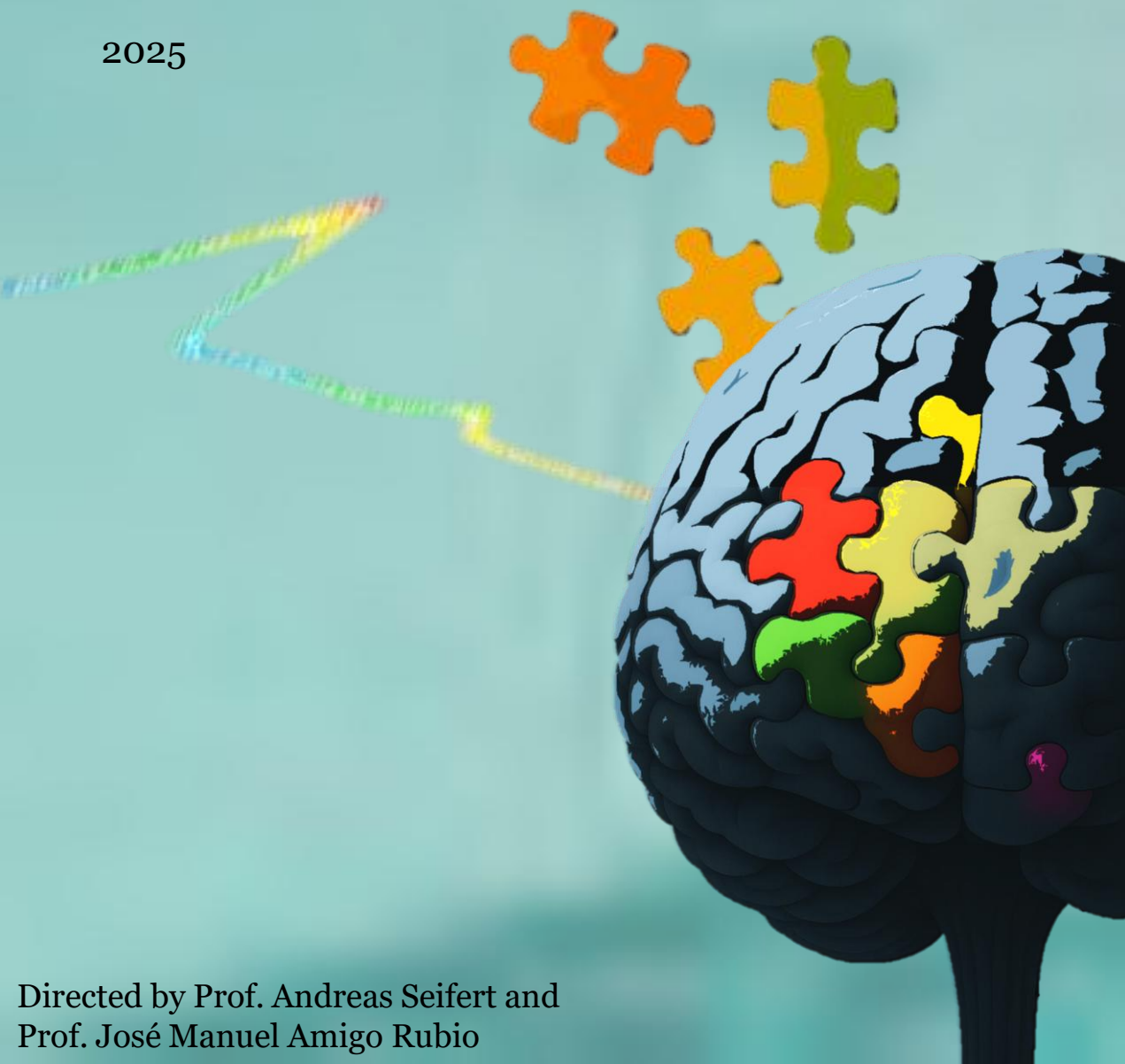
eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

ADVANCES FOR ALZHEIMER'S DISEASE DIAGNOSTICS THROUGH THE INTEGRATION OF SPECTROSCOPY AND CHEMOMETRICS

Eneko Lopez Corrillero

2025



Directed by Prof. Andreas Seifert and
Prof. José Manuel Amigo Rubio

Doctoral Thesis by Published Papers

Advances for Alzheimer's disease diagnostics through the integration of spectroscopy and chemometrics

submitted to
the Department of Analytical Chemistry
of the University of the Basque Country/Euskal Herriko Unibertsitatea

by

Eneko Lopez Corrillero

to obtain the degree
Doctor of Philosophy

supervised by
Andreas Seifert & José Manuel Amigo

Donostia – San Sebastián
2025

CIC nanoGUNE
MEMBER OF BASQUE RESEARCH & TECHNOLOGY ALLIANCE

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

"All prediction models must be validated with respect to realistic future circumstances."

– Paul Geladi, *Journal of Chemometrics*, **2010**.

Acknowledgements

After four years of intense effort, it is time to thank all of those for their generosity, patience, and wisdom. I write these lines for everyone who turned long lab days into something closer to an adventure.

Andreas, my supervisor and boss—thank you for seeing possibility where I saw only uncertainty (a word that *you* love btw). I still remember our first interview, held when I was finishing my master's degree and stepping into the unknown. A PhD was not yet in my plans, yet you opened the door and invited me in. Your steady guidance, intellectual freedom, and willingness to let me balance my current technician job with research have shaped both this thesis and the scientist I am becoming. Thank you!

José, ***Amigo***, my other supervisor and partner in this scientific roller-coaster—*A big big thank you.* By happy coincidence you were on board from the first day. From troubleshooting code to those legendary "tiger times," your advice, good humour, and friendship have been both compass and engine.

Jaione, you worked quietly in the background, while the rest of us were busy looking for Raman peaks, you were the one booking instruments, fixing paperwork tangles... Thank you for keeping the wheels turning and helping our ideas come to life.

Ralph, maybe you think you were not part of this, but everything I learnt from you during my master's made this journey possible. Thank you for always being there! United!

The Nanoengineering family—Ion, Ana, Maica, Gajendra, Mathias, Ander, Kevin, Aitor, Javi, Harun, Maitane, Li, Renata, Maryam, Honghsu, Abhishek, Abdel, and Mounir—you filled the lab and science cafés with expertise, debate, and the kind of laughter that dissolves frustration. Your fingerprints are on every figure and paragraph.

The nanoGUNE technicians et.al—César, Carlos C., Carlos G., Roger, Ane, Juanma, Nerea, Monika, Iban, and Dani—thank you for the hamaiketako times, questionable jokes, and reality checks. You proved that good science runs on caffeine and laughter in equal measure.

Collaborators beyond our walls—at IBeA, Julene, Mireia, and Kepa; the fearless CREW of IASIM 2024: Giulia, Jonan, Oier, Carlos, Reaha and Yasmin; at Vicomtech, Karen, Javier and Laura; and at Fundación CITA-Alzheimer, Maite, Pablo, and Mireia—your openness to share samples, equipment, and ideas expanded the horizon of what was possible.

Friends—you kept me connected to the world outside Raman spectra and data analysis; your insistence on weekends, padel, traveling, and badly sung karaoke ensured my work-balance and adventure alive.

Amaia, love of my life—you know every dataset, every false start, and every small victory because you lived them with me. In truth you could stand here and defend this work yourself. *Eskerrik asko maitia*, for believing in the outcome on the days I doubted the process.

To my family, and the family I gained—for tuppers filled with comfort food, for timely advice that bridged science and life, and for the unspoken promise that home is always one phone call away—this thesis is as much yours as mine. **To my sister and Grandma**— Just thank you, your proud smiles whenever you speak of my work have carried me further than you know. Thank you.

If knowledge were a dog (a Border Collie, like **Micai**), restless, and eager to chase every hint of possibility, then each of you has been the patient trainer whose whistles, gestures, and gentle corrections turned raw energy into clear direction. Thanks to your companionship and support, I know where to run and what to do. I will carry those lessons with me into every new field I enter.

Summary

The main objective of this thesis is to study the use of vibrational spectroscopy with biofluids as a liquid biopsy that is able to detect Alzheimer's disease (AD) years before the onset of cognitive symptoms. A single, label-free spectrum captures hundreds of vibrations of proteins, nucleic acids, lipids and metabolites in a matter of seconds. Therefore, the hypothesis is that these vibrations will be different in healthy people and those with AD.

However, the translation of these techniques to the clinic is challenged by three interrelated obstacles: (i) The variability between the results that different equipment produce and between different cohorts, which complicates the reproducibility of the predictive models. (ii) Model validation protocols that are sometimes too permissive and exaggerate the performance of the models. (iii) The method of measuring samples in a reproducible manner and capturing all the information, especially the most informative chemistry of the disease in question.

This thesis addresses these obstacles and describes a coherent pipeline through three interrelated studies. From an internally validated classifier for preclinical AD through a tutorial of the correct use of machine learning models, and, ending with a spatial mapping of the analyzed samples to better understand the measurement protocols to follow.

The first study demonstrates that Raman spectra, acquired in the outer ring (as stated in the literature) of dried cerebrospinal fluid (CSF) droplets, are able to differentiate cognitively normal volunteers from individuals already in the Alzheimer's continuum (in the preclinical stage) even though they are not presenting symptoms at the time of sampling. Moreover, it was shown that the algorithms were able to differentiate between them even if they belonged to different cohorts and even if the measurements were taken in different years. Forty amyloid beta (Aβ) negative controls and thirty-five asymptomatic Aβ-positive donors belonging to two different years of collection and measurement were enrolled. After Raman spectra collection and prerpocessing, a partial least squares discriminant analysis (PLS-DA) accompanied by variable selection was used as a modeling classifier.

Essentially, the model retained only a set of highly stable bands which together captured preclinical AD. When the same arquitechture was used in completely independent cohorts, the classifier achieved an internal sensitivity and specificity of 93 % and 96 %, respectively.

Research article 1 thus provides translational proof that liquid Raman biopsy holds promise to detect AD pathology years before cognitive decline, but it also exposes two vulnerabilities: model performance is sensitive to validation design, and time-consuming point-and-shoot sampling remains dependent on the operator's own knowledge of "good" measurement points.

The second article addresses the first of these vulnerabilities by dissecting how resampling schemes behave on high-dimensional collinear spectral data. Using three case studies, it is shown that k-fold random cross validation often allows replicate spectra of the same aliquot, or even the same laser spot, to be part of both training and test data. The resulting accuracy inflation can exceed up to 20 %, without modeling any real biochemical effects. This manuscript shows in a didactical manner how important the data structure is when a model is constructed and how easy it is to obtain models that look promising with wrong-designed cross-validation and external validation strategies. A comprehensive overview of the main validation strategies is shown, exemplified by three different scenarios, all of them focused on classification. These lessons are developed further in the Part 2 article now submitted. This second part gives a broader overview of the additional overfitting sources such as, improper variable selection, model selection, data augmentation, fear to publish bias, etc... It proposes an extended, step by step workflow for building trustworthy prediction models.

Even a thoroughly validated model can fail if the sampling strategy overlooks biochemical hot spots. The third study thus shifts from single-point spectra to hyperspectral Raman imaging of whole dry (1uL) droplets. About 10 000 spectra per droplet (both CSF and plasma), acquired at 25 micron resolution, were decomposed by principal component analysis, k-means clustering and multivariate curve resolution. The maps reveal a pronounced but interesting architecture: some components are concentrated at the outer edge; other metabolites migrate to inner zones; and there is another group of

components that form irregular intermediates. Several AD-associated bands identified in the first study of this thesis (1045, 1065, 1453 cm^{-1}) reach their maximum intensity not at the edge of the drop, as guided by traditional Raman in dry droplets, but within these inner zones. By sampling only guided spots at the edge of the droplet, only a biochemical part of the droplet is captured, however, with hyperspectral Raman imaging it is shown that there are other important spots in the dry droplet that should be taken into account when performing this type of measurements. Equally important is the fact that the images validate the biochemical assignments of the first study and explain previous inconsistencies in the community: studies that focused only on the drop edge naturally underestimated the other possible biomarkers.

This thesis elevates Raman spectroscopy from promising curiosity to quantitative, scalable and clinically relevant liquid biopsy, leading with translational evidence (Article 1), reinforcing it with rigorous validation (Articles 2 and 4) and, finally, illuminating the hidden topography of different types of biofluid droplets (Article 3). The three main results:

- A promising classifier that detects Alzheimer's pathology before symptoms appear
- A validation framework that is needed to follow in any spectroscopic classifier
- A spatially resolved chemical map that demonstrates past inconsistencies and guides future high-throughput protocols

Together, they form a clear path for developming vibrational spectroscopy as method for an early warning system not only for neurodegeneration, but for other diseases in which time-critical biochemical changes precede clinical alarms.

Resumen

El objetivo principal de esta tesis es evaluar la espectroscopía vibracional aplicada a biofluidos como biopsia líquida capaz de detectar la enfermedad del Alzheimer (EA) varios años antes de la aparición de los síntomas cognitivos. Un único espectro recoge, en segundos, centenares de vibraciones de proteínas, ácidos nucleicos, lípidos y metabolitos; se espera que dichas vibraciones difieran entre sujetos sanos y personas con la EA.

La traslación clínica de estas técnicas se enfrenta, sin embargo, a tres retos interrelacionados: (i) la variabilidad entre instrumentos y cohortes, que complica la reproducibilidad de los modelos predictivos; (ii) protocolos de validación a veces demasiado permisivos, que inflan el rendimiento de los modelos predictivos; y (iii) la necesidad de medir las muestras de forma reproducible y, al mismo tiempo, capturar la química más informativa de la enfermedad en cuestión.

Esta tesis aborda esos retos mediante tres estudios conectados entre sí. El primero, con un clasificador validado internamente para la detección de la EA preclínica; el segundo, una guía práctica sobre validación de modelos predictivos; y el tercero, un mapeo espacial que aclara cómo medir biofluidos.

El estudio inicial demuestra que los espectros Raman tomados en el anillo externo de una gota seca de líquido cefalorraquídeo (LCR) (tal y como indica la bibliografía relacionada) distinguen voluntarios cognitivamente normales de sujetos ya situados en el continuo de Alzheimer, aún cuando estos últimos no presentaban síntomas en el momento de la extracción. Se incluyeron cuarenta controles Aβ-negativos y treinta y cinco donantes Aβ-positivos asintomáticos de dos campañas de recogida distintas. Tras el preprocesado espectral, se empleó un modelo PLS-DA con selección de variables que identificó y conservó solo las bandas más significantes. En cohortes independientes, el clasificador alcanzó una sensibilidad del 93 % y una especificidad del 96 %.

Este primer artículo aporta una prueba de concepto traslacional, pero deja ver dos puntos a mejorar: el rendimiento depende del diseño de validación y el muestreo punto a punto, lento, sigue condicionado a la pericia del operador.

El segundo trabajo analiza esa debilidad mostrando cómo los esquemas de remuestreo actúan sobre datos espectrales de alta dimensión y alta colinealidad. Con tres casos de estudio se demuestra que la validación cruzada aleatoria permite, con frecuencia, que réplicas de espectros de la misma alíquota (o incluso medidas en la misma posición del láser) entren tanto en los modelos de entrenamiento como de prueba, inflando la exactitud hasta en un 20 %. El artículo expone de forma didáctica la importancia de respetar la jerarquía de los datos y presenta una panorámica completa de las estrategias de validación más utilizadas, centrado en modelos de clasificación.

Un modelo bien validado también puede fallar o no ser tan eficiente si el protocolo de muestreo pasa por alto zonas químicamente ricas. Por ello, el tercer estudio sustituye la medición Raman punto-a-punto por la imagen Raman hiperespectral de gotas secas (1 µL) completas. Se adquirieron unos 10 000 espectros por gota (tanto de CSF como de plasma) con una resolución de 25 µm y se analizaron mediante PCA, k-means y MCR-ALS. Los mapas muestran una arquitectura marcada: algunos componentes se concentran en el borde de la gota; otros migran hacia zonas internas; y un tercer grupo forma estructuras intermedias irregulares. Varias bandas asociadas a la EA (1045, 1065 y 1453 cm^{-1}) alcanzan su máxima intensidad no en el borde, sino en estas zonas internas. Así se explica por qué los estudios centrados solo en las mediciones en el anillo infravaloraban posibles biomarcadores.

En conjunto, esta tesis convierte la espectroscopía vibracional en una biopsia líquida cuantitativa, escalable y clínicamente relevante: primero con evidencia traslacional (Artículo 1), después con una validación rigurosa (Artículos 2 y 4) y, por último, con un mapa químico espacial que aclara inconsistencias previas y orienta futuros protocolos de mayor rendimiento (Artículo 3). Los tres logros principales son:

- Un clasificador prometedor que detecta la patología antes de los síntomas
- Un marco de validación necesario para cualquier clasificador en datos espectroscópicos
- Un mapa químico resuelto espacialmente que guía nuevos métodos de medición

En conjunto, sientan las bases para desarrollar la espectroscopía vibracional como sistema de alerta temprana, no solo en neurodegeneración, sino en otras enfermedades donde los cambios bioquímicos críticos se adelantan a los signos clínicos.

Laburpena

Tesi honen helburu nagusia espektroskopia bibrazionala biofluidotan ebaluatzea da, biopsia likido gisa, zehazki, Alzheimer gaixotasuna (AG) sintoma kognitiboak agertu baino hainbat urte lehenago detektatzeko gai den teknika gisa erabiltzeko. Espektro bakar batek, segundo gutxitan, proteinen, azido nukleikoen, lipidoen eta metabolitoen ehunka bibrazio jasotzen ditu; bibrazio horiek subjektu osasuntsuen eta AG duten pertsonen artean ezberdinak izatea espero da.

Teknika hauen translazio klinikoak, ordea, elkarri lotutako hiru erronkari egin behar die aurre: (i) tresnen eta kohorten arteko aldakortasuna, iragartzeko ereduen erreproduzigarritasuna zailtzen duena; (ii) baliozkotze protokoloak batzuetan permisiboegiak, iragartzeko ereduen errendimendua puzten dutenak; eta (iii) laginak erreproduzitzeko moduan neurtzeko eta, aldi berean, gaixotasunaren kimika informatiboena atzemateko beharra.

Tesi honek erronka horiei heltzen die, elkarri lotuta dauden hiru azterlanen bidez. Lehenengo azterlanak, AG aurreklinikoa detektatzeko barne-baliozkotutako sailkatzaile bat aurkezten du; bigarrenak, predikzio-ereduak baliozkotzeko gida praktiko bat; eta hirugarrenak, biofluidoak nola neurtu argitzen duen mapaketa espaziala.

Hasierako azterlanak erakusten duenez, likido zefalorrakideoko tanta lehor baten kanpoko eraztunean jasotako Raman espektroek (literaturak adierazten duen moduan) kognitiboki normalak diren pertsonak bereizten dituzte Alzheimer jarraituan kokatutako subjektuetatik, nahiz eta azken horiek lagina atera zitzaienean sintomarik ez izan. Laginen bilketa bi kanpaina ezberdinetan gauzatu zen, guztira Aβ negatiboko berrogei kontrol subjektu eta Aβ positiboko hogeita hamabost pertsona parte-hartu zutelarik. Espektroen aurre-prozesamenduaren ondoren, aldagai-hautapenaz baliatzen zen PLS-DA eredua erabili zen, bandarik adierazgarrienak identifikatu eta gordetzen zuena. Kohorte independenteetan, sailkatzaileak % 93ko sentsibilitatea eta % 96ko espezifikotasuna lortu zituen.

Lehen artikuluak kontzeptu translazionalaren proba bat eskaintzen du, baina hobetu beharreko bi puntu agerian uzten ditu: errendimendua balidazio-diseinuaren araberakoa da eta puntuz puntuko Raman neurketak, motel izateaz gain, ikertzailearen trebetasunaren baldintzapean dago.

Bigarren lanak, ahulezia hori aztertzen du, berlaginketa-eskemek dimentsio handiko eta kolinealtasun handiko datu espektraletan nola eragiten duten erakutsiz. Hiru kasu ezberdinetan frogatu da ausazko baliozkotze gurutzatuak aukera ematen duela, maiz, alikota bereko espektruen erreplikak (edo laser-posizio berean egindako neurketak) entrenamendu- eta proba-ereduetan sartzeko, zehaztasuna % 20raino puztuz. Artikuluak modu didaktikoan azaltzen du datuen hierarkia errespetatzearen garrantzia, eta gehien erabiltzen diren baliozkotze-estrategien panoramika osoa aurkezten du, sailkapen-ereduetan zentratuta.

Baliozkotuta dagoen eredu batek ere huts egin dezake, edo ez daiteke eraginkorrena izan, neurketa-protokoloak kimikoki aberatsak diren eremuak kontuan hartzen ez baditu. Horregatik, hirugarren azterlanean, Raman puntu-puntu neurketaren ordez, Raman tanta lehorren hiperespektral (1 µL) osoa jarri zen. Tanta bakoitzeko 10 000 espektro inguru eskuratu ziren (bai CSFkoak, bai plasmakoak) 25 µm-ko bereizmenarekin, eta PCA, k-means eta MCR-ALS bidez aztertu ziren. Mapek arkitektura nabarmen eta berezia erakutsi zuten: osagai batzuk tantaren ertzean kontzentratzen dira; beste batzuk barnealderantz migratzen dira; eta hirugarren multzo batek bitarteko egitura irregularrak eratzen ditu. AG-ri loturiko hainbat banda (1045, 1065 eta 1453 cm^{-1}) ez dute ertzetan lortzen intentsitate handiena, barneko eremu horietan baizik. Horrek azaltzen du soilik eraztunean egindako neurketetan oinarritutako ikerketek balizko biomarkadoreak gutxiesteko joera izatea.

Honela, tesi honek espektroskopia bibrazionala biopsia likido kuantitatibo, eskalagarria eta klinikoki garrantzitsu bihurtzen du: lehenik, ebidentzia translazionalarekin (1. artikulua); ondoren, baliozkotze prozesu zorrotzekin (2. eta 4. artikuluak); eta, azkenik, mapa kimiko espazial batekin, literaturako aurreko inkonsistentziak argitu eta etorkizuneko errendimendu handiagoko protokoloak bideratzen dituena (3. artikulua).

Hiru lorpen nagusiak hauek dira:

- Sintomak agertu aurretik patologia detektatzen etorkizun handiko duen sailkatzaile bat
- Datu espektroskopikoetan edozein sailkatzailerentzat beharrezkoa den balidazko esparru bat
- Neurketa-metodo berriak gidatzen dituen mapa kimiko espaziala

Oro har, espektroskopia bibrazionala alerta goiztiarreko sistema gisa garatzeko oinarriak ezartzen ditu, ez bakarrik neurodegenerazioan, baizik eta aldaketa biokimiko kritikoak zeinu klinikoak baino lehenago gertatzen diren beste gaixotasun askotan ere.

Dissemination of results

Peer-reviewed publications as first author

- Eneko Lopez, Jaione Etxebarria-Elezgarai, Maite García-Sebastián, Miren Altuna, Mirian Ecay-Torres, Ainara Estanga, Mikel Tainta, Carolina López, Pablo Martínez-Lage, Jose Manuel Amigo, Andreas Seifert. Unlocking Preclinical Alzheimer's: A Multi-Year Label-Free In Vitro Raman Spectroscopy Study Empowered by Chemometrics. *International Journal of Molecular Sciences* 25, 9:4737, (2024).

- Eneko Lopez, Jaione Etxebarria-Elezgarai, Jose Manuel Amigo, Andreas Seifert. The importance of choosing a proper validation strategy in predictive models: A tutorial with real examples. *Analytica Chimica Acta*, 1275, 341532, (2023).

- Eneko Lopez, Giulia Gorla, Jaione Etxebarria-Elezgarai, Julene Aramendia, Kepa Castro, Andreas Seifert, Jose Manuel Amigo. Hyperspectral Raman imaging to understand patterns in dried biofluids in Alzheimer's disease. *VIEW*, 20250034, (2025).

- Eneko Lopez, Giulia Gorla, Jaione Etxebarria-Elezgarai, Jose Manuel Amigo, Andreas Seifert. The importance of choosing a proper validation strategy in predictive models. Part 2: Recipes for (avoiding) overfitting. *Analytica Chimica Acta*, *submitted*.

Further publications that do not belong to this PhD thesis

- Ander Bastida, Eneko Lopez, Renata Matekalo, Andreas Seifert. Enhancing Viral Pneumonia Classification through Machine Learning: Data Fusion of FTIR and Raman Spectroscopy. *Microchemical journal*, *in review*.

- Harun Hano, Charles H. Lawrie, Beatriz Suarez, Alfredo Paredes Lario, Ibone Elejoste Echeverría, Jenifer Gómez Mediavilla, Marina Izaskun Crespo Cruz, Eneko Lopez, Andreas Seifert. Power of Light: Raman Spectroscopy and Machine Learning for the Detection of Lung Cancer. *ACS Omega* 9, 12:14084–14091, (2024).

- Jaione Etxebarria-Elezgarai, Luca Bergamini, Eneko Lopez, Maria Carmen Morant-Miñana, Jost Adam, Nerea Zabala, Javier Aizpurua, Andreas Seifert. Amplifying Sensing Capabilities: Combining Plasmonic Resonances and Fresnel Reflections through Multivariate Analysis. *Small Methods*, 8, 2301445 (2024).

- Ion Olaetxea, Hector Lafuente, Eneko Lopez, Ander Izeta, Ibon Jaunarena, Andreas Seifert, Photonic Technology for In Vivo Monitoring of Hypoxia–Ischemia. *Advanced Science*, 10, 2204834, (2022).

- Laura A. Arévalo, Stephen A. O'Brien, Eneko Lopez, Gajendra Pratap Singh, and Andreas Seifert. Design and Development of a Bimodal Optical Instrument for Simultaneous Vibrational Spectroscopy Measurements. *International Journal of Molecular Sciences* 23, 12:6834, (2022).

Previous publications

- Ion Olaetxea, Ana Valero, Eneko Lopez, Héctor Lafuente, Ander Izeta, Ibon Jaunarena, and Andreas Seifert. Machine Learning-Assisted Raman Spectroscopy for pH and Lactate Sensing in Body Fluids. *Analytical Chemistry*, 92 (20), 13888-13895, (2020).

- Jaione Etxebarria-Elezgarai, Miriam Mowat, Eneko Lopez, Carlos Rodríguez, Ion Olaetxea, and Andreas Seifert. Gaussian Beam Shaping and Multivariate Analysis in Plasmonic Sensing. *Analytical Chemistry*, 92 (24), 16236-16244, (2020).

Proceedings

- Javier Plou, Maitane Marquez, Celina-Christin Schubbe, Eneko López, Charles Lawry, Andreas Seifert. Uncovering metabolic gradients evolution using 3D-printed multiwell plates tailored to Raman spectroscopy. In *Proceedings Volume Label-free Biomedical Imaging and Sensing (LBIS) 2025.*

- Maitane Marquez, Javier Plou, Eneko López, Charles Lawry, Andreas Seifert. Real-time monitoring of metabolic gradients using a novel 3D-printed device adapted to Raman spectroscopy. In *Proceedings Volume Label-free Biomedical Imaging and Sensing (LBIS) 2025.*

- Javier Bóbeda, Jaione Etxebarria-Elezgarai, Laura Valeria Perez-Herrera, María Jesús García-González, Eduardo Alonso, Maite García-Sebastián, Miren Altuna, Mirian Ecay-Torres, Ainara Estanga, Mikel Tainta, Carolina López, Eneko Lopez, Pablo Martínez-Lage, Andreas Seifert and Karen López-Linares. Alzheimer's disease classification by artificial intelligence using microscopy images of dried human body fluids. In *InMed-24 (12th International Conference on Innovation in Medicine and Healthcare).*

- Jaione Etxebarria-Elezgarai, Eneko Lopez, Maite García-Sebastián, Miren Altuna, Mirian Ecay-Torres, Ainara Estanga, Mikel Tainta, Carolina López, Pablo Martínez-Lage, Andreas Seifert. Surface-Enhanced Raman Spectroscopy for Early Detection of Alzheimer's Disease. In *2024 International Conference on Optical MEMS and Nanophotonics (OMN).*

- Ion Olaetxea, Eneko Lopez, Ana Valero, Andreas Seifert. Determination of physiological lactate and pH by Raman spectroscopy. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).*

Glossary

Aβ amyloid-beta.

AD alzheimer's disease.

AI artificial intelligence.

ANOVA analysis of variance.

ASCA anova-simultaneous component analysis.

AT(N) A (amyloid plaques), T (tau tangles), and (N) for neurodegeneration.

ATR attenuated total reflectance.

AUROC area under the receiver operating characteristic curve.

BSA Bovin serum albumin.

CN cognitively normal.

CSF cerebrospinal fluid.

CV cross-validation.

DCDR Drop coating deposition Raman spectroscopy.

DRIFT Diffuse reflectance infrared Fourier transform.

FDG fluorodeoxyglucose.

FN false negative.

FP false positive.

FTIR fourier transform infrarred.

GASCA group-wise ASCA.

HPLC-HRMS High performance liquid chromatography-high resolution mass spectrometry.

iPLS interval partial least squares.

IR infrared.

IWG-2 international working group.

MCI mild cognitive impairment.

MCR - ALS multivariate curve resolution - alternating least squares.

ML machine-learning.

MMSE Mini-mental state examination.

MRI magnetic resonance imaging.

NfL neurofilament light.

NIA - AA national institute on aging–alzheimer's association.

NMR Nuclear magnetic resonance.

NPV negative predictive value.

PBS phosphate buffered saline.

PC principal component.

PCA principal component analysis.

PET positron emission tomography.

PLS partial least squares.

PLS-DA partial least squares discriminant analysis.

PPV positive predictive value.

RFE recursive feature elimination.

SCD subjective cognitive decline.

SERS surface-enhanced raman scattering.

SHAP shapley additive explanations.

SNR signal-to-noise ratio.

SR selectivity ratio.

SVM support vector machine.

TN true negative.

TP true positive.

VASCA variable-selection ASCA.

VIP variable importance in projection.

WHO world health organization.

Contents

Chapter 1

General introduction

1.1 Background and motivation

Alzheimer's disease (AD) is more than a medical condition which significantly affects the lives of patients, their families, and caregivers. The gradual progression of AD, starting with mild memory loss and leading to a total loss of independence, deeply affects both patients and those who care for them. Having experienced Alzheimer's disease within my own family, I have directly observed the emotional, physical, and financial challenges it creates. Witnessing a loved one slowly lose their memory, personality, and ability to recognize family members left a lasting impression on me.

AD is a neurodegenerative disease and the most common form of dementia worldwide, recognized by the World Health Organization (WHO) as a global public health priority [1]. The disease develops progressively, beginning with mild cognitive impairment and advancing to severe dementia. Currently, approximately 55 million people worldwide live with AD, and this number is expected to triple by 2050 due to an aging population [2, 3]. The growing number of cases highlights the urgent need for effective methods to diagnose and treat neurodegenerative diseases [4, 5].

Despite considerable progress in neuroscience, the early and accurate diagnosis of AD remains challenging. Traditional diagnostic methods, including positron emission tomography (PET) neuroimaging and cerebrospinal fluid (CSF) biomarker analysis, are effective but have important limitations: they are expensive, invasive, and not widely available for routine clinical practice [6]. Consequently, many individuals receive their diagnosis only after the disease has progressed and caused irreversible harm. Therefore, there is an urgent need for diagnostic tools that are both accurate and accessible.

This challenge has motivated the development of alternative diagnostic methods, especially photonic-based ones. Spectroscopic methods, including Raman and Fourier-transform infrared (FTIR) spectroscopy, show promise for non-invasive and affordable diagnosis [7, 8, 9, 10]. By analyzing the molecular composition of biofluids (such as plasma and CSF), these methods can detect biochemical changes associated with AD without invasive procedures or radiotracers. However, the complexity of spectroscopic data requires advanced computational methods, since subtle spectral variations become diagnostically meaningful only when systematically analyzed [11].

Artificial Intelligence (AI) has gained considerable attention as a rapidly evolving field with wide-ranging applications, including healthcare. Although AI often involves complex ("black-box") algorithms, it broadly refers to computational methods designed to mimic human cognition. An important part of AI is machine learning (ML), which uses data-driven methods to recognize patterns and make predictions.

In spectroscopy-based diagnostics, chemometrics (essentially an ML approach developed within analytical chemistry), has become indispensable for processing and interpreting spectral data, converting complex data into clear diagnostic categories [12]. Unlike many AI-based methods that prioritize prediction over interpretability, chemometrics relies on clear, statistically robust models that directly reflect biochemical changes. Combining spectroscopy and ML is creating new opportunities in biomarker discovery.

Based on personal experience and scientific curiosity, the integration of photonics and artificial intelligence for the early detection of AD is therefore explored in this study. The research is structured as a compilation of four publications, and is organized as follows:

- First, the application of Raman spectroscopy to cerebrospinal fluid (CSF) samples was investigated, and the feasibility of ML-based spectral classification for AD diagnosis was demonstrated.
- Second, a critical gap in the field was addressed through the publication of a tutorial on the importance of proper validation strategies in predictive modeling, ensuring that ML-driven biomarker studies yield reliable and reproducible results.
- Third, the scope was expanded by applying hyperspectral Raman imaging to both CSF and plasma samples, enabling the exploration of spatially resolved spectral patterns to further enhance diagnostic accuracy.
- Finally, a second part of the tutorial (Article 2) which addresses the importance of completely external test set which is not influenced in any step of the modeling workflow was prepared.

These studies contribute to the development of a new spectroscopy-based diagnostic method for AD that is non-invasive, affordable, and scalable. By combining biophotonics, chemometrics, and AI, this research advances the field by suggesting improved approaches for spectroscopic analysis of biofluid samples, supporting future diagnostic development.

1.2 Scope and objectives

Optical spectroscopy, in particular vibrational spectroscopy as Raman, surface-enhanced Raman scattering (SERS), and FTIR, combined with advanced ML, offer a promising strategy for the non-invasive and affordable detection of AD using biofluids such as CSF and plasma. Molecular changes caused by AD, even at very early stages, can be observed in the vibrational spectra of these accessible biofluids. By applying data-driven algorithms, it is possible to identify spectral patterns specific to AD [13]. This hypothesis is supported by evidence indicating that spectroscopic signatures in biofluids reflect AD pathology [10, 14, 15, 16, 17, 18, 19].

The goal of this Ph.D. thesis is to study diagnostic methods for AD that are simpler, safer, and less costly compared to current biomarker-based tests. Therefore, the thesis evaluates the diagnostic potential of spectroscopic methods, like Raman and FTIR, properly validates ML models, and addresses key practical issues relevant to research applications, such as sample variability, measurement standardization, and effective data-processing procedures.

Each of the four publications included in this thesis addresses a specific aspect of this general objective:

- Evaluate Raman spectroscopy for preclinical AD detection (Article 1): Demonstrate that Raman spectroscopy on CSF, combined with ML methods and feature selection, can distinguish preclinical AD patients from cognitively normal individuals. This study provides initial evidence that optical spectroscopy can detect early biochemical changes associated with AD. Specific objectives include identifying Raman spectral markers, optimizing a ML classification model based on partial least squares–discriminant analysis (PLS-DA) combined with feature selection, and performing rigorous cross-validation. A high potential marks the success of this objective for diagnostic accuracy and reproducible spectral markers associated with AD pathology [13].

- Establish robust validation strategies for predictive ML models (Articles 2 and 4): Develop practical guidelines for validating predictive models in spectroscopic diagnostics, ensuring that performance metrics accurately reflect how models will perform with new data [20]. These works help prevent common problems such as overfitting or incorrect validation procedures, supporting the broader reliability of ML in biomedical spectroscopy.

- Investigate hyperspectral Raman imaging in the analysis of biofluids (Article 3): This study [21] extends traditional Raman spectroscopy to hyperspectral imaging of biofluid samples and evaluates whether hyperspectral imaging, by revealing spatial patterns or localized biochemical signals, can provide additional diagnostic value compared to standard spectral analysis.

Collectively, these objectives explore the feasibility and reliability of photonic-based methods for AD diagnosis, connecting ML model validation procedures (Objective 2) with practical spectroscopic applications (Objectives 1 and 3).

Significance of the study

The increasing age of the global population leads to a growing number of neurodegenerative diseases. Alzheimer's disease (AD), which accounts for 60 – 70% of all dementia cases [22, 23], represents a major health and social challenge. Early and accurate diagnosis is essential, as timely detection can slow disease advancement, enhance patient outcomes, and allow patients to participate in clinical trials for new treatments [24, 25].

Most research on spectroscopy-based AD diagnostics focuses on patients who are already experiencing moderate or advanced disease symptoms. In contrast, this thesis is among the first to systematically assess biofluid spectroscopic signatures from preclinical AD cases (individuals who exhibit biomarker-confirmed pathology but no visible symptoms). Identifying AD at this very early stage could facilitate timely interventions, potentially delaying cognitive decline and improving the quality of life.

Additionally, hyperspectral Raman imaging is introduced for biofluid analysis in AD, a method that has not been extensively investigated previously. Early results indicate that spatial biochemical patterns in dried biofluid samples could offer improved diagnostic accuracy compared to traditional single-point Raman spectroscopy relying on the coffee-ring effect.

Integrating spectroscopy with ML offers a promising new direction in biomedical diagnostics [26]. Traditional AD diagnostics typically involves expensive and invasive methods. In contrast, this research demonstrates that a simple, non-invasive, and scalable alternative is achievable. The findings presented here also pave the way for future research combining multiple spectroscopic methods, such as Raman, FTIR, and SERS, to create even stronger and more reliable classification models.

In summary, this thesis advances research towards developing accessible, cost-effective diagnostic methods for neurodegenerative diseases, addressing an important need in healthcare and biomedical research.

1.3 Organization of the thesis

The first three research articles in this thesis are structured progressively, with each study building upon the findings of the previous one. The fourth article expands the focus of the second one more related to overfitting.

- Article 1

  In the first study, label-free Raman spectroscopy of dried CSF droplets is explored as a potential method for detecting preclinical AD (asymptomatic, early-stage disease). This work demonstrates that by integrating Raman spectral analysis with ML-based feature selection, subtle biochemical changes in CSF can be detected years before clinical symptoms emerge, distinguishing healthy individuals from preclinical AD with high cross-validated accuracy (up to 96%) [13].

  Raman spectroscopy provides a molecular fingerprint of CSF composition [27], and complemented with chemometric analysis identifies diagnostically relevant vibrational features associated with early AD pathology. To enhance model robustness, spectra from two independent cohorts were merged, increasing sample size and generalizability beyond previous proof-of-concept studies [13].

  This initial success raised a critical question regarding the predictive models: How can we confirm that these models generalize well beyond our specific datasets and are not simply overfitting? This concern became the main focus of the second study.

- Article 2

  The second article addresses a common issue in spectroscopy-based diagnostics: machine learning models trained on limited datasets often perform correctly initially, but have difficulty generalizing beyond the original data. To address this issue, Article 2 provides practical guidance on correctly validating chemometric models, illustrated with examples from spectroscopic data. It explains best practices, including unbiased train/test splits, cross-validation, and methods to prevent overfitting so that the models provide realistic and reliable predictions [20].

Proper validation is essential because small datasets and improper validation methods can exaggerate the performance of diagnostic models. Article 2 highlights common mistakes, such as insufficient sample sizes and biased validation methods, and offers straightforward strategies to avoid these issues. It demonstrates how improper validation can lead to misleadingly high classification performance and how correct validation ensures that models remain reliable in real-world applications.

By establishing clear validation guidelines, this tutorial not only strengthens the findings of Article 1 but also supports proper validation of the more complex hyperspectral models presented in Article 3.

- Article 3

  The third article expands spectroscopic diagnostics beyond standardized single-point spectral analysis of the coffee-ring region to hyperspectral Raman imaging. Instead of analyzing individual Raman spectra from dried CSF droplets (as in Article 1), spatial variations within biofluid samples were explored to determine whether they provide additional diagnostic information.

  Using hyperspectral Raman imaging, spectral maps are generated, where each pixel contains a Raman spectrum, which allows the analysis of biochemical distributions across CSF and blood plasma samples. Rather than merely distinguishing early-stage AD from healthy controls, this approach enables the spatial identification of specific regions within the biofluid samples where the signal-to-noise ratio (SNR) of spectral peaks linked to AD pathology is most prominent. Selecting these optimal zones can subsequently enhance the sensitivity and specificity of Raman spectroscopy for diagnostic purposes.

  Article 3 builds on key findings from the previous studies:

  – From Article 1: Chemically relevant AD biomarkers help interpret spectral-spatial patterns.

  – From Article 2: Reliable validation methods prevent overfitting also for image-based ML models.

In this study, the reliability and consistency of Raman-based biomarkers are assessed by analyzing AD spectral signatures across different biofluids. Additionally, key practical factors, such as sample preparation and instrumentation settings, that influence the potential use of Raman spectroscopy for medical diagnostics are examined.

- Article 4

  The fourth (and most recent) article takes Article 2 a step further by focusing on overfitting itself. Using six real or synthetic datasets, it shows the most common pitfalls when building chemometric models, like: biased train/test splits, information leakage during variable selection, improper data fusion and augmentation, "fear to publish" practices which hide negative results. For each of the pitfalls the article quantifies the damage and then walks the reader through the correct path.

  Article 4 turns the warnings from Article 2 into a clear, step-by-step guide which provides the safeguards required before tackling the larger hyperspectral and multimodal datasets discussed in Article 3.

In sum, this thesis shows the progress from single-spectrum analysis (Article 1) → validation principles (Articles 2 and 4) → spectral imaging (Article 3), systematically refining the spectroscopic approach for AD diagnostics. The combined findings offer a structured, data-driven evaluation of whether vibrational spectroscopy can be a practical, non-invasive tool for early AD diagnostics.

1.4 Theoretical framework

1.4.1 Alzheimer's disease: overview and diagnostics

Dementia is a general term for a group of progressive neurodegenerative syndromes that reduce cognitive function and interfere with daily life [1]. It affects over 55 million people worldwide as of 2020, with cases projected to nearly double every 20 years due to global aging [2, 22, 23].

There are many types of dementia, each with distinct underlying causes. AD is the most common form, accounting for approximately 60 – 70% of cases [1]. Other prevalent forms include vascular dementia (5 – 10%) [28], dementia with Lewy bodies (5 – 15%) [29], and frontotemporal dementia (about 5%, though higher in early-onset cases) [30, 31]. Many patients also present with mixed dementia, where multiple pathological processes coexist, such as AD with vascular damage or Lewy body pathology, particularly in older adults [1, 32]. Due to its high prevalence, AD remains the primary focus of this thesis.

AD was first described by Alois Alzheimer in 1906 and is neuropathologically characterized by two key protein aggregates: extracellular amyloid-β(Aβ) plaques and intracellular tau neurofibrillary tangles. These peptide and protein accumulations contribute to synaptic dysfunction, neuronal loss, and progressive brain atrophy, ultimately leading to cognitive decline [33, 34, 35].

The disease progresses through three overlapping clinical stages (Figure 1.1):

- **Preclinical AD**: Pathophysiological changes start years before clinical symptoms, while the individual remains cognitively normal. This stage is detectable only through biomarkers in research settings [33, 36, 37].
- **Mild cognitive impairment (MCI) due to AD**: Early cognitive difficulties emerge, most commonly affecting short-term memory, but they do not yet significantly interfere with daily activities [36].
- **Alzheimer's dementia**: Cognitive decline progresses to a severe stage, affecting daily function and independence [36].

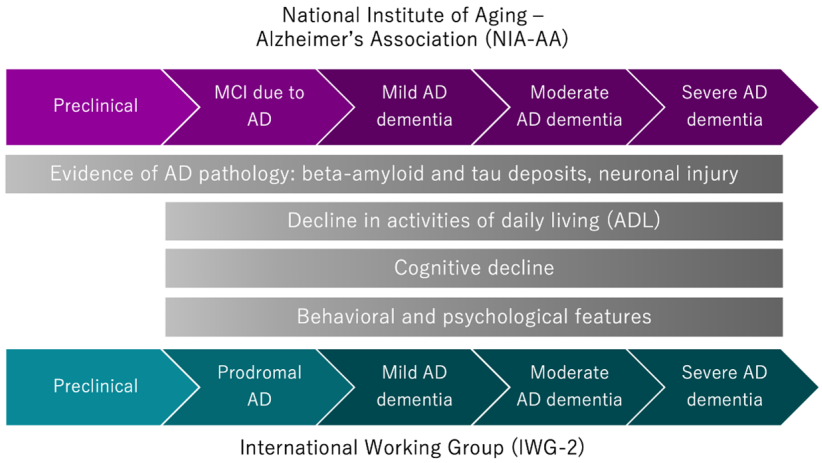


Figure 1.1: The progression of Alzheimer's disease as defined by the National Institute on Aging-Alzheimer's Association (NIA - AA) and the International Working Group (IWG-2). The figure illustrates the three overlapping clinical stages of AD: preclinical AD, mild cognitive impairment (MCI) due to AD, and Alzheimer's dementia, along with the associated pathophysiological and cognitive changes at each stage. Figure adapted from [24].

As AD progresses, individuals experience language difficulties, executive dysfunction, spatial disorientation, and behavioral changes. In its late stages, the disease leads to severe cognitive and physical decline, ultimately resulting in a loss of autonomy and the need for full-time care [38]. While clinical symptoms define these stages, modern research highlights the importance of identifying AD through biomarkers, allowing for earlier detection and potential intervention before irreversible damage occurs [39, 40, 41, 42].

Early detection and biomarkers

Early diagnosis of AD is increasingly recognized as essential for guiding patient care and advancing the development of disease-modifying therapies [39, 25, 43]. Research indicates that pathological changes associated with AD begin years (possibly decades) before clinical symptoms appear, presenting an important opportunity for early detection and intervention [6, 37]. Identifying these changes at an early stage could enable timely therapies to slow disease progression and preserve cognitive function for as long as possible.

Biomarker advancements have revolutionized early AD detection, even at preclinical stages [40, 42]. Diagnosis is now guided by a combination of clinical evaluation, cognitive testing, and biomarker assessments [33, 44, 45]. The core cerebrospinal fluid (CSF) biomarkers; decreased $A\beta_{42}$ and elevated total tau (t-tau) and phosphorylated tau (p-tau), are strongly correlated with AD pathology [46, 47, 44]. These biomarkers can be detected well before significant cognitive decline, supporting their role in early diagnosis [6, 36].

A for amyloid plaques, T for tau tangles and (N) for neurodegeneration (AT(N)) form the AT(N) research framework [33, 48, 49], who recently updated to incorporate the latest advances in biomarker research and diagnostic criteria [36], classifies biomarkers into three categories:

- **A (Amyloid pathology)**: Represents the presence of amyloid-β plaques, a hallmark of AD. It is identified through reduced CSF $A\beta_{42}$ levels or positive amyloid-PET scans [37, 34]. Amyloid pathology often appears decades before clinical symptoms, marking the earliest detectable stage of AD. While its presence is essential to define AD, it does not necessarily correlate with the severity of cognitive decline [33, 36].
- **T (Tau pathology)**: Reflects the accumulation of hyperphosphorylated tau protein, which forms neurofibrillary tangles. Tau pathology is assessed via elevated CSF p-tau or positive tau-PET scans [37]. Unlike amyloid deposition, tau pathology strongly correlates with neuronal damage and clinical symptoms. The presence of both amyloid and tau biomarkers (A+T+) is required to biologically define AD, as tau plays a central role in neurodegeneration [33, 36].
- **N (Neurodegeneration)**: Indicates neuronal injury and brain atrophy, assessed through structural MRI, FDG-PET hypometabolism, or elevated CSF t-tau [37, 50]. Neurodegeneration is not specific to AD, as it can result from other neurodegenerative conditions. However, in the context of A+T+ profiles, neurodegeneration reflects disease progression, with greater atrophy correlating with more severe cognitive impairment [48, 36].

Table 1.1: AT(N) classification of Alzheimer's disease based on amyloid (A), tau (T), and neurodegeneration (N) biomarkers.

AT(N) profiles	Biomarker category	
A-/T-(N)-	Normal AD biomarkers	
A+/T-(N)-	Alzheimer's pathologic change	**Alzheimer's continuum**
A+/T+(N)-	Alzheimer's disease	
A+/T+(N)+	Alzheimer's disease	
A+/T-(N)+	Alzheimer's change and concomitant suspected non-AD pathologic change	
A-/T+(N)-	Non-AD pathologic change	
A-/T-(N)+	Non-AD pathologic change	
A-/T+(N)+	Non-AD pathologic change	

The AT(N) framework provides a standardized approach to categorizing individuals at different stages of AD based on biomarker profiles, such as, amyloid (A), tau (T), and neurodegeneration (N), rather than relying on clinical symptoms alone [4, 36, 44]. By focusing on objective biological markers, it enables more precise staging, earlier risk identification, and improved stratification before cognitive decline becomes apparent. Table 1.1 summarizes the various profiles within the AD continuum, defined by the presence (+) or absence (−) of A, T, and N markers. Individuals with A+T+ profiles are considered to have full AD pathology, while A+T− may indicate an earlier preclinical stage. This classification also distinguishes AD from non-Alzheimer's pathologies, refining diagnostic accuracy and supporting both clinical and research efforts [24].

Despite their diagnostic utility, current CSF and imaging biomarkers have notable limitations, including invasiveness (lumbar puncture), high cost, and limited accessibility to PET imaging [35, 51]. These constraints make routine implementation challenging, particularly in primary care and resource-limited settings. As a result, research efforts are increasingly focused on identifying novel, less invasive biomarkers that can provide comparable diagnostic accuracy with greater ease of use. Emerging alternatives, such as blood-based biomarkers and spectroscopic techniques, hold promise for broader clinical application, potentially allowing for early detection in a wider population [44, 52].

Emerging blood-based biomarkers

The latest guidelines emphasize the necessity of integrating biomarker-based diagnostics into routine practice, ensuring that advancements in research translate into earlier detection and more targeted therapeutic strategies [36]. Blood-based biomarkers offer a minimally invasive alternative for AD detection, made possible by advances in ultra-sensitive assays [53, 54, 55]. Compared to CSF and PET imaging, plasma biomarkers provide a more practical and scalable approach, facilitating large-scale screening, early diagnosis, and improved access to disease-modifying treatments [43, 44, 56].

Recent research has integrated plasma biomarkers into the AT(N) framework, recognizing their ability to reflect core pathological processes in AD [43, 44]. Three main classes of blood-based biomarkers have shown strong diagnostic utility:

- **Phosphorylated tau (p-tau)**: Plasma p-tau, particularly p-tau$_{217}$ and p-tau$_{181}$, strongly correlates with tau pathology and exhibits high diagnostic accuracy, comparable to CSF and PET-based methods [43, 57]. Studies have demonstrated that plasma p-tau$_{217}$, measured via mass spectrometry effectively predicts amyloid and tau positivity, making it a promising biomarker for research and clinical application [43].
- **Amyloid-β42/40 ratio**: A reduced plasma Aβ42/40 ratio indicates amyloid pathology, predicting amyloid-PET positivity [56]. Although plasma Aβ levels exhibit greater variability than CSF, recent assay sensitivity advances have improved their reliability [43]. Combining amyloid biomarkers with p-tau enhances diagnostic accuracy, providing a more comprehensive picture of AD pathology [54].
- **Neurofilament Light (NfL)**: A marker of neuronal injury and neurodegeneration, with elevated blood levels correlating with brain atrophy and cognitive decline [58]. While NfL is not specific to AD, its increase reflects the extent of neurodegeneration and may be useful in monitoring disease progression, particularly in combination with other biomarkers [43].

1.4.2 Optical spectroscopy for AD diagnostics

Introduction to vibrational spectroscopy

Spectroscopy examines the interaction between electromagnetic radiation and matter, offering a powerful means to analyze sample composition by measuring how light is absorbed or scattered [9, 59].

In vibrational spectroscopy, incident photons interact with molecular bonds, leading to characteristic energy absorption or scattering that reveals molecular vibrational frequencies. Each molecule has a unique spectral fingerprint based on its chemical bonds and functional groups, allowing complex mixtures like biological fluids to be analyzed label-free. Importantly, these spectral signatures provide molecular-level insights into biomolecules such as lipids, proteins, and metabolites, making spectroscopy a valuable tool for biofluid analysis [7, 60].

Photonics-based spectroscopic techniques offer rapid, minimally invasive assessment of biofluids, including blood, CSF, saliva, and urine. By analyzing how light interacts with a biofluid sample, one can obtain a comprehensive biochemical profile without extensive sample preparation [9, 61]. Unlike conventional biomarker tests that target individual analytes, vibrational spectroscopy captures a holistic biochemical snapshot, which is particularly relevant for complex diseases like AD, where multiple biomarkers and pathways evolve simultaneously.

Infrared (IR) absorption detects vibrational modes of polar bonds, while Raman scattering identifies vibrational modes through changes in polarizability. These techniques provide complementary molecular information, as certain vibrational modes that are weak in IR can be strong in Raman and vice versa [59, 62]. Their combined use is particularly beneficial for analyzing biological samples containing lipids, proteins, nucleic acids, and carbohydrates [61, 63]. Spectroscopic methods have already been successfully applied to tissues, cells, and biofluids, enabling the detection of biochemical changes associated with disease [60, 64].

In particular, spectroscopy of easily accessible biofluids is gaining attention as a potential diagnostic approach for neurodegenerative diseases like AD due to its combination of sensitivity, speed, and non-destructive nature [10]. The following sections describe key vibrational spectroscopy techniques relevant to biofluid analysis and AD diagnostics, including Raman spectroscopy, Fourier-transform infrared (FTIR) spectroscopy, and surface-enhanced Raman scattering (SERS) spectroscopy .

Fourier-transform infrared (FTIR) spectroscopy

FTIR spectroscopy is a type of infrared (IR) absorption spectroscopy that employs an interferometer to obtain a broad spectral profile rapidly, enhancing signal-to-noise ratio (SNR) and throughput [9]. The different processes of IR absorption are illustrated in Figure 1.2. In IR spectroscopy, molecules absorb specific frequencies of IR light that correspond to their vibrational modes, causing transitions between vibrational energy levels. Absorption occurs when a vibrational mode involves a change in the molecule's dipole moment. Each type of bond and functional group absorbs at characteristic frequencies and produces distinct peaks in the IR spectrum.

Biofluid analysis using FTIR frequently incorporates attenuated total reflectance (ATR) technology (Figure 1.2b), which enables direct measurement of liquid or dried samples with minimal preparation [9, 65]. In ATR-FTIR, an IR beam passes through a high-refractive-index crystal (e.g., diamond or germanium) in contact with the sample. An evanescent wave penetrates a short distance (0.5 – 5 µm) into the sample, allowing spectral acquisition even for opaque or highly absorbing biofluids such as serum or plasma.

FTIR spectroscopy has been widely used for analyzing blood and other biofluids to identify disease-related biomarkers [10, 66, 67]. Notably, it has proven effective in detecting broad biochemical alterations associated with various diseases. In AD research, IR spectra of blood plasma and serum have revealed differences in protein secondary structure and lipid composition between patients and healthy controls [14, 66].

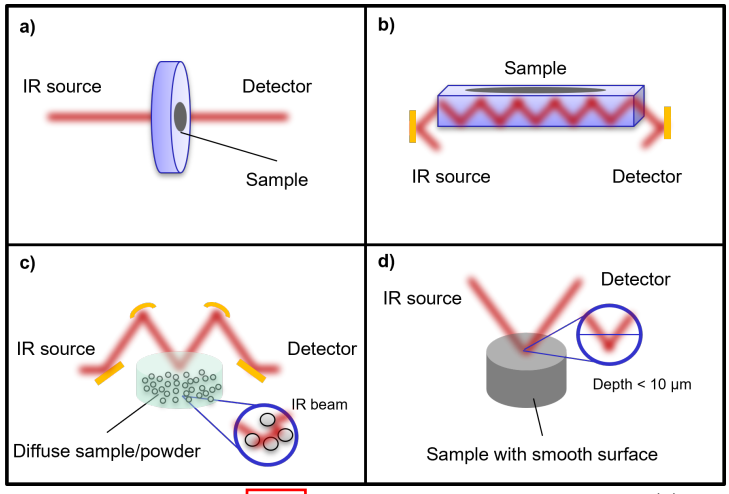


Figure 1.2: Illustration of different FTIR spectroscopy measurement modes: (a) Transmission FTIR, where IR light passes through the sample and absorption is measured; (b) ATR-FTIR, where an evanescent wave due to total internal reflection interacts with the sample surface via a high-refractive-index crystal; (c) Diffuse reflectance infrared Fourier transform (DRIFT) spectroscopy, which captures scattered IR light from powdered or inhomogeneous samples; and (d) specular reflectance FTIR, where IR light reflects directly from a smooth sample surface, useful for thin films and coatings.

Raman spectroscopy

Raman spectroscopy is based on inelastic scattering of photons, first discovered by L. Mandelstam and G. Landsberg, in 1926, before C. V. Raman and K. S. Krishnan independently observed it in 1928 [68]. When monochromatic light (usually from a laser) irradiates a sample, most photons are elastically scattered (Rayleigh scattering), but a small fraction interacts with molecular vibrations and emerges with shifted energy (frequency), this is, the Raman-scattered light. As depicted in Figure 1.3, the Raman shift (difference between incident and scattered photon energy) corresponds to a vibrational mode of the molecule. If the scattered photon loses energy (shifts to a lower frequency), the process is called *Stokes* scattering; if it gains energy (shifts to a higher frequency), it is termed *anti-Stokes* scattering, the latter being weaker. The intensity of a Raman band is proportional to the incident-light intensity and to the square of the change in molecular polarizability during the vibration [59].

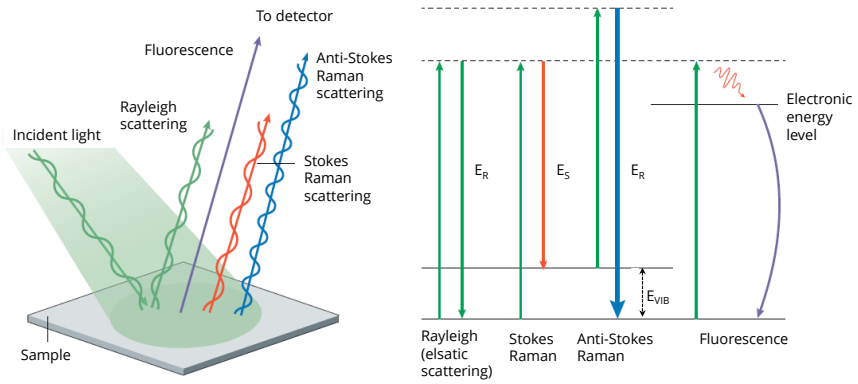


Figure 1.3: Excitation of the sample molecules with an incident laser results in different types of interactions. Momentary absorption of a photon excites the molecule from a ground state into a virtual state. Transition from this virtual state generates a new scattered photon. Incident photons might be reflected with the same energy E_R (Rayleigh or elastic scattering) or can alternatively lose E_S or gain E_{AS} some energy (Raman or inelastic scattering). Raman scattering coming from a molecule that is initially in a ground vibrational state is called Stokes Raman, while scattering coming from molecules in an excited state is known as anti-Stokes Raman. When an electronic transition is resonantly excited, partial energy loss during relaxation produces red-shifted fluorescence. Figure is adapted from [69].

The Raman scattering intensity (I_R) follows:

$$I_R \propto \nu^4 I_0 N \left(\frac{\partial \alpha}{\partial Q}\right)^2, \tag{1.1}$$

where ν is the incident frequency, I_0 the laser intensity, N the number of molecules, and $(\partial\alpha/\partial Q)$ the change in molecular polarizability during vibration.

The ν_0^4 dependence indicates that shorter wavelength (higher frequency) lasers can produce stronger Raman signals, although in practice one must balance this against sample fluorescence and photodegradation [70]. Indeed, biological samples often exhibit autofluorescence under visible excitation, which can hide the weak Raman signals; using near-infrared lasers (e.g., 785 nm or 830 nm) can reduce fluorescence but reduces Raman scattering intensity [70, 7].

A key advantage of Raman spectroscopy is its molecular specificity: each biomolecule (such as proteins, lipids or sugars) produces distinct Raman bands based on its molecular structure [63]. Depending on their initial vibrational state, molecule excitation will give raise to Stokes (loss of energy) or anti-Stokes Raman scattering (gain of energy) [63]. The Raman spectrum of a biofluid thus represents an intrinsic biochemical fingerprint of that fluid. Importantly, water has a very weak Raman signal, allowing direct analysis of aqueous solutions like blood plasma or CSF with minimal interference [71], contrary to FTIR, which is highly susceptible to the presence of water and requires specific sample handling.

The versatility of Raman spectroscopy has led to its application in various biomedical fields, including real-time monitoring of biochemical changes in stored blood products [71], species differentiation in forensic investigations [72], and quantification of clinically relevant analytes such as lactate and pH [73]. Additionally, it has been used to detect inorganic salt components in biological samples [74].

In neurological disease research, Raman spectroscopy has demonstrated potential for detecting AD-related spectral differences in biofluids such as cerebrospinal fluid and saliva [75, 17]. It has also been applied ex vivo to tissues and used intraoperatively with fiber-optic probes and Raman microscopes to distinguish brain tissue types and detect pathological protein aggregates [76, 8].

Hyperspectral Raman imaging

Despite the advantages of Raman spectroscopy, its highly localized measurement capability can be a limitation when analyzing heterogeneous biofluid samples. Variability in molecular composition across different sample regions may not be fully captured with single-point Raman measurements.

To address this limitation, hyperspectral Raman imaging offers a spatially resolved approach by collecting Raman spectra across an entire sample, generating a three-dimensional (3D) hyperspectral data cube [77].

Unlike conventional Raman spectroscopy, which acquires a single spectrum from a specific location, hyperspectral imaging maps biochemical composition with both spectral and spatial resolution, providing a more comprehensive representation of molecular distributions.

A hyperspectral Raman image can be conceptualized as a three-dimensional dataset (Figure 1.4), characterized by two spatial dimensions (X, Y) and one spectral dimension (λ) [77]. The data can be mathematically represented as:

$$I(X, Y, \lambda) = \text{Hyperspectral image}, \tag{1.2}$$

where $I(X, Y, \lambda)$ represents the intensity at spatial coordinates (x, y) for a given Raman shift λ.

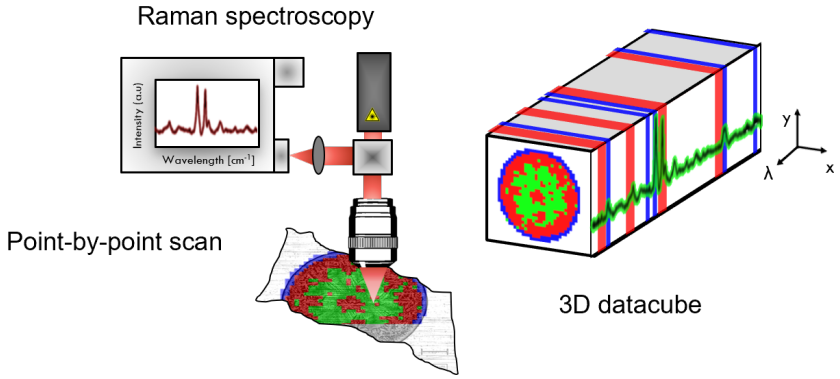


Figure 1.4: Illustration of hyperspectral Raman imaging. The technique generates a hyperspectral data cube, capturing spatial (X, Y) and spectral (λ) dimensions. The figure also depicts the point-by-point scanning method, where a focused laser sequentially collects Raman spectra at each spatial coordinate.

This multidimensional dataset enables spatially resolved biochemical analysis, which makes it particularly useful for biofluid analysis, where compositional variability can influence diagnostic interpretations [78].

Data acquisition in hyperspectral Raman imaging can be performed using different point-by-point scanning [79], as shown in Figure 1.4:

Point-by-point scanning: A focused laser sequentially scans each point in the sample, collecting individual Raman spectra at each location. While this method offers high spatial resolution, it can be time-consuming, especially for large sample areas.

Studies have demonstrated that hyperspectral Raman imaging can distinguish between different cell types, detect molecular changes associated with disease states, and monitor biochemical processes in real time [80, 7, 77, 81, 82, 83].

Surface-enhanced Raman spectroscopy (SERS)

SERS is an advanced variant of Raman spectroscopy that significantly enhances sensitivity by nanostructured metal surfaces or nanoparticles that amplify the Raman signal [84, 85]. When molecules are adsorbed or positioned near roughened metal surfaces (typically gold or silver nanostructures), the local electromagnetic field of the excitation light is intensified by localized surface plasmon resonances (LSPR) and so-called "hotspots" between plasmonic nanoparticles. This amplification results in a substantial increase in Raman scattering from nearby molecules.

The enhancement factor in SERS can range from 10^6 to 10^8 or higher, particularly for analytes positioned at regions of intense local field enhancement on the nanostructure (hotspots) [85, 84]. A widely used approximation is that the Raman intensity in SERS follows the fourth power of the local electric field amplification (the $|E|^4$ rule):

$$I_{\text{SERS}} \propto |E_{\text{local}}|^4 \tag{1.3}$$

where E_{local} represents the localized electric field near the metallic nanostructures. This means that if the local field E_{local} is M times the incident field E_0, the SERS intensity increases by approximately M^4. In addition to this electromagnetic enhancement, chemical effects such as charge transfer between the molecule and the metal surface can further enhance certain vibrational modes [85].

As a result, SERS enables detection of analytes at much lower concentrations than conventional Raman, reaching single-molecule sensitivity in some cases.

SERS has been widely explored for biosensing applications and detecting AD-related biomarkers in biofluids. Its Raman signal enhancement enables detecting low-abundance peptides, metabolites, and protein conformations linked to neurodegenerative diseases [86, 18, 87, 17].

Using a colloidal gold nanoparticle SERS assay on serum, Carlomagno *et al.* (2020) demonstrated a rapid ($< 10 \sim$ min) analysis with minimal sample preparation. Similarly, Delfino *et al.* (2020) applied SERS to tear fluids, identifying distinct spectral signatures that differentiated neurodegenerative disease patients from controls. In addition to blood and tear analysis, SERS has been utilized in saliva-based diagnostics, highlighting its adaptability for detecting disease markers across various biofluids [86, 18, 17].

A key advancement in SERS-based biosensors involves functionalizing metal surfaces with receptors such as antibodies, aptamers, or other capture molecules to selectively bind target biomarkers and improve specificity. Zhao *et al.* (2024) developed chiral gold nanostructure monolayers as SERS substrates for the detection of AD biomarkers, demonstrating improved sensitivity and specificity compared to conventional SERS platforms [88]. Similarly, Kim *et al.* (2024) introduced a surface-functionalized SERS platform combined with deep learning analysis, successfully classifying AD patient samples and showcasing the synergy between nanotechnology and artificial intelligence in next-generation diagnostics [87].

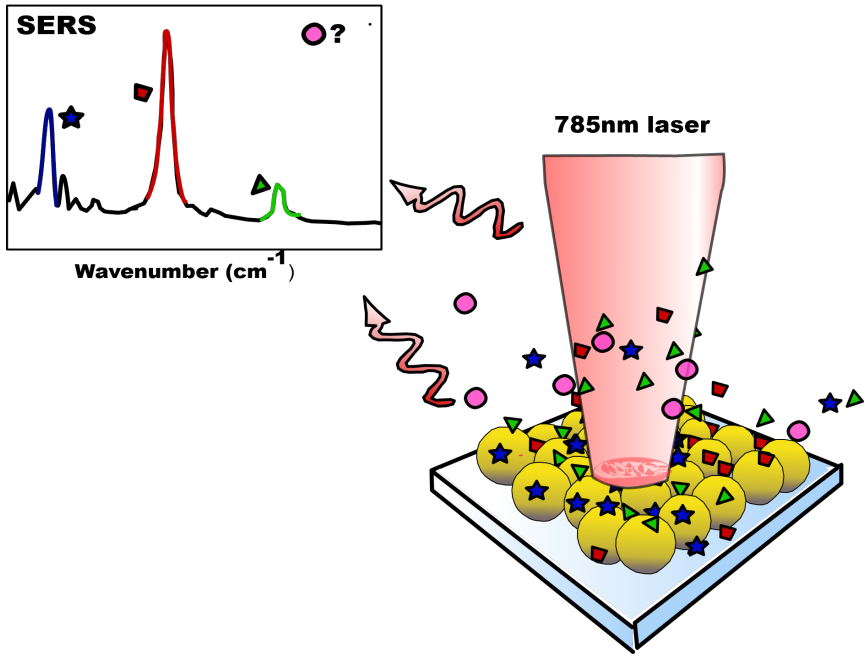


Figure 1.5: SERS spectra of biological samples are influenced by four key parameters, as illustrated in the scheme. The interaction between biomolecules and the plasmonic substrate is governed by their binding affinities, which affect signal detection. The recorded SERS signal is determined by the presence of biomolecules within the irradiated area and their intrinsic Raman cross-sections. Weakly adsorbed biomolecules or those with low Raman cross-sections (pink dots in the scheme) contribute minimally to the spectrum. Optimizing plasmonic substrates and measurement conditions is essential for enhancing SERS sensitivity in label-free applications. Adapted from [89].

While SERS offers remarkable sensitivity for disease biomarker detection, challenges remain in achieving consistent and quantitative results due to variability in substrates and experimental conditions. However, nanofabrication and substrate engineering advancements, such as waveguide-integrated SERS for controlled sampling, continue to improve reproducibility and clinical applicability [84, 89, 90, 91].

Limitations of vibrational spectroscopy for clinical diagnostics

Vibrational spectroscopic techniques, including FTIR and Raman spectroscopy, hold significant promise for clinical diagnostics due to their ability to provide detailed molecular information. However, several limitations must be addressed to fully harness their potential [92]:

- **Spectral complexity**: Biological samples are complex mixtures containing numerous molecular species, leading to overlapping vibrational bands. This overlap complicates the assignment of specific spectral features to individual biomolecules, thereby challenging accurate interpretation.
- **Autofluorescence**: In Raman spectroscopy, endogenous fluorophores present in biological tissues and fluids can emit fluorescence upon laser excitation, which often overwhelms the weaker Raman signals. Utilizing longer excitation wavelengths in the near-infrared region can mitigate this issue but may also reduce signal intensity.
- **Reproducibility issues**: Variations in sample preparation, environmental conditions, and instrument performance can lead to inconsistencies in spectral data. Specifically, in surface-enhanced Raman spectroscopy (SERS), achieving uniform enhancement is challenging due to variability in nanostructured substrates, which affects measurement reliability.
- **Computational demands**: Extracting meaningful diagnostic information from complex spectral datasets necessitates advanced chemometric and machine-learning algorithms. Developing robust models requires large, well-characterized datasets and significant computational resources.

Improving preprocessing techniques, standardizing protocols, and integrating data-driven approaches are key to overcoming these challenges. Combining vibrational spectroscopy with advanced computational methods is expected to enhance reproducibility and support its clinical translation [93].

1.4.3 Machine-learning for spectral analysis: Chemometrics

Spectra obtained from biofluids are inherently complex and high-dimensional, presenting significant challenges for analysis. ML, a subfield of AI, provides powerful tools to uncover hidden patterns in these datasets by learning from experience, say statistics, to make predictions or classifications [94, 95]. Generally, ML approaches are categorized based on the learning procedure into supervised, unsupervised, semi-supervised, and reinforcement learning, depending on the nature and availability of labeled data.

Within analytical chemistry, ML is applied through the discipline of *chemometrics*, a multidisciplinary field that integrates chemistry, statistics, mathematics, and computer science [96]. Chemometrics employs mathematical modeling and statistical techniques to extract chemically meaningful information from data, enabling models not only to predict or classify but also to reveal underlying chemical relationships [96, 97]. Fundamentally, chemometric methods prioritize explainability and interpretability, ensuring that the resulting models align with established chemical principles and remain comprehensible [98].

Data preprocessing for vibrational spectroscopy

Ensuring high-quality and consistent spectroscopic data is essential before applying machine learning, as variations in raw spectra can significantly affect analysis [99, 100]. Unpreprocessed spectra often contain noise, baseline drifts, and scattering effects, which can obscure the underlying chemical information. Implementing robust preprocessing techniques addresses these issues by removing unwanted variations [101], enhancing key spectral features, and improving the reliability of subsequent analysis [102, 103]. By eliminating irrelevant signals, such as instrument-induced artifacts or background interference, chemometric models accurately capture meaningful sample characteristics rather than distortions from extraneous factors.

Common preprocessing methods in vibrational spectroscopy include:

- **Anomalous measurements detection and removal**: outliers, caused by instrumental artifacts, sample contamination, or experimental errors, can skew analysis and degrade model performance. Principal component analysis (PCA), robust statistical measures, and Mahalanobis distance-based approaches are commonly employed to identify and eliminate anomalous spectra.

- **Cosmic ray removal**: cosmic rays introduce sharp spikes in Raman spectra, which can distort analysis [104]. Algorithms such as median filtering and derivative-based approaches effectively eliminate these anomalies while preserving spectral integrity.

- **Baseline correction**: many spectra exhibit a smooth generally non-linear background due to instrumentation or fluorescence. Baseline correction methods estimate and subtract such background signals, ensuring that the analysis focuses on actual spectral peaks rather than baseline fluctuations. Common approaches include polynomial fitting, asymmetric least squares [105], and iterative polynomial fitting [59].

- **Standarization**: intensity normalization methods, such as vector normalization and min-max scaling, adjust for variations in laser power, sample thickness, strong dynamic variations between features and instrumental differences, enabling meaningful spectral comparisons across different samples and experimental conditions.

- **Smoothing and denoising**: methods like the Savitzky-Golay filter [106] reduce noise while preserving key spectral features, improving the signal-to-noise ratio (SNR) and enhancing model accuracy and stability [107].

- **Spectral calibration and alignment**: variations in instrument calibration can cause shifts in peak positions (e.g., slight wavenumber deviations), leading to broadened peaks and reduced analytical precision. Calibration and alignment techniques correct these discrepancies, improving consistency across datasets.

In the following sections, the unsupervised and supervised chemometric methods utilized throughout this research are detailed. **Unsupervised learning** identifies patterns, structures, or clusters within unlabeled data, making it essential for exploratory data analysis and dimensionality reduction [96]. In contrast, **supervised learning** relies on labeled data to establish relationships between input features and known outcomes, making it particularly valuable for classification and regression tasks [95].

It should be noted that several additional algorithms like analysis of variance (ANOVA) simultaneous component analysis (ASCA), support vector machines (SVM), decision trees, and random forest were also tested during the course of this research. Although these methods are not directly reflected in the published articles included in this thesis, they played a valuable role in guiding fundamental decisions. Their evaluation contributed to a deeper understanding of model performance across different spectral data structures and informed the selection of the most appropriate algorithms for each specific objective.

Unsupervised methods

Unsupervised ML plays a fundamental role in exploratory data analysis within chemometrics. These methods analyze the inherent variance of complex datasets and underlying structure without the need for predefined labels or class definitions. They are essential for initially visualizing, understanding, and interpreting high-dimensional spectral data. They enable patterns to be recognized, clusters to be formed and potential outliers or batch effects to be identified.

Principal Component Analysis (PCA)

A widely used dimensionality reduction method is PCA, which projects spectral data onto orthogonal principal components (PCs), ranked by descending variance, with the first few retaining the most essential information. This transformation simplifies data visualization, improves grouping interpretation, and reduces noise while preserving key spectral patterns [108, 96].

Mathematically, PCA decomposes a data matrix $\mathbf{X}$ (with n observations and p variables) as:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E} \tag{1.4}$$

where $\mathbf{T}$ $(n \times k)$ is the matrix of scores, $\mathbf{P}$ $(p \times k)$ is the matrix of loadings, and $\mathbf{E}$ is the residual matrix not explained by the first k components.

Each column vector in $\mathbf{T}$ corresponds to the i^{th} principal component score vector, and is obtained via:

$$\mathrm{PC}_i = \mathbf{X}\boldsymbol{w}_i \tag{1.5}$$

where $\boldsymbol{w}_i$ is the i^{th} loading vector, i.e., the eigenvector associated with the i^{th} largest eigenvalue of the covariance matrix $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ (or $\mathbf{X}\mathbf{X}^{\mathrm{T}}$ depending on the convention) [97]. In other words, $\mathbf{P} = [\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k]$.

Multivariate curve resolution – alternating least squares MCR-ALS

MCR-ALS is a powerful method used to decompose complex multivariate data into pure component contributions with minimal prior knowledge of the system. MCR-ALS is mathematically similar to PCA, both decompose complex multivariate data into bilinear models. However, while PCA imposes orthogonality constraints on the components, MCR-ALS applies chemically and physically meaningful constraints, such as non-negativity, closure and unimodality, to ensure that the resolved components correspond to real chemical species [109, 110, 111]. This flexibility allows MCR-ALS to provide more interpretable solutions in chemical analysis. The bilinear model is:

$$\mathbf{X} = \mathbf{C}\mathbf{S}^{\mathrm{T}} + \mathbf{E}, \tag{1.6}$$

where $\mathbf{X}$ $(n \times p)$ is the data matrix, $\mathbf{C}$ $(n \times r)$ contains the estimated concentration profiles, $\mathbf{S}$ $(p \times r)$ contains the estimated pure component spectra, and $\mathbf{E}$ $(n \times p)$ is the residual matrix.

MCR-ALS solves this model iteratively, typically alternating between least-squares estimation of $\mathbf{C}$ and $\mathbf{S}$ while applying user-defined constraints to improve the interpretability and uniqueness of the solution.

k-means clustering

A clustering approach that partitions data into distinct clusters by minimizing intra-cluster variance, effectively grouping similar spectral data points together [96, 94]. The algorithm iteratively assigns each data point to one of k clusters based on the nearest centroid:

$$\underset{\mathbf{S}}{\operatorname{argmin}} \sum_{i=1}^{k} \sum_{\mathbf{x} \in S_i} \|\boldsymbol{x} - \boldsymbol{\mu}_i\|^2 \tag{1.7}$$

where $\mathbf{S}$ are clusters, $\boldsymbol{x}$ are data points, and $\boldsymbol{\mu}_i$ represents the cluster centroid S_i. k-means assesses natural spectral groupings. The optimal number of clusters (k) in k-means is typically determined using:

- **Elbow method**: identifies k by locating the point where adding clusters no longer significantly reduces variance.
- **Silhouette method**: evaluates cluster quality by measuring how well-separated and cohesive they are.

Supervised methods

Supervised ML methods build predictive models based on labeled datasets. Each data point in these datasets comprises input features paired with corresponding known outputs, often referred to as labels or targets. The objective is to teach a mapping function that can accurately predict the output for new, unseen inputs by generalizing from the provided examples.

Partial Least Squares - Discriminant Analysis PLS-DA

PLS-DA is an extension of partial least squares (PLS) [112], particularly suitable for supervised classification tasks in high-dimensional, collinear datasets such as Raman and FTIR spectra [98, 96].

PLS-DA constructs latent variables by maximizing the covariance between the predictor matrix $\mathbf{X}$ (spectral intensities, $(n \times p)$) and the categorical outcome vector $\mathbf{Y}$ (AD vs. control labels, $(n \times c)$). The predictive model can be expressed as:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E}, \quad \mathbf{Y} = \mathbf{U}\mathbf{Q}^{\mathrm{T}} + \mathbf{F} \tag{1.8}$$

where $\mathbf{T}(n \times r)$ and $\mathbf{U}(n \times r)$ are the latent score matrices for predictors and responses, respectively, $\mathbf{P}(p \times r)$ and $\mathbf{Q}(c \times r)$ are the loading matrices, and $\mathbf{E}$, $\mathbf{F}$ are the residual matrices. The inversion of the matrix $\mathbf{T}^{\mathrm{T}}\mathbf{T}$ is feasible since the latent variables in $\mathbf{T}$ are completely orthogonal and the number of components is quite lower than the number of observations. Both latent spaces are related by the inner relationship:

$$\mathbf{U} = \mathbf{T}\mathbf{B} + \mathbf{H} \tag{1.9}$$

where $\mathbf{B}(r \times r)$ is a diagonal matrix containing the regression coefficients linking $\mathbf{T}$ to $\mathbf{U}$ and $\mathbf{H}$ is a residual matrix. To link the latent scores to the original predictor space, a weight matrix $\mathbf{W}(p \times r)$ is defined such that

$$\mathbf{T} \;=\; \mathbf{X}\mathbf{W}. \tag{1.10}$$

Combining Eqs. (1.8)–(1.10) yields the regression coefficient matrix

$$\mathbf{B}_{\mathrm{PLS}} = \; \mathbf{W}\big(\mathbf{P}^{\mathrm{T}}\mathbf{W}\big)^{-1}\mathbf{B}\,\mathbf{Q}^{\mathrm{T}}, \tag{1.11}$$

so that model prediction is

$$\hat{\mathbf{Y}} \;=\; \mathbf{X}\mathbf{B}_{\mathrm{PLS}}. \tag{1.12}$$

Because $\hat{\mathbf{Y}}$ is continuous (typically taking values in $[0, 1]$ for binary coding or giving one score per class in one-hot coding), a class decision is obtained by:

- applying a fixed threshold (e.g. 0.5 in the binary case)
- assigning each sample to the class whose column in $\hat{\mathbf{Y}}$ has the largest value
- using latent-space distances to class centroids

Variable selection

Variable selection, also known as feature selection, is the process of identifying and keeping only the most informative variables in a dataset, while removing redundant or irrelevant ones. This step is sometimes crucial for developing robust and explainable models, especially when working with high-dimensional spectral data. By focusing on the most relevant variables, models become simpler, more accurate, and easier to interpret, allowing for a clearer understanding of the chemical information contained in the spectroscopy data.

A wide range of variable selection methods have been developed, each tailored to different modeling approaches and data characteristics [113]. Classical techniques such as interval partial least squares (iPLS) [114], variable importance in projection (VIP) [115], and recursive feature elimination (RFE) [116] remain widely used. However, numerous modern approaches, including SHapley additive explanations (SHAP) [117] and the Boruta algorithm [118] among others, have been introduced to enhance selection robustness, interpretability, and performance across high-dimensional datasets.

While modern algorithms have gained attention, classical methods such as variable importance in projection (VIP) and selectivity ratio (SR) continue to be extensively used in chemometrics, especially when used with partial least squares (PLS) models. These metrics inherently rely on fundamental concepts of variance decomposition, namely explained and residual variance, which are central to evaluating model performance and interpretability.

Explained variance in PLS refers to the proportion of the total variance in the data that is captured by the model. Each PLS component accounts for a certain amount of variance in both the predictor and response variables. The cumulative explained variance indicates how well the model represents the data. Conversely, residual variance is the portion of variance not captured by the model, residing in the residual matrices from eq 1.9 (E and F). Minimizing residual variance is essential for improving model accuracy.

VIP scores are used in PLS regression to assess the contribution of each variable to the model. The VIP score for the j^{th} variable is calculated as:

$$\mathrm{VIP}_j = \sqrt{p \cdot \frac{\sum_{a=1}^{A} w_{ja}^2 \cdot \mathrm{SSY}_a}{\mathrm{SSY}_{\mathrm{total}}}} \tag{1.13}$$

where:

p is the total number of variables, A is the number of components in the PLS model, w_{ja} is the weight of the j^{th} variable in the a^{th} component, SSY_a is the sum of squares explained by the a^{th} component, $\mathrm{SSY}_{\mathrm{total}}$ is the total sum of squares explained by all components.

Variables with VIP scores greater than 1 are generally considered significant, indicating that they contribute more than the average variable to the model. This threshold helps in distinguishing informative variables from uninformative ones.

The SR is another method used to identify important variables in PLS models. It is defined as the ratio of explained variance to residual variance for each variable:

$$\mathrm{SR}_j = \frac{\text{Explained Variance}_j}{\text{Residual Variance}_j} \tag{1.14}$$

where:

SR_j is the selectivity ratio for the j^{th} variable, $\text{Explained Variance}_j$ is the variance explained by the PLS model for the j^{th} variable, and $\text{Residual Variance}_j$ is the variance not captured by the model for the j^{th} variable. A higher SR indicates a greater contribution of the variable to the model's predictive ability. This method provides a straightforward means of ranking variables based on their explanatory power.

By employing VIP and SR in this research, we have identified and prioritized spectral variables that significantly contribute to the differentiation between AD patients and healthy controls, thereby enhancing the interpretability and performance of our chemometric models.

ANOVA-Simultaneous component analysis ASCA

This is a method which integrates PCA with analysis of variance (ANOVA) [119] and allows the separation and quantification of spectral variance due to known experimental factors (e.g., sample batches, instrumental variability) [120]. The data matrix $\mathbf{X}$ is partitioned into factor-specific matrices, each decomposed using PCA as follows:

$$\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_f + \mathbf{E} \tag{1.15}$$

where each $\mathbf{X}_f$ represents the effect associated with a particular factor. PCA is then individually applied to each $\mathbf{X}_f$, providing insight into which experimental factors predominantly influence spectral variance. ASCA improves the clarity of the spectral interpretation and ensures that the observed differences are genuinely attributed to AD rather than confounding variables [120]. ASCA variants like Variable-selection ASCA (VASCA) and Group-wise ASCA (GASCA) offer tailored approaches to address specific challenges in multivariate data analysis, thereby enhancing the robustness and interpretability of results in complex experimental designs [121, 122].

Model construction, optimization, and validation

ML algorithms learn patterns by iteratively adjusting parameters based on input data (a process known as training). The data are typically split into two subsets: a training set and a test set. Although no strict rule dictates this division, it is common to allocate about one-quarter to one-third of the total data to the test set, depending on dataset size and complexity [94, 96].

The predictive model is first developed (calibrated) using the training set, while an independent test set is used to objectively assess its performance and ability to generalize to unseen data [94, 95]. Ensuring effective generalization is crucial, as it confirms that the model captures meaningful patterns from the data rather than overfitting to noise or artifacts present in the training set.

Cross-validation (CV) is essential for model optimization and evaluation, particularly in spectral analysis, where sample sizes are often limited and datasets are high-dimensional. CV partitions the training set into k subsets (folds), iteratively using each fold for internal validation while training on the remaining $(k-1)$ folds. Model parameters and hyperparameters are fine-tuned based on performance metrics averaged across all validation folds [94, 123]. Although cross-validated models cannot be used as final predictive models, the averaged performance metrics provide a reliable estimate of the overall potential—including data acquisition, preprocessing, and subsequent ML methods.

Although widely used, improper CV strategies and especially when hierarchical data structures (e.g., repeated measures or dependent samples) are overlooked, can produce overly optimistic results and reduce the reliability of predictive models [20].

In a recent study [20], the role of validation strategies in supervised model performance was examined, particularly for spectral analysis. It explored how different validation approaches influence the reliability of predictive models and emphasized the need for external, independent datasets that reflect real-world conditions. The study also analyzed the impact of dataset size and validation schemes and shows that small sample sizes and inappropriate validation methods can lead to overly optimistic performance estimates. Additionally, the importance of maintaining sample independence whenever possible was discussed and alternative validation approaches that can be employed when complete independence is not feasible were highlighted.

Therefore, as summarized in Figure 1.6 careful consideration of data partitioning, cross-validation methods, and external validation is essential for constructing reliable, interpretable, and generalizable chemometric models.

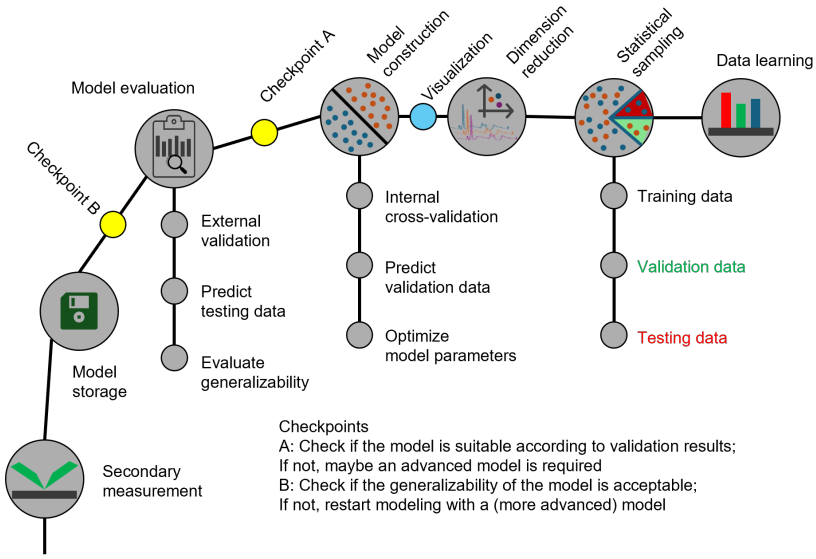


Figure 1.6: Overview of model development, validation, and evaluation. The process includes dimensionality reduction and statistical modeling, followed by internal and external validation. The final model is stored for future predictions, ensuring consistent preprocessing of new data. Adapted from [26].

Evaluation metrics for model performance:

The performance of a predictive model is assessed by its ability to classify unknown data accurately. This predictive capability is quantified using statistical metrics [124, 125]. In classification tasks, evaluation is typically performed using a confusion matrix, which serves as the basis for deriving key performance metrics, including sensitivity, specificity, accuracy, positive predictive value (PPV), negative predictive value (NPV), F-score, and the area under the receiver operating characteristic curve (AUROC).

These metrics characterize various aspects of model performance, including its ability to detect true positives (TP) and true negatives (TN), minimize false classifications (false positives (FP) and false negatives (FN)) and balance precision and recall.

Sensitivity (Recall) measures the model's ability to correctly identify positive cases, representing the proportion of actual positives that are correctly classified. It is given by

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{1.16}$$

Specificity quantifies how well the model correctly identifies negative cases, indicating the proportion of true negatives among all actual negatives. It is calculated as

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}. \tag{1.17}$$

Positive predictive value (PPV or precision) represents the proportion of predicted positive cases that are truly positive. It is defined as

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{1.18}$$

Negative predictive value (NPV) indicates the proportion of predicted negative cases that are truly negative, computed as

$$\text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}}. \tag{1.19}$$

Accuracy evaluates overall classification performance by determining the proportion of correctly classified instances among all evaluated cases. It is expressed as

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{P} + \text{N}}, \tag{1.20}$$

where P and N represent the total number of positive and negative cases, respectively.

F-score provides a balance between precision and recall by calculating their harmonic mean. It is particularly useful for imbalanced datasets, as it accounts for both false positives and false negatives. The F_1-score is defined as

$$\text{F}_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}. \tag{1.21}$$

Area under the receiver operating characteristic curve (AUROC) assesses the model's ability to distinguish between classes. It represents the probability that a randomly chosen positive instance is ranked higher than a randomly chosen negative instance. The ROC curve plots sensitivity against 1− specificity across different classification thresholds.

1.4.4 Droplet evaporation, biofluids

When a droplet of biofluid, such as CSF or blood plasma, evaporates, it undergoes complex physicochemical processes that determine the final dry droplet pattern. Understanding the physical basis of evaporation and the specific behaviors of biofluids is fundamental to interpret the results of droplet coating deposition Raman spectroscopy (DCDR). Next, the fundamental phenomena of droplet evaporation (e.g., the coffee-ring effect) and how they are modulated in complex fluids is discussed; Subsequently, the specific challenges posed by biofluid droplets (capillary and Marangoni flows, protein aggregation, phase separation, and salt crystallization) will be examined, along with the influence of interfacial and environmental conditions, the contrasting morphological characteristics of dried CSF versus plasma droplets, and, finally, the relevance of these factors to the success of DCDR-based Raman diagnostics.

Generic physics of droplet evaporation

The most common example is the drying of a spilled drop of tea or coffee on a countertop, which leaves behind a distinct outer ring of particles. This "coffee-ring" effect arises when a pinned contact line sustains a radially non-uniform evaporative flux: solvent is lost most rapidly at the rim, and the resulting capillary flow convects dispersed solutes toward the edge, producing the ring-shaped deposit [126, 127, 128, 129, 130].

In contrast, when the contact line moves inward instead of remaining pinned, the droplet can shrink more uniformly, generate multiple concentric rings through intermittent pinning, or dry without any discernible ring [131]. Consequently, contact-line dynamics (pinning versus slipping) are a primary determinant of the final deposition pattern.

Figure 1.7 shows images of evaporating droplets (blood, water on wool, and rain on glass and foliage) in panels (a–d) together with the principal dried deposit morphologies in panels (e–p): single rings, concentric rings, nearly uniform coatings, central deposits, branched spokes, and complex patterns.

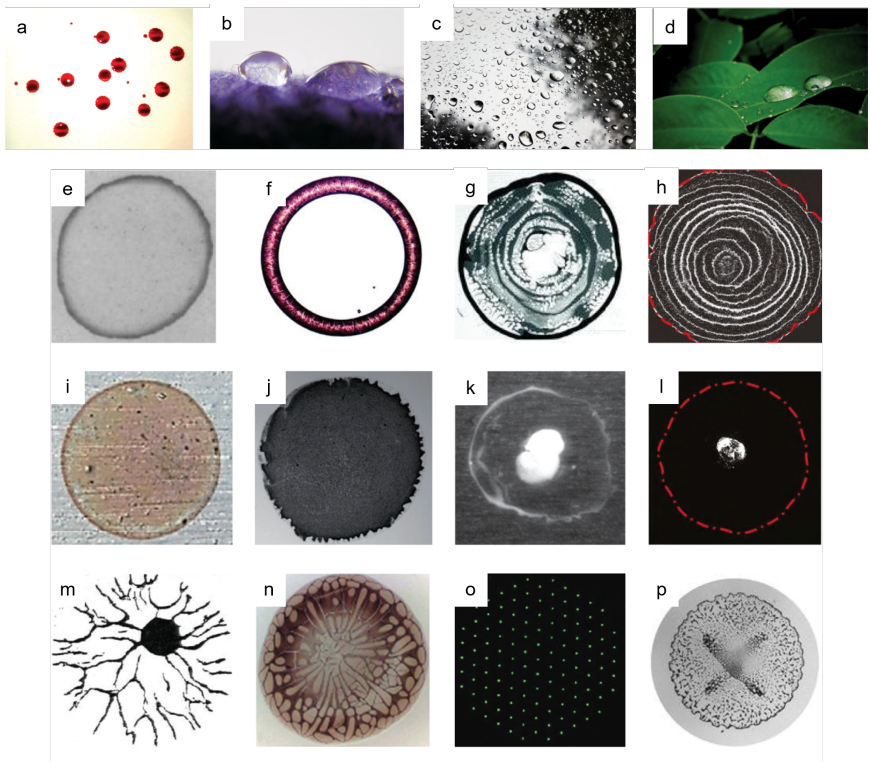


Figure 1.7: Images of evaporating sessile droplets. (a) Droplets of blood. (b) Droplets of water on a woolen fabric. (c) Raindrops on the windshield of a car. (d) Raindrops on the leaves of a plant. Examples of various deposition patterns that have been observed experimentally after a droplet containing one or more nonvolatile components completely evaporated. (e,f) Ring deposits. (g,h) Multiple or concentric ring deposits. (i,j) Uniform deposits. (k,l) Small concentrated or inner deposits. (m,n) Branched or spoke like deposits. (o,p) Patterned deposits. Panels adapted from [132]

Beyond capillary outward flow, other fundamental flows can develop. One key mechanism is the Marangoni flow, which arises from surface tension gradients along the droplet-air interface. Temperature differences or compositional gradients on the droplet surface can create Marangoni stresses that drive fluid from regions of low surface tension toward regions of higher surface tension. In a pure volatile liquid drop, evaporative cooling at the periphery increases the edge surface tension relative to the warmer interior, inducing an inward flow that counteracts the outward capillary transport.

Under certain conditions, this Marangoni circulation can even suppress the coffee-ring effect, redirecting solutes back toward the center of the drop [133]. Hu and Larson demonstrated this by showing that a suffiently strong Marangoni flow (either thermally or chemically induced) can yield a more uniform or central deposit instead of a peripheral ring [133]. In general, the final deposition pattern reflects the balance between capillary (outward) flow and Marangoni (often inward) flow, togehter with the degree of contact-line pinning. For simple single-component liquids with inert particles, a pinned contact line and negligible Marangoni effects produce the characteristic ring deposit [126], whereas significant Marangoni stresses or moving contact lines can yield more homogeneous or complex deposition patterns [133].

Evaporation dynamics in biofluid droplets

Analogously, when biofluid droplets, such as cerebrospinal fluid CSF or blood plasma, evaporate, add layers of complexity to the canonical model. They are crowded cocktails of proteins, lipids, electrolytes and metabolites whose concentrations rise by up to two orders of magnitude during drying [127, 134, 135, 136]. Their drying dynamics are therefore considerably more complex due to the presence of various biomolecules, which lead to the formation of unique patterns and microstructures [126, 127, 137].

The evaporation process can result in concentration gradients and molecular migration, significantly impacting the droplet's final structure and composition. These evaporation patterns are influenced by solute concentration and molecular interactions within the biofluid [138, 139, 132]. Also interfaces play a significant role in the evaporation process, influencing the distribution and interaction of molecules within the droplets.

In plasma droplets, the abundance of albumin and fibrinogen produces a viscoelastic interfacial "skin" within the first few seconds of drying. The skin arrests radial flow, so the coffee-ring rim becomes a thick protein band that later fractures into the characteristic polygonal "mud-flat" network [140, 141, 142].

CSF, in contrast, contains an order of magnitude less protein and a similar ionic strength to plasma; it therefore remains mobile for longer, allowing salts to supersaturate and crystallize into dendritic or "fern-like" NaCl structures in the centre while leaving only a faint protein outline at the periphery [138, 139].

The study of biofluids like CSF and blood plasma is invaluable in medical diagnostics because these are easily accessible and rich in physiological and pathological information, including crucial biomarkers [143, 144]. As each droplet dries, its chemical composition is recorded in deposits that range from coffee-ring bands to fern-like salt crystals. Recognizing these patterns and knowing where they form is therefore essential when selecting the most informative spots for spectroscopic analysis.

Drop-coating deposition Raman

Drop-coating deposition Raman spectroscopy (DCDR) takes advantage of the concentration gradients that develop as a biofluid droplet dries. As mentioned in the previous subsection, during evaporation, analytes accumulate in specific regions (most often the outer rim) forming an *in situ* solid deposit that can be probed with a Raman microscope [143, 144]. Achieving reproducible DCDR spectra requires control over four interconnected factors:

- **Capillary and Marangoni flows:** their balance decides whether the interesting molecules concentrate at the rim, in the centre or in intermediate regions.
- **Substrate properties:** surface wettability and roughness determine how strongly the droplet's contact line is pinned and therefore how pronounced the coffee-ring flow becomes.
- **Ambient conditions:** temperature and relative humidity set the evaporation rate and, through evaporative cooling, influence thermally driven Marangoni stresses.
- **Protein rheology:**, viscoelastic "skins" or phase transitions can halt internal flow and lock in compositional heterogeneities that complicate spectral interpretation.

When these variables are optimized, the droplet dries into a thin, protein rich ring that delivers high Raman signal-to-noise and low spot-to-spot variability, which are conditions that underpin the chemometric pipelines described in Section 1.4.3. If drying is not controlled, however, target biomarkers may migrate away from the chosen measurement area, increasing the risk of not performing the DCDR measurement at the appropriate region.

In this thesis, microscope cover slides covered with an aluminium foil were used to give a uniform, mildly hydrophilic surface; droplets of 1 μL were deposited inside a vacuum desiccator maintained at 40 $\pm$ 5 % relative humidity. These conditions yielded highly reproducible deposits, providing a reliable basis for the subsequent Raman measurements and multivariate analysis. Full details of the sample-preparation protocol are presented in the next section.

1.5 Sample preparation protocols

Sample preparation is a crucial first step in acquiring high-quality spectral data from biofluids, directly influencing the accuracy and reliability of diagnostic models. In the initial stage of this thesis, a standardized and reproducible protocol was established for processing human CSF and plasma samples. The protocol aimed to preserve sample quality and enhance biomarker detection without requiring separate preparations for different analytical techniques.

Sample characteristics

Thanks to a collaboration with the CITA Alzheimer Foundation, access to a well-characterized experimental database (presented in Appendix B) was possible. The members from this database underwent (i) a neurological interview, (ii) a nursing assessment, (iii) a formal neuropsychological test and (iv) a 3-T MRI protocol with structural and functional sequences.

The neurological interview captured cognitive complaints, behavioural symptoms, medical history and functional autonomy, while the nursing visit recorded anthropometrics, blood pressure, pulse, lifestyle factors and sleep quality. Neuropsychology provided z-scores for memory, language, orientation, attention, visuospatial skills and executive function, together with proxies of cognitive reserve (education, leisure activities, bilingualism, estimated IQ).

Combining the neurological and neuropsychological information yielded three syndromic groups: cognitively normal (CN), with subjective cognitive decline (SCD), and mild cognitive impairment (MCI) with no dementia cases.

MRI excluded major structural lesions and supplied quantitative metrics and AD core biomarkers were measured by conventional ELISA (Innotest). Summary of all the abovementioned:

- **Cognitive assessment**:
  - Cognitively normal (Healthy or preclinical AD)
  - Cognitively abnormal (prodromal AD)

- **CSF biomarker analysis (ATN classification system)**:
  - **A (Amyloid)**: Amyloid beta 42 peptide ($A\beta_{42}$)
  - **T (Tau)**: phosphorylated Tau (p-Tau)
  - **N (Neurodegeneration)**: Total Tau (t-Tau)

After both type of analyses, the established cut-off values for healthy subjects were defined as follows:

- p-Tau/$A\beta_{42}$ ratio: [0, 0.23]
- $A\beta_{42}$: >1,030 pg/mL
- t-Tau: <300 pg/mL
- p-Tau: <27 pg/mL

Sample preparation and processing

For plasma samples, preliminary analyses determined that specific preprocessing steps were unnecessary. However, for CSF samples, an ultrafiltration approach was implemented to concentrate biomarkers of interest.

CSF samples underwent ultrafiltration using Amicon® Ultra-0.5 centrifugal filters with a nominal molecular weight cut-off of 3 kDa [145]. This procedure yields two distinct fractions:

- **Supernatant**: Retains molecules larger than 3 kDa, predominantly proteins and aminoacids, including Aβ, t-Tau, and p-Tau, increasing their concentration by approximately 5–10 times.
- **Filtrate**: Contains molecules below 3 kDa, primarily metabolites, suitable for subsequent metabolomic analyses. This fraction will be analyzed in future work through a collaboration with the IBeA research group at UPV/EHU.

Filter cleaning procedure

Because the filters contain trace glycerol that can interfere with analysis, they are washed and rinsed following [145]. After cleaning, protein enrichment is achieved by centrifugation at 14 000 × g, 4 °C, for 30 min. The liquid passing through the membrane constitutes the filtrate fraction, which will be analysed mainly by metabolomics. The filter is then inverted and spun at 1 000 × g, 4 °C, for 2 min to recover the SNT fraction, destined for the techniques Raman, SERS and image analysis.

Filter cleaning process:

1. Add 500 µL NaOH (0.1M) to the filter; centrifuge at 14 000 g for 30 min (20 °C).
2. Invert the filter and centrifuge at 1 000 g for 2 min (20 °C).
3. Rinse three times with Milli-Q water using the above centrifugation conditions.

Filter quality control: Before processing CSF samples, filter integrity was confirmed by ensuring retention of proteins larger than 3 kDa:

1. Load 500 µL freshly prepared BSA solution (0.25 mg/mL).
2. Centrifuge at 14 000 g for 30 min (4 °C).
3. Collect filtrate; invert and collect the supernatant fraction at 1 000 g for 2 min (4 °C).
4. Deposit 1–1.5 µL droplets of each fraction on aluminum substrates, dry under vacuum (10 min), and examine under UV fluorescence microscopy (objective 5X, exposure 400 s).

As shown in the following Figure 1.8, the successful filter quality control shows fluorescence exclusively in the supernatant fraction (for both Milli-Q and PBS, Figure 1.8 c)), confirming protein retention.

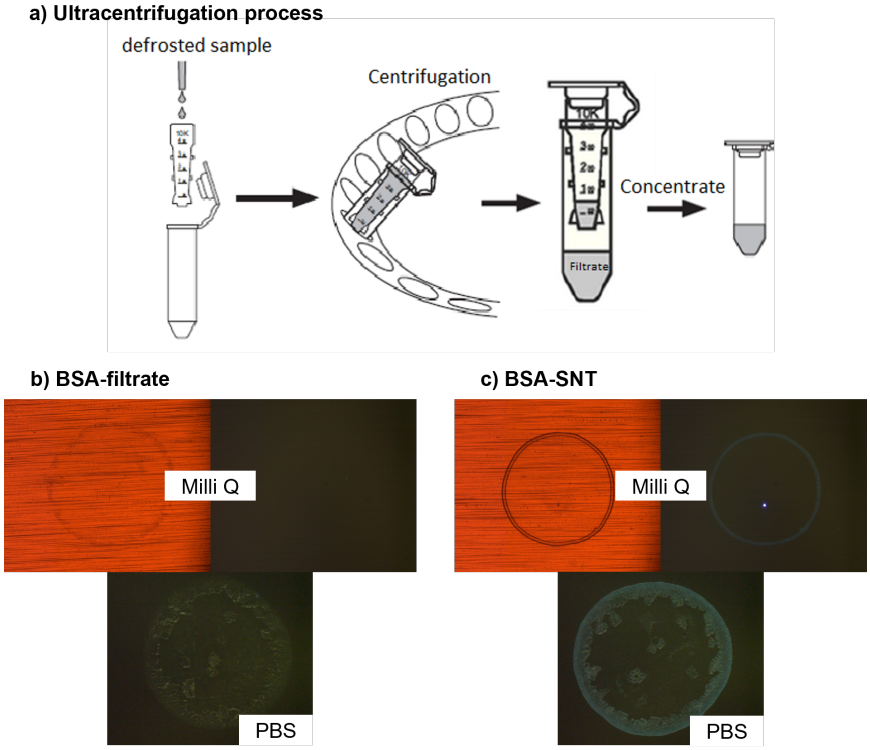


Figure 1.8: Ultracentrifugation process scheme and representative bright-field (left) and fluorescence (right) microscope images of evaporated 1 µL droplets used to verify filter performance. (a) Ultracentrifugation process steps. (b) BSA filtrate obtained after a 3 kDa centrifugal filter; (c) BSA supernatant from the same filter run. For each sample, droplets were deposited either in Milli-Q water (top row) or in PBS (bottom row), dried on microscope slides covered with aluminum foil and imaged with a ×5 microscope objective. The absence of UV-induced fluorescence in the filtrate confirms efficient protein removal, whereas the ring-shaped fluorescence in the supernatant indicates residual BSA.

CSF ultrafiltration for real samples:

Once the filter-cleaning and quality-control steps were validated with BSA (Figure 1.8), the same protocol was applied to the clinically diagnosed CSF samples. Each thawed aliquot was loaded onto a 3 kDa Amicon® Ultra-0.5 centrifugal filter unit and spun at 14 000×g (4 °C, 30 min). The run generated three fractions:

- **Native CSF** ($\sim 50\,\mu$L): reserved as an unprocessed reference.
- **Supernatant** ($\sim 75\,\mu$L): protein-enriched (~ 1.2 mg/mL); retained for Raman/SERS, FTIR and imaging studies.
- **Filtrate** ($\sim 350\,\mu$L): metabolite-rich; stored for metabolomics (collaboration with IBeA).

The starting volume of CSF is 500 µL, with a total protein concentration of 0.2 mg/mL. Ultracentrifugation yields 75 µL of supernatant, raising the protein concentration in this fraction to 1.2 mg/mL. The filtrate fraction amounts to 350 µL. A small aliquot of the native CSF is also retained before filtration.

Following ultracentrifugation, each CSF and plasma fraction was snap-frozen at - 80 ^{o}C and aliquoted into 2, 4, and 250 µL volumes selected to suit the planned experiments. Pre-aliquoting restricts every experiment to a single thaw, keeping the total number of freeze–thaw cycles below the five-cycle degradation limit and thereby preserving sample integrity for subsequent analyses.

1.6 Collaborations

Progress throughout this Ph.D. project relied on several interdisciplinary partnerships that supplied well-characterized samples, complementary expertise, and access to specialized instrumentation; together they allowed the spectroscopy–chemometrics workflow developed here to be tested far beyond its initial focus on AD.

Neurology — CITA-Alzheimer Foundation. Clinically diagnosed CSF and plasma aliquots, classified under the AT(N) biomarker framework, were provided by the CITA-Alzheimer neurology team. The precise labelling of these samples was essential for the supervised ML models evaluated in this thesis.

Artificial intelligence — Vicomtech. A collaboration with the technology centre Vicomtech examined AI-based diagnosis from bright-field microscopy images of dried CSF and plasma. Preliminary results were presented at the 12th International Conference on Innovation in Medicine and Healthcare (InMed-24, 2024):

> *Alzheimer's Disease Classification by Artificial Intelligence Using Microscopy Images of Dried Human Body Fluids*, InMed-24 (2024).

Nanoplasmonics. Within our research group, the potential of surface-enhanced Raman spectroscopy SERS with gold nanoparticles to improve AD sensitivity was assessed. Findings were published in the *Proceedings of OMN 2024*:

> *Surface-Enhanced Raman Spectroscopy for Early Detection of Alzheimer's Disease*, OMN 2024.

Oncology. Also within our research group, Raman spectroscopy combined with ML was adapted for lung-cancer diagnostics, leading to:

> *Power of Light: Raman Spectroscopy and Machine Learning for Lung Cancer Diagnostics*, *ACS Omega* (2022).

3D-printed devices. Another work within our research group, produced multiwell plates tailored for Raman spectroscopy, enabling real-time monitoring of metabolic gradients:

> *Uncovering metabolic gradients evolution using 3D-printed multiwell plates tailored to Raman spectroscopy*, in Label-Free Biomedical Imaging and Sensing (LBIS) (2025).
>
> *Real-time monitoring of metabolic gradients using a novel 3D-printed device adapted to Raman spectroscopy*, in Label-Free Biomedical Imaging and Sensing (LBIS) 2025.

Data fusion — viral pneumonia. Fusion of Fourier-transform infrared (FTIR) and Raman spectra for ML-based viral-pneumonia classification is currently under review:

> *Enhancing Viral Pneumonia Classification through Machine Learning: Data Fusion of FTIR and Raman Spectroscopy*, submitted.

Metabolomics — IBeA group. Additionally, an ongoing collaboration with the IBeA research group focuses on the metabolomic analysis of the same clinically characterized plasma and cerebrospinal fluid samples used in this thesis. This effort involves nuclear magnetic resonance (NMR) spectroscopy and high-performance liquid chromatography coupled to high-resolution mass spectrometry (HPLC-HRMS) to identify and quantify low-molecular-weight metabolites. The objective of this collaboration is to validate and potentially assign the vibrational biomarkers identified by FTIR and Raman spectroscopy to specific molecular species detected through metabolomics. This integrative approach aims to enhance the biochemical interpretability of the spectral features and strengthen the cross-platform validity of the proposed diagnostic markers.

These collaborations show the broad applicability of spectroscopy and ML methods developed in this thesis, and highlight its potential across a broad range of biomedical challenges beyond AD diagnostics.

References

[1] World Health Organization, "Dementia," 2024. Accessed: 2025-03-04.

[2] Alzheimer's Disease International, "Dementia statistics," 2024. Accessed: 2025-03-04.

[3] World Health Organization, *Global Status Report on the Public Health Response to Dementia*. Geneva: World Health Organization, 2021.

[4] C. A. Lane, J. Hardy, and J. M. Schott, "Alzheimer's disease," *European Journal of Neurology*, vol. 25, p. 59–70, Oct. 2017.

[5] M. W. Bondi, E. C. Edmonds, and D. P. Salmon, "Alzheimer's disease: Past, present, and future," *Journal of the International Neuropsychological Society*, vol. 23, p. 818–831, Oct. 2017.

[6] J. Rasmussen and H. Langerman, "Alzheimer's disease - why we need early diagnosis," *Degener Neurol Neuromuscul Dis*, vol. 9, pp. 123–130, Dec. 2019.

[7] H. J. Butler, L. Ashton, B. Bird, G. Cinque, K. Curtis, J. Dorney, K. Esmonde-White, N. J. Fullwood, B. Gardner, P. L. Martin-Hirsch, *et al.*, "Using raman spectroscopy to characterize biological materials," *Nature protocols*, vol. 11, no. 4, p. 664, 2016.

[8] G. Devitt, K. Howard, A. Mudher, and S. Mahajan, "Raman spectroscopy: An emerging tool in neurodegenerative disease research and diagnosis," *ACS Chemical Neuroscience*, vol. 9, no. 3, pp. 404–420, 2018. PMID: 29308873.

[9] M. J. Baker, J. Trevisan, P. Bassan, R. Bhargava, H. J. Butler, K. M. Dorling, P. R. Fielden, S. W. Fogarty, N. J. Fullwood, K. A. Heys, C. Hughes, P. Lasch, P. L. Martin-Hirsch, B. Obinaju, G. D. Sockalingum, J. Sulé-Suso, R. J. Strong, M. J. Walsh, B. R. Wood, P. Gardner, and F. L. Martin, "Using fourier transform IR spectroscopy to analyze biological materials," *Nature Protocols*, vol. 9, pp. 1771–1791, Aug. 2014.

[10] M. Paraskevaidi, C. L. M. Morais, K. M. G. Lima, J. S. Snowden, J. A. Saxon, A. M. T. Richardson, M. Jones, D. M. A. Mann, D. Allsop, P. L. Martin-Hirsch, and F. L. Martin, "Differential diagnosis of alzheimer's disease using spectrochemical analysis of blood," *Proceedings of the National Academy of Sciences*, vol. 114, Sept. 2017.

[11] C. L. M. Morais, K. M. G. Lima, M. Singh, and F. L. Martin, "Tutorial: multivariate classification for vibrational spectroscopy in biological samples," *Nature Protocols*, vol. 15, pp. 2143–2162, July 2020.

[12] N. M. Ralbovsky and I. K. Lednev, "Towards development of a novel universal medical diagnostic method: Raman spectroscopy and machine learning," *Chem. Soc. Rev.*, vol. 49, pp. 7428–7453, 2020.

[13] E. Lopez, J. Etxebarria-Elezgarai, M. García-Sebastián, M. Altuna, M. Ecay-Torres, A. Estanga, M. Tainta, C. López, P. Martínez-Lage, J. M. Amigo, and A. Seifert, "Unlocking preclinical alzheimer's: A multi-year label-free in vitro raman spectroscopy study empowered by chemometrics," *International Journal of Molecular Sciences*, vol. 25, no. 9, 2024.

[14] P. Carmona, M. Molina, M. Calero, F. Bermejo-Pareja, P. Martínez-Martín, and A. Toledano, "Discrimination analysis of blood plasma associated with alzheimer's disease using vibrational spectroscopy," *Journal of Alzheimer's Disease*, vol. 34, no. 4, pp. 911–920, 2013. PMID: 23302656.

[15] P. Carmona, M. Molina, E. López-Tobar, and A. Toledano, "Vibrational spectroscopic analysis of peripheral blood plasma of patients with alzheimer's disease," *Analytical and Bioanalytical Chemistry*, vol. 407, pp. 7747–7756, Oct. 2015.

[16] Y. Xu, X. Pan, H. Li, Q. Cao, F. Xu, and J. Zhang, "Accuracy of raman spectroscopy in the diagnosis of alzheimer's disease," *Front Psychiatry*, vol. 14, p. 1112615, Mar. 2023.

[17] N. M. Ralbovsky, L. Halámková, K. Wall, C. Anderson-Hanley, and I. K. Lednev, "Screening for alzheimer's disease using saliva: A new approach based on machine learning and raman hyperspectroscopy," *Journal of Alzheimer's Disease*, vol. 71, no. 4, pp. 1351–1359, 2019. PMID: 31524171.

[18] G. Cennamo, D. Montorio, V. B. Morra, C. Criscuolo, R. Lanzillo, E. Salvatore, C. Camerlingo, M. Lisitskiy, I. Delfino, M. Portaccio, and M. Lepore, "Surface-enhanced Raman spectroscopy of tears: toward a diagnostic tool for neurodegenerative disease identification," *Journal of Biomedical Optics*, vol. 25, no. 8, p. 087002, 2020.

[19] A. G. Carota, B. Campanella, R. Del Carratore, P. Bongioanni, R. Giannelli, and S. Legnaioli, "Raman spectroscopy and multivariate analysis as potential tool to follow alzheimer's disease progression," *Analytical and Bioanalytical Chemistry*, vol. 414, pp. 4667–4675, July 2022.

[20] E. Lopez, J. Etxebarria-Elezgarai, J. M. Amigo, and A. Seifert, "The importance of choosing a proper validation strategy in predictive models. a tutorial with real examples," *Analytica Chimica Acta*, vol. 1275, p. 341532, 2023.

[21] E. Lopez, G. Gorla, J. Etxebarria-Elezgarai, J. Aramendia, K. Castro, M. García-Sebastián, P. Martínez-Lage, A. Seifert, and J. M. Amigo, "Hyperspectral raman imaging to understand patterns in dried biofluids in alzheimer's disease," *VIEW*, vol. n/a, no. n/a, p. 20250034.

[22] "2023 alzheimer's disease facts and figures," *Alzheimer's & Dementia*, vol. 19, no. 4, pp. 1598–1695, 2023.

[23] "2024 alzheimer's disease facts and figures," *Alzheimer's & Dementia*, vol. 20, no. 5, pp. 3708–3821, 2024.

[24] E. L. Koedam, V. Lauffer, A. E. van der Vlies, W. M. van der Flier, P. Scheltens, and Y. A. Pijnenburg, "Early-versus late-onset alzheimer's

disease: More than age alone," *Journal of Alzheimer's Disease*, vol. 19, p. 1401–1408, Mar. 2010.

[25] A. P. Porsteinsson, R. S. Isaacson, S. Knox, M. N. Sabbagh, and I. Rubino, "Diagnosis of early alzheimer's disease: Clinical practice in 2021," *The Journal of Prevention of Alzheimer's Disease*, vol. 8, no. 3, pp. 371–386, 2021.

[26] S. Guo, J. Popp, and T. Bocklitz, "Chemometric analysis in raman spectroscopy from experimental design to machine learning–based modeling," *Nature Protocols*, vol. 16, p. 5426–5459, Nov. 2021.

[27] C. Chen, J. Qi, Y. Li, D. Li, L. Wu, R. Li, Q. Chen, and N. Sun, "Applications of raman spectroscopy in the diagnosis and monitoring of neurodegenerative diseases," *Frontiers in Neuroscience*, vol. 18, 2024.

[28] Centers for Disease Control and Prevention, "What is alzheimer's disease and related dementias?," 2024. Accessed: 2025-03-04.

[29] Lewy Body Dementia Association, "About lewy body dementia," 2024. Accessed: 2025-03-04.

[30] Stanford Health Care, "Frontotemporal dementia," 2024. Accessed: 2025-03-04.

[31] E. Canu, F. Agosta, G. Mandic-Stojmenovic, T. Stojković, E. Stefanova, A. Inuggi, F. Imperiale, M. Copetti, V. S. Kostic, and M. Filippi, "Multiparametric mri to distinguish early onset alzheimer's disease and behavioural variant of frontotemporal dementia," *NeuroImage: Clinical*, vol. 15, p. 428–438, 2017.

[32] Alzheimer's Association, "Mixed dementia," 2024. Accessed: 2025-03-04.

[33] C. R. Jack Jr., D. A. Bennett, K. Blennow, M. C. Carrillo, B. Dunn, S. B. Haeberlein, D. M. Holtzman, W. Jagust, F. Jessen, J. Karlawish, E. Liu, J. L. Molinuevo, T. Montine, C. Phelps, K. P. Rankin, C. C. Rowe, P. Scheltens, E. Siemers, H. M. Snyder, R. Sperling, Contributors, C. Elliott, E. Masliah, L. Ryan, and N. Silverberg, "Nia-aa research

framework: Toward a biological definition of alzheimer's disease," *Alzheimer's & Dementia*, vol. 14, no. 4, pp. 535–562, 2018.

[34] P. F. Høilund-Carlsen, J. R. Barrio, T. J. Werner, A. Newberg, and A. Alavi, "Amyloid hypothesis: The emperor's new clothes?," *Journal of Alzheimer's Disease*, vol. 78, p. 1363–1366, Dec. 2020.

[35] S. Engelborghs, E. Niemantsverdriet, H. Struyfs, K. Blennow, R. Brouns, M. Comabella, I. Dujmovic, W. van der Flier, L. Frölich, D. Galimberti, S. Gnanapavan, B. Hemmer, E. Hoff, J. Hort, E. Iacobaeus, M. Ingelsson, F. Jan de Jong, M. Jonsson, M. Khalil, J. Kuhle, A. Lleó, A. de Mendonça, J. L. Molinuevo, G. Nagels, C. Paquet, L. Parnetti, G. Roks, P. Rosa-Neto, P. Scheltens, C. Skårsgard, E. Stomrud, H. Tumani, P. J. Visser, A. Wallin, B. Winblad, H. Zetterberg, F. Duits, and C. E. Teunissen, "Consensus guidelines for lumbar puncture in patients with neurological diseases," *Alzheimer's amp; Dementia: Diagnosis, Assessment amp; Disease Monitoring*, vol. 8, p. 111–126, Jan. 2017.

[36] C. R. Jack, J. S. Andrews, T. G. Beach, T. Buracchio, B. Dunn, A. Graf, O. Hansson, C. Ho, W. Jagust, E. McDade, J. L. Molinuevo, O. C. Okonkwo, L. Pani, M. S. Rafii, P. Scheltens, E. Siemers, H. M. Snyder, and R. Sperling, "Revised criteria for diagnosis and staging of alzheimer's disease: Alzheimer's association workgroup," *Alzheimer's & Dementia*, vol. 20, no. 8, pp. 5143–5169, 2024.

[37] J. L. Ebenau, T. Timmers, L. M. Wesselman, I. M. Verberk, S. C. Verfaillie, R. E. Slot, A. C. van Harten, C. E. Teunissen, F. Barkhof, K. A. van den Bosch, M. van Leeuwenstijn, J. Tomassen, A. d. Braber, P. J. Visser, N. D. Prins, S. A. Sikkes, P. Scheltens, B. N. van Berckel, and W. M. van der Flier, "Atn classification and clinical progression in subjective cognitive decline: The science project," *Neurology*, vol. 95, July 2020.

[38] S. T. DeKosky and S. W. Scheff, "Synapse loss in frontal cortex biopsies in alzheimer's disease: Correlation with cognitive severity," *Annals of Neurology*, vol. 27, p. 457–464, May 1990.

[39] B. Dubois, H. Hampel, H. H. Feldman, P. Scheltens, P. Aisen, S. Andrieu, H. Bakardjian, H. Benali, L. Bertram, K. Blennow, K. Broich, E. Cavedo, S. Crutch, J. Dartigues, C. Duyckaerts, S. Epelbaum, G. B. Frisoni, S. Gauthier, R. Genthon, A. A. Gouw, M. Habert, D. M. Holtzman, M. Kivipelto, S. Lista, J. Molinuevo, S. E. O'Bryant, G. D. Rabinovici, C. Rowe, S. Salloway, L. S. Schneider, R. Sperling, M. Teichmann, M. C. Carrillo, J. Cummings, and C. R. Jack, "Preclinical alzheimer's disease: Definition, natural history, and diagnostic criteria," *Alzheimer's amp; Dementia*, vol. 12, p. 292–323, Mar. 2016.

[40] H. Zetterberg and B. B. Bendlin, "Biomarkers for alzheimer's disease—preparing for a new era of disease-modifying therapies," *Molecular Psychiatry*, vol. 26, pp. 296–308, 2021.

[41] A. S. Parul and S. Shukla, "Novel techniques for early diagnosis and monitoring of alzheimer's disease," *Expert Review of Neurotherapeutics*, vol. 25, no. 1, pp. 29–42, 2025. PMID: 39435792.

[42] H. Colvee-Martin, J. R. Parra, G. A. Gonzalez, W. Barker, and R. Duara, "Neuropathology, neuroimaging, and fluid biomarkers in alzheimer's disease," *Diagnostics*, vol. 14, no. 7, 2024.

[43] N. R. Barthélemy, G. Salvadó, S. E. Schindler, Y. He, S. Janelidze, L. E. Collij, B. Saef, R. L. Henson, C. D. Chen, B. A. Gordon, Y. Li, R. La Joie, T. L. S. Benzinger, J. C. Morris, N. Mattsson-Carlgren, S. Palmqvist, R. Ossenkoppele, G. D. Rabinovici, E. Stomrud, R. J. Bateman, and O. Hansson, "Highly accurate blood test for alzheimer's disease is similar or superior to clinical cerebrospinal fluid tests," *Nature Medicine*, vol. 30, pp. 1085–1095, Apr. 2024.

[44] J. H. Kang, M. Korecka, E. B. Lee, K. A. Q. Cousins, T. F. Tropea, A. A. Chen-Plotkin, D. J. Irwin, D. Wolk, M. Brylska, Y. Wan, and L. M. Shaw, "Alzheimer disease biomarkers: Moving from csf to plasma for reliable detection of amyloid and tau pathology," *Clinical Chemistry*, vol. 69, pp. 1247–1259, 09 2023.

[45] N. B. V. F. G. N. E. D. S. S. M. N. P. Tina Milos, Barbara Vuic and N. Pivac, "Cerebrospinal fluid in the differential diagnosis of alzheimer's

disease: an update of the literature," *Expert Review of Neurotherapeutics*, vol. 24, no. 11, pp. 1063–1079, 2024. PMID: 39233323.

[46] O. Hansson, H. Zetterberg, P. Buchhave, E. Londos, K. Blennow, and L. Minthon, "Association between csf biomarkers and incipient alzheimer's disease in patients with mild cognitive impairment: a follow-up study," *The Lancet Neurology*, vol. 5, p. 228–234, Mar. 2006.

[47] S. E. Schindler, J. D. Gray, B. A. Gordon, C. Xiong, R. Batrla-Utermann, M. Quan, S. Wahl, T. L. Benzinger, D. M. Holtzman, J. C. Morris, and A. M. Fagan, "Cerebrospinal fluid biomarkers measured by elecsys assays compared to amyloid imaging," *Alzheimer's amp; Dementia*, vol. 14, p. 1460–1469, Mar. 2018.

[48] C. R. J. Jr, D. A. Bennett, K. Blennow, M. C. Carrillo, H. H. Feldman, G. B. Frisoni, H. Hampel, and B. Dubois, "A/t/n: An unbiased descriptive classification scheme for alzheimer disease biomarkers," *Neurology*, vol. 87, no. 5, pp. 539–547, 2016.

[49] G. M. McKhann, D. S. Knopman, H. Chertkow, B. T. Hyman, C. R. Jack, C. H. Kawas, W. E. Klunk, W. J. Koroshetz, J. J. Manly, R. Mayeux, R. C. Mohs, J. C. Morris, M. N. Rossor, P. Scheltens, M. C. Carrillo, B. Thies, S. Weintraub, and C. H. Phelps, "The diagnosis of dementia due to alzheimer's disease: Recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease," *Alzheimer's amp; Dementia*, vol. 7, p. 263–269, Apr. 2011.

[50] P. Edison, H. A. Archer, R. Hinz, A. Hammers, N. Pavese, Y. F. Tai, G. Hotton, D. Cutler, N. Fox, A. Kennedy, M. Rossor, and D. J. Brooks, "Amyloid, hypometabolism, and cognition in alzheimer disease: An [11C]PIB and [18F]FDG pet study," *Neurology*, vol. 68, no. 7, pp. 501–508, 2007.

[51] P. S. Aisen, J. Cummings, C. R. Jack, J. C. Morris, R. Sperling, L. Frölich, R. W. Jones, S. A. Dowsett, B. R. Matthews, J. Raskin, P. Scheltens, and B. Dubois, "On the path to 2025: understanding the alzheimer's disease continuum," *Alzheimer's Research & Therapy*, vol. 9, p. 60, Aug. 2017.

[52] O. Hansson, A. Mikulskis, A. M. Fagan, C. Teunissen, H. Zetterberg, H. Vanderstichele, J. L. Molinuevo, L. M. Shaw, M. Vandijck, M. M. Verbeek, M. Savage, N. Mattsson, P. Lewczuk, R. Batrla, S. Rutz, R. A. Dean, and K. Blennow, "The impact of preanalytical variables on measuring cerebrospinal fluid biomarkers for alzheimer's disease diagnosis: A review," *Alzheimer's & Dementia*, vol. 14, no. 10, pp. 1313–1333, 2018.

[53] K. Blennow and H. Zetterberg, "Biomarkers for alzheimer's disease: current status and prospects for the future," *Journal of Internal Medicine*, vol. 284, p. 643–663, Aug. 2018.

[54] M. Boenink and L. van der Molen, "The biomarkerization of alzheimer's disease: From (early) diagnosis to anticipation?," in *A Pragmatic Approach to Conceptualization of Health and Disease*, pp. 141–162, Springer International Publishing, 2024.

[55] N. J. Ashton, A. Hye, A. P. Rajkumar, A. Leuzy, S. Snowden, M. Suárez-Calvet, T. K. Karikari, M. Schöll, R. La Joie, G. D. Rabinovici, K. Höglund, C. Ballard, T. Hortobágyi, P. Svenningsson, K. Blennow, H. Zetterberg, and D. Aarsland, "An update on blood-based biomarkers for non-alzheimer neurodegenerative disorders," *Nature Reviews Neurology*, vol. 16, p. 265–284, Apr. 2020.

[56] Y. Li, S. E. Schindler, J. G. Bollinger, V. Ovod, K. G. Mawuenyega, M. W. Weiner, L. M. Shaw, C. L. Masters, C. J. Fowler, J. Q. Trojanowski, M. Korecka, R. N. Martins, S. Janelidze, O. Hansson, and R. J. Bateman, "Validation of plasma Amyloid-β 42/40 for detecting alzheimer disease amyloid plaques," *Neurology*, vol. 98, pp. e688–e699, Dec. 2021.

[57] E. H. Thijssen, R. La Joie, A. Strom, C. Fonseca, L. Iaccarino, A. Wolf, S. Spina, I. E. Allen, Y. Cobigo, H. Heuer, L. VandeVrede, N. K. Proctor, A. L. Lago, S. Baker, R. Sivasankaran, A. Kieloch, A. Kinhikar, L. Yu, M.-A. Valentin, A. Jeromin, H. Zetterberg, O. Hansson, N. Mattsson-Carlgren, D. Graham, K. Blennow, J. H. Kramer, L. T. Grinberg, W. W. Seeley, H. Rosen, B. F. Boeve, B. L.

Miller, C. E. Teunissen, G. D. Rabinovici, J. C. Rojas, J. L. Dage, A. L. Boxer, and Advancing Research and Treatment for Frontotemporal Lobar Degeneration investigators, "Plasma phosphorylated tau 217 and phosphorylated tau 181 as biomarkers in alzheimer's disease and frontotemporal lobar degeneration: a retrospective diagnostic performance study," *Lancet Neurol*, vol. 20, pp. 739–752, Sept. 2021.

[58] Y. Jung and J. S. Damoiseaux, "The potential of blood neurofilament light as a marker of neurodegeneration for alzheimer's disease," *Brain*, vol. 147, pp. 12–25, Jan. 2024.

[59] P. Larkin, *Infrared and Raman spectroscopy: principles and spectral interpretation*. Elsevier, 2017.

[60] D. Naumann, *Infrared Spectroscopy of Cells, Tissues, and Biofluids*, pp. 1057–1065. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013.

[61] C. G. Atkins, K. Buckley, M. W. Blades, and R. F. Turner, "Raman spectroscopy of blood and blood components," *Applied spectroscopy*, vol. 71, no. 5, pp. 767–793, 2017.

[62] C. Farber, J. Li, E. Hager, R. Chemelewski, J. Mullet, A. Y. Rogachev, and D. Kurouski, "Complementarity of raman and infrared spectroscopy for structural characterization of plant epicuticular waxes," *ACS Omega*, vol. 4, no. 2, pp. 3700–3707, 2019.

[63] A. C. S. Talari, Z. Movasaghi, S. Rehman, and I. U. Rehman, "Raman spectroscopy of biological tissues," *Applied spectroscopy reviews*, vol. 50, no. 1, pp. 46–111, 2015.

[64] L. Marcu, S. A. Boppart, M. R. Hutchinson, J. Popp, and B. C. Wilson, "Biophotonics: the big picture," *Journal of Biomedical Optics*, vol. 23, no. 2, p. 021103, 2017.

[65] A. Rohman, A. Windarsih, E. Lukitaningsih, M. Rafi, K. Betania, and N. A. Fadzillah, "The use of FTIR and raman spectroscopy in combination with chemometrics for analysis of biomolecules in biomedical fluids: A review," *Biomedical Spectroscopy and Imaging*, vol. 8, pp. 55–71, 2019.

[66] S. Mordechai, E. Shufan, B. S. Porat Katz, and A. Salman, "Early diagnosis of alzheimer's disease using infrared spectroscopy of isolated blood samples followed by multivariate analyses," *Analyst*, vol. 142, pp. 1276–1284, 2017.

[67] M. S. K. L. Maria Paraskevaidi, Salman Karim and S. Crean, "The use of atr-ftir spectroscopy for the diagnosis of alzheimer's disease using oral buccal cells," *Applied Spectroscopy Reviews*, vol. 59, no. 8, pp. 1021–1035, 2024.

[68] C. V. Raman and K. S. Krishnan, "A new type of secondary radiation," *Nature*, vol. 121, pp. 501–502, Mar. 1928.

[69] S. Mosca, C. Conti, N. Stone, and P. Matousek, "Spatially offset raman spectroscopy," *Nature Reviews Methods Primers*, vol. 1, no. 1, pp. 1–16, 2021.

[70] D. Tuschel, "Selecting an excitation wavelength for raman spectroscopy," *Spectroscopy*, vol. 31, no. 3, pp. 14–23, 2016.

[71] C. G. Atkins, K. Buckley, D. Chen, H. G. Schulze, D. V. Devine, M. W. Blades, and R. F. Turner, "Raman spectroscopy as a novel tool for monitoring biochemical changes and inter-donor variability in stored red blood cell units," *Analyst*, vol. 141, no. 11, pp. 3319–3327, 2016.

[72] K. C. Doty and I. K. Lednev, "Differentiation of human blood from animal blood using raman spectroscopy: A survey of forensically relevant species," *Forensic science international*, vol. 282, pp. 204–210, 2018.

[73] I. Olaetxea, A. Valero, E. Lopez, H. Lafuente, A. Izeta, I. Jaunarena, and A. Seifert, "Machine learning-assisted raman spectroscopy for ph and lactate sensing in body fluids," *Analytical Chemistry*, vol. 92, no. 20, pp. 13888–13895, 2020.

[74] M. D. Fontana, K. B. Mabrouk, and T. H. Kauffmann, "Raman spectroscopic sensors for inorganic salts," *Spectrosc Prop Inorg Organomet Compd*, vol. 44, pp. 40–67, 2013.

[75] E. Ryzhikova, N. M. Ralbovsky, V. Sikirzhytski, O. Kazakov, L. Halamkova, J. Quinn, E. A. Zimmerman, and I. K. Lednev, "Raman spectroscopy and machine learning for biomedical applications: Alzheimer's disease diagnosis based on the analysis of cerebrospinal fluid," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 248, p. 119188, Mar. 2021.

[76] P. C. Ashok, M. E. Giardini, K. Dholakia, and W. Sibbett, "A raman spectroscopy bio-sensor for tissue discrimination in surgical robotics," *Journal of biophotonics*, vol. 7, no. 1-2, pp. 103–109, 2014.

[77] J. M. Amigo, ed., *Hyperspectral Imaging*, vol. 32 of *Data Handling in Science and Technology*. Elsevier, 2019.

[78] J. Marzi, L. Becker, and K. Schenke-Layland, "Advanced spectroscopic imaging techniques for tissue engineering and regenerative medicine," in *Bioimaging in Tissue Engineering and Regeneration* (A. Walter, P. Slezak, R. Mueller, G. Kerckhofs, and B. Bajoghli, eds.), Reference Series in Biomedical Engineering, ch. 7-1, Cham: Springer, 2024.

[79] S. Schlücker, M. D. Schaeberle, S. W. Huffman, and I. W. Levin, "Raman microspectroscopy: a comparison of point, line, and wide-field imaging methodologies," *Analytical Chemistry*, vol. 75, no. 16, pp. 4312–4318, 2003. PMID: 14632151.

[80] M. A. S. de Oliveira, M. Campbell, A. M. Afify, E. C. Huang, and J. W. Chan, "Hyperspectral raman microscopy can accurately differentiate single cells of different human thyroid nodules," *Biomed. Opt. Express*, vol. 10, pp. 4411–4421, Sep 2019.

[81] E. Hislop, W. Tipping, K. Faulds, and D. Graham, "Advanced surface-enhanced raman spectroscopy for molecular sensing applications," *Analytical Chemistry*, vol. 94, no. 25, pp. 8899–8908, 2022.

[82] R. Michael, A. Lenferink, G. Vrensen, E. Gelpi, R. Barraquer, and C. Otto, "Spectroscopic analysis of ocular tissues in alzheimer's disease," *Scientific Reports*, vol. 7, no. 1, p. 12345, 2017.

[83] A. Haessler, N. Jung, and M. Windbergs, "Recent advances in hyperspectral imaging for pharmaceutical and biomedical analysis," *Analytical Chemistry*, vol. 95, no. 48, pp. 17646–17653, 2023.

[84] J. Langer, D. Jimenez de Aberasturi, J. Aizpurua, R. A. Alvarez-Puebla, B. Auguié, J. J. Baumberg, G. C. Bazan, S. E. J. Bell, A. Boisen, A. G. Brolo, J. Choo, D. Cialla-May, V. Deckert, L. Fabris, K. Faulds, F. J. García de Abajo, R. Goodacre, D. Graham, A. J. Haes, C. L. Haynes, C. Huck, T. Itoh, M. Käll, J. Kneipp, N. A. Kotov, H. Kuang, E. C. Le Ru, H. K. Lee, J.-F. Li, X. Y. Ling, S. A. Maier, T. Mayerhöfer, M. Moskovits, K. Murakoshi, J.-M. Nam, S. Nie, Y. Ozaki, I. Pastoriza-Santos, J. Perez-Juste, J. Popp, A. Pucci, S. Reich, B. Ren, G. C. Schatz, T. Shegai, S. Schlücker, L.-L. Tay, K. G. Thomas, Z.-Q. Tian, R. P. Van Duyne, T. Vo-Dinh, Y. Wang, K. A. Willets, C. Xu, H. Xu, Y. Xu, Y. S. Yamamoto, B. Zhao, and L. M. Liz-Marzán, "Present and future of surface-enhanced raman scattering," *ACS Nano*, vol. 14, no. 1, pp. 28–117, 2020. PMID: 31478375.

[85] S.-Y. Ding, J. Yi, J.-F. Li, B. Ren, D.-Y. Wu, R. Panneerselvam, and Z.-Q. Tian, "Nanostructure-based plasmon-enhanced raman spectroscopy for surface analysis of materials," *Nature Reviews Materials*, vol. 1, p. 16021, Apr. 2016.

[86] C. Carlomagno, M. Cabinio, S. Picciolini, A. Gualerzi, F. Baglio, and M. Bedoni, "Sers-based biosensor for alzheimer disease evaluation through the fast analysis of human serum," *Journal of Biophotonics*, vol. 13, no. 3, p. e201960033, 2020.

[87] M. Kim, S. Huh, H. J. Park, S. H. Cho, M.-Y. Lee, S. Jo, and Y. S. Jung, "Surface-functionalized sers platform for deep learning-assisted diagnosis of alzheimer's disease," *Biosensors and Bioelectronics*, vol. 251, p. 116128, 2024.

[88] c. hao, D. Meng, W. Shi, C. Xu, Q. Wang, and H. Kuang, "Chiral gold nanostructure monolayers as sers substrates for ultrasensitive detection of enantiomer biomarkers of alzheimer's disease," *Angewandte Chemie International Edition*, vol. n/a, no. n/a, p. e202502115.

[89] J. Plou, P. S. Valera, I. García, C. D. L. de Albuquerque, A. Carracedo, and L. M. Liz-Marzán, "Prospects of surface-enhanced raman spectroscopy for biomarker monitoring toward precision medicine," *ACS Photonics*, vol. 9, no. 2, pp. 333–350, 2022.

[90] L. Vázquez-Iglesias, G. M. Stanfoca Casagrande, D. García-Lojo, L. Ferro Leal, T. A. Ngo, J. Pérez-Juste, R. M. Reis, K. Kant, and I. Pastoriza-Santos, "Sers sensing for cancer biomarker: Approaches and directions," *Bioactive Materials*, vol. 34, pp. 248–268, 2024.

[91] I. Valpapuram, P. Candeloro, M. L. Coluccio, E. I. Parrotta, A. Giugni, G. Das, G. Cuda, E. Di Fabrizio, and G. Perozziello, "Waveguiding and sers simplified raman spectroscopy on biological samples," *Biosensors*, vol. 9, no. 1, p. 37, 2019.

[92] C. M. M. S. Katharina Eberhardt, Clara Stiebing and J. Popp, "Advantages and limitations of raman spectroscopy for molecular diagnostics: an update," *Expert Review of Molecular Diagnostics*, vol. 15, no. 6, pp. 773–787, 2015. PMID: 25872466.

[93] J. M. Cameron, C. Rinaldi, S. H. Rutherford, A. Sala, A. G. Theakstone, and M. J. Baker, "Clinical spectroscopy: Lost in translation?," *Applied Spectroscopy*, vol. 76, no. 4, pp. 393–415, 2022. PMID: 34041957.

[94] G. Shobha and S. Rangaswamy, "Machine learning," in *Handbook of statistics*, vol. 38, pp. 197–228, Elsevier, 2018.

[95] I. El Naqa and M. J. Murphy, "What is machine learning?," in *machine learning in radiation oncology*, pp. 3–11, Springer, 2015.

[96] R. G. Brereton, *Chemometrics for pattern recognition*. John Wiley & Sons, 2009.

[97] B. M. Wise, N. Gallagher, R. Bro, J. Shaver, W. Windig, and R. S. Koch, "Pls toolbox 4.0," *Eigenvector Research Incorporated: Wenatchee, WA, USA*, 2007.

[98] D. Ballabio and R. Todeschini, *Multivariate Classification for Qualitative Analysis*. Elsevier, 2009.

[99] T. Bocklitz, A. Walter, K. Hartmann, P. Rösch, and J. Popp, "How to pre-process raman spectra for reliable and stable models?," *Analytica Chimica Acta*, vol. 704, no. 1, pp. 47–56, 2011.

[100] Rinnan, "Pre-processing in vibrational spectroscopy – when, why and how," *Anal. Methods*, vol. 6, pp. 7124–7129, 2014.

[101] S. J. Barton and B. M. Hennelly, "Signal to noise ratio of raman spectra of biological samples," in *Biophotonics: Photonic Solutions for Better Health Care VI*, vol. 10685, p. 106854F, International Society for Optics and Photonics, 2018.

[102] R. Gautam, S. Vanga, F. Ariese, and S. Umapathy, "Review of multidimensional data processing approaches for raman and infrared spectroscopy," *EPJ Techniques and Instrumentation*, vol. 2, no. 1, pp. 1–38, 2015.

[103] A. Mokari, S. Guo, and T. Bocklitz, "Exploring the steps of infrared (IR) spectral analysis: Pre-Processing, (classical) data modelling, and deep learning," *Molecules*, vol. 28, Sept. 2023.

[104] E. C. Muñoz, D. Ballabio, and J. M. Amigo, "Archer. a new algorithm for automatic removal of cosmic spikes and saturated pixels in hyperspectral raman spectroscopy," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 336, p. 126041, 2025.

[105] P. H. Eilers and H. F. Boelens, "Baseline correction with asymmetric least squares smoothing," *Leiden University Medical Centre Report*, vol. 1, no. 1, p. 5, 2005.

[106] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures.," *Analytical Chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.

[107] J. Luo, K. Ying, and J. Bai, "Savitzky–golay smoothing and differentiation filter for even number data," *Signal processing*, vol. 85, no. 7, pp. 1429–1434, 2005.

[108] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 2, no. 1, pp. 37–52, 1987. Proceedings of the Multivariate Statistical Workshop for Geologists and Geochemists.

[109] R. Tauler, "Multivariate curve resolution applied to second order data," *Chemometrics and Intelligent Laboratory Systems*, vol. 30, no. 1, pp. 133–146, 1995.

[110] A. de Juan and R. Tauler, "Multivariate curve resolution: 50 years addressing the mixture analysis problem – a review," *Analytica Chimica Acta*, vol. 1145, pp. 59–78, 2021.

[111] J. Jaumot, R. Gargallo, A. de Juan, and R. Tauler, "A graphical user-friendly interface for mcr-als: a new tool for multivariate curve resolution in matlab," *Chemometrics and Intelligent Laboratory Systems*, vol. 76, no. 1, pp. 101–110, 2005.

[112] M. Barker and W. Rayens, "Partial least squares for discrimination," *Journal of Chemometrics*, vol. 17, no. 3, pp. 166–173, 2003.

[113] C. M. Andersen and R. Bro, "Variable selection in regression—a tutorial," *Journal of Chemometrics*, vol. 24, no. 11-12, pp. 728–737, 2010.

[114] L. Nørgaard, A. Saudland, J. Wagner, J. P. Nielsen, L. Munck, and S. B. Engelsen, "Interval partial least-squares regression (ipls): A comparative chemometric study with an example from near-infrared spectroscopy," *Applied Spectroscopy*, vol. 54, no. 3, pp. 413–419, 2000.

[115] R. E. Shaffer, "Multi- and megavariate data analysis. principles and applications, i. eriksson, e. johansson, n. kettaneh-wold and s. wold, umetrics academy, umeå, 2001, isbn 91-973730-1-x, 533pp.," *Journal of Chemometrics*, vol. 16, no. 5, pp. 261–262, 2002.

[116] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, pp. 389–422, Jan. 2002.

[117] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, (Red Hook, NY, USA), p. 4768–4777, Curran Associates Inc., 2017.

[118] M. B. Kursa, A. Jankowski, and W. R. Rudnicki, "Boruta – a system for feature selection," *Fundamenta Informaticae*, vol. 101, no. 4, pp. 271–285, 2010.

[119] C. Bertinetto, J. Engel, and J. Jansen, "Anova simultaneous component analysis: A tutorial review," *Analytica Chimica Acta: X*, vol. 6, p. 100061, 2020.

[120] J. J. Jansen, H. C. J. Hoefsloot, J. van der Greef, M. E. Timmerman, J. A. Westerhuis, and A. K. Smilde, "Asca: analysis of multivariate data obtained from an experimental design," *Journal of Chemometrics*, vol. 19, no. 9, pp. 469–481, 2005.

[121] J. Camacho, R. Vitale, D. Morales-Jiménez, and C. Gómez-Llorente, "Variable-selection anova simultaneous component analysis (vasca)," *Bioinformatics*, vol. 39, p. btac795, 12 2022.

[122] E. Saccenti, A. K. Smilde, and J. Camacho, "Group-wise ANOVA simultaneous component analysis for designed omics experiments," *Metabolomics*, vol. 14, p. 73, May 2018.

[123] P. Refaeilzadeh, L. Tang, and H. Liu, "Cross-validation.," *Encyclopedia of database systems*, vol. 5, pp. 532–538, 2009.

[124] A. Tharwat, "Classification assessment methods," *Applied Computing and Informatics*, 2020.

[125] M. Vihinen, "How to evaluate performance of prediction methods? measures and their interpretation in variation effect analysis," in *BMC genomics*, vol. 13, pp. 1–10, BioMed Central, 2012.

[126] R. D. Deegan, O. Bakajin, T. F. Dupont, G. Huber, S. R. Nagel, and T. A. Witten, "Capillary flow as the cause of ring stains from dried liquid drops," *Nature*, vol. 389, pp. 827–829, Oct. 1997.

[127] H. Hu and R. G. Larson, "Evaporation of a sessile droplet on a substrate," *The Journal of Physical Chemistry B*, vol. 106, no. 6, pp. 1334–1344, 2002.

[128] D. Mampallil and H. B. Eral, "A review on suppression and utilization of the coffee-ring effect," *Advances in Colloid and Interface Science*, vol. 252, pp. 38–54, 2018.

[129] M. J. Hertaeg, C. Rees-Zimmerman, R. F. Tabor, A. F. Routh, and G. Garnier, "Predicting coffee ring formation upon drying in droplets of particle suspensions," *Journal of Colloid and Interface Science*, vol. 591, pp. 52–57, 2021.

[130] M. Yang, D. Chen, J. Hu, X. Zheng, Z.-J. Lin, and H. Zhu, "The application of coffee-ring effect in analytical chemistry," *TrAC Trends in Analytical Chemistry*, vol. 157, p. 116752, 2022.

[131] R. D. Deegan, "Pattern formation in drying drops," *Phys. Rev. E*, vol. 61, pp. 475–485, Jan 2000.

[132] S. K. Wilson and H.-M. Dapos;Ambrosio, "Evaporation of sessile droplets," *Annual Review of Fluid Mechanics*, vol. 55, no. Volume 55, 2023, pp. 481–509, 2023.

[133] H. Hu and R. G. Larson, "Marangoni effect reverses coffee-ring depositions," *The Journal of Physical Chemistry B*, vol. 110, no. 14, pp. 7090–7094, 2006. PMID: 16599468.

[134] Y. J. P. Carreón, M. L. Gómez-López, O. Díaz-Hernández, P. Vazquez-Vergara, R. E. Moctezuma, J. M. Saniger, and J. González-Gutiérrez, "Patterns in dried droplets to detect unfolded bsa," *Sensors*, vol. 22, no. 3, 2022.

[135] A. Pal, A. Gope, and A. Sengupta, "Drying of bio-colloidal sessile droplets: Advances, applications, and perspectives," *Advances in Colloid and Interface Science*, vol. 314, p. 102870, 2023.

[136] R. Chen, L. Zhang, D. Zang, and W. Shen, "Blood drop patterns: Formation and applications," *Advances in Colloid and Interface Science*, vol. 231, pp. 1–14, 2016.

[137] M. A. Hack, M. N. van der Linden, H. Wijshoff, J. H. Snoeijer, and T. Segers, "Ring-shaped colloidal patterns on saline water films," *Journal of Colloid and Interface Science*, vol. 673, pp. 788–796, 2024.

[138] M. E. Buzoverya, Y. P. Shcherbak, and I. V. Shishpor, "Quantitative estimation of the microstructural inhomogeneity of biological fluid facies," *Technical Physics*, vol. 59, pp. 1550–1555, Oct. 2014.

[139] M. Efstratiou, J. Christy, D. Bonn, and K. Sefiane, "The effect of substrate temperature on the evaporative behaviour and desiccation patterns of foetal bovine serum drops," *Colloids and Interfaces*, vol. 5, no. 4, 2021.

[140] B. Sobac and D. Brutin, "Structural and evaporative evolutions in desiccating sessile drops of blood," *Phys. Rev. E*, vol. 84, p. 011603, Jul 2011.

[141] F. Wang and Q. Yuan, "Evaporation-induced fractal patterns: A bridge between uniform pattern and coffee ring," *Journal of Colloid and Interface Science*, vol. 637, pp. 522–532, 2023.

[142] A. Matavž, U. Uršič, J. Močivnik, D. Richter, M. Humar, S. Čopar, B. Malič, and V. Bobnar, "From coffee stains to uniform deposits: Significance of the contact-line mobility," *Journal of Colloid and Interface Science*, vol. 608, pp. 1718–1727, 2022.

[143] J. M. Cameron, H. J. Butler, D. S. Palmer, and M. J. Baker, "Biofluid spectroscopic disease diagnostics: A review on the processes and spectral impact of drying," *J Biophotonics*, vol. 11, p. e201700299, Mar. 2018.

[144] K. A. Esmonde-White, F. W. L. Esmonde-White, M. D. Morris, and B. J. Roessler, "Characterization of biofluids prepared by sessile drop formation," *Analyst*, vol. 139, pp. 2734–2741, 2014.

[145] F. Bonnier, M. J. Baker, and H. J. Byrne, "Vibrational spectroscopic analysis of body fluids: avoiding molecular contamination using centrifugal filtration," *Anal. Methods*, vol. 6, pp. 5155–5160, 2014.

2

Chapter 2

General conclusions

2.1 Alzheimer's disease biomarkers from Raman spectroscopy

This thesis demonstrates that the combination of drop-coating deposition Raman (DCDR) spectroscopy with rigorous chemometric modeling can reveal early Alzheimer's disease (AD) pathology. In Lopez *et al.* (Int. J. Mol. Sci., 2024) dried cerebrospinal fluid (CSF) droplets from two longitudinally characterized cohorts which were acquired on the same Raman instrument in separate measurement years. After cosmic-ray removal, spectral baseline correction, and further data preprocessing, the spectra from the two independent cohorts were merged into the same matrix and analyzed by partial least-squares discriminant analysis PLS-DA with a specific variable-selection protocol. By the use of a compact set of the selected bands discrimination between cognitively normal and individuals on the preclinical AD continuum was achieved. This result came from a cross-validation which respected the patient hierarchy and returned accuracies of up to 96 %.

This study provides the first *multi-year* liquid-biopsy proof-of-principle for Raman-based AD screening; nevertheless, several limitations remain. Predictive

performance is sensitive to the cross-validation strategy, measurement success still depends on the operator's choice of sampling location, and the overall cohort size (although larger than in previous reports) remains moderate. These limitations motivate the refinements detailed in the following sections

2.2 Importance of model validation

To obtain trustworthy performance estimates, cross-validation must respect the data hierarchy: every set of replicate spectra belonging to a single patient, sample, or measurement run should be grouped within one fold rather than scattered across folds. In Lopez *et al.* (Anal. Chim. Acta, 2023) popular resampling schemes were tested on three archetypal datasets: (i) Raman spectra of COVID-19 nasopharyngeal swabs, (ii) Raman spectra for wine-origin authentication, and (iii) FTIR spectra of cerebrospinal fluid (CSF) from Alzheimer's disease (AD) and control subjects. Allowing replicate spectra from the same experimental unit to leak across folds inflated every performance metric (accuracy, sensitivity, and specificity) by up to 20 percent, whereas patient- or sample-blocked partitioning yielded estimates consistent with external validation test sets.

Two straightforward, independent statistical tests revealed how strongly data leakage across folds distorts model performance. Permutation tests failed to reject the null hypothesis when leakage was present (median $p \approx 0.20$), but produced $p < 0.01$ once replicate blocking was enforced. Hotelling T^2 control charts further flagged spectra with abnormally high leverage that clustered in the leaked folds, revealing overfitting driven by data duplication. These orthogonal checks confirm that blocking replicates during resampling is non-negotiable for honest performance assessment and are the basis for the six-item reporting checklist proposed in the paper: (1) experimental hierarchy, (2) split strategy, (3) hyper-parameter grid, (4) repeat count, (5) confidence intervals, and (6) permutation-test p-values, a minimum transparency standard for spectroscopy-based machine-learning research.

2.3 Optimizing biofluid measurement protocols

Droplet-based spectroscopy has proved its diagnostic power, but its routine use still relies on *how* and *where* each spectrum is collected. The way solutes migrate while a biofluid dries creates rich chemical gradients that a single laser spot can easily miss. Optimizing the measurement protocol therefore matters just as much as choosing the right classifier.

In Lopez *et.al VIEW*, 2025, authors move from single-point DCDR to full-drop hyperspectral Raman mapping of CSF and plasma droplets. Each 1.4 mm-diameter deposit was sampled on a 35 – 45 µm square grid, yielding ∼750 – 950 spectra per specimen. Principal-component analysis (PCA), k-means clustering, and multivariate curve resolution-alternating least squares (MCR - ALS) exposed a complex "coffee-ring" morphology: while the outer rim is protein-rich, several Alzheimer's-linked bands highlighted in [20] (1045, 1065, and 1453 cm^{-1}) reach maximal intensity in inner belts and fern-like structures in mid-zones rather than at the rim itself. Cluster-level spectra further reveal that the rim hosts at least two distinct protein signatures, contradicting the common assumption that it can be treated as a chemically homogeneous region of interest. Raman hyperspectral mapping thus becomes a two-step optimization tool:

First, the full-drop map is analyzed with unsupervised methods like PCA, k-means, and MCR - ALS to produce a high-resolution chemical map. This data-driven map shows that several disease-linked bands concentrate in zones located between the protein-rich ring and the droplet centre areas that earlier ring-based protocols overlooked. Second, the full map guides a *targeted sampling pattern*: clustered regions, selected *after* the unsupervised analysis, probes only the diagnostically rich zones in subsequent samples. This reduces acquisition time by roughly an order of magnitude and eliminates the sampling bias built into ring-only protocols.

By shifting the focus from merely protein-rich regions to those most indicative of the disease, the mapping-guided protocol delivers spectra that are more representative and reproducible while minimizing analyst workload. This are also important prerequisites for a clinical translation.

2.4 Future research directions

Photonics-based spectroscopy coupled to robust machine-learning pipelines now offers a credible route to detecting AD pathology before symptoms occur, however, several milestones remain before the approach can be adopted in clinical practice. First, extending the workflow to *multi-modal vibrational spectroscopy*, this is, acquiring Raman, FTIR and SERS spectra from the same sample under harmonized preprocessing and joint feature-selection rules. Employing the corresponding multi-block and data fusion methods. This way the models should capture complementary molecular information and raise both sensitivity and specificity. In parallel, the biomarker panel can be tried in additional biofluids like saliva and urine, to test for even less-invasive biofluids.

A second challenge is *spectroscopy–metabolomics* validation. In ongoing work with the *IBeA* group low-molecular-weight metabolites are quantified in the same clinically characterized CSF and plasma samples using untargeted NMR and HPLC-HRMS. Aligning these metabolic fingerprints with the Raman bands discovered in this thesis should assign specific molecular identities to the most discriminative spectral features and turn abstract spectra into actionable biochemical targets for therapeutic research, and supply an extra evidence for the subsequent ML model creation.

Finally, translational credibility will require larger, prospective cohorts recruited across multiple clinical centres; each trial should follow a preregistered analysis plan, be assessed with blinded external test sets, and report the six validation items listed in Section 2.2. In parallel, automating sample handling and implementing the mapping-guided targeting strategy described in Section 2.3 can shorten acquisition time without sacrificing representativeness. Taken together, these clinical and engineering advances will carry the method from laboratory feasibility to a clinically actionable diagnostic for Alzheimer's disease.

Appendix A

A

Publications as First Author

A.1 Unlocking preclinical Alzheimer's with Raman spectroscopy

Eneko Lopez, Jaione Etxebarria-Elezgarai, Maite García-Sebastián, Miren Altuna, Mirian Ecay-Torres, Ainara Estanga, Mikel Tainta, Carolina López, Pablo Martínez-Lage, Jose Manuel Amigo, Andreas Seifert. **Unlocking Preclinical Alzheimer's: A Multi-Year Label-Free In Vitro Raman Spectroscopy Study Empowered by Chemometrics**.International Journal of Molecular Sciences 25, 9:4737, (**2024**).

The manuscript was cited as ref [13]. It is accesible under

https://www.mdpi.com/1422-0067/25/9/4737

International Journal of
Molecular Sciences

MDPI

Article

Unlocking Preclinical Alzheimer's: A Multi-Year Label-Free In Vitro Raman Spectroscopy Study Empowered by Chemometrics

Eneko Lopez [1,2], Jaione Etxebarria-Elezgarai [1], Maite García-Sebastián [3], Miren Altuna [3], Mirian Ecay-Torres [3], Ainara Estanga [3], Mikel Tainta [3], Carolina López [3], Pablo Martínez-Lage [3], Jose Manuel Amigo [4,5,*] and Andreas Seifert [1,4,*]

[1] CIC nanoGUNE BRTA, 20018 San Sebasián, Spain; e.lopez@nanogune.eu (E.L.); j.etxebarria@nanogune.eu (J.E.-E.)
[2] Department of Physics, University of the Basque Country (UPV/EHU), 20018 San Sebastián, Spain
[3] Center for Research and Advanced Therapies, CITA-Alzhéimer Foundation, 20009 San Sebastián, Spain; mgarcia@cita-alzheimer.org (M.G.-S.); maltuna@cita-alzheimer.org (M.A.); mecay@cita-alzheimer.org (M.E.-T.); aestanga@cita-alzheimer.org (A.E.); mtainta@cita-alzheimer.org (M.T.); clopez@cita-alzheimer.org (C.L.); pmlage@cita-alzheimer.org (P.M.-L.)
[4] IKERBASQUE, Basque Foundation for Science, 48009 Bilbao, Spain
[5] Department of Analytical Chemistry, University of the Basque Country, 48940 Leioa, Spain
* Correspondence: josemanuel.amigo@ehu.eus (J.M.A.); a.seifert@nanogune.eu (A.S.)

Abstract: Alzheimer's disease is a progressive neurodegenerative disorder, the early detection of which is crucial for timely intervention and enrollment in clinical trials. However, the preclinical diagnosis of Alzheimer's encounters difficulties with gold-standard methods. The current definitive diagnosis of Alzheimer's still relies on expensive instrumentation and post-mortem histological examinations. Here, we explore label-free Raman spectroscopy with machine learning as an alternative to preclinical Alzheimer's diagnosis. A special feature of this study is the inclusion of patient samples from different cohorts, sampled and measured in different years. To develop reliable classification models, partial least squares discriminant analysis in combination with variable selection methods identified discriminative molecules, including nucleic acids, amino acids, proteins, and carbohydrates such as taurine/hypotaurine and guanine, when applied to Raman spectra taken from dried samples of cerebrospinal fluid. The robustness of the model is remarkable, as the discriminative molecules could be identified in different cohorts and years. A unified model notably classifies preclinical Alzheimer's, which is particularly surprising because of Raman spectroscopy's high sensitivity regarding different measurement conditions. The presented results demonstrate the capability of Raman spectroscopy to detect preclinical Alzheimer's disease for the first time and offer invaluable opportunities for future clinical applications and diagnostic methods.

Keywords: preclinical Alzheimer's; cerebrospinal fluid; vibrational spectroscopy; machine learning; PLS-DA; variable selection

Citation: Lopez, E.; Etxebarria-Elezgarai, J.; García-Sebastián, M.; Altuna, M.; Ecay-Torres, M.; Estanga, A.; Tainta, M.; López, C.; Martínez-Lage, P.; Amigo, J.M.; et al. Unlocking Preclinical Alzheimer's: A Multi-Year Label-Free In Vitro Raman Spectroscopy Study Empowered by Chemometrics. *Int. J. Mol. Sci.* **2024**, *25*, 4737. https://doi.org/10.3390/ijms25094737

Academic Editor: Ivan Bratchenko

Received: 8 April 2024
Revised: 23 April 2024
Accepted: 24 April 2024
Published: 26 April 2024

1. Introduction

Alzheimer's disease (AD) is the primary cause of cognitive impairment and represents the main prevalent neurodegenerative disorder and challenge. Characterized by a continuum, it includes a very large preclinical (PC) stage followed by a mild cognitive impairment (MCI) period, leading ultimately to a dementia stage. The importance of early AD diagnosis cannot be overestimated. Timely and successful initiation of treatment is crucial to halting the progression of the disease [1–3]. Early detection primarily involves preclinical AD (PC-AD), originally defined in the late 20th century as cognitively unimpaired individuals who exhibit AD brain lesions post-mortem. With the addition of pathologic AD markers, PC-AD now includes cases in which these markers are also present in cognitively normal

Int. J. Mol. Sci. 2024, 25, 4737

individuals [4,5]. To date, the European Academy of Neurology and other associations do not approve PC diagnosis in clinical practice. Currently, the diagnosis of MCI or dementia is made in clinical practice, and the recognition of PC stages is limited to the research context. Nevertheless, the relevance of early diagnosis cannot be underestimated, particularly given the potential of emerging treatments such as Lecanemab and Donanemab. These drugs promise to change the course of the disease, especially if they are used in the preclinical stages, as the ongoing AHEAD study to investigate the efficacy of Lecanemab in this context shows [6].

According to the International Working Group (IWG), "The diagnosis of AD is clinical-biological and requires the presence of both a specific clinical phenotype of AD and biomarker evidence of AD pathology" [7]. This statement makes the correct diagnosis of the disease even more difficult, as the disease is progressive and biomarkers change during years preceding the disease [8]. Its gradual progression covers a prolonged PC phase marked by sequential amyloid peptide and tau protein deposition, culminating in neurodegeneration preceding clinical symptomatology [7,9–11].

Although previous studies have addressed the identification of the preclinical stage, existing evidence remains limited regarding its traditional diagnostic approaches to neurological–neuropsychological assessments and the analysis of amyloid tau neurodegeneration (ATN) biomarkers (amyloid beta Aβ, phosphorylated tau and total tau) in cerebrospinal fluid (CSF) [5,12–14]. The ATN classification system classifies AD biomarkers into three groups, providing information on neuropathological changes [15–17]. AD-specific biomarkers are crucial since neurodegeneration and lesions can also occur in non-AD diseases, especially in older people with other pathologies [11,18]. Positron emission tomography (PET) imaging is used for this type of analysis; it uses radiotracers binding to Aβ or tau plaques in the brain and offers high diagnostic accuracy and localized information [19]. Another analytical route is CSF extraction by lumbar puncture and enzyme-linked immunosorbent assays (ELISA). In this case, it is possible to evaluate brain pathology and measure Aβ and tau biomarkers from the same collection. These techniques are proven, and studies have shown a strong correlation between CSF biomarkers and PET results [11,20,21]. Consequently, biomarkers are used to support the diagnosis of AD, whereas clinical diagnosis is used to identify AD severity.

The recent increase in imaging and fluid biomarkers of AD pathophysiology provides the opportunity to identify several biological stages in the preclinical phase of AD [22]. Positive Aβ and tau biomarkers can be observed in individuals without cognitive impairment (PC-AD), in those with MCI, and in those with dementia (Figure 1).

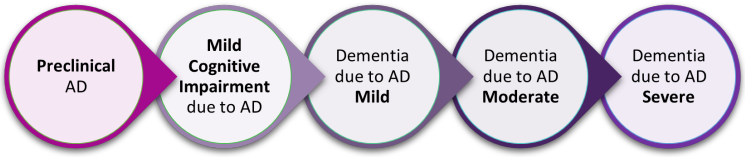


Figure 1. Alzheimer's disease continuum as defined by the International Working Group (IWG).

In the search for reliable biomarkers for neurodegenerative diseases applying novel methods, various molecular markers in tissues, biofluids, and imaging techniques are currently being investigated [23]. Although mass spectrometry (MS) and ELISA stand out as established biomarker identification and quantification techniques, they have their limitations: they are destructive, time-consuming, expensive, and require highly trained personnel [23,24].

The integration of imaging and fluid biomarkers has expanded our understanding of AD pathophysiology, offering insights into various biological stages preceding clinical symptoms. In this endeavor, novel methods are being explored, with Raman spectroscopy emerging as a holistic, more cost-effective, non-destructive, and technically less complex al-

Int. J. Mol. Sci. **2024**, *25*, 4737

ternative that overcomes the limitations of conventional approaches. Raman spectroscopy is a label-free and rapid spectroscopic method that provides chemical and structural information by detecting inelastically scattered photons [25]. Moreover, it requires minimal sample preparation, eliminates the need for additional chemicals, is non-destructive, and significantly reduces analysis time. The performance of Raman spectroscopy in the classification of AD, as reported by Xu and co-authors in a comprehensive review encompassing eight selected studies [26], demonstrates high sensitivity (0.86) and specificity (0.87), indicative of its potential for future medical diagnostics.

A drawback of Raman spectroscopy with biological samples is that Raman spectra often contain non-relevant signals superimposed onto interesting ones, along with a number of artifacts (e.g., baseline drifts) that make it difficult to directly interpret the Raman peaks obtained. Therefore, it is of utmost importance to analyze Raman spectra using sophisticated machine learning methods (chemometrics) and, in particular, in combination with variable selection methods to minimize artifacts, highlight important signals and be able to build reliable classification models, which have been investigated to diagnose AD in biofluidic samples [12,27–30]. Most of the published research is based on limited statistics with small cohorts and, therefore, compromises the accuracy of classification and additional variable selection methods, providing limited insight into the physiological origin of classification results.

In this manuscript, we put forward a new and more reliable approach by combining the datasets from two cohorts from different years, thus increasing the number of participants and enhancing the robustness and reliability of the models beyond what is commonly found in the literature. The combined dataset provides significantly larger sample sizes compared to those reported in similar proof-of-concept studies utilizing both Raman spectroscopy and surface-enhanced Raman spectroscopy for the identification of AD in bodily fluids [12,27–31]. Although Raman spectroscopy is a very sensitive technique and susceptible to changing environmental conditions, we demonstrate here that it is possible to build stable classification models with Raman datasets from different cohorts measured during different years (2022 and 2023). The focus of this study is on the early detection of AD, i.e., the classification of preclinical Alzheimer's, which has not yet been investigated in other studies. It is much more difficult to detect physiological changes in PC-AD, and we address this challenge with chemometric methods employing specific variable selection.

2. Results and Discussion

As explained in Section 3, the CSF samples were obtained from two studies performed during different years, and the Raman measurements for each of these sample sets were also performed in different years at a later time point. We refer to the corresponding datasets here as Dataset 1 and Dataset 2. Figure 2 presents the Raman spectra corresponding to Dataset 1 (light pink), Dataset 2 (dark pink), and their combined (red) form (Dataset 1 + Dataset 2), referring to the two cohorts with Raman measurements from different years. The initial focus is on variable selection for each dataset to identify molecular fingerprints indicative of PC-AD. Employing the variable selection strategy with a cross-validation of 15 random subsets and 5 iterations, Dataset 1 and Dataset 2 were reduced to 93 and 50 discriminative wavenumbers, respectively. Assembling both datasets into a unified one, the same variable selection procedure was systematically applied, discovering 213 discriminating variables. From the representative wavenumbers of each cohort, as illustrated by the dashed lines in Figure 2, distinctive and common spectral patterns emerge for the control and PC-AD groups within the Dataset 1 and Dataset 2 studies. This observation shows the potential utility of Raman spectroscopy in detecting molecular alterations associated with PC-AD. The clear differences in spectral profiles within each dataset and the presence of shared spectral features across Dataset 1, Dataset 2, and their combination indicate subtle molecular changes linked to PC-AD progression.

Int. J. Mol. Sci. **2024**, *25*, 4737

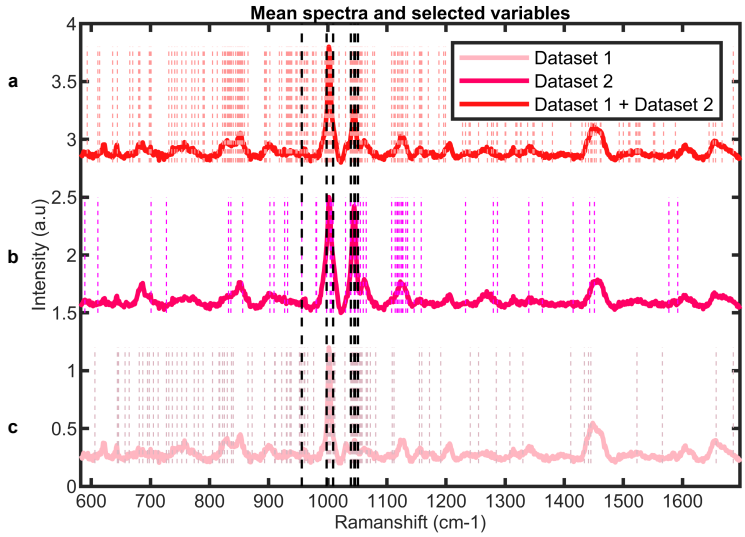


Figure 2. Averaged spectra and selected variables for the (**a**) Dataset 1 + Dataset 2 study. (**b**) Dataset 2 study. (**c**) Dataset 1 study. Dashed lines in black represent wavenumbers selected in common across all three datasets.

A more in-depth analysis of the chosen wavenumbers shows agreement with crucial peaks described in the literature as decisive for Alzheimer's discrimination. The identified peaks, among others, are outlined in Table 1.

Table 1. Characteristic molecular vibrations for discriminating PC-AD extracted from Raman spectra.

Wavenumber (cm^{-1})	Biomarkers	Description [32]
727	Nucleic acids	Phosphatidylserine, hypotaurine, guanine
956	Proteins, carbohydrates	ν_1 of the phosphate group, guanine
998		Monophosphate group
1009	Phenylalanine	Tryptophan
1039	Proteins, carbohydrates	Taurine
1045		Hypotaurine
1046		Hypotaurine, taurine
1051		Taurine
1065		Hypotaurine

In particular, the specific bands at 1045 cm^{-1} and 1065 cm^{-1} exhibit significant intensity changes that correlate with alterations in amino acids attributed to AD biomarkers such as tau proteins and Aβ42 peptides [29]. These spectral peaks may correspond to taurine and hypotaurine [32], both of which are amino acid derivatives. Taurine, known for its diverse physiological functions essential for overall health and wellbeing, acts as an osmoregulatory agent [33]. Moreover, experimental studies have shown the binding of taurine with oligomeric Aβ plaques [34], preventing the neurotoxicity of Aβ and glutamate receptor agonists, which indicates a potential interaction between taurine and key pathological features associated with AD. Interestingly, the identification of characteristic bands at 727 cm^{-1} and 956 cm^{-1} shared by both guanine and phosphatidylserine (PS) highlights potential molecular mechanisms underlying AD pathology. Guanosine, known for its neuroprotective effects, is a derivative of guanine, a nucleotide base present in DNA and RNA structures. Guanine's association with purinergic signaling and its conversion

Int. J. Mol. Sci. **2024**, *25*, 4737

to guanosine suggests a potential link between purinergic signaling pathways and AD pathology [35]. Similarly, PS, a structural component of eukaryotic membranes, plays a multifaceted role in many biological processes, including enzyme activation, apoptosis, and neurotransmission. The dysregulation of PS and other phospholipids in AD brains alters membrane viscosity and hampers essential biological processes, potentially contributing to synaptic dysfunction and neurodegeneration [36,37].

The correlations identified in the spectral analysis highlight the complex interactions of molecular components in AD pathology and emphasize the need for comprehensive research to understand its underlying mechanisms. The subsequent deeper analysis separately evaluates machine learning models for each study (Dataset 1 and Dataset 2), as shown in Table 2. The best model, which was determined for both cohorts individually, has good performance indicators. While the Dataset 1 study demonstrates comparatively more representative features and latent variables for optimal prediction (93 features and 4 LVs for Dataset 1 and 50 features and 3 LVs for Dataset 2), it still achieves commendable accuracies of up to 0.93, although slightly below the Dataset 2 model's accuracy of 0.97. Despite variations in the studies of individual years and the timeline of Raman measurements, the potential for PC-AD classification remains significant.

The decision to develop a unified model incorporating both studies is motivated by several factors: first, to comprehensively analyze both datasets with a larger sample size for an expected improved discrimination; second, to investigate whether combining data sets could improve the predictive abilities of the model by capturing a broader range of characteristics and patterns; lastly, to investigate common variables across the data sets and discovering common factors that clearly influence AD classification, which was triggered by the unified approach. In contrast to individual cohort-specific models, the unified model requires a higher complexity in terms of the number of latent variables for optimum performance, which is six compared to three and four LVs for individual sets Dataset 1 and Dataset 2. However, higher statistics did not improve the model's predictive ability because of its complexity. Of course, the differences in measurement years and cohort characteristics may have influenced these outcomes. Additionally, when considering a model with variables selected jointly for all three datasets (Dataset 1, Dataset 2, and Dataset 1 + Dataset 2), a decreased discriminative power was observed, emphasizing the impact of factors such as cohort, year, or measurement strategy. The augmentation of Raman datasets by different measurement cycles, varying measurement conditions, and/or environmental changes subject to many factors will definitely increase uncertainty in a combined dataset. However, augmentation will finally lead to a stable model that considers all uncertainty factors and can be used as a clinical predictive tool. Whether the predictive power is worse or better cannot be foreseen, as long as the inner structure of such hierarchical datasets is not fully captured.

Table 2. Figures of merit for PLS-DA models for preclinical AD prediction across different datasets. Dataset 1 and Dataset 2 individually; variable selection across the whole dataset: Dataset 1 + Dataset 2; shared variables in all datasets: Dataset 1 + Dataset 2 in common; variables occurring more than 30 times across 100 iterations for the combined dataset: Dataset 1 + Dataset 2 thr@30.

Cohort	Matrix	AUC	Accuracy	Sensitivity	Specificity	LVs	Variables
Dataset 1	40×93	0.99	0.93	0.95	0.92	4	93
Dataset 2	35×50	1.00	0.97	0.93	0.98	3	50
Dataset 1 + Dataset 2	75×213	0.98	0.93	0.91	0.94	6	213
Dataset 1 + Dataset 2 in common	75×9	0.61	0.53	0.63	0.51	2	9
Dataset 1 + Dataset 2 thr@30 [1]	75×168	0.99	0.96	0.93	0.96	6	168

[1] Variables occurring more than 30 times over the process of 100 iterations.

By extending the analysis, we employed an iterative approach for the variable selection procedure to increase discriminative information and reveal meaningful patterns in the

Int. J. Mol. Sci. **2024**, *25*, 4737

frequency analysis of selected wavenumbers. Figure 3a shows the frequency distribution of selected wavenumbers over 100 iterations, which motivated the creation of machine learning models for various frequency thresholds. Figure 3b displays heat maps of figures of merit for all the created models and highlights the optimal PLS-DA model with six latent variables and wavenumbers for frequencies exceeding 30. This model achieves a remarkable AUC of 0.99 and an accuracy of 0.96 in predicting PC-AD cases, providing a competitive model for its statistically enhanced robustness and performance comparable to single cohorts. Iterative cross-validation or jackknifing improves the discrimination information extracted from the combined data.

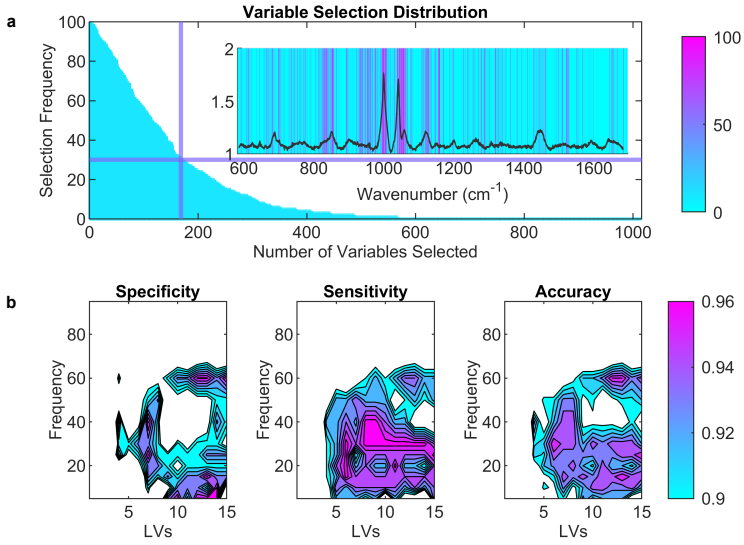


Figure 3. (**a**) Number of variables selected and frequency of each selected variable during the 100−fold iteration process. (**b**) Heat maps of specificity, sensitivity, and accuracy for each frequency threshold for the different models.

Alternatively, one could imagine constructing a model that only contains the variables that were selected together in all three data sets (Dataset 1, Dataset 2, and Dataset 1 + Dataset 2). However, the exclusive use of these selected wavenumbers, as shown in Table 1, proves to be insufficient for building machine learning models and leads to a decrease in classification performance. Various factors contribute to the selection of additional variables, especially cohort and measurement year. Figure 4 shows the model scores derived from Table 2, thereby providing a visual representation. The two-latent space depicted in the upper part of Figure 4, which includes LV1 and LV2, highlights the models' discriminative capacity for PC-AD in both studies on Dataset 1 and Dataset 2. In particular, the color variation in the labels, as observed in the lower part of Figure 4, indicates clustering between healthy and preclinical groups, whereas the studies from Dataset 1 and Dataset 2 exhibit clustering in a different direction. This explains why more variables and model latent spaces are required when combining the studies from Dataset 1 and Dataset 2 to build a more robust and predictive model for the discrimination of PC-AD.

Int. J. Mol. Sci. **2024**, *25*, 4737

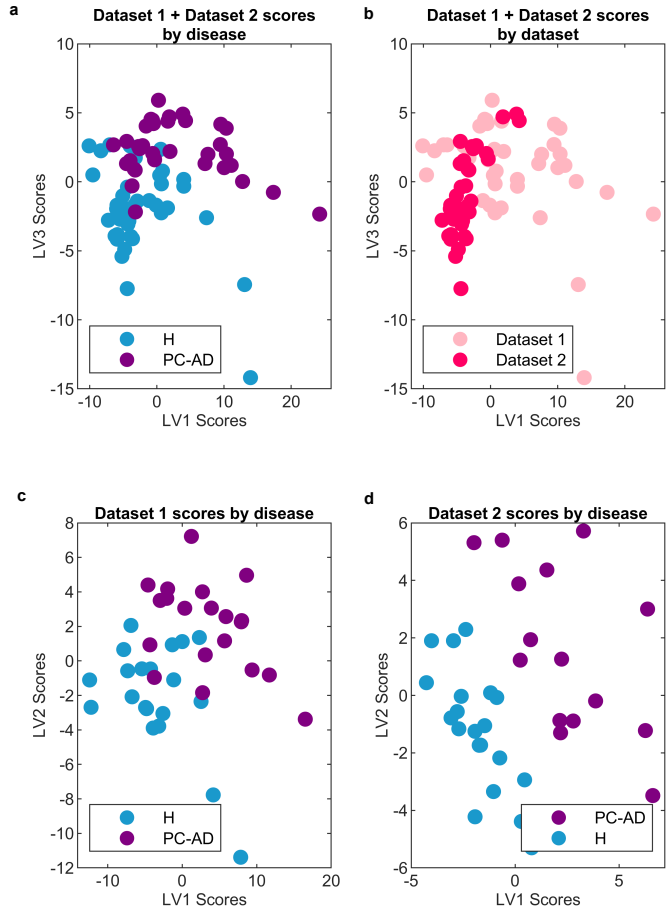


Figure 4. (**a**) Unified studies from Dataset 1 + Dataset 2 with disease−specific labels. (**b**) Unified studies from Dataset 1 + Dataset 2 with year−specific labels. (**c**) Dataset 1 study with disease−specific labels. (**d**) Dataset 2 study with disease−specific labels.

3. Materials and Methods

3.1. Dataset Creation

The samples were collected from a population-based clinical–biological cohort of adults with and without cognitive decline. They belonged to a cross-sectional study, and the recruitment of all individuals was performed by the CITA-Alzheimer Foundation. The syndromic diagnosis was established through comprehensive neurological and neuropsychological assessment, structural magnetic resonance neuroimaging (MRI), and a CSF AT(N) biomarkers study. CSF samples were obtained in 2014 and 2015 (Dataset 1) from the participants of the Gipuzkoa Alzheimer Project (GAP) [38] and from 2016 to 2018 (Dataset 2) in the DEBA study [39]. Research was conducted in accordance with the Declaration of Helsinki and approved by local Ethics Committees. A total of 75 volunteers were recruited to detect PC-AD using CSF samples. In both studies, volunteers were categorized into a

Int. J. Mol. Sci. **2024**, *25*, 4737

control group comprising healthy individuals (H) and PC-AD patients who exhibited abnormal biomarker values while maintaining normal cognitive function. The control group was defined by specific CSF analysis cutoff values (Aβ42 > 1030 pg/mL, total-Tau > 300 pg/mL, p-Tau > 27 pg/mL) to validate ATN negativity and cognitive normalcy. The inclusion of data from two separate studies increases robustness and provides an extended time range for the analytical framework, enabling a comprehensive assessment of preclinical AD detection performance. Age and gender distribution were comparable between the control and preclinical AD groups, as summarized in Table 3.

Table 3. Datasets utilized in the study: Dataset 1 corresponds to samples collected between 2014 and 2015, while Dataset 2 corresponds to a cohort with samples collected between 2016 and 2018.

	Status	Total		Male/Female	Age
Dataset 1	Healthy	20	40	10/10	59.5 ± 6.8
	Preclinical	20		10/10	65.4 ± 5.1
Dataset 2	Healthy	20	35	11/9	65.7 ± 6.1
	Preclinical	15		10/5	68.5 ± 6.2
Dataset 1 + Dataset 2	Healthy	40	75	21/19	62.7 ± 7.1
	Preclinical	35		20/15	66.7 ± 7.7

3.2. Raman Measurements and Sample Preparation

CSF samples were collected according to international consensus recommendations [21] and centrifuged immediately at 4 °C and stored at −80 °C within one hour of lumbar puncture. For Raman measurements, CSF samples underwent ultracentrifugation using an Amicon® Ultra-0.5 filter with a 3 kDa pore size, resulting in a protein-rich supernatant. Raman measurements were performed with an inVia Qontor confocal Raman microscope (Renishaw plc, Wotton-under-Edge, UK). One microliter of the CSF sample was deposited onto a microscope glass slide covered with aluminum foil to enhance the Raman signal. The droplet was dried under vacuum for 10 min before each Raman measurement. Optimal measurement conditions were defined for a balanced signal-to-noise ratio and sample preservation. Point-by-point mapping was employed, capturing 15 spectra at the ring of the dried droplet (Figure 5). Laser wavelength and output power were set to 785 nm and 73 mW, respectively; a 50×-long distance objective was selected, and 50 accumulations were performed with an exposure time of 1 s. Raman measurements from Dataset 1 samples were taken in 2022 and for the Dataset 2 samples in 2023.

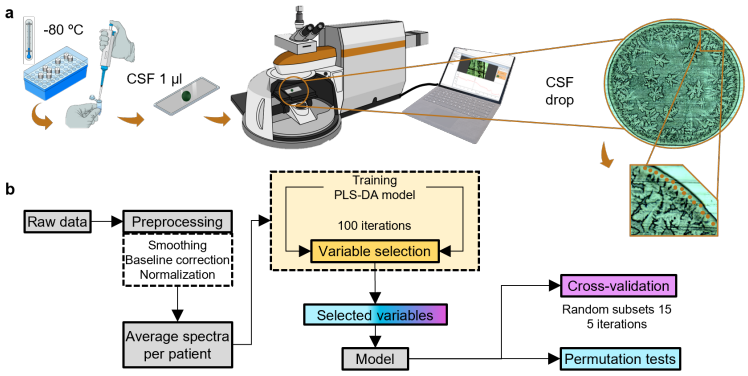


Figure 5. (**a**) Sample preparation and Raman measurement conditions. (**b**) Stages of the machine learning workflow.

Int. J. Mol. Sci. **2024**, *25*, 4737

3.3. Modeling Workflow

3.3.1. Data Preprocessing

All stages of the data analysis were carried out in MATLAB 2022a (The MathWorks, Inc., Natick, MA, USA) using in-house routines and the PLS-Toolbox (Eigenvector Research Inc., Wenatchee, WA, USA). The detailed workflow, covering feature extraction, model training, and validation, is explained in the following section. The order of these steps (Figure 5b) is essential and was carefully considered in the spectroscopic data analysis workflow. Figure 5b provides a visual summary of the intrinsic process steps of the data analysis approach. The Raman spectra underwent a meticulous pipeline for optimum preprocessing, which is important, non-trivial, and can significantly affect model performance [40]. The data collected from the two cohorts, Dataset 1 and Dataset 2, exhibit similar and distinct characteristics. To ensure meaningful comparisons and to facilitate the creation of a unified model incorporating both cohorts, a consistent preprocessing strategy was applied, starting with noise reduction and signal enhancement (using Savitzky-Golay filtering), followed by baseline correction using Whittaker's method to eliminate fluctuations and artifacts and standard normal variate (SNV) spectral scaling [41–43]. All 15 spectra per patient were first preprocessed and scaled and subsequently averaged per patient [44], as depicted in Figure 6, to obtain one representative spectrum per patient for subsequent variable selection and classification. Figure 6 displays the mean spectra for healthy (upper) and PC-AD (lower) individuals, accompanied by the standard error (shadowed area). The subtle differences between healthy and preclinical subjects are barely visible, and classification can only be achieved by proper selection of feature extraction and subsequent machine learning algorithms.

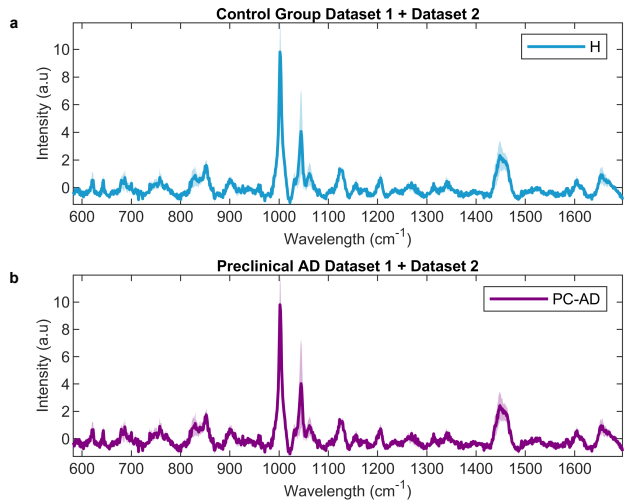


Figure 6. Preprocessed and averaged spectra per patient to unify both cohorts from Dataset 1 and Dataset 2; shaded area refers to the standard error. (**a**) Healthy or control group. (**b**) Preclinical Alzheimer's subjects.

3.3.2. Feature Extraction Methods for Identification of Discriminative Molecules in Preclinical AD's Discrimination

Feature extraction was performed using the integrated action of variable importance in projection (VIP) method and the selectivity ratio (SR). This approach uses the spectra projected into the latent space of a partial least squares discriminant analysis (PLS-DA)

Int. J. Mol. Sci. 2024, 25, 4737

model to select iterative variables due to their importance in the model [45]. The iterative process optimizes wavenumber subsets by comparing the root mean square error of cross-validation (RMSECV) values and removing those wavenumbers with the lowest influence on prediction. The approach integrates VIP and SR, progressively eliminating variables until the model no longer improves. This variable selection strategy not only elevates the importance of the feature set but also plays a crucial role in identifying relevant wavenumbers to distinguish between distinct patient cohorts. Selected wavenumbers serve as valuable indicators, which help identify molecular vibrations that contribute to the differentiation of PC-AD. To assess the reliability of the selected features, an iterative approach was implemented by running the variable selection strategy 100 times. In each iteration, 10 patients (5 from each class) from a dataset of 75 patients were randomly excluded, allowing for both interclass and intraclass variability. The procedure was then applied with the same cross-validation structure of 15 random subsets and 5 iterations, ensuring a robust assessment by excluding 10 different patients from model training and cross-validation in each iteration. Subsequently, the variable selection distribution across the 100 iterations was examined, and an additional model based on this analysis was constructed, as shown in the results section.

3.3.3. PLS-DA Model Development and Evaluation Metrics

Following the workflow from Figure 5 and using the reduced feature set from the variable selection procedure, reliable PLS-DA models were constructed. PLS-DA is a versatile multivariate classification method selected for its simplicity and effectiveness [46]. Differentiating from classical PLS regression, PLS-DA involves an additional step where a suitable threshold is applied to the computed y values. This thresholding process aids in determining the classification of a sample within a specific class. It is a powerful tool for modeling the relationship between variables and reducing data dimensionality while preserving the covariance structure, making it particularly relevant to our study. The optimization process focuses on tuning the number of latent variables (LVs). PLS-DA is well-suited for handling complex data relationships and correlated variables, making it an ideal choice for efficient machine learning [46–49]. Due to the low number of subjects in each of the individual cohorts and to ensure the robustness of the model, a cross-validation approach with random subsets was adopted, in which each data set was divided into 15 distinct subsets with 5 iterations. This resampling strategy of cross-validation included partitioning the dataset into training sets and diverse cross-validation sets to evaluate the model performance [44,50]. The performance of the PLS-DA models was systematically evaluated using standard classification metrics, including accuracy, sensitivity, specificity, and the area under the receiver operating characteristic curve (AUC-ROC). Permutation tests were also performed to determine the stability of the models. These tests involved randomizing the assignment of class labels and re-evaluating the models multiple times to assess the likelihood of obtaining similar performance by chance [44,51,52]. The permutation tests provided valuable statistical information on the reliability and stability of the developed PLS-DA models.

4. Conclusions

Our study explores the potential of Raman spectroscopy in combination with advanced chemometric methods as an innovative and far less complex technical approach for the early diagnosis of preclinical Alzheimer's disease compared to positron emission tomography, computed tomography, or CSF analysis. The priority of early detection results from the prolonged preclinical phase of Alzheimer's disease, in which medication and the development of new therapies could help slow down the pathogenesis. Traditional diagnostic modalities, such as neurological–neuropsychological assessments and biomarkers of cerebrospinal fluid, have limitations and require the exploration of alternative methods. Raman spectroscopy provides a molecular fingerprint of physiology without subjective interpretations and relatively simple technical effort. In our study, we investigated dried

Int. J. Mol. Sci. **2024**, *25*, 4737

droplets of cerebrospinal fluid and took Raman measurements at the ring of the dried structures. Our investigation, across multiple years of sampling, includes two separate studies and corroborates the potential of Raman spectroscopy to distinguish between healthy subjects and those in the preclinical stage of Alzheimer's with high accuracies reaching 0.96 in a cross-validated model. We demonstrated significant discriminative power despite variations in cohorts and measurement years. Fusing the data from both studies not only improves the robustness of the overall model but also allows for a more comprehensive assessment of the variables that play a role in the classification of preclinical Alzheimer's disease. Significant identified wavenumbers were consistent with key peaks for Alzheimer's disease reported in the literature, including amino acids found in established biomarkers such as tau proteins and Aβ42 peptides. Our study represents a significant advance in the application of Raman spectroscopy for the early detection of Alzheimer's, and we note that the influence of cohort-specific factors, including sampling and measurements at different time stamps and under different conditions, underlines the need for further research and larger datasets to capture as much uncertainty as possible, incorporate the full range of various inter and intraclass variabilities, and ultimately, provide a general robust and reliable prediction model for new, unseen data in clinical settings via external validation.

Author Contributions: Conceptualization, E.L., J.E.-E., J.M.A. and A.S.; methodology, E.L. and J.E.-E.; software, E.L. and J.M.A.; validation, J.E.-E., J.M.A. and A.S.; formal analysis, E.L.; investigation, E.L. and J.E.-E.; resources, M.G.-S., M.A., M.E.-T., A.E., M.T., C.L. and P.M.-L.; data curation, E.L.; writing—original draft preparation, E.L.; writing—review and editing, J.E.-E., J.M.A. and A.S.; visualization, E.L.; supervision, J.M.A. and A.S.; project administration, A.S.; funding acquisition, A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Basque Government (Ref. KK-2022/00001) and supported by grant CEX2020-001038-M funded by MICIU/AEI/10.13039/501100011033.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the following Ethics Committees: (1) the local Ethic and Clinical Research Committee of Gipuzkoa (protocol V3.0 and approved on 23 April 2014) and (2) the Basque Country Ethics Committee CEIm-E (protocol code PI2015153 V2_2015.11.20 and approved on 25 November 2015).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors acknowledge the generous contribution of the volunteers of the GAP and DEBA studies.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AD	Alzheimer's disease
PC	Preclinical
MCI	Mild cognitive impairment
PC-AD	Preclinical Alzheimer's disease
IWG	International working group
ATN	Amyloid, tau, and neurodegeneration
Aβ	Amyloid beta
CSF	Cerebrospinal fluid
PET	Positron emission tomography
ELISA	Enzyme-linked immunosorbent assays
MS	Mass spectrometry
MRI	Magnetic resonance imaging

GAP	Gipuzkoa Alzheimer project
H	Healthy individuals
SNV	Standard normal variate
VIP	Variables in projection
SR	Selectivity ratio
PLS-DA	Partial least squares discriminant analysis
RMSECV	Root mean square error of cross-validation
LV	Latent variable
AUC-ROC	Area under the receiver operating characteristic curve
PS	Phosphatidylserine

References

1. Pais, M.; Martinez, L.; Ribeiro, O.; Loureiro, J.; Fernandez, R.; Valiengo, L.; Canineu, P.; Stella, F.; Talib, L.; Radanovic, M.; et al. Early diagnosis and treatment of Alzheimer's disease: New definitions and challenges. *Braz J. Psychiatry* **2020**, *42*, 431–441. [CrossRef]
2. Park, J.; Jang, S.; Gwak, J.; Kim, B.C.; Lee, J.J.; Choi, K.Y.; Lee, K.H.; Jun, S.C.; Jang, G.J.; Ahn, S. Individualized diagnosis of preclinical Alzheimer's Disease using deep neural networks. *Expert Syst. Appl.* **2022**, *210*, 118511. [CrossRef]
3. Mayeux, R. Alzheimer's Disease Biomarkers—Timing Is Everything. *N. Engl. J. Med.* **2024**, *390*, 761–763. [CrossRef]
4. Hubbard, B.M.; Fenton, G.W.; Anderson, J.M. A quantitative histological study of early clinical and preclinical Alzheimer's disease. *Neuropathol. Appl. Neurobiol.* **1990**, *16*, 111–121. [CrossRef]
5. Dubois, B.; Hampel, H.; Feldman, H.H.; Scheltens, P.; Aisen, P.; Andrieu, S.; Bakardjian, H.; Benali, H.; Bertram, L.; Blennow, K.; et al. Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria. *Alzheimer's Dement.* **2016**, *12*, 292–323. [CrossRef]
6. Rafii, M.S.; Sperling, R.A.; Donohue, M.C.; Zhou, J.; Roberts, C.; Irizarry, M.C.; Dhadda, S.; Sethuraman, G.; Kramer, L.D.; Swanson, C.J.; et al. The AHEAD 3-45 Study: Design of a prevention trial for Alzheimer's disease. *Alzheimer's Dement.* **2022**, *19*, 1227–1233. [CrossRef]
7. Dubois, B.; Villain, N.; Frisoni, G.B.; Rabinovici, G.D.; Sabbagh, M.; Cappa, S.; Bejanin, A.; Bombois, S.; Epelbaum, S.; Teichmann, M.; et al. Clinical diagnosis of Alzheimer's disease: Recommendations of the International Working Group. *Lancet Neurol.* **2021**, *20*, 484–496. [CrossRef]
8. Jia, J.; Ning, Y.; Chen, M.; Wang, S.; Yang, H.; Li, F.; Ding, J.; Li, Y.; Zhao, B.; Lyu, J.; et al. Biomarker Changes during 20 Years Preceding Alzheimer's Disease. *N. Engl. J. Med.* **2024**, *390*, 712–722. [CrossRef]
9. Ashton, N.J.; Brum, W.S.; Di Molfetta, G.; Benedet, A.L.; Arslan, B.; Jonaitis, E.; Langhough, R.E.; Cody, K.; Wilson, R.; Carlsson, C.M.; et al. Diagnostic Accuracy of a Plasma Phosphorylated Tau 217 Immunoassay for Alzheimer Disease Pathology. *JAMA Neurol.* **2024**, *81*, 255–263. [CrossRef]
10. Aisen, P.S.; Cummings, J.; Jack, C.R.; Morris, J.C.; Sperling, R.; Frölich, L.; Jones, R.W.; Dowsett, S.A.; Matthews, B.R.; Raskin, J.; et al. On the path to 2025: Understanding the Alzheimer's disease continuum. *Alzheimer's Res. Ther.* **2017**, *9*, 60. [CrossRef]
11. Jack, C.R., Jr.; Bennett, D.A.; Blennow, K.; Carrillo, M.C.; Dunn, B.; Haeberlein, S.B.; Holtzman, D.M.; Jagust, W.; Jessen, F.; Karlawish, J.; et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's Dement.* **2018**, *14*, 535–562. [CrossRef]
12. Ralbovsky, N.M.; Halámková, L.; Wall, K.; Anderson-Hanley, C.; Lednev, I.K. Screening for Alzheimer's Disease Using Saliva: A New Approach Based on Machine Learning and Raman Hyperspectroscopy. *J. Alzheimer's Dis.* **2019**, *71*, 1351–1359. [CrossRef]
13. Long, J.M.; Coble, D.W.; Xiong, C.; Schindler, S.E.; Perrin, R.J.; Gordon, B.A.; Benzinger, T.L.S.; Grant, E.; Fagan, A.M.; Harari, O.; et al. Preclinical Alzheimer's disease biomarkers accurately predict cognitive and neuropathological outcomes. *Brain* **2022**, *145*, 4506–4518. [CrossRef]
14. Milà-Alomà, M.; Ashton, N.J.; Shekari, M.; Salvadó, G.; Ortiz-Romero, P.; Montoliu-Gaya, L.; Benedet, A.L.; Karikari, T.K.; Lantero-Rodriguez, J.; Vanmechelen, E.; et al. Plasma p-tau231 and p-tau217 as state markers of amyloid-β pathology in preclinical Alzheimer's disease. *Nat. Med.* **2022**, *28*, 1797–1801. [CrossRef]
15. Jack, C.R., Jr.; Bennett, D.A.; Blennow, K.; Carrillo, M.C.; Feldman, H.H.; Frisoni, G.B.; Hampel, H.; Jagust, W.J.; Johnson, K.A.; Knopman, D.S.; et al. A/T/N: An unbiased descriptive classification scheme for Alzheimer disease biomarkers. *Neurology* **2016**, *87*, 539–547. [CrossRef]
16. Koychev, I.; Jansen, K.; Dette, A.; Shi, L.; Holling, H. Blood-Based ATN Biomarkers of Alzheimer's Disease: A Meta-Analysis. *J. Alzheimer's Dis.* **2021**, *79*, 177–195. [CrossRef]
17. Alcolea, D.; Delaby, C.; Muñoz, L.; Torres, S.; Estellés, T.; Zhu, N.; Barroeta, I.; Carmona-Iragui, M.; Illán-Gala, I.; Santos-Santos, M.Á.; et al. Use of plasma biomarkers for AT(N) classification of neurodegenerative dementias. *J. Neurol. Neurosurg. Psychiatry* **2021**, *92*, 1206–1214. [CrossRef]
18. Porsteinsson, A.P.; Isaacson, R.S.; Knox, S.; Sabbagh, M.N.; Rubino, I. Diagnosis of Early Alzheimer's Disease: Clinical Practice in 2021. *J. Prev. Alzheimer's Dis.* **2021**, *8*, 371–386. [CrossRef]

19. Jack, C.R., Jr.; Knopman, D.S.; Jagust, W.J.; Petersen, R.C.; Weiner, M.W.; Aisen, P.S.; Shaw, L.M.; Vemuri, P.; Wiste, H.J.; Weigand, S.D.; et al. Tracking pathophysiological processes in Alzheimer's disease: An updated hypothetical model of dynamic biomarkers. *Lancet Neurol.* **2013**, *12*, 207–216. [CrossRef]
20. Blennow, K.; Mattsson, N.; Schöll, M.; Hansson, O.; Zetterberg, H. Amyloid biomarkers in Alzheimer's disease. *Trends Pharmacol. Sci.* **2015**, *36*, 297–309. [CrossRef]
21. Duits, F.H.; Martinez-Lage, P.; Paquet, C.; Engelborghs, S.; Lleó, A.; Hausner, L.; Molinuevo, J.L.; Stomrud, E.; Farotti, L.; Ramakers, I.H.; et al. Performance and complications of lumbar puncture in memory clinics: Results of the multicenter lumbar puncture feasibility study. *Alzheimer's Dement.* **2016**, *12*, 154–163. [CrossRef]
22. Gonzalez-Ortiz, F.; Kirsebom, B.E.; Contador, J.; Tanley, J.E.; Selnes, P.; Gísladóttir, B.; Pålhaugen, L.; Suhr Hemminghyth, M.; Jarholm, J.; Skogseth, R.; et al. Plasma brain-derived tau is an amyloid-associated neurodegeneration biomarker in Alzheimer's disease. *Nat. Commun.* **2024**, *15*, 2908. [CrossRef]
23. Lausted, C.; Lee, I.; Zhou, Y.; Qin, S.; Sung, J.; Price, N.D.; Hood, L.; Wang, K. Systems approach to neurodegenerative disease biomarker discovery. *Annu. Rev. Pharmacol. Toxicol.* **2013**, *54*, 457–481. [CrossRef]
24. Wang, H.; Dey, K.K.; Chen, P.C.; Li, Y.; Niu, M.; Cho, J.H.; Wang, X.; Bai, B.; Jiao, Y.; Chepyala, S.R.; et al. Integrated analysis of ultra-deep proteomes in cortex, cerebrospinal fluid and serum reveals a mitochondrial signature in Alzheimer's disease. *Mol. Neurodegener.* **2020**, *15*, 43. [CrossRef]
25. Kuhar, N.; Sil, S.; Verma, T.; Umapathy, S. Challenges in application of Raman spectroscopy to biology and materials. *RSC Adv.* **2018**, *8*, 25888–25908. [CrossRef]
26. Xu, Y.; Pan, X.; Li, H.; Cao, Q.; Xu, F.; Zhang, J. Accuracy of Raman spectroscopy in the diagnosis of Alzheimer's disease. *Front. Psychiatry* **2023**, *14*, 1112615. [CrossRef]
27. Paraskevaidi, M.; Morais, C.L.M.; Halliwell, D.E.; Mann, D.M.A.; Allsop, D.; Martin-Hirsch, P.L.; Martin, F.L. Raman Spectroscopy to Diagnose Alzheimer's Disease and Dementia with Lewy Bodies in Blood. *ACS Chem. Neurosci.* **2018**, *9*, 2786–2794. [CrossRef]
28. Carlomagno, C.; Cabinio, M.; Picciolini, S.; Gualerzi, A.; Baglio, F.; Bedoni, M. SERS-based biosensor for Alzheimer disease evaluation through the fast analysis of human serum. *J. Biophotonics* **2020**, *13*, e201960033. [CrossRef]
29. Ryzhikova, E.; Ralbovsky, N.M.; Sikirzhytski, V.; Kazakov, O.; Halamkova, L.; Quinn, J.; Zimmerman, E.A.; Lednev, I.K. Raman spectroscopy and machine learning for biomedical applications: Alzheimer's disease diagnosis based on the analysis of cerebrospinal fluid. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2021**, *248*, 119188. [CrossRef]
30. Cialla-May, D.; Krafft, C.; Rösch, P.; Deckert-Gaudig, T.; Frosch, T.; Jahn, I.J.; Pahlow, S.; Stiebing, C.; Meyer-Zedler, T.; Bocklitz, T.; et al. Raman Spectroscopy and Imaging in Bioanalytics. *Anal. Chem.* **2022**, *94*, 86–119. [CrossRef]
31. Khristoforova, Y.; Bratchenko, L.; Bratchenko, I. Raman-Based Techniques in Medical Applications for Diagnostic Tasks: A Review. *Int. J. Mol. Sci.* **2023**, *24*, 15605. [CrossRef]
32. Pezzotti, G. Raman spectroscopy in cell biology and microbiology. *J. Raman Spectrosc.* **2021**, *52*, 2348–2443. [CrossRef]
33. Chesney, R.W. Taurine: Its biological role and clinical implications. *Adv. Pediatr.* **1985**, *32*, 1–42. [CrossRef]
34. Jang, H.; Lee, S.; Choi, S.L.; Kim, H.Y.; Baek, S.; Kim, Y. Taurine Directly Binds to Oligomeric Amyloid-β and Recovers Cognitive Deficits in Alzheimer Model Mice. In *Taurine 10*; Lee, D.H., Schaffer, S.W., Park, E., Kim, H.W., Eds.; Springer: Dordrecht, The Netherlands, 2017; pp. 233–241.
35. Lanznaster, D.; Tasca, C.I. Targeting the guanine-based purinergic system in Alzheimer's disease. *Neural Regen. Res.* **2017**, *12*, 212–213.
36. Ma, X.; Li, X.; Wang, W.; Zhang, M.; Yang, B.; Miao, Z. Phosphatidylserine, inflammation, and central nervous system diseases. *Front. Aging Neurosci.* **2022**, *14*, 975176. [CrossRef]
37. Xu, Z.J.; Li, Q.; Ding, L.; Shi, H.H.; Xue, C.H.; Mao, X.Z.; Wang, Y.M.; Zhang, T.T. A comparative study of the effects of phosphatidylserine rich in DHA and EPA on Aβ-induced Alzheimer's disease using cell models. *Food Funct.* **2021**, *12*, 4411–4423. [CrossRef]
38. Ecay-Torres, M.; Estanga, A.; Tainta, M.; Izagirre, A.; Garcia-Sebastian, M.; Villanua, J.; Clerigue, M.; Iriondo, A.; Urreta, I.; Arrospide, A.; et al. Increased CAIDE dementia risk, cognition, CSF biomarkers, and vascular burden in healthy adults. *Neurology* **2018**, *91*, e217–e226. [CrossRef]
39. Tainta, M.; Iriondo, A.; Ecay-Torres, M.; Estanga, A.; de Arriba, M.; Barandiaran, M.; Clerigue, M.; Garcia-Sebastian, M.; Villanua, J.; Izagirre, A.; et al. Brief cognitive tests as a decision-making tool in primary care. A population and validation study. *Neurologia* **2022**, *in press*. [CrossRef]
40. Guo, S.; Popp, J.; Bocklitz, T. Chemometric analysis in Raman spectroscopy from experimental design to machine learning-based modeling. *Nat. Protoc.* **2021**, *16*, 5426–5459. [CrossRef]
41. Savitzky, A.; Golay, M.J.E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* **1964**, *36*, 1627–1639. [CrossRef]
42. Whittaker, E.T. On a New Method of Graduation. *Proc. Edinb. Math. Soc.* **1922**, *41*, 63–75. [CrossRef]
43. Randolph, T.W. Scale-based normalization of spectral data. *Cancer Biomark* **2006**, *2*, 135–144. [CrossRef]
44. Lopez, E.; Etxebarria-Elezgarai, J.; Amigo, J.M.; Seifert, A. The importance of choosing a proper validation strategy in predictive models. A tutorial with real examples. *Anal. Chim. Acta* **2023**, *1275*, 341532. [CrossRef]
45. Inc, E.R. Selectvars. Available online: https://wiki.eigenvector.com/index.php?title=Selectvars (accessed on 8 March 2018).

46. Ballabio, D.; Consonni, V. Classification tools in chemistry. Part 1: Linear models. PLS-DA. *Anal. Methods* **2013**, *5*, 3790–3798. [CrossRef]
47. Barker, M.; Rayens, W. Partial least squares for discrimination. *J. Chemom.* **2003**, *17*, 166–173. [CrossRef]
48. Gottfries, J.; Blennow, K.; Wallin, A.; Gottfries, C.G. Diagnosis of dementias using partial least squares discriminant analysis. *Dementia* **1995**, *6*, 83–88. [CrossRef]
49. Christin, C.; Hoefsloot, H.C.; Smilde, A.K.; Hoekman, B.; Suits, F.; Bischoff, R.; Horvatovich, P. A Critical Assessment of Feature Selection Methods for Biomarker Discovery in Clinical Proteomics. *Mol. Cell. Proteom.* **2013**, *12*, 263–276. [CrossRef]
50. Refaeilzadeh, P.; Tang, L.; Liu, H. Cross-Validation. In *Encyclopedia of Database Systems*; Liu, L., Özsu, M.T., Eds.; Springer: Boston, MA, USA, 2009; pp. 532–538.
51. Rudolph, P.E. *Permutation Tests. A Practical Guide to Resampling Methods for Testing Hypotheses*; Springer Series in Statistics; Springer: Berlin/Heidelberg, Germany; New York, NY, USA, 1994; 228p, ISBN 3-540-94097-9.
52. van der Voet, H. Comparing the predictive accuracy of models using a simple randomization test. *Chemom. Intell. Lab. Syst.* **1994**, *25*, 313–323. [CrossRef]

A.2 The importance of model validation

Eneko Lopez, Jaione Etxebarria-Elezgarai, Jose Manuel Amigo, Andreas Seifert. **The importance of choosing a proper validation strategy in predictive models: A tutorial with real examples**. Analytica Chimica Acta, 1275, 341532, (**2023**).

The manuscript was cited as ref [20]. It is accesible under

https://www.sciencedirect.com/science/article/pii/S0003267023007535?via%3Dihub

Analytica Chimica Acta 1275 (2023) 341532

Contents lists available at ScienceDirect

Analytica Chimica Acta

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

The importance of choosing a proper validation strategy in predictive models. A tutorial with real examples

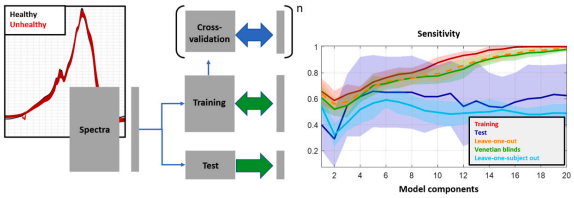
Eneko Lopez [a,b], Jaione Etxebarria-Elezgarai [a], Jose Manuel Amigo [c,d,*], Andreas Seifert [a,c,**]

[a] CIC NanoGUNE BRTA, Tolosa Hiribidea 76, San Sebastián, 20018, Spain
[b] Department of Physics, University of the Basque Country (UPV/EHU), San Sebastián, 20018, Spain
[c] IKERBASQUE, Basque Foundation for Science, Plaza Euskadi, 5, Bilbao, 48009, Spain
[d] Department of Analytical Chemistry, University of the Basque Country, Barrio Sarriena S/N, Leioa, 48940, Spain

HIGHLIGHTS

- We highlight the importance of cross-validation and external test set in prediction.
- Model performance is not having best figures of merit in training but in testing.
- Cross-validation in small datasets can deliver misleading models.
- Calibration and validation must consider the inner and hierarchical data structure.
- If independency in samples is not guaranteed, perform several validation procedures.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Validation
Cross-validation
PLS-DA
Resampling
Permutation test
Jackknife
Bootstrap

ABSTRACT

Machine learning is the art of combining a set of measurement data and predictive variables to forecast future events. Every day, new model approaches (with high levels of sophistication) can be found in the literature. However, less importance is given to the crucial stage of validation. Validation is the assessment that the model reliably links the measurements and the predictive variables. Nevertheless, there are many ways in which a model can be validated and cross-validated reliably, but still, it may be a model that wrongly reflects the real nature of the data and cannot be used to predict external samples. This manuscript shows in a didactical manner how important the data structure is when a model is constructed and how easy it is to obtain models that look promising with wrong-designed cross-validation and external validation strategies. A comprehensive overview of the main validation strategies is shown, exemplified by three different scenarios, all of them focused on classification.

* Corresponding author. Department of Analytical Chemistry, University of the Basque Country, Barrio Sarriena S/N, Leioa, 48940, Spain.
** Corresponding author. CIC NanoGUNE BRTA, Tolosa Hiribidea 76, San Sebastián, 20018, Spain.
E-mail addresses: josemanuel.amigo@ehu.eus (J.M. Amigo), a.seifert@nanogune.eu (A. Seifert).

https://doi.org/10.1016/j.aca.2023.341532
Received 28 February 2023; Received in revised form 13 June 2023; Accepted 14 June 2023
Available online 17 June 2023

1. The importance of assuring the reliability of a model

1.1. The model

Artificial Intelligence in natural sciences, commonly known as Chemometrics, was first introduced in 1972 by Svante Wold and Bruce Kowalsky, being defined as the discipline that uses mathematical, statistical, and other derived methods employing formal logic to (a) design or select optimal measurement procedures and experiments and (b) provide maximum relevant chemical information by analyzing chemical data [1]. Machine Learning (ML) is the part of Chemometrics devoted to establishing a mathematical connection between a set of measured data and predictive variables [2]. Despite the outstanding achievements of ML, its correct implementation sometimes remains unclear. There is a dangerous trend in the literature towards building more sophisticated algorithms, mostly validated in specific, often limited, analytical situations with no clear definition of the data structure [2] rather than giving importance to the three elements that make an algorithm successful, that are, as we will show during this manuscript, the data, the reference values, and the utmost important validation steps (cross-validation and external validation).

Given a set of M samples characterized by N independent variables $\mathbf{X}$ ($M \times N$) and a property $\boldsymbol{y}$ ($M \times 1$), the simplest model is the one that establishes a linear correlation between $\mathbf{X}$ and $\boldsymbol{y}$,

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{b} + \boldsymbol{e}, \tag{1}$$

where $\boldsymbol{b}$ is the regression vector calculated differently depending on the algorithm, and $\boldsymbol{e}$ is the vector ($M \times 1$) containing the residuals. That relationship can be linear or even non-linear and with varying complexity. This manuscript is putting its emphasis on classification models. Specifically, we focus on the well-known Partial Least Squares-Discriminant Analysis (PLS-DA) due to its high applicability in different classification scenarios. PLS-DA is a multivariate classification method [3,4] that has become popular in the field of chemometrics for well over two decades [5,6]. One of the main advantages of PLS-DA over other classification approaches is the apparent simplicity in optimizing the model, where only one parameter needs to be tuned: the proper number of latent variables (LVs). Obviously, that selection depends on the complexity of the data and the different preprocessing steps (including variable selection, if necessary) that must be done. PLS-DA allows us to handle highly correlated variables by searching for the orthogonal directions (LVs) of maximum covariance with the **Y**-block while explaining the relevant sources of data variability [3]. PLS-DA is based on classical PLS regression followed by the selection of a proper threshold imposed on the calculated $\boldsymbol{y}$ to assess the belonging of a sample to a certain class [3,7]. The main difference with its regression version is that $\boldsymbol{y}$ is composed of two values, one denoting samples that do not belong to the class (e.g. 0) and another one for samples belonging to the class (e.g. 1).

1.2. Reliability of a model

It is well-known among the data analysis community that to test the reliability of a model, it must be trained and, consequently, tested. The concept of model training is the process in which a model learns to infer a function from a collection of training data. Therefore, given a set of M spectra measured at N variables $\mathbf{X}$ ($M \times N$), the aim is to find the best vector $\boldsymbol{b}$ ($N \times 1$) that fits $\boldsymbol{y}$ ($M \times 1$) with the least possible error. That is, the **training or calibration model** generates an inferred vector $\boldsymbol{b}$ with $\mathbf{X}$ and $\boldsymbol{y}$, as in Equation (1) (Fig. 1). This $\boldsymbol{b}$ can be fitted until achieving a perfect fit, where the fitting error is zero. However, we need to count on the fact that there will never exist such perfection since the data used in $\mathbf{X}$ and $\boldsymbol{y}$ will contain part of the signal that is completely uncorrelated, has inherent noise and small artifacts (i.e., uncertainty in the reference values $\boldsymbol{y}$) and will move away our model from the ideal situation. Moreover, the classification models are sensitive to chance-correlation, when the number of samples is much smaller than the number of variables and, consequently, bears the risk of overfitting [8].

Regardless of the main model strategy (regression or classification) and the data parameters optimization (preprocessing, scaling and variables/features selections), the main purpose of any model is to be able to predict with reliability in completely new samples that have not been part of the construction of the model. In other words, the best model is not the one offering the best statistics in the training stage but the most reliable one for new data. A model is defined to be reliable when it fulfils the following statistical assessments [8].

1) Stability: The model does not change significantly after removing several samples.

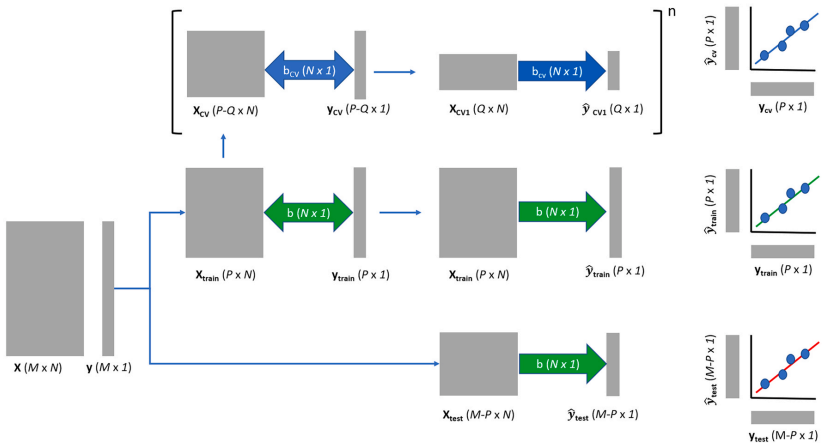


Fig. 1. Schematic flowchart of training, testing (external validation) and cross-validating a dataset for developing a predictive model.

2) Significance: The response of the model is not due to random chance or overfitting.
3) Predictability: The model is able to predict correctly (i.e., with high figures of merit).
4) Generalizability: The model can be used to predict new samples.

The fulfilment of these 4 assessments will guarantee that the model is reliable enough to be considered a valid model and, therefore, the user can be confident with it. Sadly, in order to be 100% sure of the statistical significance of the calculations, we should count with 100% of plausible measurements for each scenario. And acknowledging that this is impossible, Sir Ronald Fisher already recognized the tremendous help of empirically-generated sampling distributions as a valid approximation [9]. This is what is normally known as resampling strategies.

1.3. Resampling in brief

Resampling is the generation of multiple random subsets in a model to check its reliability by calculating confidence intervals of the four statistical assessments previously mentioned. Different resampling strategies have been proposed in the literature with slightly different purposes. In this manuscript, we will emphasize the importance of splitting the data between training and test set (generalizability), different cross-validation strategies (predictability), jackknifing and bootstrapping (stability) and permutation test (significance) [8].

All resampling strategies assume that every data subsample should reflect the population structure from which it came [9]. Here we normally must face another crucial issue: the samples are normally taken under different circumstances (e.g., seasonal changes, different devices, time changes, etc.), and, moreover, the samples are usually measured in replicates (or triplicates). Therefore, the independence of a sample is compromised, if it turns out that there is a certain implicit structure in it. This is particularly relevant when there are few samples available, or when the samples are strongly structured, compromising any statistical results obtained.

1.4. Targets of this tutorial

Considering validation as a holistic term, this manuscript shows how important a reliable validation procedure is in three different datasets. The first dataset suffers from a lack of samples and has strongly structured data. The second one contains enough samples to construct reliable models. These two cases deal with medical data, where obtaining robust and sufficient spectral information is complex due to the nature of the samples. The third case is a classic example of classification in food sciences.

The manuscript will show different validation strategies to assess the reliability of the model in a didactical manner. All strategies will be thoughtfully explained, and the key references will be given, such that the reader can always revisit them for more advanced insights. Special interest will be given to validation strategies where with the "wrongly/apparently" adequate cross-validation strategy, "wrongly/apparently" extremely good results can be obtained.

The tutorial is divided into ten sections. The first one settles the basis and the scope of this tutorial. The second section is devoted to explaining the main validation strategies. The third section presents the datasets and defines the different validation strategies for analyzing them. The fourth, fifth and sixth sections will present the results obtained for the three datasets. Section 7 will discuss the results obtained with permutation tests. One peculiarity of PLS-DA is that regression vectors and particular statistics (like the Hotelling T^2 and residuals) can be used as extra statistical parameters to assess the reliability of the models. This will be treated in section eight. In the end, several major conclusions, together with guidelines, will be given.

2. Validation strategies. Test set and resampling

2.1. Test set

As said before, the main aim when creating a model is to assure that the model contains the precise information to be able to predict completely new samples (generalizability). Therefore, splitting the dataset into a training (calibration) and a test (validation) set is, arguably, the most important step when validating a model. Now, our real model $\boldsymbol{b}$ will be constructed with $\mathbf{X}_{\mathbf{train}}$ ($P \times N$) and $\boldsymbol{y}_{\mathbf{train}}$ ($P \times 1$), where P indicates a sub-set of M, and tested with $\mathbf{X}_{\mathbf{test}}$, of dimensions ($M$-$P \times N$), which have not been used in the training stage before, and, moreover, $\boldsymbol{y}_{\mathbf{test}}$ (M-$P \times 1$) is perfectly known. $\mathbf{X}_{\mathbf{test}}$ and $\boldsymbol{y}_{\mathbf{test}}$ are part of what is commonly known as external validation or test set (Fig. 1). This way, the final training model can be adjusted and tuned with the test data to only contain the relevant part. With the model calculated, a new set of $\hat{y}_{\mathbf{train}}$ can be calculated and compared with the original $\boldsymbol{y}_{\mathbf{train}}$. In the same way, the model can be used to calculate a new set of $\hat{y}_{\mathbf{test}}$ to compare with the original $\boldsymbol{y}_{\mathbf{test}}$.

Splitting the data set $\mathbf{X}$ and $\boldsymbol{y}$ into training ($\mathbf{X}_{\mathbf{train}}$, $\boldsymbol{y}_{\mathbf{train}}$) and test ($\mathbf{X}_{\mathbf{test}}$, $\boldsymbol{y}_{\mathbf{test}}$) depends on several features of the dataset: The purpose (regression or classification), the structure of the data, the total number of samples in $\mathbf{X}$ and, specifically in classification challenges, the number of samples that each class contains. Sample balance refers to an equal representation of each class, while an imbalanced scenario occurs when certain class(es) are overrepresented. While the latter scenario may provide insights into real-world challenges, it can introduce potential complications such as bias towards overrepresented classes, compromising the generalization capabilities of the model. In any case, the reader must always have in mind that the test split must represent the same variability as the training dataset to obtain reliable estimates of the model's true predictive ability [3]. Many strategies can be followed in the literature based on clusters or similarities [10], of which three of the most popular ones are explained subsequently.

Kennard-Stone: Kennard-Stone algorithm is, arguably, one of the most popular methods for splitting the data into $\mathbf{X}_{\mathbf{train}}$ and $\mathbf{X}_{\mathbf{test}}$ [11]. It is based on a sequential selection of samples from $\mathbf{X}$ into $\mathbf{X}_{\mathbf{test}}$ covering uniformly the variance of $\mathbf{X}$. The method starts by selecting the two samples that have the largest distance between them (using any distance metric like Euclidean distance). This procedure ensures the selection of samples that lie on the boundary of the variance space of $\mathbf{X}$. Then, the next sample to add to $\mathbf{X}_{\mathbf{test}}$ is the one with the greatest distance from the already selected samples. This step is repeated sequentially until a specific number of samples is selected. The main advantage of the Kennard-Stone method is the simplicity of its implementation since it only depends on distances and, therefore, the variance of $\mathbf{X}$. Nevertheless, an important point to consider is that the splitting is made without considering the variance (or range) of $\boldsymbol{y}$. This is especially relevant for regression (quantitative) models and the reader is encouraged to ascertain that both training and test sets are spanning the same space in $\mathbf{X}$ and $\boldsymbol{y}$.

Random test-set validation: Randomly splitting a dataset into a calibration and external validation set is frequently used but unsuitable for imbalanced data. This system does not provide insight into how the observations are spread across the dataset. For example, if binary classification is performed (two classes), the training model may be built with 100% of individuals of one class, leaving the individuals of the other class for the external validation test set out.

Validation across dataset structure: This is a much more appropriate validation method. The calibration model is created by considering the data belonging to one structured sample as a unified set. The best example is to consider that all replicates of the same sample should be a unified set. However, in cases where the dataset is small, the model is influenced by the subjects with which the model was created, depending on the intraclass variability of the data. Some validation sets might not be properly predicted in these cases because the model was

unreliable.

Regarding the last two methods, a confusing point is that they can be repeated several times by changing the set of samples placed in training and testing, thus, inducing the possibility of calculating statistics like the standard deviation of the models (in the optimization parameters and the predictions) and, therefore, approaching some jackknifing sort of validation (see further sections).

Different approaches can be found in the literature as to which methodology must be used. In general, when the number of samples is relatively low, methods based on similarity or distances (like Kennard-Stone) are advisable to be used. Random methods may tend to produce imbalanced training and testing sets. On the other hand, when the number of samples is relatively high, any of the methodologies previously mentioned are expected to provide representative training and testing datasets.

Following the previous risk, another important consideration should be made for classification-related challenges, which is the fact of splitting the data where the classes have significantly different number of samples (imbalanced data). This must be done with special care, always keeping the proportion of the classes in both training and testing datasets [3].

Another important issue is when to split the dataset, before or after preprocessing. This decision must be made depending on two points: the type of splitting methodology and the type of preprocessing. Considering the methodologies, random splitting and structure-wise splitting do not depend on the quality of the data but on the position of the samples in the matrix. Therefore, applying row-wise preprocessing methods (e.g., SNV, smoothing, or derivatives) before or after the splitting operation is irrelevant. Nevertheless, Kennard-Stone (and other methodologies) aims to be representative of the variance in the data. Accordingly, the use of row-wise preprocessing methods may have an impact on the algorithm's selection of samples. Since row-wise preprocessing inherently removes sampling artifacts from the data, it may therefore be advisable to perform preprocessing before splitting the data.

Generally, special care should be taken with certain preprocessing methods that use information extracted from the whole dataset, such as multiplicative scatter correction (MSC), or methods that use column-operations, such as mean centering or autoscaling. In that case, the prediction model will be influenced by the testing data set, and regardless of the splitting methodology, the following four steps are advisable.

1) Preprocess the whole $\mathbf{X}$ data with row-wise methods that are independent of sample to sample.
2) Split the data with the chosen method.
3) If needed, preprocess $\mathbf{X}_{\mathbf{train}}$ with row-wise preprocessing methods that depend on the whole $\mathbf{X}_{\mathbf{train}}$ set (e.g., MSC). At the same time, preprocess $\mathbf{X}_{\mathbf{train}}$ and $\mathbf{Y}_{\mathbf{train}}$ with column-wise operations (e.g., mean centering and autoscaling). Consider that the preprocessing parameters, such as mean and standard deviation in the case of autoscaling, will be part of the parameters of your model.
4) Once the training model has been developed, apply the same preprocessing parameters of step 3 on the $\mathbf{X}_{\mathbf{test}}$ and $\mathbf{Y}_{\mathbf{test}}$ datasets and then apply the model $\boldsymbol{b}$.

2.2. Resampling strategies

There is an intermediate step between training and testing that is commonly not seen in modern literature of ML outside Chemometrics but can answer some of the concerns arising when the dataset is split. This important step is resampling. Conceptually speaking, resampling implies that the $\mathbf{X}_{\mathbf{train}}$ and $\boldsymbol{y}_{\mathbf{train}}$ samples will be divided into subsets of $\mathbf{X}_{\mathbf{cv}}$ (P-$Q \times N$) and $\boldsymbol{y}_{\mathbf{cv}}$ (P-$Q \times 1$) following different approaches in n repeated loops depending on how the data is structured. Then, new submodels (i.e., new regression vectors $\boldsymbol{b}_{\mathbf{cv}}$) will serve for predicting new values of $\boldsymbol{y}_{\mathbf{cv1}}$ ($Q \times 1$) with the samples $\mathbf{X}_{\mathbf{cv1}}$ ($Q \times N$) left out in each iteration (Fig. 1). As said previously, this operation is repeated n times to assure that all the samples have been resampled at least once, thus obtaining a value of $\hat{y}_{\mathbf{cv}}$ ($P \times 1$) to be compared with the original vector $\boldsymbol{y}_{\mathbf{train}}$, as shown in Fig. 1.

With the development of fast computing, many different resampling methods have appeared, of which the most important ones are jackknifing/cross-validation, bootstrapping and randomization/permutation tests [9]. These methods were developed at different times with subtle different goals. Here we will provide the main practical features, encouraging the readers to visit the supplied references for further details.

At this point, the reader must strongly consider one feature of the resampling methods. The main feature is that when a model is constructed, $\boldsymbol{b}$ remains invariant in the training and testing steps, regardless of the repetitions of the training–test splitting. With resampling, this does not happen since $\hat{y}_{\mathbf{cv}}$ is calculated with a similar $\boldsymbol{b}$ model but with less information ($\boldsymbol{b}_{\mathbf{cv}}$) since it is calculated with a subset of the training samples (Fig. 1). This fact is a warranty that the model is tested at multiple levels of stratification, gaining a greater understanding of the reliability of the model, and exposing $\boldsymbol{b}$ to slight variations that could not be considered otherwise. Another important feature is that resampling will test the values of $\boldsymbol{y}_{\boldsymbol{train}}$, thus having an extra indication of the quality (in terms of sensitivity to small changes) of the model.

2.2.1. Cross-validation

The most subtle difference between resampling methods is, probably, the one between jackknifing and cross-validation. Both split the calibration data into some subsets following the same strategies. Nevertheless, while cross-validation [12] focuses on evaluating the model performance of $\boldsymbol{b}_{\mathbf{cv}}$ (thus calculating some statistics of the left-out samples and assessing the predictability of the model), jackknifing [13,14] focuses on evaluating the stability of the model $\boldsymbol{b}_{\mathbf{cv}}$ itself with the samples used to construct $\boldsymbol{b}_{\mathbf{cv}}$ and calculating statistical parameters, as e.g., standard deviation or hypothesis tests, to assess the stability of the model.

Regarding cross-validation, several strategies can be followed, which differ in how the sample sets are selected. Selecting the correct strategy is crucial, and one must consider how the samples in $\mathbf{X}_{\mathbf{train}}$ are arranged together with the hierarchical structure of the data [10]. Normally, the samples are arranged into groups based on background information about their origin. Typical arrangements are: Considering the same operator and instrumental setups (repeatability), across different instrumental settings, operators, reagents, etc. (reproducibility), sampling site and time, across treatment/origin (year, raw material, batch, etc.), among others, which also fall mostly under the category reproducibility. Some of the most popular cross-validation strategies are the four subsequently listed.

Leave-one-out cross-validation: a single observation is removed from the dataset, and the analysis is run on the rest of the data. This process is repeated for each observation in the dataset, allowing the model to be tested on different validation datasets. Recent studies in the field state that leave-one-out cross-validation (LOO-CV) can be used to avoid overfitting/underfitting when utilizing ML models [15,16]. We will see further in the results section that the previous statement is not universally applicable and illustrates the significance of considering the characteristics of the data when determining an appropriate validation method. Despite its effectiveness, this technique can be prone to overfitting if not used correctly, e.g., when the dataset structure contains subject replicates, and hence, not advisable in that case.

Venetian blinds cross-validation: Each test set is determined by selecting every ith object in the dataset, starting at objects numbered 1 through i.

Cross-validation across data structure: This cross-validation is ideal when the data structure is known, and the observations are not randomly distributed. Examples include subject replicates or any other structure-influencing parameters on the data, for example seasonality,

location, operator, etc.

Random subsets: In this cross-validation, $\mathbf{X_{train}}$ is split multiple times randomly with a certain number of samples randomly chosen. The splitting can be done by including the data structure in the random selection.

Deciding on the type of cross-validation is sometimes difficult, and moreover, it is extremely dependent on the data structure. Let's take as an example LOO-CV and Venetian blinds. Depending on how the subsets of the training set are formed, one can consider individual replicates or samples (all replicates of a single sample) as the units that must be divided among the subsets. LOO-CV can have the option to leave out all datasets that belong to the same subject/sample (LOSO-CV). Likewise, Venetian blinds could do exactly the same when the blind is defined by the replicates of a sample, being extremely important that the blind dataset is well structured and organized.

One advisable procedure is the implementation of multiple rounds of cross-validation [8,17], as in the train-test split, to test their performances and decide which strategy is the one offering a better understanding of the structure of the $\mathbf{X_{train}}$ and its linkage with $\boldsymbol{y}_{\mathbf{train}}$, and not which strategy is the one offering the best figures of merit. This has the advantage of calculating statistics like the standard deviation, which will be extremely helpful when assessing the predictability of the model.

2.2.2. Jackknifing, bootstrapping and randomization/permutation tests

Together with cross-validation, Jackknifing, bootstrapping, and randomization/permutation tests are the three most important resampling methods found in the literature. These three methods were developed at different times with slightly different goals [9] and ways of applying them, creating certain confusion and misunderstanding among the scientific community.

These methods are also based on the creation of subsets of $\mathbf{X_{train}}$. Nevertheless, important differences can be found with cross-validation. The first and most important one is that these methods focus on the statistical assessment of the model, instead of assessing the validity of the predictions of the samples left out in each iteration. Another difference is the way of resampling $\mathbf{X_{train}}$ [18–20].

Jackknife: Jackknife [9,13,21,22] classically followed the leave-one-out method previously seen in cross-validation to remove one sample from $\mathbf{X_{train}}$ and assess the stability of the model parameters with statistical tests. For example, when the PLS-DA model is created, another indication of its performance is the stability of the regression matrix $\mathbf{B}$ previously mentioned, where regression coefficients with high absolute values would be promising variables for interpreting the classification model. Nevertheless, high values of, for instance, the variance of a regression coefficient along the different repetitions would indicate the instability of that regression coefficient, transforming jackknifing into a valid strategy for variable selection [22] as well as model stability assessment. Nowadays, Jackknife can be applied following more reliable resampling schemes than leave-one-out, representing a test that can be perfectly performed simultaneously with cross-validation.

Bootstrap methods [8,9,21,23,24] are a family of methods were the resampling of $\mathbf{X_{train}}$ is done in such a way that the $\mathbf{X_{cv}}$ generated has the same dimensions of $\mathbf{X_{train}}$ but it is done with replacement. The concept of replacement is equivalent to sampling from an infinite-sized population; therefore, samples are allowed to be "repeated" in each iteration to form $\mathbf{X_{cv}}$.

Lastly, randomization/permutation tests [25] are a family of resampling methods aimed at assessing the significance of the outcome of a model by calculating the probability that the outcome comes by chance, i.e. from a random assignation of the model. Permutation tests involve random reordering of the values in the $\boldsymbol{y}_{\mathbf{train}}$, recalculating the model and examining whether that random "shuffling" of only the $\boldsymbol{y}_{\mathbf{train}}$ (class assignments in our case) still provides an acceptable outcome. Moreover, any cross-validation strategy can be applied to the new permuted dataset. This operation is repeated many times, and depending on the size of the samples, a good starting point could be 100 times. The results obtained in calibration and cross-validation are then compared with those obtained with the model with no permutations using different hypothesis tests, like the Wilcoxon Test, among others [26,27].

2.3. The omnipresent challenge of test sets and resampling

The definition and understanding of the origin, meaning, and structure of the data are crucial for generating a reliable model in the training stage. This is essential for any supervised, semi-supervised, or reinforcement learning, regardless of the final aim of the model (regression or classification) [28,29].

As said before, the performance of the model in predicting new data depends largely on the quality of the data used in the training set. Therefore, if the dataset is small-sized, resampling is an extremely useful tool, and it might be preferred over splitting the dataset into calibration and external validation sets. If the data is of high quality, and the resampling process is done properly, the predictions should be almost optimal. However, if the data used in the training set is of low quality (high uncertainties and artifacts) or the randomization procedure is not conducted properly, the predictions can be quite poor. This means that the appropriateness of the samples used to build the model will be confirmed by the quality of the figures of merit, but only when the model is tested to avoid generating too optimistic models that might jeopardize the predictability of the model.

Some suggestions can be found in the bibliography [10], where external validation should be used for datasets greater than 50 samples, while resampling is the better option for smaller to medium datasets. This advice must be taken as a recommendation because the strategy finally depends on the intra-class and inter-class relations of data, or with other words on all types of data variability and uncertainty. The final definition of how large a dataset should be will depend on the quality of $\mathbf{X}$ and $\boldsymbol{y}$ and their correlation.

As defined earlier, small datasets do not converge to a realistic result that easily. The two main reasons for this are the difficulty in obtaining a large representative cohort of samples and the fact that normally, in classification problems, obtaining balanced classes is not the common situation. As an example for medical data, high-risk disease patients are normally the minority group, hampering the possibility of increasing the cohort of healthy subjects [10]. When the class imbalance is too large, it becomes difficult to build a classification model using conventional ML algorithms [30]. In these cases, alternatives like bootstrapping, permutation tests and jackknifing are the strategies to follow, apart from cross-validation.

Certainly, splitting the data into training and testing can affect the reliability of the model. Moreover, the weight of wrongly classified samples over the total number of samples will drastically decrease the goodness of the figures of merit studied. As pointed out beforehand, relying on resampling strategies sounds like the sole alternative to constructing a valid model in these cases. Nevertheless, there is a dangerous trend in solely applying cross-validation and looking for the cross-validation strategy that gives the best figures of merit, without considering the data structure, and risking the trueness of the model.

2.4. Figures of merit in classification, Hotelling T^2 and Q residuals

Among all the figures of merit that are proposed in the literature for assessing the performance of the model [3], sensitivity, specificity, precision and accuracy are normally the most common ones because they rely in a straightforward way on the four basic measures that can be easily counted and form the confusion matrix, which are true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), as pointed out in Table 1. For binary models (i.e., models containing two classes), these parameters are calculated only once, where the positive class is the class of interest (e.g., disease), and the negative class is the control class (e.g., healthy controls) [31].

Table 1
Definition of some figures of merit.

	Formula	Definition
Sensitivity	$\frac{TP}{TP+FN}$	Ratio of true positives to total (actual) positives in the data
Specificity	$\frac{TN}{TN+FP}$	Ratio of true negatives to total (actual) negatives in the data
Precision	$\frac{TP}{TP+FP}$	Ratio of true positives to total predicted positives
Accuracy	$\frac{TP+TN}{TP+FP+TN+FN}$	Ratio of correctly classified samples to all samples

One property of a projection algorithm like PLS (and its classification variant PLS-DA used in this manuscript) [32,33] is the possibility of ascertaining how well the samples are projected into the space defined by the model (i.e., the space defined by the new latent variables calculated). This is normally checked by the calculation of the distance of the sample to the center of the model (Hotelling T^2, closely related to the leverage) and the residual of that sample in the model (Q) [1,32]. These two parameters are usually displayed in a combined plot (T^2 vs. Q) that is very useful (and very dangerous if not used properly) to check whether a sample belongs to the model (low Q and T^2 concerning the rest of the samples) or might be considered as an outlier (high Q and T^2 with respect to the rest of the samples). For assessing that "weight" of the sample, boundaries are normally established at a certain level of confidence (normally 95%), and then the samples are evaluated. This manuscript will not discuss the suitability of T^2 and Q for outlier detection. We will focus on using them as a very helpful tool to detect samples that behave in a very unstable manner in the model and might bias the final results of it.

As mentioned before, T^2 and Q can also account for a holistic confidence level for the researcher that a sample belongs to the model and can also be predicted with that confidence level. When a sample is used for calibration, it can have values of T^2 and Q that are under the confidence level for the model. Nevertheless, suppose that the sample is validated or cross-validated (with any of the above-mentioned resampling strategies) and its T^2 and/or Q values increase beyond the confidence level, then it might be an indication that the sample is biasing the model (drastic changes in T^2), or does not really belong to the model (drastic changes in Q), or both things at the same time (drastic changes in T^2 and Q).

3. Datasets under study, resampling strategies and software

3.1. Small medical dataset. Alzheimer's disease assessed with FTIR spectroscopy

This dataset is a small study of 20 individuals diagnosed with preclinical Alzheimer's Disease (PC-AD) and 19 individuals diagnosed as healthy (HI). Cerebrospinal fluid (CSF) was measured in triplicates using Fourier-transform infrared spectroscopy (FTIR) in attenuated total internal reflectance (ATR) within the range of 2000–4000 cm^1. Each sample was measured three times to be statistically significant and correspond to a final data matrix X with dimensions (117 × 1036) (Fig. 2a).

3.2. Large medical dataset. COVID-19 detection with Raman spectroscopy

A second medical dataset, but larger than the first one, is presented. This dataset corresponds to a study for developing a method to diagnose coronavirus disease 2019 (COVID-19) using Raman spectroscopy. For our analysis, a total of 103 samples belonging to 53 confirmed subjects and 50 healthy controls were used. Each sample contained three replicates, giving a final data matrix of dimensions **X** (309 × 900) (Fig. 2b). The full dataset and further details can be found in the supplied references [34].

3.3. Small dataset in food science. Wine authentication with NIR

The third dataset presented in this manuscript comes from the food science field. The dataset results from infrared spectroscopy measurements for distinguishing two different types of wine. The data was acquired from 37 bottles of authenticated single-variety wine with 19 bottles of Cabernet Sauvignon and 18 of Syrah grape variety. Three subsamples were measured from each bottle (Fig. 2c), yielding a dataset of dimensions **X** (111 × 235). The dataset and further information can be found in Ref. [29].

The data preprocessing stage is one of the main cornerstones in developing predictive models. In this manuscript, we only present the results obtained with the best preprocessing method obtained for the three cases under study. For this, all three datasets were subjected to standard normal variate (SNV) transformation to remove offset and correct for global intensity variations (Fig. 2). Moreover, the spectral matrix **X** was mean-centered and the predictor ***y*** was autoscaled. This

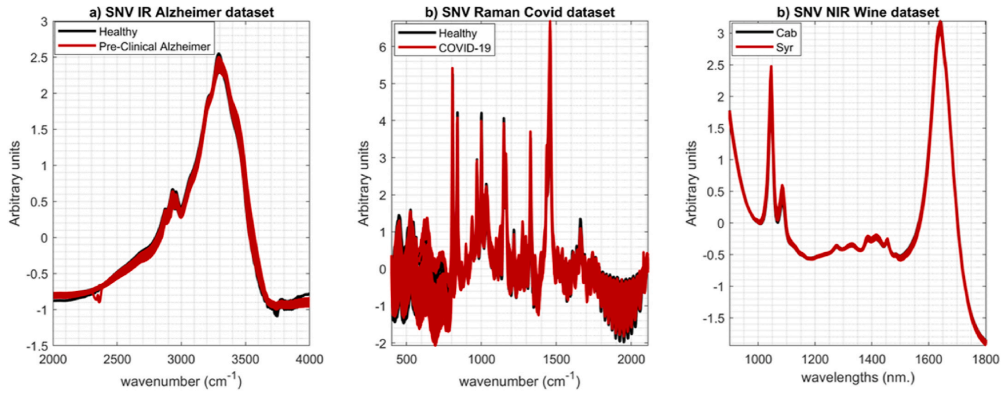


Fig. 2. a) FTIR spectra for Alzheimer's disease. b) Raman spectra for COVID-19. c) NIR spectra for wine classification.

step involves translating the origin of the coordinate system to the center of gravity of the data to avoid uncertainties of the mean variance.

3.4. Strategies of resampling and external validation

Fig. 3 shows the flowchart of the analysis used for the three datasets.

Splitting the dataset between training and test: The datasets were split so that 80% of the data was used for training and 20% for test (validation). Two strategies were followed (Fig. 3).

1. Bearing in mind that the samples are measured three times, the whole sample was left in one of the sets. A sample must be considered a subject in the first 2 cases and a bottle in the third case.
2. Completely random split of the data. This procedure is repeated 10 times in each strategy and the standard deviation of the results calculated to have sufficient statistics.

Resampling methods: The training dataset was submitted to different resampling strategies. The following ones were applied in the cross-validation stages (Fig. 4).

- Venetian blinds cross-validation by selecting every 10th spectrum.
- Leave-one-spectrum-out, accounting for the classical leave-one-out cross validation.
- Leave-one-subject-out cross-validation, meaning that one sample (three replicates) was left out in each cross-validation iteration.

Since several iterations were done splitting the training and test set, the cross-validation strategies were applied in each one of the iterations. This allows us to calculate the standard deviation of the figures of merit previously mentioned. Moreover, two more tests were performed: Permutation tests with 100 iterations with the implementation of the Wilcoxon Test [35], and the Jackknife test by calculating the standard deviation of the regression vectors with each iteration in each cross-validation procedure.

3.5. Software

All analyses were done in MATLAB 2022a (The MathWorks, Inc.), using in-house routines and the PLS Toolbox (Eigenvector Research Inc., WA, USA).

4. Small medical dataset

4.1. When cross-validation went extremely well but the model is extremely wrong

A total of 48 spectra from healthy patients and 45 spectra from Alzheimer's-diagnosed patients is used in the training model, which belongs to 80% of the dataset. Fig. 5 represents the evolution of the performance of the figures of merit for calibration/training (red), validation/test (blue), leave-one-out cross-validation (green), 10-fold Venetian blinds (dashed orange), and leave-one-subject-out cross-validation (cyan), as a function of the number of LVs.

Looking at the results obtained for training, the first perception is that the classification model could be considered optimal, but complex in terms of the selection of LVs. For the sake of simplicity, only the results obtained for the class Healthy are shown here since the results obtained for the other class are directly correlated in a binary class problem (the specificity of one class is the sensitivity of the other and vice versa). For acceptable results, which we assume as 0.9, accounting for 90%, four to five LVs are required for specificity and accuracy. However, in the case of sensitivity and precision, a larger number of LVs is needed, increasing the complexity of the model and, consequently, the regression vector.

The results obtained for training/calibration are as expected because the more LVs are included in the model, the better the performance, since more spectral information is included in the model, or generally speaking, the entire uncertainty (dispersion of data) available is considered. Nevertheless, the important fact is not to include as much spectral information in the model as possible, but to include the spectral information (through the LVs) that is, indeed, correlated with the property that is to be classified.

When looking at the results of leave-one-out cross-validation (green), it can be seen how similar all the figures of merit behave compared to the ones in the calibration set. This result might look apparently optimal since it demonstrates that when a sample is left out of the calibration set, the model does not change substantially; accordingly, a good sample prediction will be obtained. However, this applies only when all samples (spectra) are completely independent. The same situation is obtained when the model is cross-validated using a 10-fold Venetian blinds cross-validation strategy (dashed orange line). In this case, trying to include more samples in the cross-validation procedure, one can be tempted just to apply this type of cross-validation. Again, the figures of merit are equally good as with leave-one-out cross-validation and the calibration sets.

In this specific dataset, each patient was measured three times. It is quite common to take several measurements from one subject using

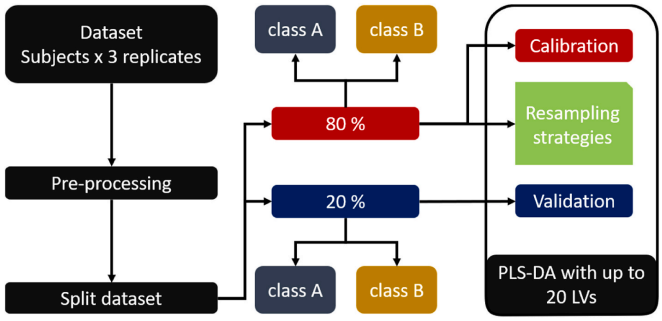


Fig. 3. Flowchart of the data analysis. The central part represents the splitting procedure between the training set (80% of balanced data) and the test set (20%). LV: latent variable.

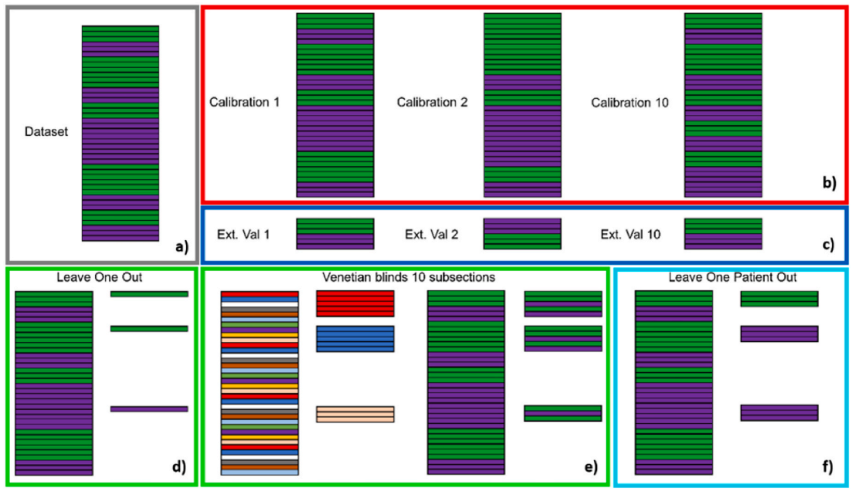


Fig. 4. Validation strategies. a) Initial dataset. b) Calibration dataset scheme for the different iteration loops. c) External validation test set for the different iteration loops. d) Leave-one-out cross-validation structure. e) Venetian Blinds with 10 blinds cross-validation structure. f) Leave-one-subject-out cross-validation structure. Depending on the class, each row of X, painted in either purple or green, represents one measurement. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

A2

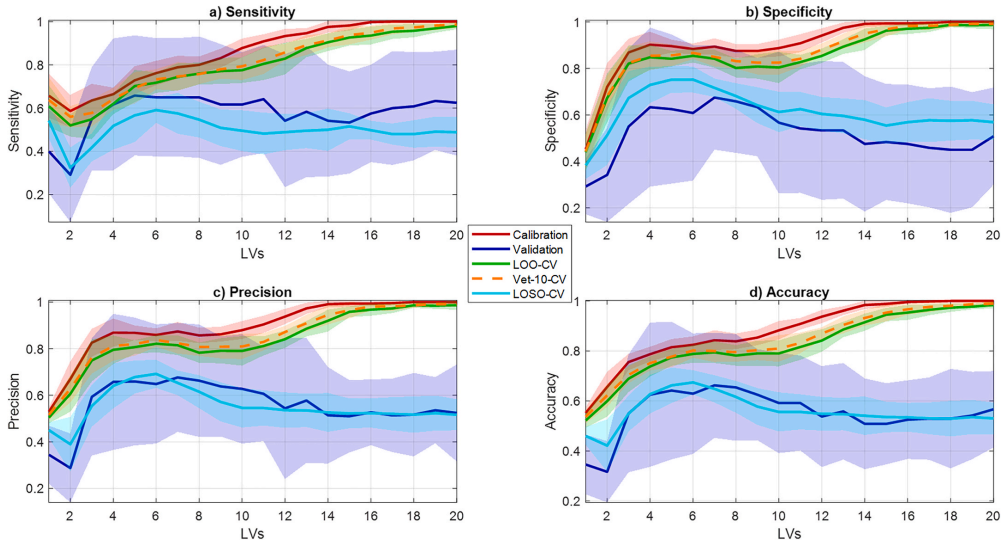


Fig. 5. Evolution of figures of merit (standard deviation shaded) with the number of latent variables for calibration, different cross-validation and external test validation procedures: a) sensitivity, b) specificity, c) precision, d) accuracy. Curves based on FTIR data for Alzheimer's classification.

vibrational spectroscopy to capture the intra-variability of replicates in the data. Consequently, those three samples should not be considered independent. If those three samples from the same subject are not left out together during cross-validation, the model will still contain information from that very patient. Therefore, the results will definitely be overoptimistic and will not reflect the true variance (inter-variability) between patients.

Considering the hierarchical structure of the data, leave-one-subject-

out cross-validation was performed. The corresponding results are far from ideality (cyan curves in Fig. 5). The best performance is achieved for models with 6 LVs, with a performance of 75%, which in classification-related issues in medical data is not an acceptable result. This enormous difference in the results compared to previous models highlights an enormous instability in the classes. In the case of external validation (blue line in Fig. 5), the situation is even worsening: whilst the mean values of the figures of merit of external validation and leave-one-subject-out cross-validation are very similar, the standard deviation of the 10 runs covers a range between 0.3 and 0.8.

An extra analysis was performed to confirm the previous results, focusing only on the external validation test set. In this specific case, and particularly to prove the importance of the dataset knowledge, the dataset is randomly split into 80% training and 20% external validation set from the very beginning, assuming that all samples (spectra) are independent, resulting in a loss of the knowledge of the inner structure of the dataset (Fig. 6).

Considering the mean value of the figures of merit obtained for both calibration (red) and external validation (blue), it can be observed that the model evolves positively towards an improvement in the quality of the model performance (Fig. 6). This might indicate that, effectively, the classification model could be considered as a good model with 4–5 LVs, since the mean values of the figures of merit for calibration and validation are equally good and stable. Nevertheless, one important aspect is the behavior of the standard deviation of the figures of merit of the validation. They present a high variability range, indicating certain instability in the model. However, this range is much smaller than the range presented when the external validation was made considering the data structure, i.e., each subject is measured in triplicate, as seen in Fig. 5. In the complete random generation of the validation set, it might occur that some of the replicates for the same subject still remain in the training set, giving over-optimistic results.

4.2. An alternative way: averaging of spectra belonging to one subject

To statistically improve the data and, particularly, to eliminate intra-patient variability, an appropriate approach can consist in averaging spectra per subject. Averaging is a common practice and can, to a certain extent, provide better, i.e., more reliable, results. A possible drawback of this practice is that the models are now developed with fewer "samples" than before. In our case, a total of 16 mean spectra of healthy patients and 15 mean spectra from Alzheimer's-diagnosed patients will be used in the training model, which belongs to the 80% of the averaged dataset. Data splitting is again performed randomly, thereby excluding 20% of a balanced number of patients. In the end, 4 subjects from healthy control and 4 Alzheimer's-diagnosed patients are left out for the external validation test set.

Due to the low number of samples, the procedure of randomly splitting the dataset between the training and the external test set is repeated 10 times to enhance the statistics sufficiently. The obtained results can be seen in Fig. 7.

As expected, the figures of merit obtained for the training stage are of high quality. However, again, one has to prove the quality by cross-validation and external validation. The results obtained for leave-one-out cross-validation and 10-fold Venetian blinds cross-validation are of special interest compared to the previous results. Using averaged spectra for each subject, the curves follow the same trend with the number of LVs as in the leave-one-subject-out cross-validation and external validation. The only little differences that can be observed are a result of the random selection in each of the 10 iterations. All the figures of merit of the cross-validation approaches follow the same trend as the figures of merit of the external validation, highlighting, once again, that the cross-validation is inherently linked to the inter and intra-patient variability of the data.

Again, one has to emphasize the extremely high variance observed in the evolution of the figures of merit obtained with the external test, showing strong instability due to the relatively small dataset and high

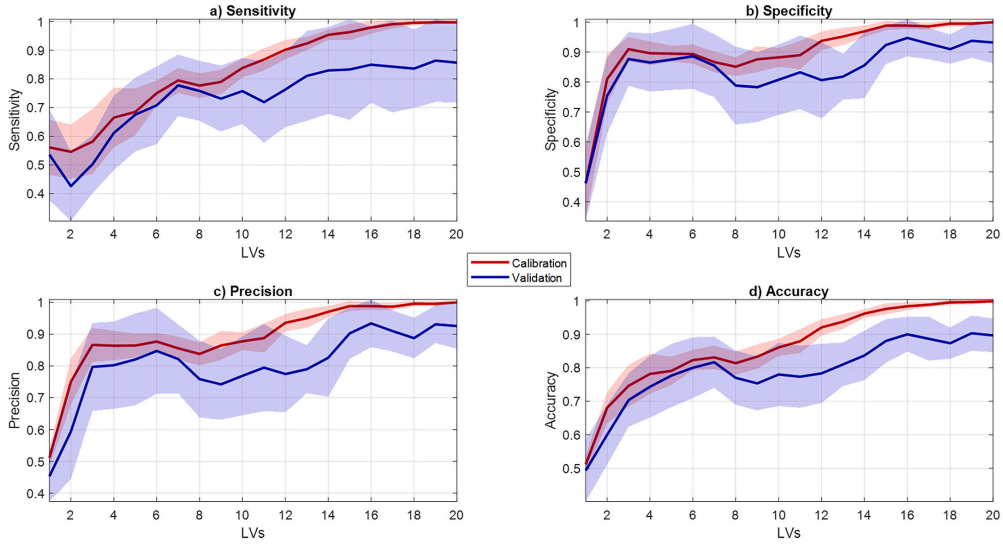


Fig. 6. Evolution of figures of merit (standard deviation shaded) with the number of latent variables for calibration (red) and external validation (blue) without considering the structure of the data: a) sensitivity, b) specificity, c) precision, d) accuracy. Curves based on FTIR data for Alzheimer's classification. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

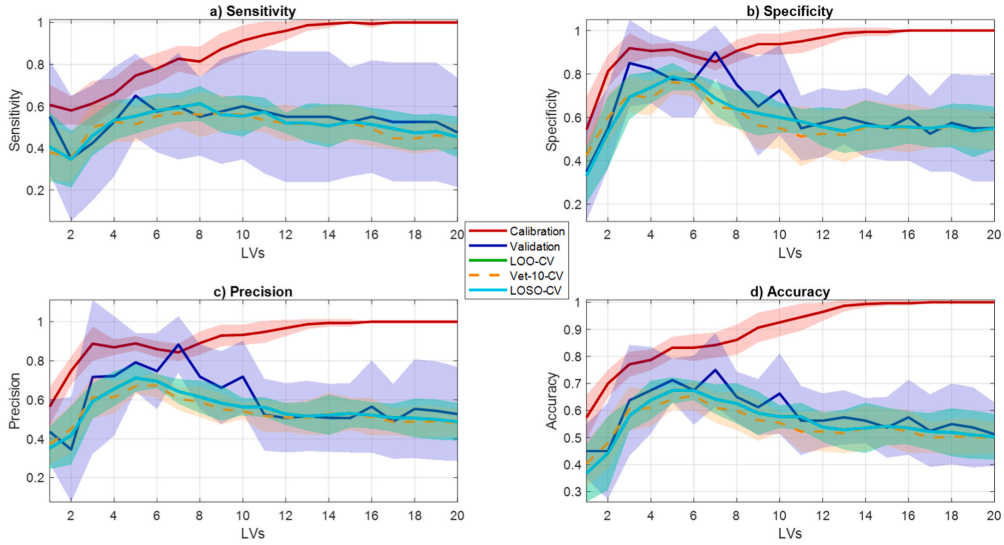


Fig. 7. Evolution of figures of merit (standard deviation shaded) with the number of latent variables for calibration, validation and different cross-validation strategies in the dataset with the average spectrum for each subject: a) sensitivity, b) specificity, c) precision, d) accuracy. Curves based on FTIR data for Alzheimer's classification.

A2

spectral variability. Consequently, the classification models are prone to overfit the data in all circumstances.

5. Large medical dataset. COVID-19 detection with Raman spectroscopy

The class distribution in this case is 53 ill subjects vs. 50 healthy. Following the same strategy as in the previous section, Fig. 8 represents the evolution of the performance of the accuracy for calibration/training (red), validation/test (blue), leave-one-out cross-validation (LOO-CV, green), 3-fold Venetian blinds (VET 3-CV, dashed orange), and leave-one-subject-out cross-validation (Lcyan) as a function of the number of LVs at the same three different scenarios. For the sake of simplicity, we have only included the data for accuracy. We have modified the number of blinds to show again the importance of how the data is structured. One could think that using multiple blinds of the number of replicates (3) is sufficient to certify the acceptability of the cross-validation. However, this is not dependent on the number of replicates but on how the sample replicates are distributed over the dataset.

Whole dataset with replicates: Looking at the results from Fig. 8a, we can observe that with 11 LVs, the training model could achieve 100% accuracy, whereas cross-validated models will never achieve performances above 95%, which for classification-related issues in medical data is already an acceptable result. However, this scenario again shows

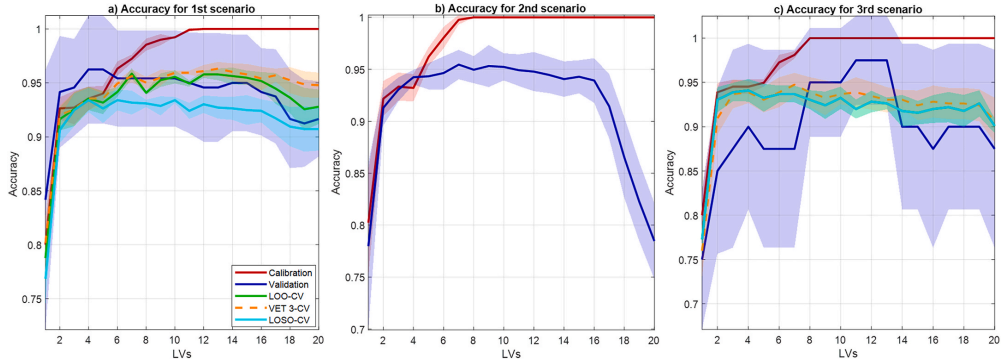


Fig. 8. Evolution of the accuracy (standard deviation shaded) with the number of latent variables for calibration, validation and different cross-validation strategies in: a) the whole dataset, b) the dataset without considering the structure of the data, c) the dataset with the average spectrum for each subject. Curves based on Raman data for COVID-19 classification.

how similarly LOO-CV and VET 3-CV behave and how the hierarchical structure of the data is not considered compared to the cyan plotted curve, where individual subject distribution over the matrix is considered, and all the spectra belonging to the same subject are left out for cross-validation (LOSO-CV). In this latter case, 4 LVs would be enough to define the best model, which mirrors the external validation set maximum performance also with 4 LVs.

Random splitting of the dataset between training and external validation set: With a larger dataset like this one, we could be encouraged to split the dataset from the very beginning into training (80% of the data) and external validation (20% of the data) sets [3] as depicted in Fig. 8b, assuming independency of all spectra and losing the knowledge of the inner structure of the dataset. After 10 iterations of this random splitting, we can observe that models with 8 LVs give the maximum performance for training and validation sets. It also shows how models with more than 15 LVs include noise and get worse performance.

Averaging the spectra belonging to one subject: Fig. 8c shows the best performance in all the cross-validations and validation cases for 4 LVs. This result goes along with the one obtained for LOSO-CV from the first scenario. In this case, each sample is independent, which is why all cross-validation methods behave similarly. This proves also that any cross-validation approaches would be valid in cases where the inner structure per subject is not contained (i.e., when averaged samples form it). However, the standard deviation of the external validation set is really high (from 5 iterations). This is the drawback of using less data, then models are less robust in the predictive performance.

6. Small dataset in food science. Wine authentication with IR

The last example includes 37 samples (without replicates), of which 19 bottles belong to Cabernet Sauvignon and 18 to Syrah. Fig. 9 presents the evolution of the performance of the accuracy for calibration/training (red), validation/test (blue), leave-one-out cross-validation (green), 3-fold Venetian blinds (dashed orange), and leave-one-subject-out cross-validation (cyan) as a function of the number of LVs at the same three different scenarios.

Whole dataset with replicates: The same effect observed in the two previous datasets is clearly represented in Fig. 9a, where with 9 LVs the training model could achieve 100% of accuracy, and the same happens to cross-validated models with LOO-CV and VET 3-CV. A completely different behavior is observed when the hierarchical structure of the data is considered (LOSO-CV). In this latter case, 9 LVs give model accuracy of around 80%, mirroring the external validation set performance also with 9 LVs. This also shows how overoptimistic results could be achieved by the other wrong model cross-validation approaches.

Random splitting of the dataset between training and external validation set: In a small dataset like this one, we would be tempted to randomly split the dataset from the very beginning into training (80% of the data) and external validation (20% of the data) sets as depicted in Fig. 9b. It can be observed how training and validation sets behave very similarly and achieve perfect performances with 9 LVs. This is a serious mistake and yields again overoptimistic results.

Averaging the spectra belonging to one subject: The third scenario confirms the previous statements and shows an alternative method for model validation when the dataset is small and contains sample replicates. Fig. 9c shows models developed with the averaging of the spectra per subject where the best performance in all the cross-validations and validation cases is given for 9 LVs. This result goes along with the one obtained for LOSO-CV from the first scenario. However, we again can observe the high standard deviation of the external validation set (from 5 iterations). Again, we can conclude that the drawback of using less data is that the created models are less reliable.

7. Permutation tests

Profiting from the different resampling strategies used in the three datasets studied above, permutation tests were also applied to the $\mathbf{X}_{\mathbf{train}}$ for each iteration to study the probability whether the result obtained is overfitted or a product of chance/correlation (significance of the model). To test the probability, the Wilcoxon test was performed.

The results obtained for the Alzheimer's disease dataset (Fig. 10a) are extremely clear, demonstrating, once again, how important the data structure is when a predictive model is constructed. After 5 LVs, LOO-CV and VET 3-CV show a p-value below 0.05, indicating that those results significantly differ from random predictions. Nevertheless, once again, the LOSO-CV strategy shows a high probability for those results to be nonsignificant.

A very different situation is obtained for the other two datasets, where it can be observed that regardless of the validation strategy, there is a high confidence that the results obtained with the proper number of LVs are significant. Note that in the second case (COVID-19) there might appear a discrepancy for 16 LVs. However, looking at the absolute values of the p-value, one can notice that we are talking about very small numeric values (2×10^{-13}) that are very close to the numeric precision of the computing software).

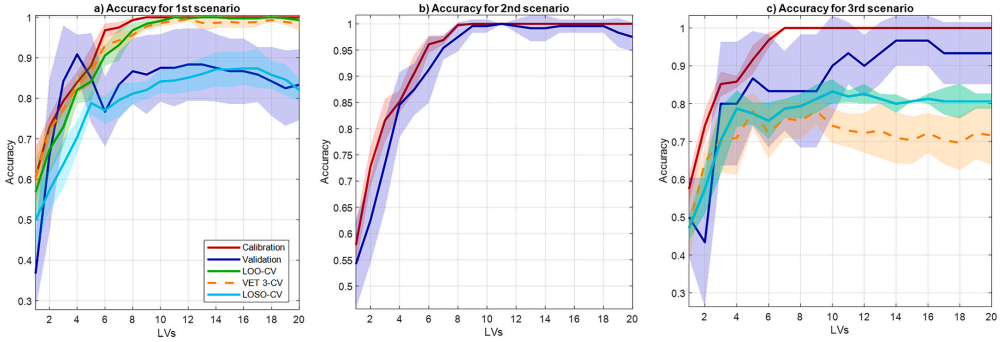


Fig. 9. Evolution of the accuracy (standard deviation shaded) with the number of latent variables for calibration, validation and different cross-validation strategies in: a) the whole dataset, b) the dataset without considering the structure of the data, c) the dataset with the average spectrum for each subject. Curves based on IR data for wine classification.

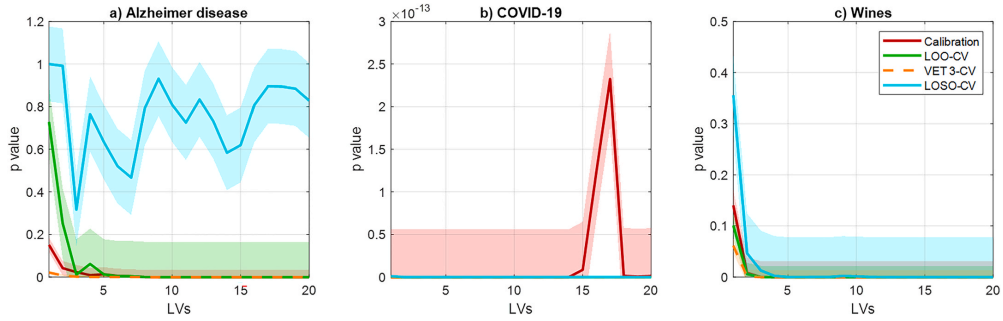


Fig. 10. Evolution of the p-value (standard deviation shaded) with the number of LVs for the three datasets. Calibration in red. The resampling strategies tested are LOO-CV (green), VET 3-CV (orange) and LOSO-CV (cyan). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

8. Hotelling T^2 and residuals

Due to the complexity and the strong influence of the data structure, here we will only present the results for the Alzheimer's disease dataset. Two cases have been studied: Models with 5 LVs with LOO-CV and LOSO-CV (Fig. 11 a and b); and 14 LVs with LOO-CV and LOSO-CV (Fig. 11 c and d). All cases show the projection of the samples in calibration (red) and cross-validation (black). For the sake of simplicity, the same sample has been highlighted in green in all the subplots, with a circle in red for its projection in calibration and black for cross-validation.

In general, we can say that a sample belongs to the model if it is within the boundaries for the Hotelling T^2 and the residual (dashed green lines in the figure). In the first case, when the calibration model has been calculated with 5 LVs (Fig. 11 a and b), we can see how the highlighted sample has a quite high T^2 and a low residual for calibration. This could indicate that this sample has special importance in the model, subject to its quite low residual. The projection of that sample in LOO-CV

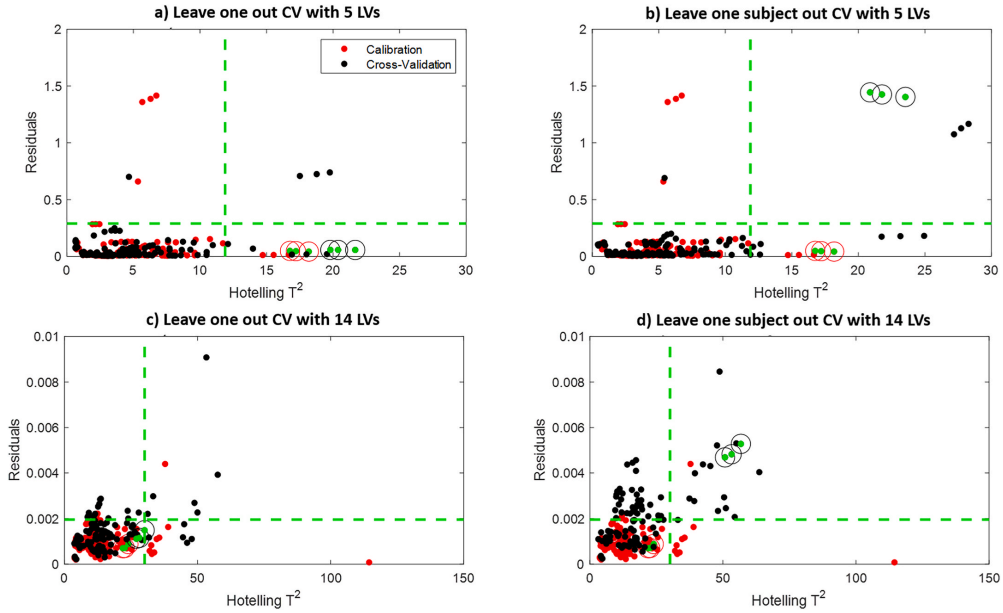


Fig. 11. Hotelling T^2 versus Q (residuals) for different cases in the Alzheimer's disease dataset. a) 5 LVs and LOO-CV. b) Five LVs and LOSO-CV, c) Fourteen LVs and LOO-CV, d) Fourteen LVs and LOSO-CV. The same sample has been highlighted in green in the four subplots. Also, its projection in calibration (red) and cross-validation (black) has been highlighted. Green dashed lines correspond to the T^2 and Q calculated at 95% confidence level. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

for 5 LVs appears to be very close to the projection in calibration (Fig. 11a). Nevertheless, when that sample is cross-validated with LOSO-CV, we can see how the residual and the T^2 values drastically increased (Fig. 11b), indicating that the sample might not belong to that model.

When the model is built with 14 LVs and LOO-CV is applied, both projections for the sample appear to be within the expected confidence levels (Fig. 11c). In this sense, including too many components may give the impression that the sample looks normal (Fig. 11c), while it is clearly outside the confidence limits if the model uses the correct size of 5 LVs instead of 14. Once again, when the dataset structure is considered, and LOSO-CV is performed, the cross-validation projection of that sample appears far from the confidence levels (Fig. 11d).

Even not being an absolute indicator of the reliability of the model, T^2 and Q can offer extremely interesting information about the behavior of particular samples when they are used in different stages of the validation of the model.

9. Some notes on the use of classification models and regression vectors. Jackknife estimation

One can observe a dangerous trend, particularly in medical diagnostics, in using PLS-DA (or other classification approaches) as a data reduction or visualization tool before applying further advanced classification algorithms. In this regard, the regression vector obtained in the classification stage is an important component. The regression vector is a key parameter calculated in every classification procedure. It gives weight to specific variables that are important for the classification and reduces the importance of counterproductive variables. However, we must not forget that the regression vector is calculated in the calibration stage. This means that when any resampling strategy offers poor results, the regression vector just gives zero relevant information. Returning to the first dataset we have calculated, the standard deviation of the regression vector follows a jackknifing strategy. Fig. 12 represents the regression vectors considering 5 and 14 LVs calculated for the 10 random runs in the case where three replicates formed each subject. The number of LVs was chosen following the observations made before. 5 LVs were chosen because the model arrives at an apparent stability, while 14 LVs were chosen to compare the regression vector of a clearly overfitted model with the previous one.

The first outstanding remark is that even though a good classification of this dataset into the two proposed classes is not feasible, the regression vectors calculated in the 10 runs are very stable (low standard deviation). This finding might indicate that our model is indeed quite acceptable. Nevertheless, we must always bear in mind that the regression vectors are just merely fitted parameters on the training data. The fact of leaving some samples out does probably not have a high impact on the regression vector but on the residuals of the model.

Another important remark is that the figures of merit of the calibration set do not say anything about good classification, not even if the figures of merit are excellent. Those figures are just an indication that the model correctly predicts the same samples that have been used to construct it. Therefore, this should never be taken as an optimal indication of the performance of the model.

10. Final remarks

There is an observable trend in the scientific community towards overusing certain machine learning algorithms, rapidly applying methods without properly evaluating whether the utilized method is appropriate. One of these trends is solely applying leave-one-out cross-validation just because it is the validation strategy that gives the best figures of merit possible without considering the data structure while risking the reliability of the model. Another one is assuming that the used algorithms and libraries are well-built, without considering that certain steps, especially in validation, might not be the optimal ones for our dataset. Or the riskiest one, making a classification model like PLS-DA (or orthogonalized PLS-DA – OPLS-DA) without resampling with the aim of just looking at the calibration outcome and check in the scores and loadings for any "first impression" of the classification. These trends should be severely avoided, and as researchers, we should wonder what

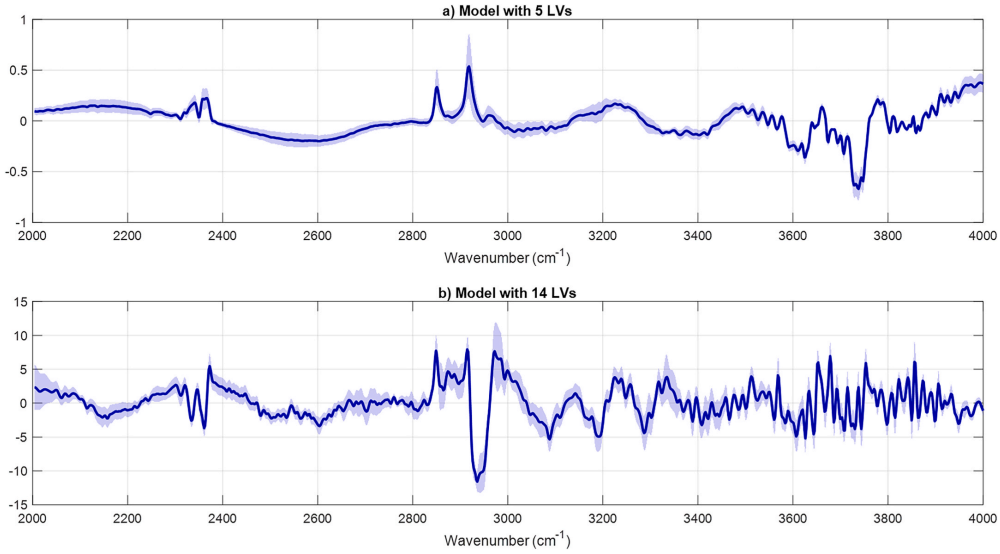


Fig. 12. Mean regression vectors (solid line) and the standard deviation (shaded area) of the calibration model, considering a) 5 LVs, and b) 14 LVs.

we are trying to achieve with our regression and classification models: To obtain good figures of merit or to obtain figures of merit that teach us the reliability of our classification and the independence from any block/structure scheme that we might have in our dataset.

Regarding the three dataset examples in this manuscript, the main conclusion is that, even though some cross-validation strategies offered reliable figures of merit, it is impossible to develop a trustable and reliable classification model to classify subjects with Alzheimer's disease using the FTIR spectra from dried cerebrospinal fluid shown in this tutorial. However, and this is extremely important to note, this conclusion should not be interpreted as an absolute statement. We assume that the dataset does not represent the entire variability expected for this disease and the healthy control; in other words, it is very likely that the statistics is still too low. A similar example is shown with a small dataset from a completely different field, the wines dataset. The classifications obtained were good, but more samples are needed to increase the confidence level of the classification. A good counter example of the previous statements is the COVID-19 dataset, a much larger dataset for the medical field where results were acceptable.

While PLS-DA demonstrates a remarkable capacity to identify features that differentiate classes, it is evident that datasets containing a substantial number of features are separable, and many separating hyperplanes are mere "noise". Fortunately, the implementation of proper resampling strategies can easily identify when PLS-DA is being applied ineffectively. We have remarked that various factors play a role in determining the appropriate type of resampling method for a specific dataset, such as the structure of the dataset, the total number of observations and variables, and the variance of the replicates in the dataset. The inner structure of data that describes both the intra (within replicates and within one class) and inter-variability (between the classes) is of utmost importance for developing appropriate classification models.

Using an intelligent criterion for data splitting can ensure the preservation of diversity within the dataset in both the training and testing subsets. Depending on the specific modelling objectives, this is crucial in selecting an appropriate validation strategy. Such findings have significant implications in chemometrics, especially when applied to data where the number of variables exceeds largely the number of samples.

In a more general framework, using cross-validation exclusively with small datasets is a risky adventure that is not advisable. Acknowledging the difficulty in certain research areas of collecting samples, a wrong cross-validation strategy can lead to completely overinterpreted results. Nevertheless, this drawback should not be used as a valid argument for still proposing models whose performance is poor.

Nowadays, we must place high demands on our models but always keep in mind that the best model is not the model that best fits the calibration data, or delivers the best results in cross-validation or, more generally spoken, resampling. The model that is stable and allows us to predict new samples with high reliability is the preferred one.

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.

Data availability

Data will be made available on request.

Acknowledgement

The authors want to thank the reviewers for the fruitful revision made. It helped to improve the quality of the manuscript to a great extent. This work was funded by the Basque Government (Elkartek 2022 (nG22): KK-2022/00001); further financial support by grant CEX2020-001038-M funded by MCIN/AEI/10.13039/501100011033. The authors acknowledge Fundación CITA-Alzheimer for providing the biosamples and Aitor Echeverria Ibarbia for doing the FTIR measurements.

References

[1] D.L. Massart, B.G.M. Vandeginste, J.M.C. Buydens, S. de Jong, P.J. Lewi, J. Smeyers-Verberke, L.M.C. Buydens, S. De Jong, J. Smeyers-Verbeke, Handbook of chemometrics and qualimetrics. https://doi.org/10.1016/S0922-3487(97)80056-1, 1997.
[2] J.M. Amigo, Data mining, machine learning, deep learning, chemometrics definitions, common points and trends (spoiler alert: VALIDATE your models!), Brazilian J. Anal. Chem. 8 (2021) 22–38.
[3] D. Ballabio, V. Consonni, Classification tools in chemistry. Part 1: linear models. PLS-DA, Anal. Methods (2013), https://doi.org/10.1039/c3ay40582f.
[4] D. Ruiz-Perez, H. Guan, P. Madhivanan, K. Mathee, G. Narasimhan, So you think you can PLS-DA? BMC Bioinf. 21 (2020).
[5] G. J, B. K, W. A, C.G. Gottfries, Diagnosis of dementias using partial least squares discriminant analysis, Dementia (1995) 83–88.
[6] C. Christin, H.C.J. Hoefsloot, A.K. Smilde, B. Hoekman, F. Suits, R. Bischoff, P. Horvatovich, A critical assessment of feature selection methods for biomarker discovery in clinical proteomics, Mol. Cell. Proteomics 12 (2013) 263–276.
[7] M. Barker, W. Rayens, Partial least squares for discrimination, J. Chemom. 17 (2003) 166–173, https://doi.org/10.1002/cem.785.
[8] C.M. Rubingh, S. Bijlsma, E.P.P.A. Derks, I. Bobeldijk, E.R. Verheij, S. Kochhar, A. K. Smilde, Assessing the performance of statistical validation tools for megavariate metabolomics data, Metabolomics 2 (2006) 53–61, https://doi.org/10.1007/s11306-006-0022-6.
[9] J.L. Rodgers, The bootstrap, the jackknife, and the randomization test: a sampling taxonomy, Multivariate Behav. Res. 34 (1999) 441–456, https://doi.org/10.1207/S15327906MBR3404_2.
[10] F. Westad, F. Marini, Validation of chemometric models - a tutorial, Anal. Chim. Acta 893 (2015) 14–24, https://doi.org/10.1016/j.aca.2015.06.056.
[11] R.W. Kennard, L. Stone, Computer aided design of experiments, Technometrics 11 (1969) 137–148, https://doi.org/10.2307/1266770.
[12] P. Refaeilzadeh, L. Tang, H. Liu, Cross-Validation, Encycl. Database Syst., 2009, pp. 532–538.
[13] B. Efron, The jackknife, the bootstrap and other resampling plans, Soc. Ind. Appl. Mathem. (1982), https://doi.org/10.1137/1.9781611970319.
[14] B. Thompson, Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications, American Psychological Association, 1995.
[15] T.G. Rio, G. Larios, B. Marangoni, S.L. Oliveira, C. Cena, C.A. do N. Ramos, FTIR spectroscopy with machine learning: a new approach to animal DNA polymorphism screening, Spectrochim. Acta Part A Mol. Biomol. Spectrosc. 261 (2021), 120036.
[16] T. Franca, D. Goncalves, C. Cena, ATR-FTIR spectroscopy combined with machine learning for classification of PVA/PVP blends in low concentration, Vib. Spectrosc. (2022), 103378.
[17] E. Szymańska, E. Saccenti, A.K. Smilde, J.A. Westerhuis, Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies, Metabolomics 8 (2012) 3–16, https://doi.org/10.1007/s11306-011-0330-3.
[18] K.H. Esbensen, P. Geladi, Principles of Proper Validation: use and abuse of re-sampling for validation, J. Chemom. 24 (2010) 168–187.
[19] A.L. Pomerantsev, O.Y. Rodionova, Procrustes Cross-Validation of short datasets in PCA context, Talanta 226 (2021), 122104, https://doi.org/10.1016/j.talanta.2021.122104.
[20] S. Kucheryavskiy, S. Zhilin, O. Rodionova, A. Pomerantsev, Procrustes cross-validation—a bridge between cross-validation and independent validation sets, Anal. Chem. 92 (2020) 11842–11850, https://doi.org/10.1021/acs.analchem.0c02175.
[21] L. Reichwein Zientek, B. Thompson, Applying the bootstrap to the multivariate case: bootstrap component/factor analysis, Behav. Res. Methods 39 (2007) 318–325.
[22] H. Martens, M. Martens, Modified Jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR), Food Qual. Prefer. 11 (2000) 5–16, https://doi.org/10.1016/S0950-3293(99)00039-7.
[23] Ł. Smaga, Bootstrap methods for multivariate hypothesis testing, Commun. Stat. Simulat. Comput. 46 (2017) 7654–7667, https://doi.org/10.1080/03610918.2016.1248573.
[24] H. Babamoradi, F. Van Den Berg, Å. Rinnan, Bootstrap based confidence limits in principal component analysis - a case study, Chemometr. Intell. Lab. Syst. 120 (2013) 97–105, https://doi.org/10.1016/j.chemolab.2012.10.007.
[25] P. Good, Permutation tests, Springer New York, New York, NY, https://doi.org/10.1007/978-1-4757-2346-5, 1994.
[26] H. van der Voet, Comparing the predictive accuracy of models using a simple randomization test, Chemometr. Intell. Lab. Syst. 25 (1994) 313–323, https://doi.org/10.1016/0169-7439(94)85050-X.
[27] E.V. Thomas, Non-parametric statistical methods for multivariate calibration model selection and comparison, J. Chemom. 17 (2003) 653–659, https://doi.org/10.1002/cem.833.
[28] S. Arlot, A. Celisse, A survey of cross-validation procedures for model selection, Stat. Surv. 4 (2010), https://doi.org/10.1214/09-SS054.

[29] E.K. Kemsley, M. Defernez, F. Marini, Multivariate statistics: considerations and confidences in food authenticity problems, Food Control 105 (2019) 102–112, https://doi.org/10.1016/j.foodcont.2019.05.021.
[30] M.M. Rahman, D.N. Davis, Addressing the class imbalance problem in medical datasets, Int. J. Mach. Learn. Comput. (2013) 224–228.
[31] C.L.M. Morais, K.M.G. Lima, M. Singh, F.L. Martin, Tutorial: multivariate classification for vibrational spectroscopy in biological samples, Nat. Protoc. 15 (2020) 2143–2162.
[32] H. Martens, T. Næs, Multivariate calibration, in: Chemometrics, Springer Netherlands, Dordrecht, 1984, pp. 147–156, https://doi.org/10.1007/978-94-017-1026-8_5.
[33] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics, in: Chemom. Intell., Lab. Syst., 2001, pp. 109–130, https://doi.org/10.1016/S0169-7439(01)00155-1.
[34] G. Yin, L. Li, S. Lu, Y. Yin, Y. Su, Y. Zeng, M. Luo, M. Ma, H. Zhou, L. Orlandini, D. Yao, G. Liu, J. Lang, An efficient primary screening of COVID-19 by serum Raman spectroscopy, J. Raman Spectrosc. 52 (2021) 949–958, https://doi.org/10.1002/jrs.6080.
[35] F. Wilcoxon, Individual comparisons by ranking methods, Biometrics Bull. 1 (1945) 80, https://doi.org/10.2307/3001968.

Dr. Andreas Seifert is Ikerbasque Research Professor at CIC nanoGUNE in San Sebastián, Spain. His research includes photonics/plasmonics combined with nanotechnology and artificial intelligence, primarily for applications in medical diagnostics, food quality control and environmental monitoring, reflected in more than 100 journal publications and proceedings, and 12 patents. Most prestigious awards are the 'Rudolf Kingslake Medal and Prize 2013' and the 'Carl Zeiss Innovation Award 2006'. https://www.nanogune.eu/en/research/groups/nanoengineering

Eneko Lopez is a mechanical engineer with more than 5 years of experience in optomechanical assemblies and works as laboratory engineer in the Nanoengineering Group at CIC nanoGUNE (San Sebastián, Spain). He holds a MSc in Smart Systems, specialized in microsystems from the Furtwangen University (Germany) and is currently pursuing his PhD career. His current research includes the application of chemometrics to spectroscopic data for the development of medical diagnostic tools, which is reflected in his 5 last publications.

Jaione Etxebarria is a postdoctoral researcher at CIC nanoGUNE in San Sebastián, Spain. She has extensive experience in microfluidics and biomedical microsystems for Lab-on-a-chip (LOC) and Point-of-Care (POC) rapid analysis devices. Her current research aims at developing new medical diagnostic tools using nanotechnology and photonics/plasmonics combined with artificial intelligence. She has authored more than 30 scientific papers in high-impact journals and proceedings and 2 patents, and she has been awarded with two best oral presentations for Postdoctoral Scientists, sponsored by Biosensors MPDI and IkerGazte.

Dr. José Manuel Amigo Research Professor. IKERBASQUE. Basque Foundation for Sciences. Bilbao. Spain; and Distinguished Professor at the Department of Analytical Chemistry. University of Basque Country. Spain. Current research interests include hyperspectral image analysis and the application of Chemometrics. He has authored more than 180 publications and has given more than 60 conferences and courses at international meetings. He received the "2014 Chemometrics and Intelligent Laboratory Systems Award" and the "2019 Tomas Hirschfeld Award". He is editor of the book "Hyperspectral Imaging. Volume 32. Elsevier. ISBN: 9780444639776. In Data Handling in Science and Technology.

A.3 Dried Biofluid Hyperspectral Analysis for Alzheimer's

Eneko Lopez, Giulia Gorla, Jaione Etxebarria-Elezgarai, Julene Aramendia, Kepa Castro, Maite García-Sebastián, Pablo Martínez-Lage, Andreas Seifert, Jose Manuel Amigo. **Hyperspectral Raman imaging to understand patterns in dried biofluids in Alzheimer's disease**. VIEW, 20250034, (**2025**).

The manuscript was cited as ref [21]. It is accesible under

http://doi.org/10.1002/VIW.20250034

Received: 11 December 2024 | Revised: 29 April 2025 | Accepted: 7 May 2025

DOI: 10.1002/VIW.20250034

RESEARCH ARTICLE

VIEW

Hyperspectral Raman imaging to understand patterns in dried biofluids in Alzheimer's disease

Eneko Lopez[1,2] | **Giulia Gorla**[3] | **Jaione Etxebarria-Elezgarai**[1] | **Julene Aramendia**[3] | **Kepa Castro**[3] | **Maite García-Sebastián**[4] | **Pablo Martínez-Lage**[4] | **Andreas Seifert**[1,5] | **Jose Manuel Amigo**[3,5]

[1]CIC NanoGUNE BRTA, San Sebastián, Spain

[2]Department of Physics, University of the Basque Country (UPV/EHU), San Sebastián, Spain

[3]Department of Analytical Chemistry, University of the Basque Country (UPV/EHU), Leioa, Spain

[4]Center for Research and Advanced Therapies, CITA-Alzhéimer Foundation, San Sebastián, Spain

[5]IKERBASQUE, B, asque Foundation for Science, Bilbao, Spain

Correspondence
Andreas Seifert, CIC NanoGUNE BRTA, San Sebastián 20018, Spain.
Email: a.seifert@nanogune.eu

Jose Manuel Amigo, Department of Analytical Chemistry, University of the Basque Country (UPV/EHU), Leioa 48009, Spain.
Email: josemanuel.amigo@ehu.eus

Funding information
MICIU/AEI/10.13039/501100011033, Grant/Award Number: CEX2020-001038-M; University of the Basque Country; Agencia Estatal de Investigación, Grant/Award Number: CEX2020-001038-M

Abstract

Raman spectroscopy is an ideal technique for identifying biomarkers in biological fluids. Raman measurements in biofluids are typically carried out based on drop coating deposition Raman spectroscopy, whereby the external ring formed by the so-called coffee ring effect is analyzed. It is well known that molecules, including proteins, generally migrate to the outer edges of the drop, resulting in molecular accumulation. In principle, this phenomenon simplifies the decision regarding where the Raman measurements should be carried out. However, relying solely on this effect might overlook important spectral information at other zones in the sample. Every biofluid is different, so important information can be lost by only focusing on the outer ring of the droplet. Here, the compositional distribution of the droplets is analyzed using hyperspectral Raman mapping to examine the entire biofluid droplet. With this approach, the spectral heterogeneity in the entire droplet is evaluated so that optimal areas for Raman measurements can be determined. The droplet is analyzed using principal component analysis to identify regions of maximum variance, *k*-means to group pixels by similarity, and multivariate curve resolution for each cluster to obtain

Abbreviations: AD, Alzheimer's disease; AT(N), amyloid, tau, and neurodegeneration; Aβ, amyloid beta; CSF, cerebrospinal fluid; DCDR, drop coating deposition Raman; ELISA, enzyme-linked immunosorbent assays; H, healthy individuals; MCR-ALS, multivariate curve resolution by alternating least squares; MRI, magnetic resonance imaging; MS, mass spectrometry; PC, principal component; PCA, principal component analysis; PC-AD, preclinical Alzheimer's disease; PD-AD, prodromal Alzheimer's disease; ROI, regions of interest; SNR, signal-to-noise ratio; SNV, standard normal variate.

the spectra of unique components that may highlight specific compounds in particular regions of the liquid. Although time consuming, hyperspectral Raman mapping helps identify areas with relevant information in the dried droplet for further systematic and detailed measurements. Important spectral features, as biomarkers of Alzheimer's disease, were found to be more pronounced in zones other than the outer coffee ring.

KEYWORDS
Alzheimer's disease, biofluids, biomarkers, drop coating deposition, hyperspectral imaging, Raman spectroscopy

1 | INTRODUCTION

1.1 | Study of biofluids by Raman spectroscopy

The study of biofluids is invaluable in medical diagnostics since they are easily accessible and rich in physiological and pathological information, including important biomarkers.[1,2] Compared to other analytical techniques, Raman spectroscopy has been widely employed to analyze biofluids and has demonstrated its potential to reveal label-free and detailed molecular information with high specificity and sensitivity.[3–13] This vibrational spectroscopic technique detects inelastically scattered light when the biofluid is irradiated with light, preferably with a laser, allowing for the observation of molecular vibrations in a powerful, non-invasive, and precise manner.

The Raman measurements in biofluids are often performed following the so-called drop coating deposition Raman (DCDR) technique. In DCDR, a small amount of the biofluid to be analyzed is deposited onto a substrate and evaporated until it is completely dry. In previous studies, DCDR spectroscopy was used, whereby the Raman signal of the measurements was taken from the outer ring region of dried drops from biofluids where the protein concentration is highest.[14–18] This is largely attributed to the coffee ring effect, where protein accumulation and, hence, the signal-to-noise ratio (SNR) are enhanced in the outermost regions of the droplet. However, relying solely on this effect overlooks the inherent complexity of the biofluid drying process and redistribution of the constituents.

1.2 | Coffee ring effect

The simple evaporation of liquid droplets occurs on many occasions every day. The most common example is drying a spilt drop of tea or coffee on a countertop, which leaves behind a distinct pattern with an outer ring of higher concentrated particles. This well-known phenomenon is called the "coffee ring" effect and is caused by capillary flows during evaporation that transports suspended particles to the periphery of the droplet.[19–22] Analogously, when biofluid droplets, such as cerebrospinal fluid (CSF) or blood plasma, evaporate, they are subject to complex interfacial and molecular dynamics[23–26] due to the presence of different macromolecules such as proteins or cells and salts, which lead to the formation of unique patterns and structures.[19,23,27] Therefore, several types of patterns appear in drying droplets as the result of the fluid composition and evaporation process. These patterns include cracks, dendrites, fractals, inclusions, crystals, fern-like structures, and specifically the famous coffee ring structure.[28,29]

The evaporation process induces concentration gradients and molecular migration that profoundly affect the final structure and composition of the droplet. These patterns are strongly influenced by the concentration of solutes and the molecular interactions within the biofluid[30,31] and indicate a specific physiological state.

The biofluid itself plays a significant role in the evaporation process, as solute molecules migrating to the droplet–air interface can alter molecular interactions, leading to phase separation or the formation of concentration gradients.[31,32]

1.3 | Alzheimer's disease

The pattern formation is of particular interest in biomedical applications such as medical diagnostics, where the understanding of detailed molecular composition of biofluids can aid in the detection of diseases such as Alzheimer's disease (AD),[14,33–35] which is the most common neurodegenerative disorder and a worldwide leading cause of dementia in individuals over 65 years. AD continuum is a progressive disease that develops gradually and is characterized by the accumulation of amyloid-β plaques

and misfolded tau proteins, leading to neurotoxicity and brain atrophy, particularly in regions critical for memory and cognition.[36–38] The progression continues with symptoms such as memory loss and mild cognitive impairment and finally with total incapacitation.[39] Therefore, early diagnosis is crucial as it can significantly help in finding new treatment options to improve patient outcomes and slow the progression of the disease.[40]

Common AD diagnostic techniques such as the enzyme-linked immunosorbent assay (ELISA) and mass spectrometry (MS) have specific limitations. MS, while capable of simultaneously analyzing multiple biomarkers with high sensitivity and specificity, requires extensive sample preparation and complex data analysis and is costly and time consuming.

Emerging research focuses on identifying biomarkers in less invasive biofluids such as blood, urine, and saliva. Such approaches promise earlier AD detection, potentially decades before symptoms appear, by tracking metabolic alterations associated with the disease.[26,27] Identifying these biomarkers could transform AD diagnosis and provide new insights into disease progression and potential treatment strategies.[41]

1.4 | Analysis of the entire droplet

The observed heterogeneity in the dried droplets suggests that by limiting the analysis to the ring region, valuable molecular information from other dried droplet regions could be neglected.[24,42] Analyzing the entire droplet will deliver a comprehensive understanding of the molecular distribution and composition. This holistic approach provides more information about the heterogeneous nature of the sample and enables the identification of biomarkers that may be unevenly distributed in the droplet.

Despite the advantages of Raman spectroscopy, its highly localized measurement capability can be a limitation when accounting for the full complexity of biofluid samples. Hyperspectral Raman imaging emerges as a measurement method that can account for both the spatial and spectral information within a sample and is therefore well suited for the investigation of complex biological samples.[42–46] A hyperspectral Raman image can be conceptualized as a three-dimensional (3D) array (Figure 1C), characterized by three dimensions: two spatial dimensions (x and y) and one spectral dimension (λ).[47] There are studies that show that hyperspectral Raman imaging can distinguish between different types of cells, detect molecular changes associated with disease states, and monitor biochemical processes in real time.[18,48–50]

1.5 | Scope of this study

In this work, we go one step further and study the Raman signal of two different biofluids (CSF and plasma) with hyperspectral Raman imaging.[44,45,47,48] Particular attention is paid to the spatial distribution of the composition of the droplets reflected in the Raman signal, which is due to the fluid- and physiology-specific pattern formation. We demonstrate that the different patterns differ not only microscopically, but also in their inherent Raman signal. We use the power of hyperspectral Raman imaging in combination with chemometrics to better understand the spatially distributed information of dried biofluids, using samples of two completely different biofluids from Alzheimer's patients. Principal component analysis (PCA), k-means, and multivariate curve resolution-alternating least squares (MCR-ALS) can provide the hidden Raman information coming from the different biomolecules in both biofluids and help determine different biomarkers across the different formed zones of the dried biofluids.

The objective of this study is to determine how constituents of body fluids are distributed in dried droplets and to find regions of high enrichment of disease-specific biomarkers—in our case for AD. The identification of elevated concentrations of specific biomarkers will enhance SNR and consequently provide a potential tool for classification and future medical diagnostics based on a straightforward and rapid technology that does not require special sample handling.

2 | MATERIALS AND METHODS

2.1 | Human samples

CITA-Alzheimer Foundation (San Sebastian, Spain) recruited participants for this study from a population-based clinical and biological cohort, which included adults with and without cognitive decline from a cross-sectional study.[51] Syndromic diagnosis was established through comprehensive neurological and neuropsychological assessments, structural magnetic resonance imaging, and CSF amyloid, tau, and neurodegeneration (AT(N)) biomarkers analysis.

Research was conducted in accordance with the Declaration of Helsinki and approved by local Ethics Committees.[51] Volunteers were categorized into three different groups: a control group of healthy individuals, a preclinical Alzheimer's disease (PC-AD) group and a prodromal Alzheimer's disease (PD-AD) group. The PC-AD group consisted of individuals who had abnormal biomarker values but maintained normal cognitive

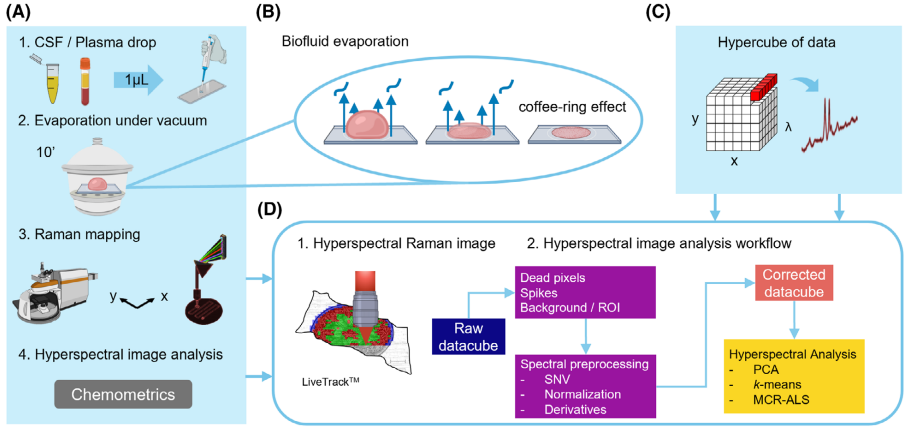


FIGURE 1 Schematic of the entire process chain from sampling to data analysis. (A) Schematic overview of entire process chain. (B) Biofluid evaporation process. (C) Hypercube of data. (D) Hyperspectral Raman image acquisition and data processing workflow.

function, whereas the PD-AD group exhibited abnormal cognitive function apart from abnormal biomarker values. The control group was defined by specific CSF cutoff values (Aβ42 > 1030 pg/mL, total-tau < 300 pg/mL, and p-tau < 27 pg/mL) to ensure AT(N) negativity and cognitive normalcy. The ELISA immunoassay was used to measure CSF biomarkers, including Aβ42, total-tau, and p-tau, to determine the biomarker values for group categorization.

A total of 15 participants were selected from whom CSF and blood plasma samples were collected for further analysis. Of these, five were diagnosed with preclinical AD, five with prodromal AD, and five participants represented the healthy controls (H). Each group comprised five plasma and five CSF samples, with the cohort consisting of 12 males and 18 females with an average age of 65.2 ± 2.3 years. CSF and blood samples were collected following international consensus recommendations as described in ref.[52] Blood was filled in 10 mL EDTA tubes and centrifuged at 4°C to separate the plasma from the cellular fraction. Both fractions were aliquoted and stored at –80°C in less than 1 h upon blood collection. CSF samples were centrifuged immediately at 4°C and stored at −80°C within 1 h of lumbar puncture. A summary of the demographic and clinical parameters of the study participants is provided in Table 1.

2.2 | General procedure for DCDR

Before doing the hyperspectral Raman spectroscopy measurements, CSF samples were defrosted and underwent ultracentrifugation using an Amicon Ultra-0.5 filter with a 3 kDa pore size, resulting in a protein-rich supernatant. This step was performed to isolate proteins such as amyloid and tau, based on their size and molecular weight, for the biomarker targeting with the DCDR spectroscopy measurements.[53] Blood plasma samples from the same participants were thawed, but not centrifuged, to investigate and compare the outputs of both biofluids.

As depicted in Figure 1A, 1 µL of the CSF or blood plasma sample was deposited onto a pre-cleaned microscope glass slide under stable laboratory conditions (inside a laminar flow cabinet at room temperature ~23°C and relative humidity 35%–45%), which was covered with aluminum foil to enhance the Raman signal. Apart from enhancing the Raman signal, the use of a hydrophobic substrate ensured that the droplet remained confined and uniform in shape, which is crucial for a reproducible drop deposition process and subsequent Raman measurements.[5,6] The droplet was deposited inside a laminar flow cabinet and dried under a desiccator for 10 min before the Raman measurements (Figure 1B).

2.3 | Hyperspectral Raman imaging measurements

The hyperspectral Raman images from CSF and plasma samples were acquired using the Renishaw inViaTM Qontor confocal Raman microscope. A diode laser beam with an excitation wavelength of 785 nm was focused on the sample with a power of 73 mW. A circular measurement grid with a diameter of 1400 µm was defined in each

TABLE 1 Demographic and clinical parameters of the participants from this study.

Gender	Years of education	Age (years)	Diagnostic	ABETA (pg/mL)	PTAU (pg/mL)	TTAU (pg/mL)	APOE genotype
Female	15	64.3	PD-AD	465.5	24.13	290.6	4/4
Female	15	65.2	H	1574	17.51	213.6	3/3
Female	16	63.1	PC-AD	329.4	12.54	184.4	3/4
Male	8	62.1	PD-AD	706.8	42.41	443.3	3/2
Male	17	64.8	H	1086	13.9	172.1	3/3
Male	14	65.6	PC-AD	558	27	312	3/4
Female	12	64.2	PD-AD	538.2	15.4	184.4	3/3
Female	9	63.1	H	1508	20.11	255.7	3/3
Female	6	65.7	PC-AD	567	51	520	3/2
Male	12	67.0	PD-AD	634.7	29.4	281.3	3/4
Male	12	67.5	PC-AD	501.2	38.4	285	3/4
Male	12	67.9	H	1355	14.48	190.1	3/3
Female	6	70.8	PD-AD	528.2	25.24	243.8	3/2
Female	15	61.5	PC-AD	554.4	26.34	276.1	3/3
Female	8	65.6	H	1281	15.77	201.1	3/4

Abbreviations: ABETA, amyloid β42 protein concentration; ApoE, Apolipoprotein E genotype; H, healthy controls; PC-AD, preclinical Alzheimer's disease; PD-AD, prodromal Alzheimer's disease; PTAU, phosphorylated tau protein concentration; TTAU, total tau protein concentration.

droplet, resulting in 750–950 spectra per sample, which required approximately 14 h of measurement time per sample. A piezo-electrically driven scanning stage moved the sample through the laser focus in a Cartesian raster pattern on the defined circular grid. The step size between two raster points was 35–45 µm. For each measurement position, one spectrum was recorded with an acquisition time of 1 s and 50 accumulations. The backscattered light was collected through the same microscope objective and diffracted using a grating with 1200 lines/mm with a spectral resolution of 2 cm^{-1}.

Raman spectroscopy is very sensitive to surface height differences. To assess the surface topography of the dried biofluid droplets, 3D profilometry was therefore performed using a non-contact optical profilometer (Bruker Contour GT-I). This technique captures high resolution, 3D surface maps by scanning the droplet surface with an optical sensor, which measures variations in height across the sample. The obtained 3D height maps, shown in Figure 2B, allowed the identification of regions with significant surface height heterogeneity, which could affect subsequent Raman measurements. The analysis of surface roughness and thickness variation was critical for selecting the measurement mode of our Raman instrument. Accurate focusing is essential to maximize the Raman signal and achieve high spatial resolution. Variations in sample height can lead to defocusing, resulting in decreased signal intensity, loss of spectral resolution, and potential misinterpretation of the molecular information. To maintain optimal focus throughout the measurements, the LiveTrack feature of the WIRE software was used to adapt to the different surface morphologies and topologies of plasma and CSF samples. This real-time focusing system continuously adjusts the sample focus to accommodate surface height variations and ensures consistent and accurate Raman signal acquisition during the 14 h of data acquisition per sample. Sample integrity is thus maintained, and high-quality spectral data are obtained, which is particularly important when the surface morphology of the biofluids significantly influences the measurement results.

2.4 | Hyperspectral Raman image analysis

All stages of the data analysis were carried out in MATLAB 2022a (The MathWorks, Inc.) using in-house routines and the HYPER-Tools[54] free graphical user interface for hyperspectral and multispectral image analysis. Figure 1D illustrates the overall process and provides a visual summary of the inherent process steps of the data analysis approach.

Several data preprocessing steps were applied before multivariate data analysis to ensure robust and accurate hyperspectral images. The individual Raman spectra collected from the two different types of samples, CSF and plasma, exhibit partly similar but also distinct characteristics. Data preprocessing includes corrections for dead pixels, cosmic rays, and spikes, which are standard

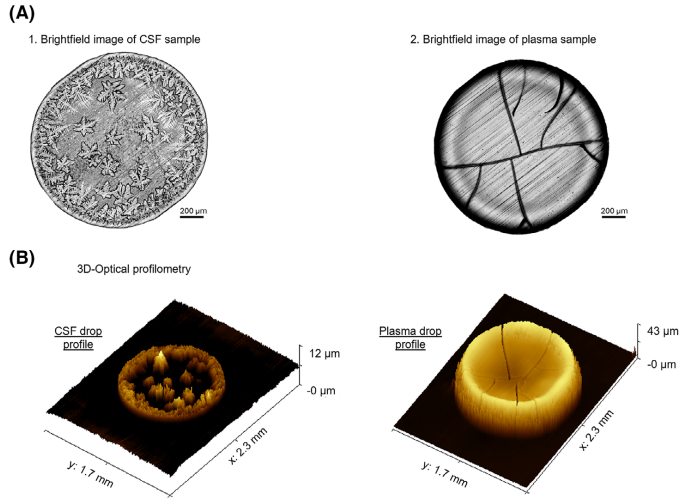


FIGURE 2 (A) Microscopy images of dried droplets from cerebrospinal fluid (CSF) and plasma. (B) Height profiles of dried droplets from CSF (left) and plasma (right) measured by 3D optical profilometry.

routines in Raman signal processing. Additionally, common preprocessing techniques such as baseline correction and normalization were applied to the raw data. For this purpose, a baseline correction was performed to eliminate fluctuations and artifacts by means of asymmetric least squares based on Eilers' work (order = 2, smoothness = 1000, asymmetry = 0.0001),[55] followed by spectral scaling via standard normal variate[53] and offset addition in order to be able to profit from the nonnegativity constraint in further MCR-ALS applications.

As the measurement range extended beyond the droplet boundaries, it was necessary to remove irrelevant points that could bias the analysis; this is also addressed as the definition of the region of interest (ROI).

In subsequent steps, the hyperspectral images from human plasma and CSF samples were subjected to unsupervised analyses using several multivariate methods, including PCA, *k*-means clustering, and MCR-ALS.[56,57] These methods were employed to reduce the dimensionality of the dataset by capturing the variance in the data, isolating individual spectral components, and identifying spectral patterns across the biofluid droplets with the goal of assessing their potential for discriminating Alzheimer's samples.

PCA was applied to decompose the hyperspectral data into orthogonal components, capturing the maximum variance within the dataset. By reducing the dimensionality of the data to a low number of principal components (the ones explaining the highest amount of variance), PCA enables the identification of significant spectral features and highlights regions of spectral heterogeneity within the biofluid samples.[58] *k*-Means clustering was used to group regions of the hyperspectral image based on spectral similarity. This method groups the pixels into clusters, each of which represents zones with similar spectral characteristics.[59] Applying *k*-means, makes it possible to visualize and map zones of similar chemical composition across the droplet, with the aim of identifying areas of interest for further Raman analysis. The number of clusters was determined through the elbow method, which calculates the sum of square distances between the centroids and each point of the cluster. At a certain number k of clusters, the curve of summed distances makes a kink and continues more or less linearly with little slope. This is the point of the optimum numbers of clusters.[60] Despite its power, *k*-means does not provide a comprehensive elucidation of the chemical compounds. Each cluster is created as an association of similar pixels due to its closeness in the variable space (using, for instance, the Euclidean Distance). To elucidate the major chemical compounds in each cluster, it is necessary to apply a curve resolution method such as MCR-ALS.

MCR-ALS is a powerful method used to decompose complex multivariate data into pure component contributions with minimum prior knowledge of the system.[61] In hyperspectral images, MCR-ALS performs a bilinear decomposition of the unfolded hyperspectral image $\mathbf{X}(MN \times \lambda)$, where M and N are the spatial dimensions of the pixels

as indicated in Equation (1):

$$\mathbf{X} = \mathbf{C}\mathbf{S}^{\mathrm{T}} + \mathbf{E}, \tag{1}$$

where **X** is the unfolded hyperspectral image with dimensions M^*N pixels and λ wavenumbers, **C** ($M^*N \times F$) contains the relative concentration profiles for each F component that can be refolded into the relative concentration surfaces $M \times N$. The superscript T indicates the transpose of the **S** matrix with the pure signals, while **E** ($MN \times \lambda$) is the matrix of residues.

This decomposition is done under optional constraints such as non-negativity by optimizing an initial estimation of the pure spectra $\mathbf{S}^{\mathrm{T}}$ ($F \times \lambda$) and concentration profiles **C** ($MN \times F$), knowing beforehand the number of components F. The algorithm alternates between updating the relative concentration surfaces and spectral profiles using least-squares regression while enforcing the chosen constraints, allowing for the resolution of overlapping signals in hyperspectral imaging by minimizing the error (Equation 2):

$$\min_{S_0} \|\mathbf{X} - \mathbf{C}\mathbf{S}^{\mathrm{T}}\|_2^2 \tag{2}$$

where $\min_{S_0} \|\|$ indicates the minimum when initial estimations of the spectral profiles are given to start the iterations and $\|\|_2^2$ indicates the Frobenius norm. MCR-ALS is widely applied in analytical chemistry, particularly in spectral unmixing, reaction monitoring, and hyperspectral image analysis, offering a flexible, data-driven approach to resolving mixed spectroscopic information.[62,63]

3 | RESULTS AND DISCUSSION

3.1 | Surface and interface topography

Figure 2A shows microscopy images with typical patterns formed by the deposition of CSF and blood plasma. Although the coffee ring effect can be observed in both sample types, the distribution and type of constituents are clearly different. The images from 3D optical profilometry (Figure 2B) reveal that the height profiles of the two biofluids differ significantly despite identical initial volumes, identical substrate material, and same drying process. CSF droplets exhibit a maximum height of approximately 15 μm, while plasma droplets can reach up to 50 μm. This disparity is likely due to differences in fluid composition and evaporation dynamics. Different patterns can be recognized in the CSF and blood plasma samples after evaporation. In the CSF samples, the center of the dried drop (Figure 2A, left) tends to contain fewer proteins, forming thin fern-like patterns, while denser clusters of macromolecules accumulate as one moves away from the center.[23,26]

The physical evaporation process and associated pattern formation is important for hyperspectral image analysis. The pattern formation is driven by a variety of factors that control the Marangoni flows, as surface properties of the substrate (wettability), surface tensions of the droplets, "particles" inside the droplet (proteins, lipids, etc.), salts, and the viscosity and volume of the droplet. When a CSF drop evaporates on a hydrophobic surface, the protein that is the least soluble of all the dissolved components and has the highest concentration starts to deposit at the periphery of the drop. This deposition continues until the ring is completely formed, and only a minimal amount of solution remains in the center of the drop. At this point, the fern patterns begin to form. From our mapped spectral data and from literature,[64] it is known that the ferns contain sodium chloride, urea, and a small amount of proteins. The last regions to dry are the edges of the fern-like structures, where the patches of sodium bicarbonate are deposited. These microscopic and surface morphology observations showed a diversity of pattern formations, which led to the decision to employ hyperspectral Raman imaging to target zones that are characteristic for identifying AD.

The pattern formation in plasma samples follows a different process, which is attributed to a very distinct composition. Plasma mostly consists of ~93% water-carrying ions, nutrients, and ~7% proteins.[25] Upon closer inspection of the plasma samples (Figure 2, right-hand side), the fern-like patterns cannot be seen. Instead, various cracks appear, as also reported by Zhang et al.[26] who suggested several reasons for the segregation of different components in a solution during drying on a hydrophobic surface. These include kinetic effects, thermodynamics, and various solubilities and concentrations of the individual components. Their observations align with our experimental data, indicating that solubility and concentration significantly affect the segregation and repartition of molecules during deposition.

In contrast to CSF samples, and as previously noted, the elevated height in the coffee ring was particularly pronounced in plasma samples. This observation indicates a higher concentration of proteins and macromolecules in the peripheral ring formed during the drying process of plasma.

3.2 | Preliminary study of the most characteristic wavenumbers for AD

A detailed hyperspectral analysis of CSF and human plasma samples was performed to characterize the spatial

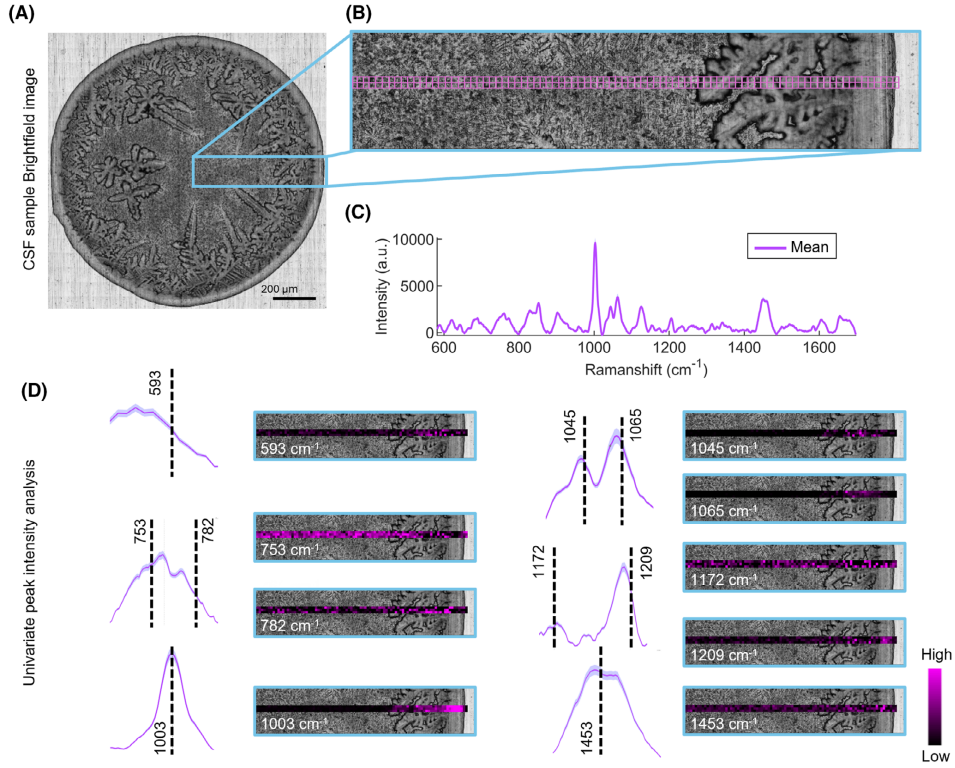


FIGURE 3 (A) Dried cerebrospinal fluid (CSF) sample with fern-like patterns. (B) Zoom in on the mapped area in (A). (C) Averaged Raman spectrum of the mapped rectangular area. (D) Univariate analysis of characteristics peaks for Alzheimer's disease (AD) discrimination from Table 2.[11,65]

and spectral heterogeneity in dried biofluid samples. First, a preliminary examination of a single dried CSF drop was carried out to assess the heterogeneity of the sample and to investigate the spatial distribution of the intensities of the AD biomarker peaks. Our approach was motivated by a detailed measurement of a dried CSF droplet, as illustrated in Figure 3.

While we assumed for simplicity that outward-driven forces generate patterns that are characteristic within a given radius, we acknowledge that the inherent variability in biofluid drying dynamics leads to structural microscale differences in each droplet due to Marangoni flows and other drying effects.[23–26] However, it seems logical that the Raman spectra of these microscale structures would be representative regardless of their location. This assumption allowed us to perform an initial analysis with a manageable number of measurement points (270 points) to identify general trends in the spatial distribution of AD-related biomarkers. For this purpose, a Raman image extending from the center to the edge was performed. The image has a rectangular shape with three rows and 90 columns with a measuring grid of 11 µm spacing, as shown in the zoomed area of Figure 3B. Figure 3C shows the average spectrum of all 270 points. A preliminary analysis on those wavenumbers that had been identified in prior research as potentially discriminative for AD (Table 2)[11,65] is conducted here. The resulting analysis (Figure 3D) reveals clear pattern distributions in both the microscopy image and the corresponding Raman spectra. Remarkably, not all peaks that are characteristic of AD appear more intense at the ring—a novel observation in dried CSF samples. The most intense peaks, such as the one at 1003 cm^{-1}, show that the signal emanates more strongly from the edge of the ring. However, other intense peaks, like those at 1045

TABLE 2 Potentially discriminative Raman peaks for Alzheimer's disease.[11,65]

Peak position ± 5 (cm^{-1})	Band assignment
593	Tryptophan, fructose, citric acid
753	Tryptophan, valine, phosphoenolpyruvate, phosphatidylinositol, symmetric stretching vibration of phosphate
782	Citric acid, tryptophan, coenzyme A, riboflavin, histidine, proline, lactose
1003	Phenylalanine, acetoacetate, serine, n-acetyl-glucosamine, β-carotene, glutamate, β-D-glucose, lactose, fatty acids
1045	Glycine, proline, fructose, glutathione, glutamate, acetyl coenzyme A, guanine, tryptophan, tyrosine, hypotaurine, taurine
1065	Arginine, fatty acids, galactosamine, histidine, valine
1172	Tyrosine, acetoacetate, histidine, glutathione, proline, fatty acids
1209	Tyrosine, acetyl coenzyme A, citric acid, tryptophan, fatty acids, n-acetylglucosamine, β sheet
1453	Valine, proline, lauric acid, tryptophan, fatty acids, fructose-6-phosphate, lactose, fructose, galactosamine, uracil, glutathione

and 1065 cm^{-1}, do not show such a high intensity, especially at the edge. Instead, their signal comes mainly from the structures formed adjacent to the edge. The same pattern is observed for the peak at 1453 cm^{-1}. Our following extensive analysis of pattern distribution is based on these particular findings that have not been investigated in detail in other publications.

To further validate the conclusions of the simplified hyperspectral Raman map, a more comprehensive Raman mapping of another entire CSF droplet was performed, as shown in Figure 4. The surface map confirmed the results obtained in the previous univariate analysis, demonstrating that the characteristic AD peaks are distributed across the sample, with some peaks showing higher intensities at the edge and others in adjacent or inner regions. Specifically, the same intensity distribution patterns were observed for those peaks characteristic of AD. For instance, the peak at 1003 cm^{-1} still exhibits a stronger intensity at the ring of the droplet, while other peaks, such as those at 1045, 1065, and 1453 cm^{-1}, show more intense signals in structures adjacent to the ring and in inner regions, rather than directly at the edge. This analysis corroborates our initial hypothesis and demonstrates that specific zones within the droplet exhibit higher peak intensities, confirming the heterogeneous distribution of biomarkers in the dried CSF sample.

This analysis reveals that the characteristic AD peaks are not localized but distributed over the entire sample. Motivated by these findings, we extended our analysis to an integrated dataset of dried CSF and plasma biofluids. Although this intensity-based analysis provides a straightforward overview of the spatial distribution of AD-related peaks, a more detailed hyperspectral analysis using chemometric techniques is required to understand the biochemical complexity of the sample fully.

3.3 | Hyperspectral data analysis

Initially, PCA was employed to differentiate between CSF and plasma samples in order to better understand the spectral variations characteristic of each biofluid. Following this, isolated *k*-means clustering analyses was performed on each biofluid type to investigate specific spatial and spectral differences. For plasma, clustering helped identify a more homogeneous peak distribution within the dried drops. In contrast, for CSF, *k*-means clustering was combined with MCR-ALS analysis to further resolve pure spectral components within each cluster.

3.3.1 | Hyperspectral analysis of human samples from CSF and plasma

The inclusion of both CSF and plasma from the same participants allows us to perform a comparative analysis, which provides more information about the biochemical differences and similarities between these biofluids and the localization of these differences. The outcome makes routine Raman measurements more effective.

As shown in Figure 5, PCA reveals that the two biofluids differ topographically and spectrally, which is particularly

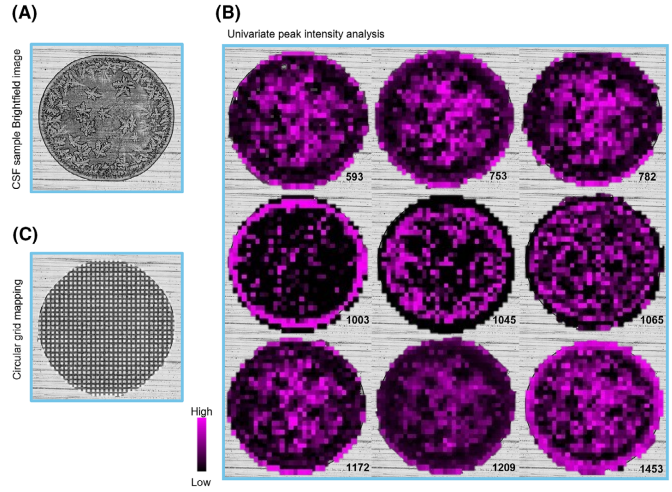


FIGURE 4 Hyperspectral Raman imaging of one complete cerebrospinal fluid (CSF) droplet with univariate analysis performed on characteristic wavenumbers associated with Alzheimer's disease (AD) from Table 2.[11,65] (A) Microscopic image of the CSF droplet. (B) Univariate analysis of peaks characteristic of AD (in cm^{-1}). (C) Circular grid for hyperspectral Raman mapping.

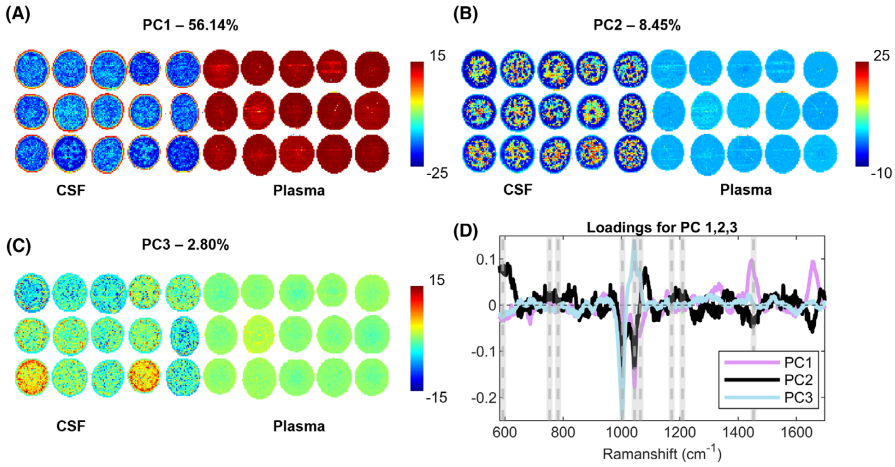


FIGURE 5 Principal component analysis (PCA) analysis with three principal components to demonstrate the spectral difference in both biofluids. (A) Results from the first principal component. (B) Second principal component. (C) Third principal component. (D) Loadings from each of the principal components (PCs) showing the characteristic spectral peaks associated with Alzheimer's from Table 2.[11,65]

pronounced in principal component 1 (PC1) and the other PCs. Here, we rely on the cumulative explained variance to determine an appropriate number of principal components. The signal from plasma samples seems homogeneous across the entire droplet, indicating a uniform distribution of molecular components. Surprisingly, PC1 is strongly and equally expressed in both the plasma samples and the outer ring of CSF droplets, indicating

TABLE 3 Comparative summary of the diagnostics characteristics of cerebrospinal fluid (CSF) and human plasma biomarkers for Alzheimer's disease (AD) detection.[66–69]

Biofluid	Advantages	Disadvantages	Diagnostic potential
CSF	Direct measurement of central nervous system biomarkers ($A\beta_{42}$, p-tau, t-tau); gold standard for brain pathology correlation	Invasive lumbar puncture; procedural risks; limited use for repeated or population-wide testing	Very high diagnostic accuracy (AUC 0.85–0.95); ideal for definitive diagnosis but impractical for screening or serial monitoring
Plasma	Minimally invasive; easily accessible; scalable for widespread screening; recent assays (especially p-tau217) match CSF accuracy	Lower biomarker concentrations; potential peripheral confounding; requires strict assay standardization	Recent plasma biomarkers (e.g., p-tau217) achieve diagnostic performance comparable to CSF (AUC 0.9–0.96); highly promising for non-invasive AD detection
CSF + plasma	Integrates central (CSF) and peripheral (plasma) information; could capture complementary aspects of AD pathology	Higher complexity and cost; need for standardization	May enhance diagnostic robustness, particularly in early/preclinical stages; valuable for personalized diagnostics but requires further research validation

Abbreviation: AUC, Area under the Curve.

a similar molecular composition in these specific zones. Interestingly, characteristic AD peaks from Table 2 were identified in PC2 in CSF samples, which were absent in the plasma samples. PC3 still shows the difference between both biofluids and additionally some internal patterns among the CSF samples.

In addition to these spectral differences revealed by PCA, it is important to consider the broader diagnostic characteristics of CSF and plasma as biofluids for AD detection. Table 3 summarizes the main advantages, limitations, and diagnostic potential of CSF and plasma biomarkers. While CSF biomarkers remain the reference standard for confirming Alzheimer's pathology,[66] recent advances in plasma biomarker assays, particularly plasma p-tau217, have demonstrated diagnostic performance comparable to CSF-based methods.[67–69]

As shown in Table 3, plasma biomarkers offer a less invasive and more scalable alternative to CSF analysis, with promising diagnostic accuracy that could enable broader clinical implementation.[68] Nevertheless, technical considerations, such as lower biomarker concentrations and the need for strict pre-analytical control, remain critical for plasma-based testing. Combining CSF and plasma biomarker profiles could be particularly valuable for early or preclinical AD detection, as it provides complementary information on disease progression and increases diagnostic robustness.[66]

3.3.2 | Hyperspectral analysis of human plasma

The results obtained with *k*-means are depicted in Figure 6. The appropriate number of clusters is three, which was determined by the elbow method as described above. Figure 6 shows a very similar pattern to the PC1 results of the PCA model. This indicates that the two biofluids differ not only in topography but also in their spectral profiles, suggesting that individual analyses of each biofluid can provide distinct compositional characteristics.

The patterns in Figure 6A, based on *k*-means clustering, show that the analysis of the plasma dataset alone provides much less homogeneity than originally assumed when both sample types, plasma and CSF, were merged into one dataset. Upon visual inspection of the different clusters, it can be observed that the protein concentration and the SNR become higher in the ring region of the droplets. This result is consistent with existing literature and reinforces the idea that protein aggregation during the drying process tends to concentrate at the periphery of the droplet.

These results suggest that the variability in human plasma samples poses a challenge for simple clustering of Alzheimer's biomarkers. However, the consistent observation of an elevated protein concentration in the ring region indicates the potential to effectively differentiate between plasma samples from Alzheimer's patients and healthy control, particularly with statistically relevant datasets.

3.3.3 | Hyperspectral analysis of human CSF

In view of the heterogeneity observed in the PCA analysis of the CSF samples, *k*-means clustering was applied. Figure 6B shows that three predominant clusters were identified. CSF samples exhibit distinct differences in cluster centroids and spectra, indicating a variable molecular composition that can potentially enhance AD classification; this issue has already been discussed in Section 3.3.1.

Previous studies have found that the predominant peaks characteristic for AD closely match those found in cluster III of our analysis.[11,65] This observation suggests that

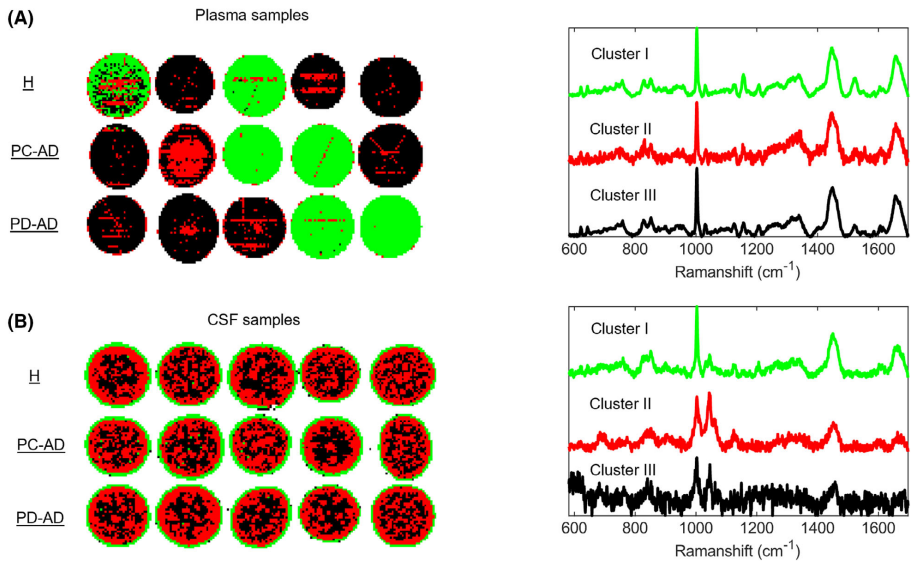


FIGURE 6 Hyperspectral Raman images analyzed by *k*-means. Each color represents one cluster and right part of the figure represents the corresponding cluster centroids. (A) Plasma samples and (B) cerebrospinal fluid (CSF) samples.

spectral features within cluster III may hold particular relevance in distinguishing AD-related molecular signatures when chemometric models are applied. In addition, the integration of data from both the ring region and cluster I promises to provide a more complete and comprehensive assessment of biofluid composition, potentially enhancing the ability to identify spectral patterns associated with AD. However, the variance between patients is greater than the variance observed across the different regions of the biofluid, so it is necessary to increase the dataset in a statistically relevant manner to further confirm this assumption.

To further investigate the composition of the clusters from the CSF samples, MCR-ALS was applied to the previously identified clusters in the CSF samples, extracting two pure main components for cluster I and three pure main components for clusters II and III (Figure 7). This analysis revealed significant spectral complexity, with the spatial distribution of pure components showing an overlap of spectral similarities between clusters. This finding indicates that neighboring areas within different clusters share comparable spectral profiles. In practical terms, this may require Raman measurements in areas such as the peripheral ring—including regions classified differently by clustering algorithms—capturing signals from chemically distinct but spectrally overlapping components.

In cluster I (Figure 7A), which is associated with the outer ring of the CSF, MCR-ALS revealed two distinct pure spectra, indicating structural and compositional complexity within this region. These spectra differ from each other, with one sharing similarity to a primary component in cluster II (Figure 7B). In particular, a peak at 1045 cm^{-1}, typically observed on the inner side of the dried drop (cluster II), occasionally appears within the outer ring in certain samples, indicating significant spectral variability within this peripheral zone.

In prior studies, the outer ring was often considered a homogenous Raman ROI, as it was assumed to have a high SNR. However, our results challenge this assumption, showing that different zones within the ring can contain at least two distinct spectral components. The lack of a precise definition for the "ring" in the literature may lead to misinterpretation or potential biases when regions with different spectral profiles are analyzed as a single unit.

The first component from cluster II (Figure 7B) shows a completely different spectrum, characterized, among others, by peaks at approximately 1045 and 1065 cm^{-1}. Interestingly, these peaks also appear in the second component of cluster III (Figure 7C) at some specific localized regions. These Raman bands have been associated with vibrational modes that are typical for protein structures, such as the ring breathing mode of phenylalanine (~1003

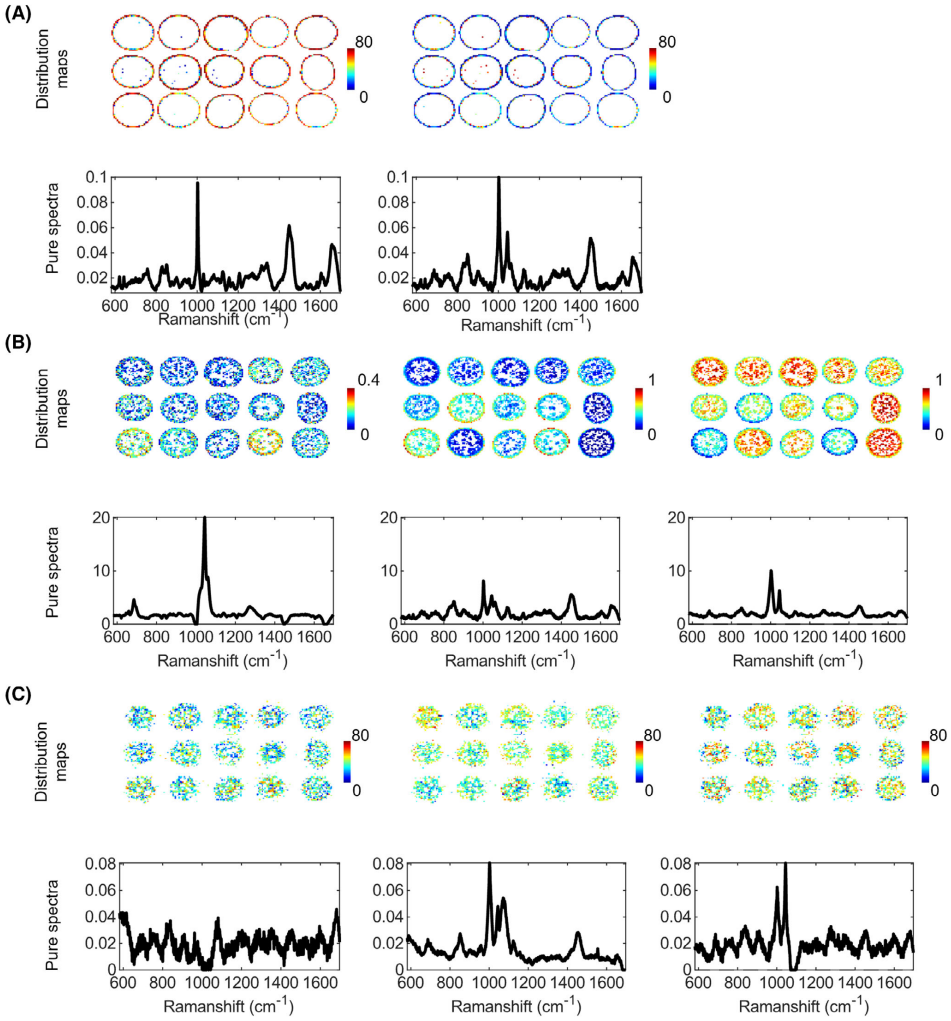


FIGURE 7 Multivariate curve resolution-alternating least squares (MCR-ALS) results for the different clusters obtained by *k*-means from the cerebrospinal fluid (CSF) samples. The distribution maps of each of the components are presented at the top, and pure spectra for each of the components are presented at the bottom of each of the figures. (A) Cluster I, (B) cluster II, and (C) cluster III.

cm^{-1}) and C–H bending/C–O stretching of methionine and phenylalanine (~1045–1065 cm^{-1}).[70,71] This suggest that the observed spectral features may be partially related to the secondary structure of proteins or to aggregation phenomena.

This outcome highlights the importance of a detailed characterization of the ring and each cluster in biofluid analysis, particularly in studies focused on the identification of disease-specific biomarkers or pathological spectral patterns. MCR-ALS not only enables the identification of primary components within clusters but also highlights zones of high spectral similarity that are not discerned by unsupervised techniques such as *k*-means. Therefore, a rigorous interpretation of the data from each cluster is crucial

to avoid unintentional grouping of spectrally distinct components, as the results demonstrate substantial variability within the ring, with some spectra comparable to those of adjacent clusters.

4 | CONCLUSION

In the presented study, hyperspectral Raman imaging was used to analyze droplets from different dried biofluids and evaluate corresponding molecular compositional differences, specifically in CSF and blood plasma. The results help us understand the complex dynamics involved in the drying process of these biofluids. Analyzing these dried patterns through hyperspectral Raman imaging clarifies where optimal measurements can be taken and highlights the most informative regions within the droplet for medical diagnostics, here demonstrated in the context of AD, but can be easily extrapolated to other diseases.

Before performing the hyperspectral Raman measurements, surface topography analysis already indicated how the different biofluids exhibited distinct morphologies. To further investigate and map the spectral distribution across the entire sample, hyperspectral Raman imaging was employed. This allowed us to explore the molecular composition in different regions of the droplet, particularly in areas where specific morphological patterns—such as coffee ring structures, fern-like patterns, and other complex formations—were observed in the dried CSF droplets. The samples were subsequently analyzed using various chemometric techniques, including PCA, *k*-means clustering, and multivariate curve resolution-alternating least squares, which confirmed that the spectral and morphological patterns differed significantly between the biofluids.

Interestingly, an exploratory analysis provides important information on the spatial distribution of the molecular signals in the biofluid droplets, particularly when comparing the two fluids. While similar studies often rely on the coffee ring effect to perform Raman measurements on the outer ring of dried biofluids—where protein accumulation and Raman SNR are typically higher—we observed a significant differentiation between the two biofluids. Specifically, the uniformity of the Raman signal across the entire plasma droplet validates the coffee ring-based measurement procedure for these samples. However, in the case of CSF, and particularly in the context of AD, our results show that the characteristic peaks associated with the disease are not limited to the outer ring, but also occur in other regions of the droplet. Apparently, spectral data particularly from the cluster with very little structure in the central and peripherical region (cluster III in the text above) appear to enhance the distinction of AD-related spectral features. Furthermore, integrating data from both the ring region of the dried droplet and that specific cluster may offer a more comprehensive diagnostic approach by capturing a wider range of molecular signatures. This discovery implies that focusing exclusively on the outer ring may overlook important molecular information present in other zones of the CSF droplet. Therefore, we hypothesize that a more comprehensive analysis that includes Raman measurements in multiple regions of the dried droplet will lead to the development of more robust and discriminatory models for Alzheimer's detection, and likewise for the detection of other pathological cases. This hypothesis highlights the need for a refined approach in Raman spectroscopy studies of CSF and presumably other non-invasive biofluids such as saliva, urine, etc., to fully capture the molecular complexity associated with Alzheimer's or other pathologies and diseases.

While hyperspectral Raman imaging provided rich spatial and compositional information in this study, we acknowledge that a cross-check with other analytical reference techniques would further strengthen the assignment of Raman spectral features to specific AD biomarkers (e.g., Aβ42, p-tau). Although such biochemical correlation was beyond the scope of the present work, future studies combining Raman imaging with spatially resolved biochemical validation are essential to bridge the gap between spectral patterns and established molecular markers.

In conclusion, this study demonstrates that drop-coating deposition Raman hyperspectral imaging improves the biochemical analysis of biofluid droplets and provides a detailed observation of the molecular composition and distribution of the compounds present in the biofluid. We have shown that the diagnostic capacity can be improved by optimizing the measurement conditions with respect to the type of biofluid and measurement location within the dried droplet. In the context of AD, such advances are particularly important for early detection, which is nowadays critical for effective intervention.

AUTHOR CONTRIBUTIONS

Eneko Lopez: Conceptualization; methodology; software; formal analysis; investigation; data curation; visualization; and writing—original draft. **Giulia Gorla**: Conceptualization; methodology; software; validation; data curation; and writing—review and editing. **Jaione Etxebarria-Elezgarai**: Conceptualization; methodology; and writing—review and editing. **Julene Aramendia**: Methodology; investigation; and writing—review and editing. **Kepa Castro**: Writing—review and editing; and supervision. **Maite García-Sebastián**: Resources. **Pablo Martínez-Lage**: Resources. **Andreas Seifert**:

Conceptualization; validation; writing—review and editing; supervision; project administration; and funding acquisition. **Jose Manuel Amigo**: Conceptualization; software; validation; data curation; writing—review and editing; and supervision. All authors read and approved the submitted version.

ACKNOWLEDGMENTS

This work was financially supported by grant CEX2020-001038-M funded by MICIU/AEI/10.13039/501100011033. Open access funding was provided by the University of the Basque Country.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

ORCID

Jose Manuel Amigo https://orcid.org/0000-0003-1319-1312

REFERENCES

1. J. M. Cameron, H. J. Butler, D. S. Palmer, M. J. Baker, *J. Biophotonics* **2018**, *11*, e201700299.
2. K. A. Esmonde-White, F. W. L. Esmonde-White, M. D. Morris, B. J. Roessler, *Analyst* **2014**, *139*, 2734.
3. E. Kočišová, M. Procházka, L. Vaculčiaková, *Appl. Spectrosc.* **2015**, *69*, 939.
4. A. Kuižová, M. Přikryl, M. Procházka, E. Kočišová, *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2021**, *262*, 120109.
5. C. Ortiz, D. Zhang, Y. Xie, A. E. Ribbe, D. Ben-Amotz, *Anal. Biochem.* **2006**, *353*, 157.
6. E. Kočišová, A. Kuižová, *Appl. Spectrosc. Rev.* **2024**, *59*, 717.
7. A. Kuižová, E. Kočišová, *J. Raman Spectrosc.* **2023**, *54*, 694.
8. E. Kočišová, A. Kuižová, M. Procházka, *Analyst* **2024**, *149*, 3276.
9. J. Filik, N. Stone, *Analyst* **2007**, *132*, 544.
10. P. Šimáková, E. Kočišová, M. Procházka, *J. Raman Spectrosc.* **2013**, *44*, 1479.
11. E. Lopez, J. Etxebarria-Elezgarai, M. García-Sebastián, M. Altuna, M. Ecay-Torres, A. Estanga, M. Tainta, C. López, P. Martínez-Lage, J. M. Amigo, A. Seifert, *Int. J. Mol. Sci.* **2024**, *25*, 4737.
12. I. Olaetxea, H. Lafuente, E. Lopez, A. Izeta, I. Jaunarena, A. Seifert, *Adv. Sci.* **2023**, *10*, 2204834.
13. H. Hano, C. H. Lawrie, B. Suarez, A. Paredes Lario, I. Elejoste Echeverría, J. Gómez Mediavilla, M. I. Crespo Cruz, E. Lopez, A. Seifert, *ACS Omega* **2024**, *9*, 14084.
14. M. P. dos Santos, K. M. G. Lima, *Appl. Spectrosc. Rev.* **2023**, *58*, 869.
15. D. Cialla-May, C. Krafft, P. Rösch, T. Deckert-Gaudig, T. Frosch, I. J. Jahn, S. Pahlow, C. Stiebing, T. Meyer-Zedler, T. Bocklitz, I. Schie, V. Deckert, J. Popp, *Anal. Chem.* **2022**, *94*, 86.
16. P. Giamougiannis, C. L. M. Morais, R. Grabowska, K. M. Ashton, N. J. Wood, P. L. Martin-Hirsch, F. L. Martin, *Anal. Bioanal. Chem.* **2021**, *413*, 911.
17. E. Buchan, M. Hardy, P. de Carvalho Gomes, L. Kelleher, H. O. M. Chu, P. G. Oppenheimer, *Appl. Spectrosc. Rev.* **2022**, *59*, 277.
18. H. J. Butler, L. Ashton, B. Bird, G. Cinque, K. Curtis, J. Dorney, K. Esmonde-White, N. J. Fullwood, B. Gardner, P. L. Martin-Hirsch, M. J. Walsh, M. R. McAinsh, N. Stone, F. L. Martin, *Nat. Protoc.* **2016**, *11*, 664.
19. R. D. Deegan, O. Bakajin, T. F. Dupont, G. Huber, S.R. Nagel, T.A. Witten. *Nature* **1997**, *389*, 827.
20. M. Yang, D. Chen, J. Hu, X. Zheng, Z.-J. Lin, H. Zhu, *TrAC, Trends Anal. Chem.* **2022**, *157*, 116752.
21. D. Mampallil, H. B. Eral, *Adv. Colloid Interface Sci.* **2018**, *252*, 38.
22. M. J. Hertaeg, C. Rees-Zimmerman, R. F. Tabor, A. F. Routh, G. Garnier, *J. Colloid Interface Sci.* **2021**, *591*, 52.
23. H. Hu, R. G. Larson, *J. Phys. Chem. B* **2002**, *106*, 1334.
24. Y. J. P. Carreón, M. L. Gómez-López, O. Díaz-Hernández, P. Vazquez-Vergara, R. E. Moctezuma, J. M. Saniger, J. González-Gutiérrez, *Sensors* **2022**, *22*, 1156.
25. A. Pal, A. Gope, A. Sengupta, *Adv. Colloid Interface Sci.* **2023**, *314*, 102870.
26. R. Chen, L. Zhang, D. Zang, W. Shen, *Adv. Colloid Interface Sci.* **2016**, *231*, 1.
27. M. A. Hack, M. N. van der Linden, H. Wijshoff, J. H. Snoeijer, T. Segers, *J. Colloid Interface Sci.* **2024**, *673*, 788.
28. F. Wang, Q. Yuan, *J. Colloid Interface Sci.* **2023**, *637*, 522.
29. A. Matavž, U. Uršič, J. Močivnik, D. Richter, M. Humar, S. Čopar, B. Malič, V. Bobnar, *J. Colloid Interface Sci.* **2022**, *608*, 1718.
30. M. E. Buzoverya, Y. P. Shcherbak, I. V. Shishpor, *Tech. Phys.* **2014**, *59*, 1550.
31. M. Efstratiou, J. Christy, D. Bonn, K. Sefiane, *Colloids Interfaces* **2021**, *5*, 43.
32. E. M. Wilfong, Y. Kogiso, S. Muthukrishnan, T. Kowatz, Y. Du, A. Bowie, J. H. Naismith, C. M. Hadad, E. J. Toone, T. L. Gustafson, *J. Am. Chem. Soc.* **2011**, *133*, 11515.
33. N. M. Ralbovsky, L. Halámková, K. Wall, C. Anderson-Hanley, I. K. Lednev, *J. Alzheimer's Disease* **2019**, *71*, 1351.
34. Y. Xu, X. Pan, H. Li, Q. Cao, F. Xu, J. Zhang, *Front. Psychiatry.* **2023**, *14*, 1112615, https://doi.org/10.3389/fpsyt.2023.1112615 (accessed: Mar 2023).
35. N. M. Ralbovsky, G. S. Fitzgerald, E. C. McNay, I. K. Lednev, *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2021**, *254*, 119603.
36. M. H. Baig, K. Ahmad, G. Rabbani, I. Choi, *Front Aging Neurosci.* **2018**, *10*, 21.
37. U. C. Müller, T. Deller, M. Korte, *Nat. Rev. Neurosci.* **2017**, *18*, 281.
38. R. Sengoku, *Neuropathology* **2020**, *40*, 22.
39. R. J. Castellani, R. K. Rolston, M. A. Smith, *Dis. Mon.* **2010**, *56*, 484.
40. Y. Seol, S. Ki, H. L. Ryu, S. Chung, J. Lee, H. Ryu, *Front Mol. Neurosci.* **2020**, *13*, 593724.
41. F. Han, J. Chen, A. Belkin-Rosen, Y. Gu, L. Luo, O. M. Buxton, X. Liu, the A. D. N. Initiative, *PLoS Biol.* **2021**, *19*, e3001233.
42. P. Slezak, R. Mueller, G. Kerckhofs, B. Bajoghli, A. Walter, *Bioimaging in Tissue Engineering and Regeneration: Advanced Microscopy and Preclinical Imaging*, 1st ed., Springer International Publishing, Cham **2023**.
43. M. A. Singh, D. Kumar, K. Mishra, *Raman Spectroscopy: Advances and Applications*, 1st ed., Springer Nature Singapore, Singapore **2024**.
44. H. Noothalapati, K. Iwasaki & T. Yamamoto, *Anal. Sci.* **2017**, *33*, 15.
45. F. Masia, A. Karuna, P. Borri, W. Langbein, *J. Raman Spectrosc.* **2015**, *46*, 727.

46. X. Zhao, K. Liu, K. Gao, W. Li, *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1.
47. J. M. Amigo, *Data fusion methodology and applications*, 1st ed., Elsevier Series in Data Handling in Science and Technology, Vol. 31, Elsevier **2019**.
48. E. W. Hislop, W. J. Tipping, K. Faulds, D. Graham, *Anal. Chem.* **2022**, *94*, 8899.
49. R. Michael, A. Lenferink, G. F. J. M. Vrensen, E. Gelpi, R. I. Barraquer, C. Otto, *Sci. Rep.* **2017**, *7*, 15603.
50. A. Haessler, N. Jung, M. Windbergs, *Anal. Chem.* **2023**, *95*, 17646.
51. M. Tainta, A. Iriondo, M. Ecay-Torres, A. Estanga, M. de Arriba, M. Barandiaran, M. Clerigue, M. Garcia-Sebastian, J. Villanua, A. Izagirre, J. Saldias, A. Aramburu, J. Taboada, J. Múgica, A. Barandiaran, A. Arrospide, J. Mar, P. Martinez-Lage, *Neurologia* **2022**, *39*, 781.
52. F. H. Duits, P. Martinez-Lage, C. Paquet, S. Engelborghs, A. Lleó, L. Hausner, J. L. Molinuevo, E. Stomrud, L. Farotti, I. H. G. B. Ramakers, M. Tsolaki, C. Skarsgård, R. Åstrand, A. Wallin, M. Vyhnalek, M. Holmber-Clausen, O. V. Forlenza, L. Ghezzi, M. Ingelsson, E. I. Hoff, G. Roks, A. De Mendonça, J. M. Papma, A. Izagirre, M. Taga, H. Struyfs, D. A. Alcolea, L. Frölich, M. Balasa, L. Minthon, et al., *Alzheimer's Dement.* **2016**, *12*, 154.
53. S. Guo, J. Popp, T. Bocklitz, *Nat. Protoc.* **2021**, *16*, 5426.
54. N. Mobaraki, J. M. Amigo, *Chemom. Intell. Lab. Syst.* **2018**, *172*, 174.
55. P. Eilers, H. Boelens, Leiden University Medical Centre Report **2005**, *1*, 5.
56. A. de Juan, in *Hyperspectral Imaging* (Ed: J. M. Amigo), Elsevier **2019**.
57. E. C. Muñoz, F. Gosetti, D. Ballabio, S. Andò, O. Gómez-Laserna, J. M. Amigo, E. Garzanti, *Microchem. J.* **2023**, *190*, 108655.
58. S. Wold, K. Esbensen, P. Geladi, *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37.
59. X. Jin, J. Han, in *Encyclopedia of Machine Learning* (Eds: C. Sammut, G. I. Webb), Springer US, Boston, MA **2010**.
60. R. Todeschini, D. Ballabio, V. Termopoli, V. Consonni, *Chemom. Intell. Lab. Syst.* **2024**, *251*, 105117.
61. A. de Juan, R. Tauler, *Anal. Chim. Acta* **2021**, *1145*, 59.
62. R. Tauler, *Chemom. Intell. Lab. Syst.* **1995**, *30*, 133.
63. *Hyperspectral Imaging* (Ed: J. M. Amigo), Elsevier **2019**.
64. E. I. Pearce, A. Tomlinson, *Ophthalmic Physiol. Opt.* **2000**, *20*, 306.
65. E. Ryzhikova, N. M. Ralbovsky, V. Sikirzhytski, O. Kazakov, L. Halamkova, J. Quinn, E. A. Zimmerman, I. K. Lednev, *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2021**, *248*, 119188.
66. R. Ossenkoppele, G. Salvadó, S. Janelidze, A. Pichet Binette, D. Bali, L. Karlsson, S. Palmqvist, N. Mattsson-Carlgren, E. Stomrud, J. Therriault, N. Rahmouni, P. Rosa-Neto, E. M. Coomans, E. van de Giessen, W. M. van der Flier, C. E. Teunissen, E. M. Jonaitis, S. C. Johnson, S. Villeneuve, T. L. S. Benzinger, S. E. Schindler, R. J. Bateman, J. D. Doecke, V. Doré, A. Feizpour, C. L. Masters, C. Rowe, H. J. Wiste, R. C. Petersen, C. R. Jack, et al., P.-A. R. Group, *Nat. Aging* **2025**, *5*, 1.
67. N. J. Ashton, A. Puig-Pijoan, M. Milà-Alomà, A. Fernández-Lebrero, G. García-Escobar, F. González-Ortiz, P. R. Kac, W. S. Brum, A. L. Benedet, J. Lantero-Rodriguez, T. A. Day, J. Vanbrabant, E. Stoops, E. Vanmechelen, G. Triana-Baltzer, S. Moughadam, H. Kolb, P. Ortiz-Romero, T. K. Karikari, C. Minguillon, J. J. Hernández Sánchez, I. Navalpotro-Gómez, O. Grau-Rivera, R. María Manero, V. Puente-Periz, R. de la Torre, J. Roquer, J. L. Dage, H. Zetterberg, K. Blennow, et al., *Alzheimer's Dement.* **2022**, *19*, 1913.
68. N. R. Barthélemy, G. Salvadó, S. E. Schindler, Y. He, S. Janelidze, L. E. Collij, B. Saef, R. L. Henson, C. D. Chen, B. A. Gordon, Y. Li, R. La Joie, T. L. S. Benzinger, J. C. Morris, N. Mattsson-Carlgren, S. Palmqvist, R. Ossenkoppele, G. D. Rabinovici, E. Stomrud, R. J. Bateman, O. Hansson, *Nat. Med.* **2024**, *30*, 1085.
69. S. Palmqvist, S. Janelidze, Y. T. Quiroz, H. Zetterberg, F. Lopera, E. Stomrud, Y. Su, Y. Chen, G. E. Serrano, A. Leuzy, N. Mattsson-Carlgren, O. Strandberg, R. Smith, A. Villegas, D. Sepulveda-Falla, X. Chai, N. K. Proctor, T. G. Beach, K. Blennow, J. L. Dage, E. M. Reiman, O. Hansson, *J. Am. Med. Assoc.* **2020**, *324*, 772.
70. D. Mrđenović, B. F. Combes, R. Ni, R. Zenobi, N. Kumar, *ACS Chem. Neurosci.* **2024**, *15*, 78.
71. K. Sofińska, S. Seweryn, K. Skirlińska-Nosek, J. Barbasz, E. Lipiec, *Nanoscale* **2024**, *16*, 5294.

How to cite this article: E. Lopez, G. Gorla, J. Etxebarria-Elezgarai, J. Aramendia, K. Castro, M. García-Sebastián, P. Martínez-Lage, A. Seifert, J. M. Amigo, *VIEW.* **2025**, 20250034.
https://doi.org/10.1002/VIW.20250034

Appendix B

Supplementary information

B.1 Patient information

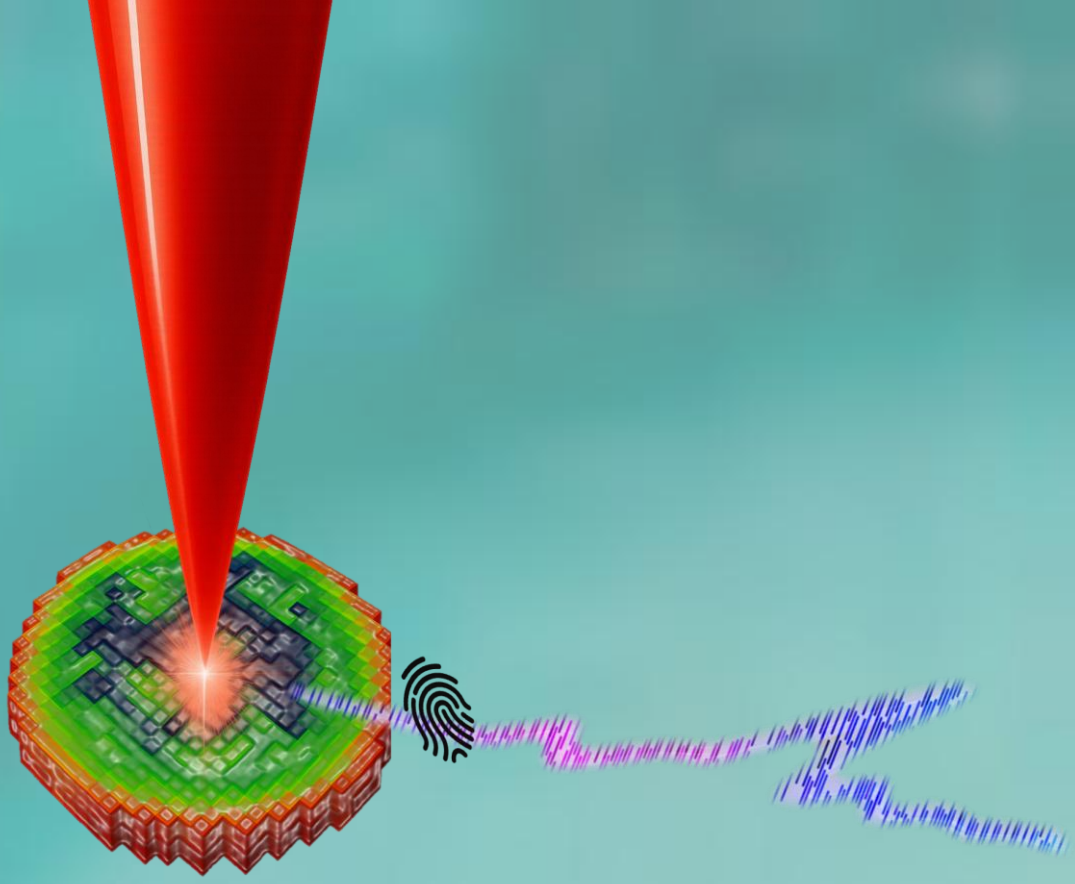
Tables B1.1 and B1.2 summarise the demographic and clinical characteristics of the study participants whose bio-fluid samples were analyzed in this thesis. Table B1.1 lists the cohorts used for sample preparation and Raman and FTIR experiments measurement; Table B1.2 details the subset used for the measurements and subsequent analyses from this PhD thesis.

For each group, the tables report mean of age, sex distribution, mini-mental state examination (MMSE) score, amyloid status, and, where available, tau protein status and ApoE ε4 genotype carrier frequency. These variables provide context for interpreting the spectroscopic findings and for comparing the present cohorts with those in other biomarker studies.

Gender	Education (years)	Age (years)	Clinical diagnosis	Alzheimer's diagnosis	Aβ42 (pg/ml)	p-Tau (pg/ml)	t-Tau (pg/ml)	Aβ40 (pg/ml)	Aβ42/Aβ40	ApoE genotype	ApoE carrier
Female	10	66	NORMAL	PRECLIN	285,49	80	622,44	6619,08	0,043	3/3	0
Female	17	73	NORMAL	PRECLIN	345,49	112	795,79	7985,52	0,043	3/2	0
Female	8	73	NORMAL	PRECLIN	356,70	72	598,23	7602,42	0,047	3/4	1
Female	16	68	NORMAL	PRECLIN	380,39	63	585,40	7950,26	0,048	4/4	1
Male	15	67	NORMAL	PRECLIN	239,59	51	355,91	4795,29	0,050	3/2	0
Male	9	58	NORMAL	PRECLIN	346,91	54	450,55	6568,08	0,053	3/4	1
Female	6	63	NORMAL	PRECLIN	405,55	102	689,25	7148,79	0,057	3/2	0
Female	16	60	NORMAL	PRECLIN	201,93	31	238,73	3495,24	0,058	3/4	1
Male	17	61	NORMAL	PRECLIN	324,14	41	260,55	5104,81	0,063	3/4	1
Female	16	75	NORMAL	PRECLIN	490,79	67	479,24	7609,92	0,064	3/3	0
Male	11	70	NORMAL	PRECLIN	376,28	39	258,96	5635,30	0,067	3/4	1
Male	15	62	NORMAL	PRECLIN	431,28	50	422,98	6418,63	0,067	3/4	1
Female	10	67	NORMAL	PRECLIN	390,18	55	363,61	5757,64	0,068	3/4	1
Male	14	62	NORMAL	PRECLIN	426,27	39	267,27	6040,11	0,071	3/4	1
Male	17	56	NORMAL	PRECLIN	362,53	29	250,97	4801,13	0,076	3/3	0
Female	17	66	NORMAL	PRECLIN	445,37	49	356,85	5813,14	0,077	3/4	1
Female	20	60	NORMAL	PRECLIN	489,94	38	355,27	6333,79	0,077	3/2	0
Male	8	67	NORMAL	PRECLIN	710,20	73	487,79	9043,03	0,079	3/3	0
Male	16	68	NORMAL	PRECLIN	681,43	69	553,85	8561,57	0,080	3/3	0
Male	12	67	NORMAL	PRECLIN	539,60	47	363,26	6487,89	0,083	3/3	0
Female	9	59	NORMAL	PRECLIN	335,05	25	157,44	4003,25	0,084	3/4	1
Male	23	64	NORMAL	NORMAL	552,93	46	370,72	6391,50	0,087	3/3	0
Female	8	78	NORMAL	NORMAL	417,63	33	196,43	4801,20	0,087	3/3	0
Male	12	61	NORMAL	NORMAL	471,42	32	289,97	5411,08	0,087	3/3	0
Male	15	57	NORMAL	NORMAL	499,88	26	165,45	5421,54	0,092	3/3	0
Male	11	56	NORMAL	NORMAL	611,95	30	213,79	6330,77	0,097	3/3	0
Male	11	53	NORMAL	NORMAL	454,37	32	246,30	4650,44	0,098	3/3	0
Male	16	65	NORMAL	NORMAL	454,65	33	235,40	4557,07	0,100	3/3	0
Female	22	64	NORMAL	NORMAL	833,04	51	338,86	8061,45	0,103	3/3	0
Female	12	59	NORMAL	NORMAL	527,75	37	255,14	5067,19	0,104	3/3	0
Female	12	61	NORMAL	NORMAL	658,93	42	272,14	6311,27	0,104	3/3	0
Female	13	53	NORMAL	NORMAL	696,62	48	334,48	6624,69	0,105	3/2	0
Female	14	57	NORMAL	NORMAL	752,61	40	262,24	7090,42	0,106	3/3	0
Male	18	64	NORMAL	NORMAL	600,15	46	302,56	5620,02	0,107	3/3	0
Female	16	62	NORMAL	NORMAL	611,81	36	246,66	5602,38	0,109	3/3	0
Male	13	66	NORMAL	NORMAL	587,73	38	279,42	5308,38	0,111	3/3	0
Male	20	57	NORMAL	NORMAL	850,04	51	377,18	7656,97	0,111	3/3	0
Female	17	49	NORMAL	NORMAL	480,85	24	168,68	4312,59	0,111	3/4	1
Female	17	60	NORMAL	NORMAL	740,53	49	414,80	6535,89	0,113	3/3	0
Male	18	46	NORMAL	NORMAL	835,11	47	328,75	6930,27	0,121	3/3	0

Gender	Education (years)	Age (years)	Clinical diagnosis	Alzheimer's diagnosis	Aβ42 (pg/ml)	p-Tau (pg/ml)	t-Tau (pg/ml)	Aβ42/Aβ40	ApoE genotype	ApoE carrier
Male	19	70	NORMAL	NORMAL	1519,00	16,58	211,30	0,01	3/3	0
Female	14	78	NORMAL	NORMAL	1163,00	12,13	159,80	0,01	3/3	0
Female	15	80	NORMAL	NORMAL	1158,00	13,72	174,70	0,01	3/3	0
Male	13	68	NORMAL	NORMAL	1194,00	14,49	175,10	0,01	3/2	0
Female	16	63	NORMAL	PRECLIN	329,40	12,54	184,40	0,04	3/4	1
Male	17	65	NORMAL	NORMAL	1086,00	13,90	172,10	0,01	3/3	0
Male	8	62	DCL	PRODROMAL	706,80	42,41	443,30	0,06	3/2	0
Female	10	74	NORMAL	PRECLIN	586,40	15,88	189,80	0,03	3/4	1
Female	15	65	NORMAL	NORMAL	1574,00	17,51	213,60	0,01	3/3	0
Female	12	64	DCL	PRODROMAL	538,20	15,40	184,40	0,03	3/3	0
Female	11	60	NORMAL	NORMAL	1461,00	15,20	194,10	0,01	3/3	0
Male	16	76	DCL	PRODROMAL	531,00	16,54	193,40	0,03	3/3	0
Male	12	68	NORMAL	NORMAL	1355,00	14,48	190,10	0,01	3/3	0
Female	20	71	NORMAL	PRECLIN	288,30	20,45	212,40	0,07	3/4	1
Male	15	62	NORMAL	NORMAL	1088,00	11,49	152,20	0,01	3/2	0
Female	16	61	NORMAL	NORMAL	1166,00	14,20	179,90	0,01	3/3	0
Male	12	67	DCL	PRODROMAL	634,70	29,40	281,30	0,05	3/4	1
Male	9	72	NORMAL	PRECLIN	837,70	27,52	294,10	0,03	3/3	0
Female	8	58	NORMAL	NORMAL	1431,00	21,21	258,00	0,01	3/3	0
Male	9	60	NORMAL	PRECLIN	661,00	30,00	316,00	0,05	3/4	1
Male	14	66	NORMAL	PRECLIN	558,00	27,00	312,00	0,05	3/4	1
Male	8	69	NORMAL	NORMAL	1416,00	15,04	181,50	0,01	3/3	0
Male	12	64	NORMAL	NORMAL	1186,00	13,01	161,30	0,01	3/3	0
Female	10	61	NORMAL	NORMAL	1298,00	9,52	126,80	0,01	3/3	0
Female	10	69	QSM	PRECLIN	485,30	40,90	378,40	0,08	3/3	0
Female	9	63	NORMAL	NORMAL	1508,00	20,11	255,70	0,01	3/3	0
Female	8	59	NORMAL	NORMAL	1238,00	13,65	170,40	0,01	3/3	0
Male	20	74	NORMAL	NORMAL	1425,00	18,18	237,10	0,01	3/3	0
Female	12	66	NORMAL	NORMAL	1133,00	13,12	171,90	0,01	3/4	1
Male	18	59	NORMAL	NORMAL	1220,00	11,38	154,60	0,01	3/2	0
Female	15	64	DCL	PRODROMAL	465,50	24,13	290,60	0,05	4/4	1
Female	15	62	NORMAL	PRECLIN	554,40	26,34	276,10	0,05	3/3	0
Female	8	76	QSM	PRECLIN	659,60	35,07	368,20	0,05	3/4	1
Female	8	66	NORMAL	NORMAL	1281,00	15,77	201,10	0,01	3/4	1
Female	8	77	NORMAL	PRECLIN	350,00	37,00	342,00	0,11	3/4	1
Male	15	69	NORMAL	PRECLIN	311,20	21,30	234,50	0,07	3/2	0
Female	16	78	NORMAL	PRECLIN	920,00	32,00	340,00	0,03	3/3	0
Female	15	58	NORMAL	PRECLIN	685,60	19,81	242,40	0,03	3/3	0
Female	6	66	NORMAL	PRECLIN	567,00	51,00	520,00	0,09	3/2	0
Female	6	71	DCL	PRODROMAL	528,20	25,24	243,80	0,05	3/2	0
Male	12	68	NORMAL	PRECLIN	501,20	38,40	285,00	0,08	3/4	1
Female	4	91	DCL	PRODROMAL	660,90	32,65	326,00	0,05	3/3	0
Female	8	81	DCL	PRODROMAL	565,70	35,09	379,40	0,06	3/3	0
Female	10	76	DCL	PRODROMAL	234,20	25,46	235,80	0,11	3/4	1
Female	10	84	DCL	PRODROMAL	469,40	32,31	338,20	0,07	3/3	0
Male	8	78	DCL	PRODROMAL	481,20	20,76	226,70	0,04	3/3	0
Male	7	68	DCL	PRODROMAL	555,60	24,37	271,00	0,04	3/3	0
Female	8	73	DCL	PRODROMAL	425,40	22,64	206,60	0,05	3/3	0
Female	6	79	DCL	PRODROMAL	280,10	10,28	126,60	0,04	3/3	0
Female	8	68	DCL	PRODROMAL	569,60	14,51	170,00	0,03	-	-
Male	21	74	DCL	PRODROMAL	297,80	11,90	129,20	0,04	3/3	0

B1



eman ta zabal zazu

Universidad del País Vasco

Euskal Herriko Unibertsitatea

CIC nanoGUNE

MEMBER OF BASQUE RESEARCH & TECHNOLOGY ALLIANCE