AD spectra was observed to be in the range of 80–84%. These results demonstrate a potential for improvement to build a discriminative model with high performance. The best performance was achieved by the network with 10 neurons in the hidden layer (84% sensitivity and 84% specificity); further increase of the number of neurons in the hidden layer did not improve the results. It is important to determine the optimal number of neurons in the hidden layer to successfully model complex decision boundaries while achieving satisfactory generalization capabilities. The obtained numbers demonstrate the potential of Raman spectroscopy to identify AD using CSF.

The instability in classification accuracy that was observed can be attributed to two main aspects of the modeling process. The first aspect is the intrinsic properties of the experimental dataset, such

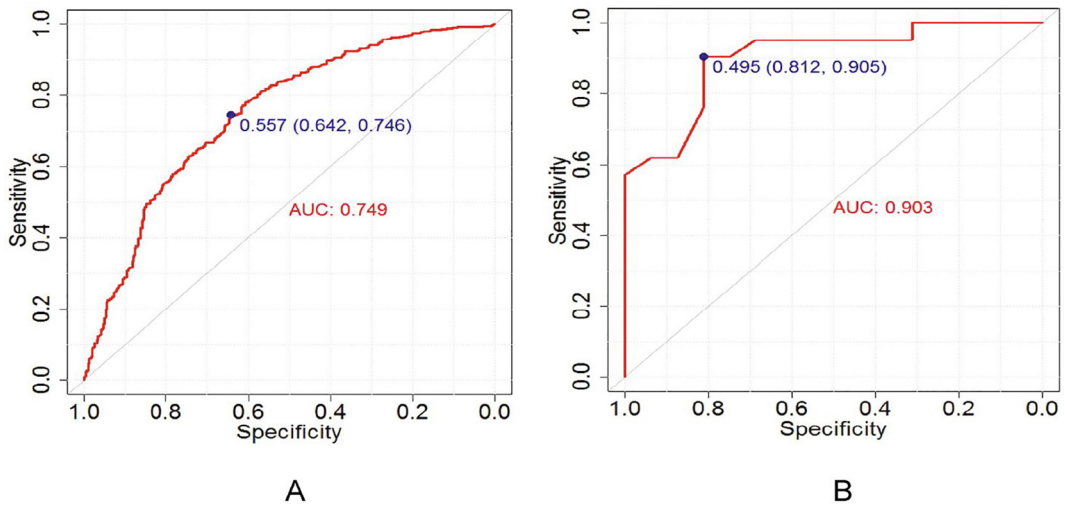


Fig. 4. The ROC curves for the SVM-DA classifier for classification of AD subjects including thresholds based on probabilities for (A) each spectrum and (B) for each subject.

Table 3
Performance of four ANN architectures for discrimination of AD from HC spectroscopic CSF data.

Model #	Network structure	HC Sensitivity/ Specificity, %	AD Sensitivity/ Specificity,%
1	5-10-1	81/83	84/84
2	5-50-1	82/83	80/84
3	5-100-1	76/79	82/81
4	5-200-1	86/83	83/84

as the limited size of the Raman dataset, the destructive contribution of noise to spectra, and the intrinsic heterogeneity of CSF samples, which is exacerbated by the applied drying process and does not guarantee that every recorded AD CSF Raman spectrum will depict at least some evidence of the disease. Furthermore, it is also important to mention that certain pre-analytical factors influence the stability of cerebrospinal fluid content including differences in collection procedure, time between sampling and centrifugation, length of storage before use, and number of freeze/thaw cycles for each sample [52,53]. Standardized CSF sampling, handling, and storage are essential to ensure successful transfer of new diagnostic assays from the research environment to the clinical laboratory [53,54]. With that being said, the achieved level of 84% of accurately classified single Raman spectra for validation datasets is truly a remarkable result. The second aspect to be considered is a propensity of ANN algorithms to fall into local minima, which yield a reduced generalization ability[55,56] as well as lead to overtraining of the model [57]. SVM-based algorithms do not tend to converge on local minima and often outperform alternative classification methods, including ANN [56]. However, in this study, the performance parameters of the best calculated SVM-DA model for individual spectra classification were slightly lower than those of the ANN model. Based on these points, the ANN model with the highest validated classification accuracy can be deemed acceptable. There is a great incentive to further investigate the potential of the method to identify AD using Raman spectroscopic analysis of CSF.

4. Conclusions

Raman spectroscopy has successfully been applied for solving various disease diagnostic problems in the past and has shown a great potential for use as a future clinical tool. While other studies have focused on analyzing different body fluids for detecting AD, this study utilized CSF as it is perhaps the most biologically relevant body fluid for detecting AD. The combination of Raman spectroscopy and machine learning is shown here as a valuable method for AD evaluation and one which can potentially aid in early AD detection. Spectral bands and regions identified by Genetic Algorithm in the measured Raman spectra represent specific spectroscopic differences between healthy and diseased CSF. These differences can be capitalized on by multivariate statistics to differentiate CSF of Alzheimer's patients from healthy donors. AD subjects were separated from healthy subjects with high levels of accuracy (84%), which is in the range of CSF tau and $A\beta_{42}$ protein analyses currently used in clinical laboratories [58]. This proof-of-concept study shows great success using a small dataset; however, it is important to note more detailed studies with a larger subject population size will allow for further evaluation of the reliability of this method.

CRediT authorship contribution statement

Elena Ryzhikova: Methodology, Data curation, Writing - review & editing. **Nicole M. Ralbovsky:** Methodology, Writing - review & editing. **Vitali Sikirzhytski:** Data curation. **Oleksandr Kazakov:** Data curation. **Lenka Halamkova:** Data curation. **Joseph Quinn:** Methodology. **Earl A. Zimmerman:** Methodology. **Igor K. Lednev:** Conceptualization, Methodology, Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The work was supported by a research grant from the Layton Aging and Alzheimer's Center, Portland, OR (NIA-AG008017) and by the State University of New York Technology Accelerator Fund. N.M.R. was supported by NIH grant T32 545 GM13206.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.saa.2020.119188.

References

[1] J.R. Baena, B. Lendl, Raman spectroscopy in chemical bioanalysis, Curr. Opin. Chem. Biol. 8 (5) (2004) 534–539.
[2] C. Krafft, Bioanalytical applications of Raman spectroscopy, Anal. Bioanal. Chem. 378 (1) (2004) 60–62.
[3] N.M. Ralbovsky, I.K. Lednev, Towards development of a novel universal medical diagnostic method: Raman spectroscopy and machine learning, Chem. Soc. Rev. 49 (2020) 7428–7453, https://doi.org/10.1039/d0cs01019g.
[4] H.M. Schipper, C.S. Kwok, S.M. Rosendahl, D. Bandilla, O. Maes, C. Melmed, D. Rabinovitch, D.H. Burns, Spectroscopy of human plasma for diagnosis of idiopathic Parkinson's disease, Biomarkers Med. 2 (3) (2008) 229–238.
[5] G. Basar, U. Parlatan, S. Seninak, T. Gunel, A. Benian, I. Kalelioglu, Investigation of Preeclampsia Using Raman Spectroscopy, Spectrosc.: Int. J. 27 (4) (2012) 239–252.
[6] N.M. Ralbovsky, I.K. Lednev, Raman spectroscopy and chemometrics: A potential universal method for diagnosing cancer, Spectrochim. Acta Part A Mol. Biomol. Spectrosc. 219 (2019) 463–487.
[7] N.M. Ralbovsky, I.K. Lednev, Raman hyperspectroscopy shows promise for diagnosis of Alzheimer's, Biophotonics 4 (25) (2018) 33–37.
[8] N.M. Ralbovsky, P. Dey, A. Galfano, B.K. Dey, I.K. Lednev, Diagnosis of a model of Duchenne muscular dystrophy in blood serum of mdx mice using Raman hyperspectroscopy, Sci. Rep. 10 (1) (2020) 11734.
[9] N.M. Ralbovsky, I.K. Lednev, Analysis of individual red blood cells for Celiac disease diagnosis, Talanta 221 (2021) 121642.
[10] N.M. Ralbovsky, L. Halámková, K. Wall, C. Anderson-Hanley, I.K. Lednev, Screening for Alzheimer's disease using saliva: a new approach based on machine learning and Raman hyperspectroscopy, J. Alzheimers Dis. 71 (4) (2019) 1351–1359.
[11] E. Ryzhikova, O. Kazakov, L. Halamkova, D. Celmins, P. Malone, E. Molho, E.A. Zimmerman, I.K. Lednev, Raman spectroscopy of blood serum for Alzheimer's disease diagnostics: specificity relative to other types of dementia, J. Biophotonics 8 (7) (2015) 584–596.
[12] E. Ryzhikova, N.M. Ralbovsky, L. Halámková, D. Celmins, P. Malone, E. Molho, J. Quinn, E.A. Zimmerman, I.K. Lednev, Multivariate statistical analysis of surface enhanced raman spectra of human serum for Alzheimer's disease diagnosis, Appl. Sci. 9 (16) (2019) 3256.
[13] A.S. Association, Alzheimer's disease facts and figures. Alzheimer's & Dementia, 14 (3) (2018) 367–429.
[14] R. Brookmeyer, D.A. Evans, L. Hebert, K.M. Langa, S.G. Heeringa, B.L. Plassman, W.A. Kukull, National estimates of the prevalence of Alzheimer's disease in the United States, Alzheimer's & Dementia 7 (1) (2011) 61–73.
[15] P.R. Solomon, C.A. Murphy, Early diagnosis and treatment of Alzheimer's disease, Expert Rev. Neurother. 8 (5) (2008) 769–780.
[16] K.G. Yiannopoulou, S.G. Papageorgiou, Current and future treatments for Alzheimer's disease, Therapeutic Adv. Neurol. Disorders 6 (1) (2013) 19–33.
[17] A. Ward, S. Tardiff, C. Dye, H.M. Arrighi, Rate of conversion from prodromal Alzheimer's disease to Alzheimer's dementia: a systematic review of the literature, Dementia Geriatric Cognit. Disorders Extra 3 (1) (2013) 320–332.
[18] M.R. Farlow, Etiology and pathogenesis of Alzheimer's disease, Am. J. Health-Syst. Pharm. 55 (Suppl 2) (1998) S5–S10.
[19] C. Rosén, O. Hansson, K. Blennow, H. Zetterberg, Fluid biomarkers in Alzheimer's disease - current concepts, Mol. Neurodegener. 8 (2013) 20.
[20] K. Blennow, H. Zetterberg, The application of cerebrospinal fluid biomarkers in early diagnosis of Alzheimer disease, Med. Clin. North Am. 97 (3) (2013) 369–376.

[21] N. Mattsson, H. Zetterberg, O. Hansson, N. Andreasen, L. Parnetti, M. Jonsson, S. K. Herukka, W.M. van der Flier, M.A. Blankenstein, M. Ewers, K. Rich, E. Kaiser, M. Verbeek, M. Tsolaki, E. Mulugeta, E. Rosén, D. Aarsland, P.J. Visser, J. Schröder, J. Marcusson, M. de Leon, H. Hampel, P. Scheltens, T. Pirttilä, A. Wallin, M.E. Jönhagen, L. Minthon, B. Winblad, K. Blennow, CSF biomarkers and incipient Alzheimer disease in patients with mild cognitive impairment, J. Am. Med. Assoc. 302 (4) (2009) 385–393.
[22] M.S. Henry, A.P. Passmore, S. Todd, B. McGuinness, D. Craig, J.A. Johnston, The development of effective biomarkers for Alzheimer's disease: a review, Int. J. Geriatric Psychiatry 28 (4) (2013) 331–340.
[23] S. Engelborghs, N. Le Bastard, The impact of cerebrospinal fluid biomarkers on the diagnosis of Alzheimer's disease, Mol. Diagn. Therapy 16 (3) (2012) 135–141.
[24] A. Anoop, P.K. Singh, R.S. Jacob, S.K. Maji, CSF Biomarkers for Alzheimer's Disease Diagnosis, Int. J. Alzheimer's Dis. 2010 (606802) (2010).
[25] S. Engelborghs, Clinical indications for analysis of Alzheimer's disease CSF biomarkers, Revue Neurologique 169 (10) (2013) 709–714.
[26] N. Mattsson, K. Blennow, H. Zetterberg, Inter-laboratory variation in cerebrospinal fluid biomarkers for Alzheimer's disease: united we stand, divided we fall, Clin. Chem. Lab. Med. 48 (5) (2010) 603–607.
[27] N. Mattsson, CSF biomarkers in neurodegenerative diseases, Clin. Chem. Lab. Med. 49 (3) (2011) 345–352.
[28] M. Bjerke, E. Portelius, L. Minthon, A. Wallin, H. Anckarsäter, R. Anckarsäter, N. Andreasen, H. Zetterberg, U. Andreasson, K. Blennow, Confounding factors influencing amyloid Beta concentration in cerebrospinal fluid, Int. J. Alzheimer's Dis. 2010 (986310) (2010).
[29] A.N. Fonteh, J. Chiang, M. Cipolla, J. Hale, F. Diallo, A. Chirino, X. Arakaki, M.G. Harrington, Alterations in cerebrospinal fluid glycerophospholipids and phospholipase A_2 activity in Alzheimer's disease, J. Lipid Res. 54 (10) (2013) 2884–2897.
[30] A.N. Fonteh, M. Cipolla, J. Chiang, X. Arakaki, M.G. Harrington, Human Cerebrospinal Fluid Fatty Acid Levels Differ between Supernatant Fluid and Brain-Derived Nanoparticle Fractions, and Are Altered in Alzheimer's Disease, PLoS ONE 9 (6) (2014) e100519.
[31] C. Humpel, Identifying and validating biomarkers for Alzheimer's disease, Trends Biotechnol. 29 (1) (2011) 26–32.
[32] E. Martínez-Morillo, O. Hansson, Y. Atagi, G. Bu, L. Minthon, E.P. Diamandis, H. M. Nielsen, Total apolipoprotein E levels and specific isoform composition in cerebrospinal fluid and plasma from Alzheimer's disease patients and controls, Acta Neuropathol. 127 (5) (2014) 633–643.
[33] M. Paraskevaidi, C.L. Morais, D.E. Halliwell, D.M. Mann, D. Allsop, P.L. Martin-Hirsch, F.L. Martin, Raman spectroscopy to diagnose Alzheimer's disease and dementia with Lewy bodies in blood, ACS Chem. Neurosci. 9 (11) (2018) 2786–2794.
[34] C. Carlomagno, M. Cabinio, S. Picciolini, A. Gualerzi, F. Baglio, M. Bedoni, SERS-based biosensor for Alzheimer disease evaluation through the fast analysis of human serum, J. Biophotonics 13 (3) (2020) e201960033.
[35] X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.C. Sanchez, M. Muller, pROC: an open-source package for R and S+ to analyze and compare ROC curves, BMC Bioinf. 12 (2011) 77.
[36] B.M. Wise, N.B. Gallagher, R. Bro, J.M. Shaver, W. Windig, R.S. Koch, D. OÕSullivan, PLS_Toolbox 6 for use with Matlab, Software, Eigenvector Research Inc, Wenatchee, WA, 2011.
[37] Z.M. Zhang, S. Chen, Y.Z. Liang, Baseline correction using adaptive iteratively reweighted penalized least squares, Analyst 135 (5) (2010) 1138–1146.
[38] Z. Ramadan, D. Jacobs, M. Grigorov, S. Kochhar, Metabolic profiling using principal component analysis, discriminant partial least squares, and genetic algorithms, Talanta 68 (5) (2006) 1683–1691.
[39] A. Niazi, R. Leardi, Genetic algorithms in chemometrics, J. Chemom. 26 (6) (2012) 345–351.
[40] A.C. Baclig, T.C. Bakker Schut, G.M. O'Regan, A.D. Irvine, W.H.I. McLean, G.J. Puppels, P.J. Caspers, Possibilities for human skin characterization based on strongly reduced Raman spectroscopic information, J. Raman Spectrosc. 44 (3) (2013) 340–345.
[41] D.L. Tong, A.C. Schierz, Hybrid genetic algorithm-neural network: Feature extraction for unpreprocessed microarray data, Artif. Intell. Med. 53 (1) (2011) 47–56.
[42] D. Mantzaris, G. Anastassopoulos, A. Adamopoulos, Genetic algorithm pruning of probabilistic neural networks in medical disease estimation, Neural Netw. 24 (8) (2011) 831–835.
[43] C.-L. Huang, C.-J. Wang, A GA-based feature selection and parameters optimization for support vector machines, Expert Syst. Appl. 31 (2) (2006) 231–240.
[44] B.H. Menze, B.M. Kelm, D. Heck, M.P. Lichy, F.A. Hamprecht, Machine-Based Rejection of Low-Quality Spectra and Estimation of Brain Tumor Probabilities from Magnetic Resonance Spectroscopic Images, in: H. Handels, J. Ehrhardt, A. Horsch, H.-.P. Meinzer, T. Tolxdorff (Eds.), Bildverarbeitung für die Medizin 2006, Springer, Berlin Heidelberg, 2006, pp. 31–35.
[45] J. Bueno, L. Halámková, A. Rzhevskii, I.K. Lednev, Raman microspectroscopic mapping as a tool for detection of gunshot residue on adhesive tape, Anal. Bioanal. Chem. 410 (28) (2018) 7295–7303.
[46] M.H. Smith, The amino acid composition of proteins, J. Theor. Biol. 13 (1966) 261–282.
[47] C. Cortes, V. Vapnik, Support-Vector Networks, Mach. Learn. 20 (3) (1995) 273–297.
[48] J.R. Hands, P. Abel, K. Ashton, T. Dawson, C. Davis, R.W. Lea, A.J.S. McIntosh, M.J. Baker, Investigating the rapid diagnosis of gliomas from serum samples using infrared spectroscopy and cytokine and angiogenesis factors, Anal. Bioanal. Chem. 405 (23) (2013) 7347–7355.
[49] B. Yan, Y. Li, G. Yang, Z.-N. Wen, M.-L. Li, L.-J. Li, Discrimination of parotid neoplasms from the normal parotid gland by use of Raman spectroscopy and support vector machine, Oral Oncol. 47 (5) (2011) 430–435.
[50] E.R. DeLong, D.M. DeLong, D.L. Clarke-Pearson, Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach, Biometrics 44 (3) (1988) 837–845.
[51] E.R. Malinowski, Factor analysis in chemistry, (3rd edition). . John Wiley & Sons, Ltd., (2002) pp. 10–23.
[52] J. Klener, K. Hofbauerova, A. Bartos, J. Ricny, D. Ripova, V. Kopecky Jr., Instability of cerebrospinal fluid after delayed storage and repeated freezing: a holistic study by drop coating deposition Raman spectroscopy, Clin. Chem. Lab. Med. 52 (5) (2014) 657–664.
[53] A.H. Simonsen, J.M.C. Bahl, P.B. Danborg, V. Lindstrom, S.O. Larsen, A. Grubb, N. H.H. Heegaard, G. Waldemar, Pre-analytical factors influencing the stability of cerebrospinal fluid proteins, J. Neurosci. Methods 215 (2) (2013) 234–240.
[54] S.D. Pearson, D.A. Ollendorf, J.A. Colby, Biomarker tests for the diagnosis of Alzheimer's disease: Generating evidence to inform insurance coverage determinations, Alzheimer's & Dementia 9 (6) (2013) 745–752.
[55] B. Cheng, D.M. Titterington, Neural Networks: A Review from a Statistical Perspective, Statist. Sci. 9 (1) (1994) 2–30.
[56] B.S. Everitt, G. Dunn, Applied Multivariate Data Analysis, Second Edition ed., John Wiley & Son Ltd, 2001, pp. 248–268.
[57] L. Breiman, Bagging predictors, Mach. Learn. 24 (2) (1996) 123–140.
[58] K. Blennow, H. Hampel, CSF markers for incipient Alzheimer's disease, Lancet Neurol. 2 (10) (2003) 605–613.