Classification of Alzheimer's disease in a mixed clinical cohort using biofluid Raman spectroscopy

George Devitt[1,2,3,4,*], **Sofia Michopoulou**[1,4], **Latha Kadalayil**[3,4], **Niall Hanrahan**[3], **Angus Prosser**[1,4], **Boyd Ghosh**[1,4], **Amrit Mudher**[2,4], **Christopher Kipps**[1,4,5,6,‡], **Sumeet Mahajan**[3,4,‡]

[1]University Hospital Southampton NHS Foundation Trust, Southampton, UK
[2]School of Biological Sciences University of Southampton, Southampton, UK
[3]School of Chemistry, University of Southampton, Southampton, UK
[4]The Interdisciplinary Dementia and Ageing Centre (IDeAC), Southampton
[5]Dept. Neurology, Wessex Neurological Centre, University Hospital Southampton Foundation Trust
[6]Clinical Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton, UK

* Corresponding author
[‡] Authors contributed equally

Abstract

Introduction There is a clinical unmet need for scalable, accessible and objective tests for dementia. Raman spectroscopy (RS) is a laser-based optical method that can rapidly provide chemically rich information ('spectral biomarkers') from biofluids but its utility for Alzheimer's disease (AD) diagnosis has not been rigorously established.

Methods We measured cerebrospinal fluid (CSF) samples from a mixed clinical cohort of patients (N=143) using RS. Machine-learning algorithms were trained, optimized and evaluated on Raman spectra to classify AD from non-AD.

Results AD was classified with 93% accuracy and spectral biomarkers were identified and primarily assigned to protein-derived aromatic amino acids. These spectral biomarkers directly correlated with pathological CSF biomarker concentrations.

Conclusions The feasibility of applying our simple, holistic and label-free spectral biomarker approach to dementia diagnosis was demonstrated. Compared to current and emerging methods, RS does not require sophisticated or specialized labs and is reagentless potentially offering unprecedented scalability and accessibility.

1 Background

Clinical diagnosis of Alzheimer's disease (AD) is often inconsistent with post-mortem neuropathology at autopsy, which is necessary for a definitive AD diagnosis (1-3). Reported accuracies of clinical AD diagnosis vary widely (4), in part due to diagnostic inaccuracies and patient heterogeneity. To address this, there has been a move from a symptomatic definition of AD towards a biological definition based on pathological biomarkers, including amyloid-β (Aβ) and tau protein biofluid levels (3). However, there is still a need to develop more accessible and objective biomarkers that are highly sensitive and specific to AD, whilst retaining stratification capability over other clinically overlapping neurodegenerative diseases (NDDs). Such biomarkers could improve AD diagnosis rate and accuracy, as well as patient cohort selection and stratification in drug trials. Furthermore, patient selection for AD disease-modifying therapies like Lecanemab (5) underscores the importance of identifying those most likely to benefit and tailoring interventions accordingly.

Current biomarkers for AD include structural and functional imaging such as magnetic resonance imaging (MRI) and single-photon emission computed tomography (SPECT)/positron emission tomography (PET), respectively. Biofluid markers are routinely utilized, particularly cerebrospinal fluid (CSF) markers for amyloid, tau and neurodegeneration (ATN), specifically amyloid-β-42 (Aβ42, A) phospho-tau 181 (p-tau, T), and total tau (t-tau, N). Several novel biomarkers are emerging as candidates to further support AD diagnosis, including plasma p-tau 217 (3, 5), as well as omics-based (meta-genomics, proteomics, metabolomics) biomarkers (6), and tau-specific PET ligands (7). A simpler yet powerful solution may be provided by optical spectroscopy. Raman spectroscopy (RS) is an optical technique that can provide highly selective label-free readouts within seconds and has the potential for portability at scale (8). RS is label-free and reagentless and can be carried out with a small footprint using portable devices in a range of healthcare settings, potentially without the need of skilled operators. Biofluids can be analysed directly to provide an unbiased and holistic chemical fingerprint based on the vibrational modes of the molecules within the sample. The RS-based chemical fingerprinting approach represents the collective characteristics of the biofluid and contrasts with the detection of a specific molecule or analyte (9). Proof-of-concept studies have demonstrated that RS may have clinical utility for identification of AD in CSF, but small patient sample sizes of 19 (10) and 37 (11) currently limit any confidence in clinical validity. Additionally, these studies used healthy controls as comparators to AD patients, which does not reflect real-world diagnostic scenarios as it is unlikely that 'healthy' people will seek an AD diagnosis. A more appropriate control group will be patients who sought dementia diagnoses due to cognitive complaints but were deemed to have a condition other than AD. It is therefore still not established whether RS can detect AD in a clinical cohort, representative of the clinical setting or population, with appropriate statistical power.

To address this critical gap, we utilized a mixed clinical cohort of patients (N=143) who reported cognitive complaints and for whom diagnosis was not initially clear, so CSF samples were taken by lumbar puncture for ATN biomarker analysis. We hypothesized that RS analysis of CSF can provide unique fingerprints of AD-specific pathology/neurodegeneration enabling differentiation between AD and non-AD patients even in a heterogenous cohort. We tested AD classification of Raman spectra using machine-learning (ML) algorithms. From the Raman spectra we extracted spectral biomarkers specific to AD and assessed their association with CSF ATN biomarkers by correlation. Our results demonstrate an important step towards the clinical translation of Raman spectroscopy for supporting AD diagnosis in the future.

Table 1. CSF sample cohort			
Clinical diagnosis	Age (mean ± SD)	Sex (%M / %F)	Number of patients
Non-AD	68 ± 11	71 / 29	76
AD	64 ± 9	43 / 57	67
Total	66 ± 10	65 / 35	143

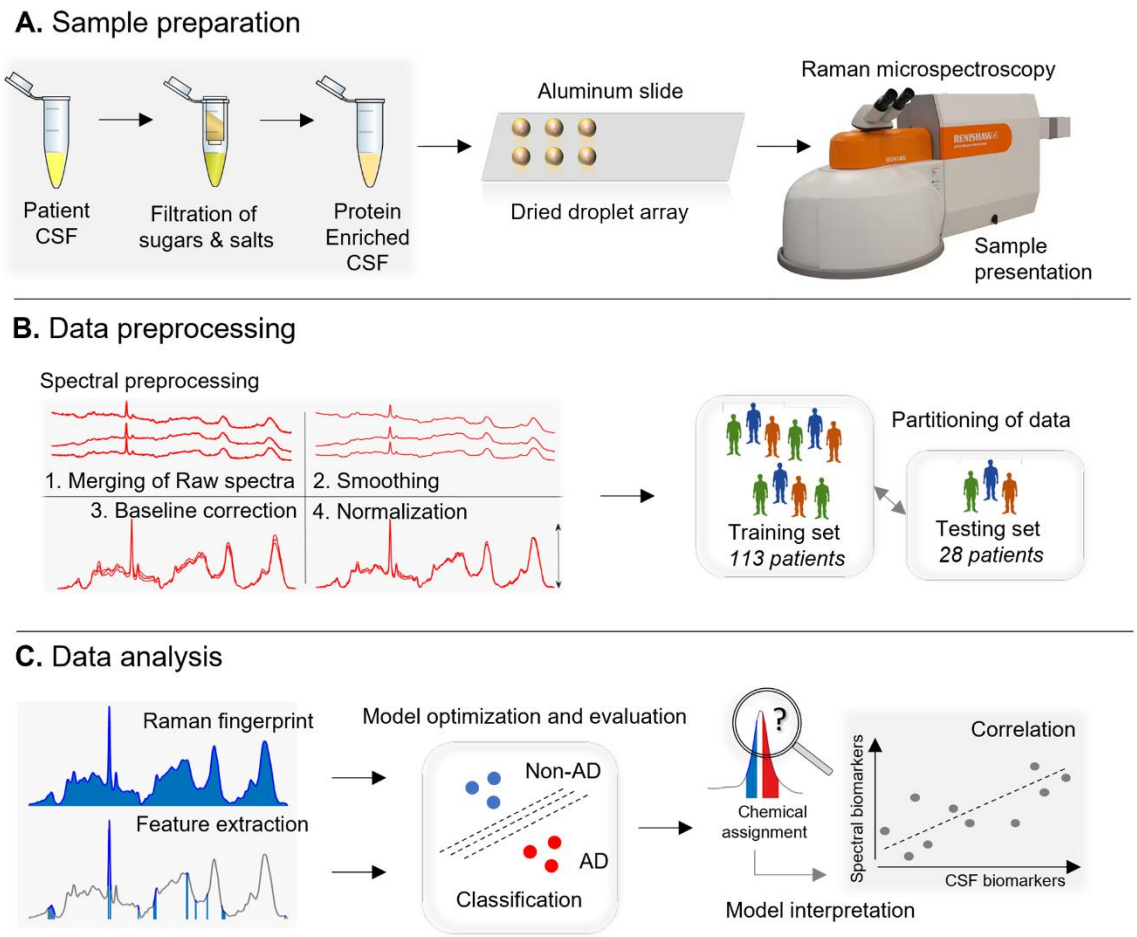


Figure 1. Schematic of workflow

***A**. Sample preparation for Raman spectroscopy. **B**. Preprocessing of Raman spectra for analysis. **C**. Analytical methods for classification of Raman spectra and identification of spectral biomarkers.*

2 Methods

CSF samples

The current study utilized 143 CSF samples collected from patients with cognitive complaints referred to Wessex Cognitive Disorders Clinic at University Hospital Southampton NHS Foundation Trust between 2014 and 2021 for diagnostic lumbar puncture to assess neurodegenerative pathology. Inclusion criteria for the current study were referral for cognitive complaints with query of dementia, availability of AD CSF biomarker results, and availability of CSF sample for Raman analysis. The cohort was divided into two groups including AD and non-AD (Table 1) based on clinical diagnosis that used a combination of clinical test scores, brain imaging, and CSF biomarker results. CSF biomarker concentrations for Aβ42, total tau, and p-tau 181 were measured by the UKAS

accredited Neuroimmunology and CSF laboratory at the National Hospital for Neurology and Neurosurgery at Queen Square London using the INNOTEST (Fujirebio) enzyme linked immunosorbent assays (ELISA) system until 2020, and the Lumipulse G (Fujirebio) chemiluminescent immunoassay since 2020 (each of which has consistent results (12)). The AD group included different AD variants (amnestic, speech, visual), as well as AD with other chronic comorbidities. Diagnoses in the non-AD group spanned a range of clinical conditions including dementia caused by other NDDs, as well as developmental and chronic conditions not related to neurodegeneration. A full list of clinical diagnoses can be found in Supplementary Table 1. Samples were collected as part of two studies that were approved by the Research Ethics Committee (REC:20/NW/0222 and REC:15/SC/0231). Written informed consent was provided at the time of the lumbar puncture by participants or their next of kin, including permission for the storage of excess CSF and use in research studies.

Sample handling and storage

A minimum of 2 ml of CSF was collected from each participant by lumbar puncture and centrifuged within 30 min in Starstedt polypropylene tubes. 500µL were aliquoted in an Elkay polypropylene tube, frozen at –70°C and sent for clinical diagnostic analysis to the Neuroimmunology and CSF laboratory at University College London Hospitals NHS Foundation Trust (UCLH). Remaining sample was frozen and stored at –80°C. For this research study, these samples were thawed, aliquoted (40 µl), and flash frozen in liquid nitrogen, and stored at –80°C locally. Samples were anonymized and randomized so that researchers were blinded to each participant's underlying condition during sample preparation and spectral preprocessing.

Sample preparation for Raman spectroscopy

CSF samples were thawed in room temperature (RT) water and vortexed well. CSF salts were exchanged for double-distilled H_2O (ddH_2O) using a 10 kDa MWCO filter (Vivaspin) as described below. Filters were first washed thoroughly by two 4-minute spins in 500µl ddH_2O at 12,000 x g in a tabletop microcentrifuge at 20°C. Filtrate, collected in the bottom compartment, was discarded. 40µl CSF was then added to the top part of the filter and diluted to 500µl with ddH_2O and centrifuged for 10 min at 12,000 x g, 20°C. This process of dilution and filtration was repeated three times. The desalted CSF was then transferred to a fresh tube and vortexed well. 1µl of the desalted CSF was deposited onto an aluminium coated (1000Å) microscope slide (SUBSTRATA, ANGSTROM Engineering®) and allowed to dry in a small, vented container at RT for 30 min.

Raman spectroscopy

A Renishaw inViaTM Qontor microscope system was used for Raman spectroscopy. Briefly, the Raman system was calibrated to the 520 cm^{-1} reference peak of the internal silicon substrate prior to each experiment. The charge-coupled device (CCD) detector and spectrometer slit areas were aligned using the auto align function and the laser spot was manually aligned to the centre of the crosshairs using the camera. The desalted dried CSF droplets were located and brought into focus using a Leica DM 2500-M bright field microscope and an automated 100 nm-encoded XYZ stage. Spectral data were collected, and parameters were determined using Renishaw WIRE5.5 software. The samples were excited using a 785nm laser at 50% power focused through a Leica 50x long working distance objective (numerical aperture = 0.5). Laser power at the sample was approximately 6.25 mW. Each Raman spectrum was collected for 30s, consisting of two 15s measurement acquisitions.

Thirty spectra were collected from the dried droplets. Only the bottom half of each droplet was measured, as seen through the camera. All raw spectra consisted of 1015 variables and detected using a Peltier cooled CCD (1024 pixels × 256 pixels) after dispersion through a 1800L/mm diffraction

grating with a range of 659cm^{-1}–1761cm^{-1} with a resolution of 1.09cm^{-1}. As multiple acquisitions were acquired per spectrum, cosmic rays were removed manually using the WIRE5.5 zap function after each spectral measurement. Background spectra from the aluminium surface were measured in 3 roughly evenly spaced locations around the measured region of the dried droplet using equivalent Z distances as for each sample measurement. All spectra were background-subtracted using an average of the three spatially independent spectra of the aluminium surface for each CSF droplet.

Spectral preprocessing

Preprocessing and multivariate analysis were performed using the IRootLab plugin (0.15.07.09-v) for MATLAB® R2023a (13). High-frequency noise was removed using the Haar-wavelet denoising function with 6 decomposition levels. A fifth-order polynomial was used to remove fluorescence background. The ends of each spectrum were then anchored to the axis using the rubberband-like function. Spectral intensity normalization was applied using vector normalisation and spectra were standardized for outlier removal by principal component analysis (PCA). Identified spectral outliers were removed from the dataset for each patient. PCA was also applied at the level of patient resulting in the removal of 2 patient's spectra from downstream analysis due to chemical contamination of sample (likely from blood during lumbar puncture).

Spectral biomarker feature extraction

The Mann-Whitney U-Test per Raman shift of the spectrum was used to identify the variables that were most different between Non-AD and AD groups using all 3139 spectra in the training dataset. 10 spectral regions were identified, and the area of each region (A) was calculated to reduce the data from each region into a single variable. These 10 variables were used for classification (see classification). Chemical assignment of each region was based on the literature values.

Classification

Samples remaining after the removal of outliers (n=141) were randomly split into a training set (113 patients, 80%) and a testing set (28 patients, 20%) whilst retaining class balance. This resulted in 3139 independent Raman spectra in the training dataset and 766 independent Raman spectra in the testing dataset. Model hyperparameters are shown in Supplementary Table 2. For classification of the complete Raman fingerprint, two different types of classification models were utilized; a convolutional neural network (CNN) and a linear support-vector machine (SVM) via the MATLAB® Classification Learner application. For extracted features (see spectral biomarker feature extraction) a bagged decision tree model was utilized. Feature reduction using PCA (for the CNN and SVM only) and misclassification costs were optimized manually through iterations, followed by hyperparameter tuning using default optimizable settings for each model. Specifically, 5 separate models were generated for the training dataset using a Bayesian optimization procedure of 20 iterations of hyperparameter optimization with 'expected improvement per second plus' for each of the models. The models were trained and validated using the training data and 5 k-fold cross validation. The best optimized model, of the 5 models, was then exported to MATLAB® to retrain a further 10 times, each time shuffling cross-validation groups, to assess model variability by SEM (standard error of the mean). Each of the 10 retrained models were evaluated on the unseen patient spectra (testing set) for classification of the unseen data into 2 distinct groups (Non-AD and AD). Classified spectra were grouped per patient to provide a classification score (%) for each patient based on the number of spectra that correctly classified the patient into AD or non-AD groups. Averaged scores were used to construct the final receiver-operating characteristic (ROC) curves to evaluate the robustness of the classification model generated.

Feature importance

Global LIME (local interpretable model-agnostic explanations) and partial dependence analysis were

performed using the MATLAB® Classification Learner application explain function. Global LIME was performed using all 3139 spectra in the training dataset using a tree model and a kernel width of 0.5. Partial dependence maxima were extracted from the highest score from the partial dependence plot for each feature.

Partial correlation analysis

Partial correlation analysis was performed using IBM SPSS Statistics (v30) software. Adjustments were made for age and sex to find associations between classifier output scores and diagnostic scale of the patient's PLM status (Paris-North Lille Montpellier biomarker scoring system), as well as between extracted features (spectral biomarkers) and log-transformed CSF biomarker levels, specifically Aβ42, total tau, and p-tau levels. The cutoffs for positive results used were; Aβ42 <680 pg/m (ptau >56 pg/ml, total tau >355 pg/ml). Total tau biomarker results were not available for 7 patients, so these patients were removed from partial correlation analysis for total tau. Two-sided $P < 0.05$ was considered statistically significant. We used False Discovery Rate (FDR) analysis to control for multiple comparisons for the extracted features and CSF biomarkers, accounting for 30 comparisons in total. Specifically, the Benjamini-Hochberg method was used with an FDR <5% considered statistically significant. Scatter plots were used to illustrate the correlations.

3 Results

3.1 AD classification using the Raman fingerprint

We are interested in developing a rapid and sensitive diagnostic method for Alzheimer's disease (AD) using biofluid Raman spectroscopy. The aim of the present study is to assess the accuracy at which Raman spectroscopy can classify AD using CSF samples from a mixed clinical cohort. To address this, we measured Raman spectra of 143 CSF samples that were collected by lumbar puncture acquired from a cross-section of patients with neurological symptoms and stored in the University of Southampton Tissue Bank. The study cohort consisted of 67 AD patients and 76 non-AD patients (Table 1) who were diagnosed clinically by neurologists.

We have previously shown that proteins at physiological concentrations can be measured directly by Raman spectroscopy when samples are dried in droplets ≤1 µl using a method known as droplet deposition Raman spectroscopy (DDRS, (14, 15)). To utilize this method, salts and sugars were removed from CSF samples by centrifugal filtration before air drying on an aluminum surface in an array for DDRS. 30 Raman spectra were collected from each dried CSF droplet using a near-infrared laser (785 nm). Samples from up to 6 patients were thawed, desalted, dried and measured per experiment in a blinded fashion. The experimental and analytical workflow for the present study is depicted in Figure 1.

Raman spectra collected from CSF samples were preprocessed using established routines (15-17) and spectral outliers were removed using principal component analysis (PCA, see methods). The CSF Raman fingerprint is expectedly dominated by peaks that can be attributed to proteins (Figure 2A), including the characteristic phenylalanine ring-breathing mode at 1003 cm^{-1}, the extended protein amide III region between 1200-1350 cm^{-1}, and the protein amide I region between 1600-1700 cm^{-1}. Interfering signals from salts and from sugars were removed by the filtration step and were subsequently undetectable by RS (Supplementary Figure 1). 2 CSF samples in the N=143 cohort were identified as outliers (Supplementary Figure 2) and these samples had notably darker colored solution, likely caused by blood contamination during lumbar puncture. Raman spectra from these 2 samples was subsequently excluded from the study analysis resulting in a final spectral dataset from 141 CSF samples.

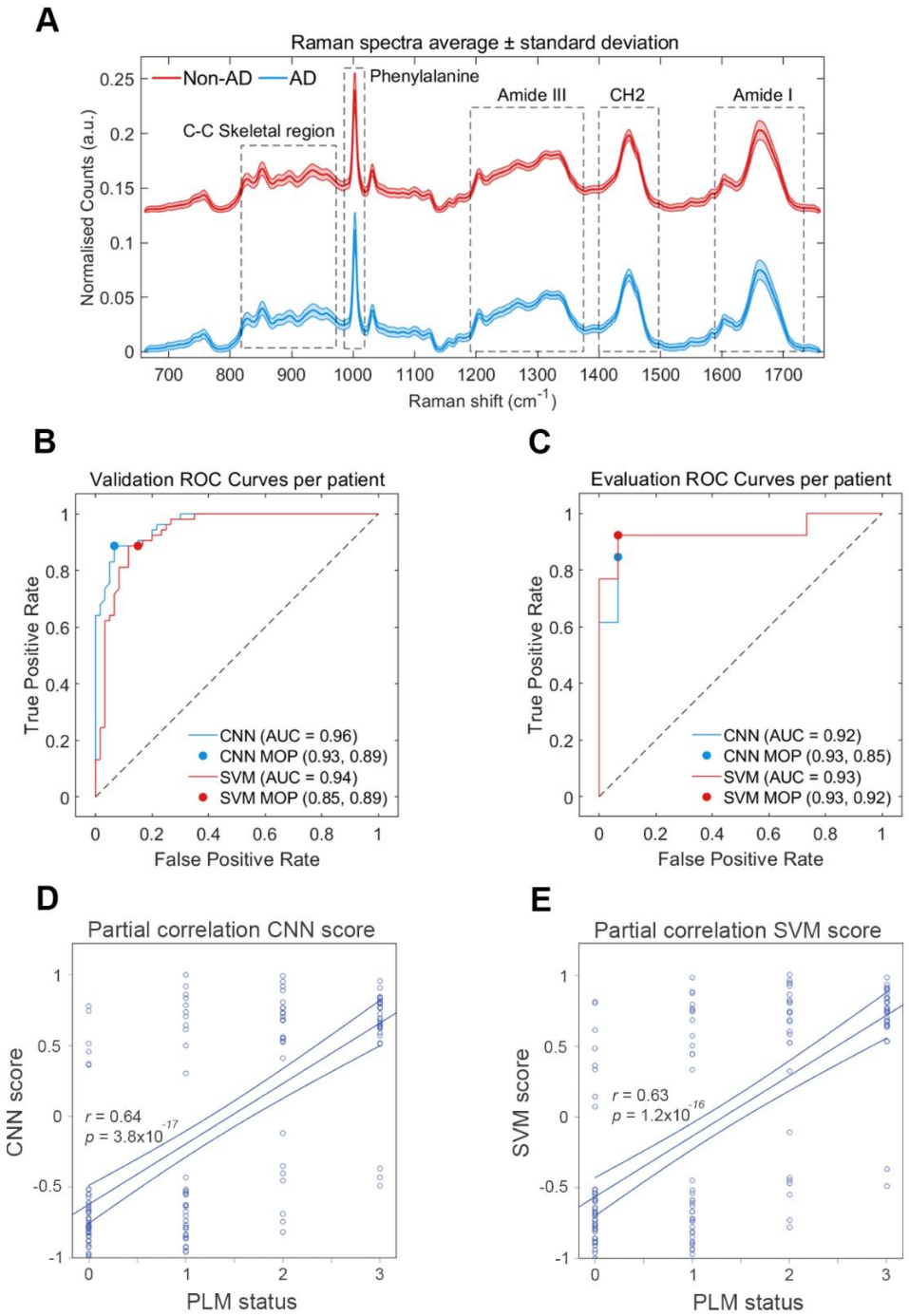


Figure 2. AD classification using the Raman fingerprint

A. *Average Raman spectra collected from CSF samples of 141 patients of whom 66 received a clinical diagnosis of Alzheimer's disease (AD, blue trace) and 75 received a clinical diagnosis other than AD (non-AD, red trace). Key regions of the Raman fingerprint are annotated.* ***B-C.*** *Receiver operating curve (ROC) analysis was performed for validation of the training data (**B**) and evaluation of the testing data (**C**). Raman spectra were separated into training and testing data at an 80:20 ratio at the level of patient. CNN (blue traces) and SVM (red traces) models were trained and optimized on the training data and evaluated on the testing data. Model operating points (MOP) representing sensitivity and specificity cutoffs for each model are depicted as a point on each curve.* ***D-E****. Partial corelation analysis of PLM biomarker status and Raman classification score for each patient CSF sample for CNN (**D**) and SVM (**E**) models controlling for age and sex. 95% confidence intervals are depicted as minor lines on each side of the regression lines.*

As the Raman fingerprint is a superposition of signals, that is, combined chemical information from many molecules, we first wanted to assess classification without limiting the spectral range. To reduce the feature space, we retained 95% of the variance extracted from the whole Raman fingerprint using PCA to train binary (AD vs non-AD) ML models with integrated feature selection algorithms to classify the spectra. The Raman spectra were first partitioned into a training set consisting of spectra from 80% of the CSF samples (113 samples, 3139 spectra), and a testing set consisting of spectra from 20% of the CSF samples (28 samples, 766 spectra). The training set was used to train, validate, and optimize the ML model, whilst the testing set, unseen by the model during the optimization process, was used for model evaluation. We selected two different model architectures – support vector machine (SVM) and convolutional neural network (CNN). The SVM model is an all-purpose classifier that relies on data transformation using the kernel method before linear separation of classes, whilst the CNN can better learn unknown and complex features in datasets for classification but has a higher computational cost. ROC (receiver operating characteristics) curves and the area under the curves (AUC) were used to evaluate the efficacy of the models visually and quantitatively (Figure 2B-C).

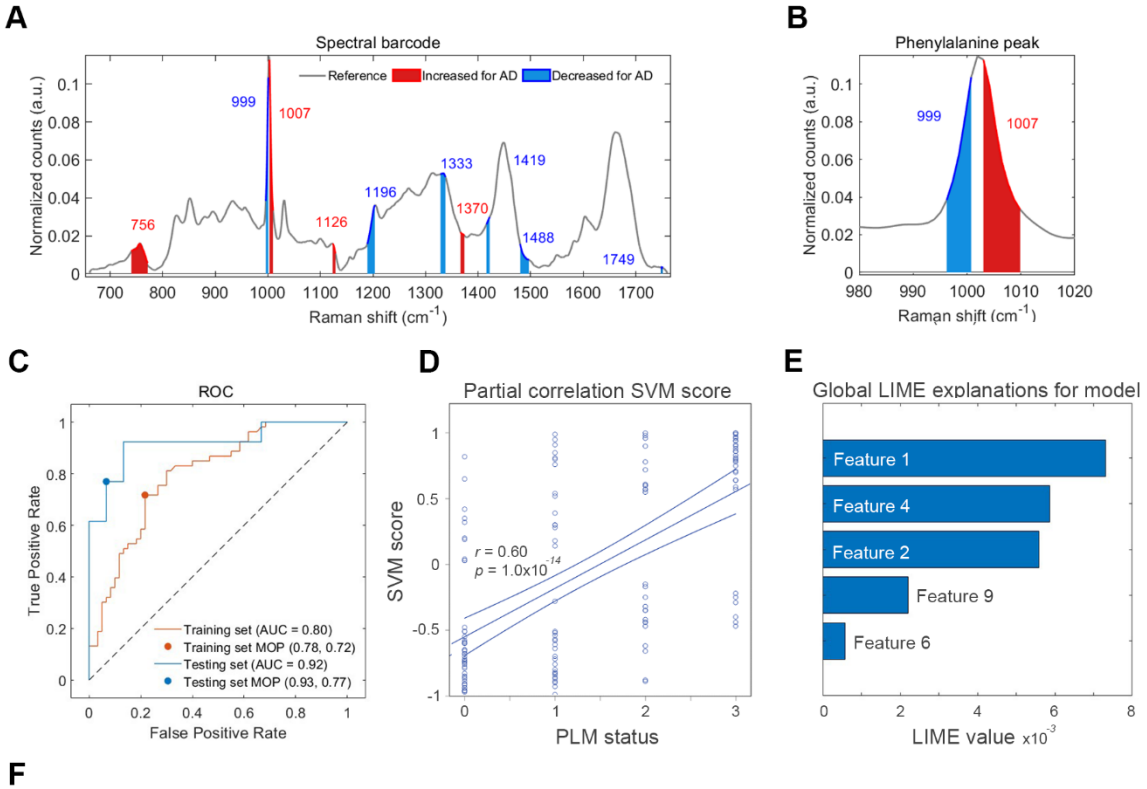
We trained our ML model using all the spectra collected for each CSF sample in the training set after outlier removal to augment our data and improve model training (11). This was followed by the post-hoc binning of classification scores for each spectrum at the level of sample. Each CSF sample was scored based on the number of correctly classified spectra, i.e. if 21/30 spectra for a given sample were correctly classified based on the clinical diagnosis, then the score for that sample would be 0.7 on a scale of 0-1, or 70%. The ROC curves for the optimized CNN and SVM model on the training set is shown in Figure 2B (cross-validation: AUC for CNN = 0.96, for SVM = 0.94). Both CNN and SVM showed similar performance and generalized well to the study cohort as demonstrated by the ROC curves for the testing set (evaluation: AUC for CNN = 0.92, for SVM = 0.93).

We wanted to understand how closely our classification models corelated with Aβ42, p-tau and total tau (A, T, N) biological markers used for *ante-mortem* Alzheimer's disease diagnosis. To do this, we utilized the PLM scale (18) that is used to classify AD based on the presence of these 3 CSF biomarkers; where positivity for all 3 biomarkers indicates very high, 2 biomarkers indicates high, 1 biomarker indicates low and none of the three biomarkers indicates very low likelihood of AD. We correlated the classifier output score for each sample with the corresponding PLM biomarker status (presence of 0, 1, 2, or all 3 of the biomarkers). We controlled for age and sex in each case and observed a positive correlation between the CNN score and PLM biomarker status ($r = 0.64$, $p < 0.001$, Figure 2D), and between the SVM score and PLM biomarker status ($r = 0.63$, $p < 0.001$, Figure 2E).

3.2 Interpretable AD classification using Raman spectral barcode

After demonstrating that AD can be accurately classified using the CSF Raman fingerprint in combination with an optimized CNN or SVM ML model we wanted to identify the regions of the Raman spectrum that were most important for AD classification. This will provide spectral biomarkers to allow the assignment of chemical information and the correlation of spectral biomarkers with ATN biomarker concentrations. This was not possible for the CNN and SVM models as they use integrated feature selection algorithms based on PC scores and are therefore 'black-box' in nature and difficult to interpret.

To develop a more interpretable model for AD classification, we utilized a simple feature selection method developed previously (19) to generate a 'spectral barcode' for AD (Figure 3A). The 10



F

Feature importance									
Feature (U-test rank)	Raman spectral region (cm^{-1})	Chemical assignment	Global LIME score	Global LIME rank	Partial dependence maximum	Partial dependence rank	Partial dependence association	PLM correlation	PLM correlation rank
1	1003 – 1010	Phenylalanine	$7.3x10^{-3}$	1	0.58	7	Positive	0.22*	4
2	996 – 1001	Phenylalanine	$5.9x10^{-3}$	3	0.75	1	Negative	-0.20*	5
3	740 – 771	Tryptophan	-	-	0.57	8	Positive	0.25**	1
4	1328 – 1338	Tyrosine	$5.9x10^{-3}$	2	0.65	2	Negative	-0.16*	7
5	1123 – 1129	C-N skeletal	-	-	0.57	9	Positive	0.20	6
6	1480 – 1496	Histidine	$5.8x10^{-3}$	5	0.61	4	Negative	-0.24*	2
7	1189 – 1203	Tyrosine	-	-	0.58	6	Negative	-0.09	9
8	1416 – 1422	Tryptophan	-	-	0.59	5	Negative	-0.15	8
9	1366 – 1374	Tyrosine	$2.2x10^{-3}$	4	0.64	3	Positive	0.08	10
10	1747 – 1751	Ester	-	-	0.56	10	Negative	-0.22*	3
Multiple adjusted *P* values representing significant correlations are indicated with asterisks: * = $P < 0.05$, ** = $P < 0.01$									

Figure 3. Interpretable AD classification using Raman spectral barcode

A. *Raman spectral barcode generated from identified spectral features evaluated using the Mann-Whitney U-test per Raman shift on the training dataset between non-AD and AD spectra. The average reference spectrum is shown in grey and spectral biomarker regions are shown in red (increased area in AD mean spectrum relative to non-AD) and blue (decreased area in AD mean spectrum relative to non-AD).* ***B.*** *Spectrum enhanced for clarity of the phenylalanine peak including feature 1 (red) and feature 2 (blue).* ***C.*** *Receiver operating curve (ROC) analysis was performed for validation of the training data (orange curve) and evaluation of the testing data (blue curve). Raman spectra were separated into training and testing data at an 80:20 ratio at the level of patient. The bagged decision tree model was trained and optimized on the training data and evaluated on the testing data. Model operating points (MOP) representing sensitivity and specificity cutoffs for each model are depicted as a point on each curve.* ***D.*** *Partial corelation analysis of PLM biomarker status and Raman classification score for each patient CSF sample for the bagged decision tree model controlling for age and sex. 95% confidence intervals are depicted as minor lines on each side of the regression lines.* ***E.*** *Global LIME (local interpretable model-agnostic explanations) analysis to determine feature importance to model classification.* ***F.*** *Table summarising each feature based on analyses. Adjusted P values are displayed after false detection rate (FDR) analysis for 10 correlations using the Benjamini-Hochberg method with FDR <5% considered statistically significant, *P < 0.05, **P < 0.01.*

features extracted from the training data are shown as bars on the Raman spectrum in Figure 3A. Area under the curve was calculated for each spectral region, with red bars representing an increase in feature area in AD spectra relative to non-AD spectra, and blue bars representing a decrease in feature area in AD spectra relative to non-AD spectra. We named the features #1 - #10 and ranked them based on their statistical significance between non-AD and AD spectra, with feature 1 being most significant. Features 1 and 2, which are both related to the frequency of the Phenylalanine peak ~1003 cm^{-1}, were the most significant features between non-AD and AD groups and are shown in isolation for clarity in Figure 3B. We trained a bagged decision tree model for AD classification based on these 10 spectral features or 'spectral biomarkers' and observed acceptable accuracy for the optimized model for spectral classification at the level of sample for the training data (AUC = 0.80) and the testing data (AUC = 0.92). The increased accuracy during evaluation suggests that sample populations within the training and testing sets have different distribution based on these extracted spectral features. There was positive correlation between the decision tree classification score and PLM biomarker status for each patient ($r = 0.60$, $p < 0.001$) (Figure 3D and Table 2), which was only slightly weaker than for the CNN ($r = 0.64$) and SVM ($r = 0.63$) models despite using only 10 features while the latter use all of the variables in the Raman fingerprint.

We assessed the importance of spectral features to the decision tree model using multiple approaches. First, we applied global LIME (local interpretable model-agnostic explanations) analysis, which estimated that feature 1 A(1003 cm^{-1} - 1010 cm^{-1}) assigned to phenylalanine (20), feature 4 A(1328 cm^{-1} -1338 cm^{-1}), assigned to tyrosine (21), and feature 2 A(996 cm^{-1} -1001 cm^{-1}) assigned to phenylalanine (20), contributed the most to AD classification based on an average local feature importance for all spectra (Figure 2E). This is in line with the initial statistical analysis used to generate the spectral barcode as features were numbered in order of statistical significance (feature 1 – feature 10). We next calculated the partial dependence (PD) of each feature to show their individual effect on model prediction (Figure 3F). The top 3 features based on maximum PD score were feature 2 A(996 cm^{-1} -1001 cm^{-1}), feature 9 A(1366 cm^{-1} - 1374 cm^{-1}), assigned to tyrosine (21), and feature 4 A(1328 cm^{-1} -1338 cm^{-1}). The partial dependency plot for each feature revealed a linear relationship between each spectral biomarker and prediction score (Supplementary Figure 3). For each method of analysis feature 2 A(996 cm^{-1} -1001 cm^{-1}), specifically arising from phenylalanine ring-breathing vibrations, was represented within the top 3 features for feature importance and is therefore likely to be important for prediction of AD. Feature 2 A(996 cm^{-1} -1001 cm^{-1}) also correlated with PLM status ($r = -0.20$, $P = 0.03$), although the strongest correlation was observed for feature 3 A(740 cm^{-1} -771 cm^{-1}, $r = 0.25$, $P = 0.009$), which represents a skeletal bond vibration from tryptophan (22). Of the 10 features, 4 further features were correlated with PLM status, including feature 1 A(996 cm^{-1} - 1001 cm^{-1}), feature 4 A(1328 cm^{-1} -1338 cm^{-1}), feature 6 A(1480 cm^{-1} - 1496 cm^{-1}), assigned to histidine (23), and feature 10 A(1747 cm^{-1} - 1751 cm^{-1}), assigned to esters (24).

Figure 4A shows the partial correlation matrix between the 10 extracted features (intensity area, A) and each of the 3 CSF biomarker concentrations (pg/ml) after adjusting for age and sex. After false discovery rate (FDR) correction using the Benjamini Hochberg adjustment, 4 signifant associations were observed of the 30 tested. Feature 4 A(1328 cm^{-1} -1338 cm^{-1}, tyrosine) negatively correlated with p-tau concentration ($r = 0.299$, $P = 0.005$), feature 3 A(740 cm^{-1} -771 cm^{-1}, tryptophan) postively correlated with p-tau concentration ($r = 0.298$, $P = 0.005$), and total tau concentration ($r = 0.259$, $P = 0.023$), and feature 6 A(1480 cm^{-1} - 1496 cm^{-1}, histidine) positively correlated with Aβ42 concentration ($r = 0.257$, $P = 0.020$). Scatter plots depicting the correlations between CSF spectral biomarkers and CSF ATN markers are shown in Figure 4B-E respectively.

A

	Aβ42	p-tau	Total tau
Feature 1 (1003 cm^{-1} – 1010 cm^{-1})	*r* = -0.159 *P* = 0.189	*r* = 0.145 *P* = 0.201	*r* = 0.126 *P* = 0.201
Feature 2 (996 cm^{-1} – 1001 cm^{-1})	*r* = 0.132 *P* = 0.244	*r* = -0.121 *P* = 0.277	*r* = -0.112 *P* = 0.281
Feature 3 (740 cm^{-1} – 771 cm^{-1})	*r* = 0.002 *P* = 0.981	***r* = 0.298** ***P* = 0.005**	***r* = 0.259** ***P* = 0.010**
Feature 4 (1328 cm^{-1} – 1338 cm^{-1})	*r* = 0.036 *P* = 0.720	***r* = -0.299** ***P* =0.005**	*r* = -0.145 *P* = 0.103
Feature 5 (1123 cm^{-1} – 1129 cm^{-1})	*r* = -0.194 *P* = 0.103	*r* = 0.094 *P* = 0.341	*r* = 0.076 *P* = 0.341
Feature 6 (1480 cm^{-1} – 1496 cm^{-1})	***r* = 0.257** ***P* = 0.015**	*r* = -0.201 *P* = 0.103	*r* = -0.158 *P* = 0.188
Feature 7 (1189 cm^{-1} – 1203 cm^{-1})	*r* = 0.103 *P* = 0.316	*r* = -0.047 *P* = 0.650	*r* = -0.051 *P* = 0.430
Feature 8 (1416 cm^{-1} – 1422 cm^{-1})	*r* = 0.091 *P* = 0.347	*r* = -0.107 *P* = 0.315	*r* = -0.105 *P* = 0.281
Feature 9 (1366 cm^{-1} – 1374 cm^{-1})	*r* = 0.031 *P* = 0.743	*r* = 0.122 *P* = 0.277	*r* = 0.133 *P* = 0.236
Feature 10 (1747 cm^{-1} – 1751 cm^{-1})	*r* = 0.102 *P* = 0.316	*r* = -0.158 *P* =0.189	*r* = -0.089 *P* = 0.201

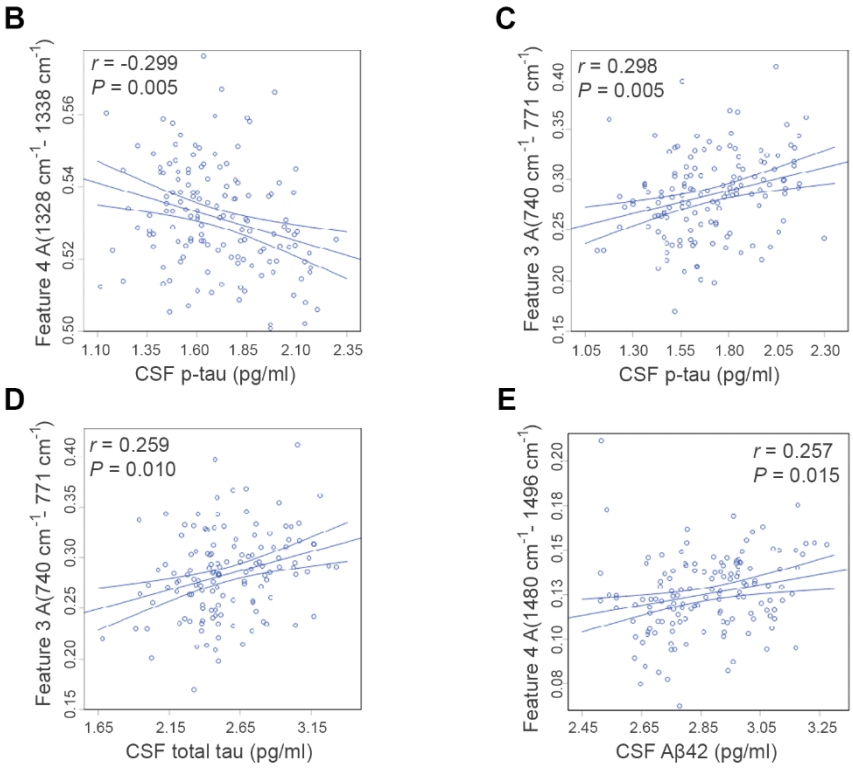


Figure 4. Feature correlation with ATN biomarker concentration

Correlations between extracted Raman features (area within spectral range) and log transformed CSF antibody biomarker concentrations (Aβ42, p-tau, total tau). The partial correlation coefficients (r) were adjusted for age and sex. Adjusted P values are displayed after false detection rate (FDR) analysis for 30 correlations using the Benjamini-Hochberg method with FDR <5% considered statistically significant. ***A.*** *Heatmap matrix for all 30 partial correlations displaying adjusted partial correlation coefficient (r) in blue for positive correlations and red for negative correlations, Adjusted P values are shown for each correlation: *P < 0.05, **P < 0.01.* ***B-E.*** *Scatter plots of extracted Raman features, and log transformed CSF antibody biomarker levels for statistically significant correlations.*

Table 2. Correlation for classification score per model corrected for age and sex					
Model architecture	Model input variables	Correlations			
		PLM biomarker status	A biomarker concentration	T biomarker concentration	N biomarker concentration
CNN	PCs from Fingerprint	0.64*	-0.47*	0.55*	0.41*
SVM	PCs from Fingerprint	0.63*	-0.43*	0.54*	0.41*
Tree	Extracted features	0.60*	-0.42*	0.40*	0.33*
* Indicates corrected P values representing significant correlations at P < 0.001					

4 Discussion

4.1 Summary

In this study, we conducted Raman spectroscopy (RS) analysis of CSF samples from patients in a mixed clinical cohort. Our main aim was to assess the utility of RS for the classification of AD in a population with high clinical heterogeneity and thus reflecting the real-world. ML models were first trained to identify AD using Raman spectral fingerprints collected from 113 CSF samples and then evaluated on previously unseen spectra collected from 28 CSF samples. The optimized models achieved 93% classification accuracy with 26 of 28 patients correctly classified as AD or non-AD during model evaluation. We next extracted features from the Raman fingerprint and used these as spectral biomarkers to develop an interpretable classification model for AD. These spectral biomarkers explained the majority, but not all, of the classification from the Raman fingerprint. Spectral biomarkers for AD primarily arose from protein-derived aromatic amino acid vibrations, the intensity of which directly correlated with biological CSF markers for AD, including Aβ42, phospho-tau 181 (p-tau), and total tau concentrations. Whilst these results support the utility of RS in aiding the clinical diagnosis of AD, larger cross-sectional studies are required to assess whether such high levels of discriminative accuracy are retained at a population level.

4.2 Clinical translation of Raman Spectroscopy for AD diagnosis

To date, two small proof-of-concept studies have used RS approaches to detect AD in human CSF. RS in combination with artificial neural networks and SVM discriminant analysis was used to identify AD with 84% accuracy in a cohort of 37 patients (11), whilst another study utilized a surface-enhanced variant of RS (SERS) in combination with a CNN to identify AD with 92% accuracy in a cohort of 19 patients (10). Despite small sample sizes, these studies provided initial evidence that AD signatures could be detected in CSF using RS. However, distinguishing AD from healthy controls is unlikely to be useful outside of population level screening, for which extracting CSF by lumbar puncture is not appropriate. CSF is often taken for ATN antibody biomarker analysis as part of the diagnostic pathway for AD (3) and RS has the potential to supplement such antibody markers and aid diagnostic decisions as demonstrated in this study. The advantages include; 1. RS is label-free and does not require antibodies or dyes, 2. RS requires very little sample with 40 µl CSF used for filtration and then 1 µl used for RS measurement, 3. RS measurement and classification can be automated and performed in seconds.

Larger proof-of-concept studies have demonstrated that RS can be used to detect AD from healthy controls in blood serum, N=56 (25), N=48 (26), and plasma, N=47 (27), (28). Although none of these studies met all three of the following criteria; interpretable classification, a testing dataset for model evaluation, or n>20. Blood is a much more accessible biofluid for both diagnostics and screening and contains far higher protein levels (60-80mg/ml) than CSF (0.15-6mg/ml), also making it more

amenable for RS analysis. It remains unknown whether RS can accurately classify AD using blood samples from a mixed clinical cohort as demonstrated for CSF in the present study. As well as limited clinical studies, a current barrier to the translation of RS is the size and cost of lasers and spectrometers. In the present study, we used a research-grade system, but system portability is possible (29), allowing large reductions in cost.

4.3 Limitations

For the study cohort, the first limiting factor is sample size (n=141), which is relatively small considering the clinical heterogeneity observed within the population. Further to this, confirmation of AD diagnosis by post-mortem neuropathology was not possible as the work was conducted *ante-mortem*. Therefore, mixed pathologies cannot be ruled out and it is unknown how these may impact the Raman fingerprint. Importantly, rates of accuracy for clinical diagnosis of AD vary depending on disease stage and has been observed at 83% for patients with a clinical diagnosis of probable AD (4). In the present study, pathological ATN biomarkers and in many cases SPECT imaging were used to provide a robust clinical diagnosis for each patient to maximize diagnostic confidence. A subset of the sample cohort has also been classified in a previous research study (N=105) using AD-specific immunogenic inflammatory markers that correlated with total tau and p-tau levels (30), giving further confidence in clinical diagnosis. We also noted that the sex of the sample population was skewed towards male patients (65%). We addressed this by controlling for sex in all partial correlation analyses.

For the sample preparation, CSF samples were filtered using a microcentrifuge and spin filter to remove interference from sugars and salts. Whilst this simplified both measurement and analysis, such sample processing may hinder clinical translation, with cheaper and simpler solutions required such as one-step capillary separation (31). Whilst our correlation analyses suggest that variation in classification score has a molecular basis, it is possible that this can also be impacted by measurement heterogeneity. Much research effort is being directed towards the development of quality control or standards of practice for spectroscopic measurement pipelines (32).

For the analytical method, we required 'black box' classifiers for the most accurate classification of AD samples. These algorithms are prone to overfitting and their lack of interpretability reduces confidence in their application (33). To offset these problems, we further trained an interpretable classification model (34) based on features extracted from the Raman spectrum, i.e. spectral biomarkers. Whilst an increase in error rate was observed for classification, this approach enabled us to assign specific spectrochemical information to AD classification. Whilst these spectral biomarkers may not relate precisely to the features selected by the black-box classifiers, they provide an interpretable explanation of class identity. This enabled us to correlate spectral biomarkers to biological CSF markers resulting in 4 associations between 3 spectral biomarkers and all 3 CSF biomarkers (Aβ42, p-tau, and total tau). It is important to note that peaks in the Raman spectrum are not necessarily independent and can be correlated. They are a composite or *fingerprint* of the vibrations arising from all molecules in a given mixture and are therefore akin to an 'omic signature. We therefore propose that 'spectromic' signatures should be considered holistically during classification, and interpretable classification should be used to uncover associations with known biomarkers supporting mechanistic understanding hence enhancing confidence in classification.

4.4 Conclusion

Our study is the first to show the feasibility of applying the simple, holistic and label-free (unbiased) spectral biomarker approach using RS for dementia diagnosis on a real-world cohort. In conclusion, our study demonstrates that Raman spectroscopy can be used to classify AD with high sensitivity and

specificity using CSF samples from a mixed clinical cohort. Importantly, samples from patients with other neurodegenerative diseases and long-term chronic conditions were clearly distinguished from patients clinically diagnosed with AD supporting the fact that RS provides a biochemical fingerprint, which will be disease-specific because the pathological processes are distinct in different diseases. We showed that Raman spectral biomarkers arising from protein-derived aromatic amino acids primarily explained spectral differences between the AD and non-AD CSF, likely due to proteomic changes specific to AD. Whilst the application of RS to dementia diagnosis remains a research field in relative infancy, our study demonstrates utility in a real-world population and suggests that there is potential for future clinical translation after larger and more diverse clinical studies. Finally, as this method is holistic and does not rely on a specific analyte, the full potential of Raman spectroscopy can be realized if spectral biomarkers are identified that are specific to other diseases that cause dementia including but not limited to vascular dementia (VaD), frontotemporal dementia (FTD) and dementia with Lewy bodies (DLB). Compared to current and emerging methods including blood-based biomarkers, the Raman spectroscopy approach neither requires sophisticated instrumentation or specialized labs nor reagents and can be implemented in a variety of clinical and healthcare settings without the need of skilled operators potentially offering unprecedented scalability and accessibility that can transform dementia diagnosis.

Acknowledgements

The authors give thanks to Dr. Joe Chouhan and Professor Jessica Teeling for their support with sample preparation and the National Academy for Health Research (NIHR) Research Support Service for assistance with study design. Dr. George Devitt is funded through an Alzheimer's Research UK (ARUK) Research Fellowship for this research project (ARUK-RF2022B-010) and has also received research funding from the Institute for Life Sciences (IfLS) at the University of Southampton (UoS). The authors of this study also acknowledge research funding from the Engineering and Physical Sciences Research Council (EP/T020997/1, EP/V038036/1) and the National Institute for Health Research (NIHR Southampton Antimicrobial Resistance Laboratory). CMK is supported by the NIHR Southampton Biomedical Research Centre (BRC).

Conflict of interest statement

The authors declare that they have no conflict of interest. Author disclosures are available in the supporting information

Consent statement

Informed consent was obtained from all patients, and this study was performed in accordance with the ethical standards as laid down in the Declaration of Helsinki. Samples were collected as part of two studies that were approved by the Research Ethics Committee (REC:20/NW/0222 and REC:15/SC/0231). The cohort represents a typical population of a coastal city and a wider referral network comprising rural and urban areas. As such, the results should generalize to a real-world clinical diagnostic setting.

References

1. Lee JC, Kim SJ, Hong S, Kim Y. Diagnosis of Alzheimer's disease utilizing amyloid and tau as fluid biomarkers. Exp Mol Med. 2019;51.
2. Markesbery WR. Neuropathological criteria for the diagnosis of Alzheimer's disease. Neurobiol Aging. 1997;18(4):S13-S9.
3. Jack Jr CR, Andrews JS, Beach TG, Buracchio T, Dunn B, Graf A, et al. Revised criteria for diagnosis and staging of Alzheimer's disease: Alzheimer's Association Workgroup. Alzheimers & Dementia. 2024;20(8):5143-69.
4. Beach TG, Monsell SE, Phillips LE, Kukull W. Accuracy of the Clinical Diagnosis of Alzheimer Disease at National Institute on Aging Alzheimer Disease Centers, 2005-2010. J Neuropath Exp Neur. 2012;71(4):266-73.
5. Barthelemy NR, Salvado G, Schindler SE, He YX, Janelidze S, Collij LE, et al. Highly accurate blood test for Alzheimer's disease is similar or superior to clinical cerebrospinal fluid tests. Nat Med. 2024;30(3):1085-95.
6. Hampel H, Nisticò R, Seyfried NT, Levey AI, Modeste E, Lemercier P, et al. Omics sciences for systems biology in Alzheimer's disease: State-of-the-art of the evidence. Ageing Res Rev. 2021;69.
7. Pontecorvo MJ, Devous MD, Navitsky M, Lu M, Salloway S, Schaerf FW, et al. Relationships between flortaucipir PET tau binding and amyloid burden, clinical diagnosis, age and cognition. Brain. 2017;140:748-63.
8. Devitt G, Howard K, Mudher A, Mahajan S. Raman Spectroscopy: An Emerging Tool in Neurodegenerative Disease Research and Diagnosis. Acs Chemical Neuroscience. 2018;9(3):404-20.
9. Kong K, Kendall C, Stone N, Notingher I. Raman spectroscopy for medical diagnostics - From in-vitro biofluid assays to in-vivo cancer detection. Adv Drug Deliver Rev. 2015;89:121-34.
10. Yu XK, Srivastava S, Huang S, Hayden EY, Teplow DB, Xie YH. The Feasibility of Early Alzheimer's Disease Diagnosis Using a Neural Network Hybrid Platform. Biosensors-Basel. 2022;12(9).
11. Ryzhikova E, Ralbovsky NM, Sikirzhytski V, Kazakov O, Halamkova L, Quinn J, et al. Raman spectroscopy and machine learning for biomedical applications: Alzheimer's disease diagnosis based on the analysis of cerebrospinal fluid. Spectrochim Acta A Mol Biomol Spectrosc. 2021;248:119188.
12. Keshavan A, Wellington H, Chen ZB, Khatun A, Chapman M, Hart M, et al. Concordance of CSF measures of Alzheimer's pathology with amyloid PET status in a preclinical cohort: A comparison of Lumipulse and established immunoassays. Alzh Dement-Dadm. 2020;12(1).
13. Trevisan J, Angelov PP, Scott AD, Carmichael PL, Martin FL. IRootLab: a free and open-source MATLAB toolbox for vibrational biospectroscopy data analysis. Bioinformatics. 2013;29(8):1095-7.
14. Kocisová E, Kuizová A, Procházka M. Analytical applications of droplet deposition Raman spectroscopy. Analyst. 2024;149(12):3276-87.
15. Devitt G, Rice W, Crisford A, Nandhakumar I, Mudher A, Mahajan S. Conformational Evolution of Molecular Signatures during Amyloidogenic Protein Aggregation. ACS Chemical Neuroscience. 2019;10(11):4593-611.
16. Devitt G, Crisford A, Rice W, Weismiller HA, Fan ZY, Commins C, et al. Conformational fingerprinting of tau variants and strains by Raman spectroscopy. Rsc Advances. 2021;11(15):8899-915.
17. Devitt G, Johnson PB, Hanrahan N, Lane SIR, Vidale MC, Sheth B, et al. Mechanisms of SARS-CoV-2 Inactivation Using UVC Laser Radiation. Acs Photonics. 2023;11(1):42-52.
18. Lehmann S, Dumurgier J, Schraen S, Wallon D, Blanc F, Magnin E, et al. A diagnostic scale for Alzheimer's disease based on cerebrospinal fluid biomarker profiles. Alzheimers Res Ther. 2014;6(3).
19. Devitt G, Hanrahan N, Ramirez M, Mudher A, Mahajan S. A Novel Spectral Barcoding and Classification Approach for Complex Biological Samples using Multi-excitation Raman Spectroscopy (MX-Raman). ChemRxiv. 2024.
20. Hernández B, Pflüger F, Kruglik SG, Ghomi M. Characteristic Raman lines of phenylalanine analyzed by a multiconformational approach. J Raman Spectrosc. 2013;44(6):827-33.

21. Hernández B, Coïc YM, Pflüger F, Kruglik SG, Ghomi M. All characteristic Raman markers of tyrosine and tyrosinate originate from phenol ring fundamental vibrations. J Raman Spectrosc. 2016;47(2):210-20.
22. Hirakawa AY, Nishimura Y, Matsumoto T, Nakanishi M, Tsuboi M. Characterization of a Few Raman Lines of Tryptophan. J Raman Spectrosc. 1978;7(5):282-7.
23. Takeuchi H. Raman structural markers of tryptophan and histidine side chains in proteins. Biopolymers. 2003;72(5):305-17.
24. Rothschild KJ, Asher IM, Anastassakis E, Stanley HE. Raman spectroscopic evidence for two conformations of uncomplexed valinomycin in the solid state. Science. 1973;182(4110):384-6.
25. Paraskevaidi M, Morais CLM, Halliwell DE, Mann DMA, Allsop D, Martin-Hirsch PL, et al. Raman Spectroscopy to Diagnose Alzheimer's Disease and Dementia with Lewy Bodies in Blood. ACS Chemical Neuroscience. 2018;9(11):2786-94.
26. Ryzhikova E, Kazakov O, Halámková L, Celmins D, Malone P, Molho E, et al. Raman spectroscopy of blood serum for Alzheimer's disease diagnostics: specificity relative to other types of dementia. J Biophotonics. 2015;8(7):584-96.
27. Carmona P, Molina M, Calero M, Bermejo-Pareja F, Martinez-Martin P, Toledanog A. Discrimination Analysis of Blood Plasma Associated with Alzheimer's Disease Using Vibrational Spectroscopy. Journal of Alzheimers Disease. 2013;34(4):911-20.
28. Habartová L, Hrubesová K, Syslová K, Vondrousová J, Fisar Z, Jirák R, et al. Blood-based molecular signature of Alzheimer's disease via spectroscopy and metabolomics. Clin Biochem. 2019;72:58-63.
29. Hubbard TJE, Shore A, Stone N. Raman spectroscopy for rapid intra-operative margin analysis of surgically excised tumour specimens. Analyst. 2019;144(22):6479-96.
30. Michopoulou S, Prosser A, Kipps C, Dickson J, Guy M, Teeling J. Biomarkers of Inflammation Increase with Tau and Neurodegeneration but not with Amyloid-beta in a Heterogenous Clinical Cohort. J Alzheimers Dis. 2022;89(4):1303-14.
31. Guo WJ, Hansson J, van der Wijngaart W. Synthetic Paper Separates Plasma from Whole Blood with Low Protein Loss. Anal Chem. 2020;92(9):6194-9.
32. Guo SX, Popp J, Bocklitz T. Chemometric analysis in Raman spectroscopy from experimental design to machine learning-based modeling. Nat Protoc. 2021;16(12):5426-+.
33. Wongvibulsin S, Wu KC, Zeger SL. Improving Clinical Translation of Machine Learning Approaches Through Clinician-Tailored Visual Displays of Black Box Algorithms: Development and Validation. Jmir Med Inf. 2020;8(6).
34. Hassija V, Chamola V, Mahapatra A, Singal A, Goel D, Huang KZ, et al. Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence. Cogn Comput. 2024;16(1):45-74.