Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 328 (2025) 125474

Contents lists available at ScienceDirect

Spectrochimica Acta Part A:
Molecular and Biomolecular Spectroscopy

journal homepage: www.journals.elsevier.com/spectrochimica-acta-part-a-molecular-and-biomolecular-spectroscopy

Rapid diagnosis and severity scale of post-COVID condition using advanced spectroscopy

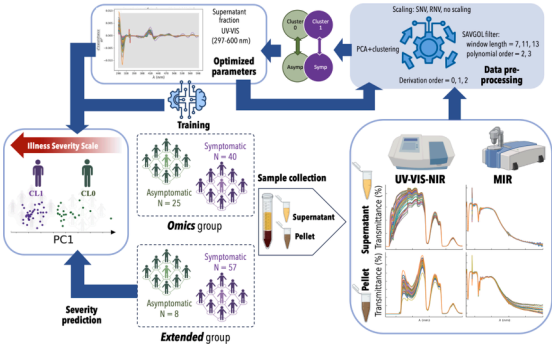
Paula Antelo-Riveiro [a,b], Manuel Vázquez [c], María Jesús Domínguez-Santalla [d], Emilio Rodríguez-Ruiz [e,f,g], Ángel Piñeiro [b,*], Rebeca Garcia-Fandino [a,*]

[a] Department of Organic Chemistry, Center for Research in Biological Chemistry and Molecular Materials, Santiago de Compostela University, CIQUS, Spain
[b] Soft Matter & Molecular Biophysics Group, Department of Applied Physics, Faculty of Physics, University of Santiago de Compostela, Spain
[c] Department of Analytical Chemistry, Faculty of Veterinary, Campus Terra, University of Santiago de Compostela, 27002 Lugo, Spain
[d] Internal Medicine Department, University Clinic Hospital of Santiago de Compostela (CHUS), Galician Public Health System (SERGAS), Santiago de Compostela, Spain
[e] Intensive Care Medicine Department, University Clinic Hospital of Santiago de Compostela (CHUS), Galician Public Health System (SERGAS), Santiago de Compostela, Spain
[f] Simulation, Life Support & Intensive Care Research Unit of Santiago de Compostela (SICRUS), Health Research Institute of Santiago de Compostela (IDIS), Santiago de Compostela, Spain
[g] CLINURSID Research Group, University of Santiago de Compostela, Santiago de Compostela, Spain

HIGHLIGHTS

- UV–VIS spectroscopy enables rapid, cost-effective PCC severity diagnosis.
- Machine learning enhances spectral analysis, distinguishing symptomatic PCC patients.
- The 297–600 nm range correlates strongly with biochemical markers of PCC.
- Spectroscopy demonstrates high agreement with clinical and proteomic data.
- Promising tool for portable, point-of-care PCC diagnostics.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Post-COVID Condition (PCC)
UV–VIS-NIR-MIR spectroscopy
Machine learning
Rapid diagnosis
Biochemical monitoring

ABSTRACT

The COVID-19 pandemic has resulted in a persistent health challenge known as Post-COVID Condition (PCC), characterized by symptoms lasting at least three months after the initial SARS-CoV-2 infection and potentially persisting for several years. While studies on PCC using lipidomics and proteomics have been conducted, these methods are costly and time-consuming. The comprehensive analysis of UV–VIS–NIR–MIR spectroscopy is explored here as an alternative for the rapid and cheap diagnosis and quantification of the severity of PCC. Blood samples from 65 PCC patients, previously analyzed in lipidomic and proteomic studies, along with samples from 65 new patients, were examined to develop a model that quantifies the severity of PCC based solely on spectrophotometric data. Significant spectral variability was observed in the UV–VIS region, particularly between 297 and 600 nm, correlating strongly with patient symptoms. Unsupervised clustering algorithms in this spectral

* Corresponding authors.
E-mail addresses: angel.pineiro@usc.es (Á. Piñeiro), rebeca.garcia.fandino@usc.es (R. Garcia-Fandino).

https://doi.org/10.1016/j.saa.2024.125474
Received 27 August 2024; Received in revised form 17 November 2024; Accepted 20 November 2024
Available online 24 November 2024

region effectively differentiated between asymptomatic and symptomatic patients, achieving a Jaccard similarity score of 0.667 when compared with clinical symptom classifications. Comparative analysis with proteomic and lipidomic studies indicated that UV–VIS spectroscopy captures clinically relevant biochemical information. The results of the model developed in this work to quantify the severity of PCC demonstrated robustness with new patient data, underscoring the method's potential as a rapid, non-invasive, and cost-effective diagnostic tool. This study highlights the strengths of spectroscopic techniques, suggesting their suitability for widespread clinical application in diagnosing and monitoring PCC, and emphasizes the need for further refinement and integration of these methods into healthcare practice, particularly for their potential implementation in portable devices.

1. Introduction

The COVID-19 pandemic has left in its wake a complex and persistent health challenge known as Post-COVID Condition (PCC), also referred to as long-COVID or Post-Acute Sequelae of COVID-19 (PASC) [1–5]. As defined by the World Health Organization, PCC is characterized by symptoms that persist or develop at least three months after the initial SARS-CoV-2 infection, potentially lasting for years and significantly impacting patients' quality of life [6].

The prevalence and severity of PCC have become increasingly apparent as the pandemic has progressed. Studies indicate that 35 %–60 % of COVID-19 survivors experience symptoms such as fatigue and dyspnea in the months following infection, with variations depending on the follow-up period [7]. While estimates vary, it is widely believed that at least 10 % of people who have contracted COVID-19 globally suffer from PCC [8]. Recent *meta*-analyses have further revealed the chronic nature of this condition. Approximately 30 % of patients still experience post-COVID symptoms two years after the initial infection [9]. The most prevalent long-term symptoms include fatigue and cognitive disorders, followed by pain, psychological disturbances, and sleep problems. These findings highlight PCC's potential to cause prolonged disability and reduced quality of life, underscoring the urgent need for comprehensive research and improved diagnostic methods. Notably, PCC affects a broad demographic, including individuals aged 36–50 years who experienced only mild acute illness, suggesting that the long-term impact of COVID-19 extends beyond the severity of the initial infection [10].

To date, omics-based methods, such as proteomics and lipidomics, have played a crucial role in elucidating the complexities of the acute phase of COVID-19, offering detailed molecular insights into the immediate effects of the infection [11–20]. However, research focusing on the long-term changes related to PCC is considerably less developed, with limited identification of the specific lipidomic and proteomic disruptions responsible for the persistent symptoms [20–27]. Our group recently conducted separate lipidomic and proteomic studies on the same cohort of PCC patients, providing complementary insights into the molecular basis of this condition. The lipidomic analysis revealed significant changes in lipid subclasses, including increased lysophosphatidylglycerols and phosphatidylethanolamines, and decreased lysophosphatidylcholines [28]. On the other hand, proteomic investigations using SWATH-MS technology identified dysregulations in the complement system and thromboinflammatory pathways, with distinct profiles between symptomatic and asymptomatic PCC patients [29]. These findings contribute to the understanding of PCC and suggest potential biomarkers for its diagnosis and monitoring.

While these omics approaches provide valuable insights, their clinical implementation is hindered by the need for costly instrumentation, complex sample preparation, and specialized technical expertise. Consequently, there is a pressing need for more accessible methods to diagnose and monitor PCC in clinical settings.

Spectroscopic techniques have emerged as promising tools for the rapid and cost-effective analysis of biofluids in various medical applications [30]. These methods, particularly in the visible and near-infrared (VIS-NIR) range, offer several advantages in disease diagnosis and monitoring. They provide immediate information about the overall composition of biological samples, potentially revealing metabolic fingerprints associated with specific conditions [31,32]. The strengths of spectroscopic approaches lie in their relative affordability compared to omics techniques, their potential for implementation in portable devices, and their ability to offer non-invasive and rapid diagnostic solutions [33]. These characteristics enhance their accessibility and clinical utility, making them particularly attractive for widespread application in healthcare settings. When combined with advanced data analysis techniques, spectroscopic methods have the potential to offer robust diagnostic tools for complex conditions like PCC. Machine Learning (ML) algorithms are particularly valuable in this context, as they can detect subtle signal changes, potentially improving diagnostic accuracy [34,35]. These computational methods can analyze spectral features to differentiate between disease and control cases without necessarily identifying specific biomarkers. This offers a novel approach to disease detection.

In the specific context of COVID-19, recent studies have demonstrated the feasibility of various spectroscopic techniques for diagnosis [36–38]. When combined with chemometrics tools and ML algorithms, these approaches have shown promise in identifying SARS-CoV-2 infection, achieving notable levels of accuracy, sensitivity, and specificity [39]. The application of these techniques to PCC represents a logical extension of their use in acute COVID-19 diagnosis. In this study, the synergy between spectroscopy and ML algorithms was leveraged to develop a novel, efficient, and accessible method for PCC diagnosis and severity quantification, with the potential to transform the management of this complex condition. UV–VIS-NIR-MIR spectroscopy combined with ML was employed to analyze blood samples from PCC patients (Fig. 1). This approach aims to develop a predictive model and severity scale based solely on spectrophotometric data. The cohort includes 65 patients previously analyzed by lipidomic and proteomic studies, along with 65 additional new patients. The potential of this approach extends beyond the current study, offering possibilities for rapid, point-of-care diagnosis and monitoring of PCC. The ultimate aim of this study is to facilitate broader access to PCC diagnosis and potentially enable more timely and targeted interventions for those suffering from long-term effects of COVID-19.

2. Materials and methods

2.1. Recruitment of patients

The analyzed plasma samples from recovered COVID-19 patients (n = 130) come from a cohort of patients enrolled in a PCC clinical study for the medical monitoring of symptoms evolution (Table S1). 65 of these patients were previously evaluated in lipidomic [28] and proteomic [29] studies, referred to as the ***Omics group***. This group was analyzed to provide a comprehensive comparison across multiple analytical techniques and to leverage existing detailed data. The remaining 65 patients, categorized as the ***Extended group***, were used to assess findings and ensure the robustness of the analysis across a broader population. This division into two groups—Omics for detailed comparison with molecular data, and Extended for model validation—was designed to assess both the model's development and its generalizability

across a broader PCC patient population. All the patients had been infected with SARS-CoV-2 virus, from early 2020 to middle 2021 and diagnosed by RT-PCR assay targeting viral RNA from nasal swab samples and/or by IgG detection assay in blood. Blood samples were collected at different stages after the first symptoms, but only once for each patient. According to the PCC monitoring, every patient has a medical evaluation of the symptoms 90 days and 9 months after the infection. Moreover, another evaluation is set for the day of the blood draw. All the samples were collected within the same geographical region (Galicia, Spain), following standardized procedures for temperature, clotting tube, and clotting time. The blood samples were obtained through the CHUS Biobank [40] under a protocol approved by the Santiago-Lugo Committee of Ethics and Clinical Research (2021/079). The methods were carried out in accordance with the approved guidelines. The CHUS Biobank complies with the quality management, traceability and biosecurity, set out in Spanish Law 14/2007 of Biomedical Research and in Royal Decree 1716/2011. The study was conducted according to the Declaration of Helsinki [41]. All study subjects provided written informed consent.

The patients were divided into several categories based on their clinical characteristics and timing of their symptoms, as detailed in Table S1. These categories include symptoms experienced on the day of blood extraction (*Ext*), 90 days after diagnosis (*90d*), and 9 months after diagnosis (*9m*). Additional categories include combined symptomatic status across these three time points (*Ext* + *90d* + *9m*), symptoms during the COVID infection (*COVID*), hospitalization status during the infection (*Hosp*), previous medication (*Prev medication*), previous diseases (*Prev diseases*), medication during COVID infection (*COVID medication*), and overall COVID status (*COVID status*).

2.2. Blood sample processing

Blood samples were collected from each patient, with 5–8 mL of whole blood drawn into EDTA vacutainers. After collection, these samples were centrifuged at 1600 rpm for 10 min at a consistent temperature of 23 °C. The centrifugation process effectively separated the blood into two main fractions. The top layer, predominantly plasma (supernatant), was meticulously extracted for further analysis. The remaining portion, consisting of both the intermediate and the lower layers, was collectively processed as a separate component. This residual portion (pellet), rich in blood cells and other elements, was then preserved at −80 °C alongside the plasma (supernatant) samples, until they were analyzed.

2.3. Spectroscopic measurement procedure for supernatant and pellet samples

The UV–VIS–NIR transmittance spectra of each plasma sample were measured with Spectrophotometer Jasco V670 (Jasco Inc., Hachioji, Tokyo, Japan), employing a cuvette made of Quartz Suprasil 300 with a 1 mm light path (Hellma GmbH & Co. KG, Mulheim, Germany). The transmittance spectra were acquired at 1 nm intervals in the 190–2500 nm range, giving a total of 2310 wavelength channels. The UV (ultraviolet) region corresponds to the 190–340 nm range, VIS (visible) to 340–780 nm and NIR (near infrared) to 780–2500 nm. The transmittance spectra in the MIR (mid-infrared) region were also recorded with Spectrophotometer Jasco FTIR-4X coupled with ATR PRO 4X (Jasco Inc., Hachioji, Tokyo, Japan). The transmittance spectra in the MIR region were measured in the 399–4000 cm^{-1} range (2500–25000

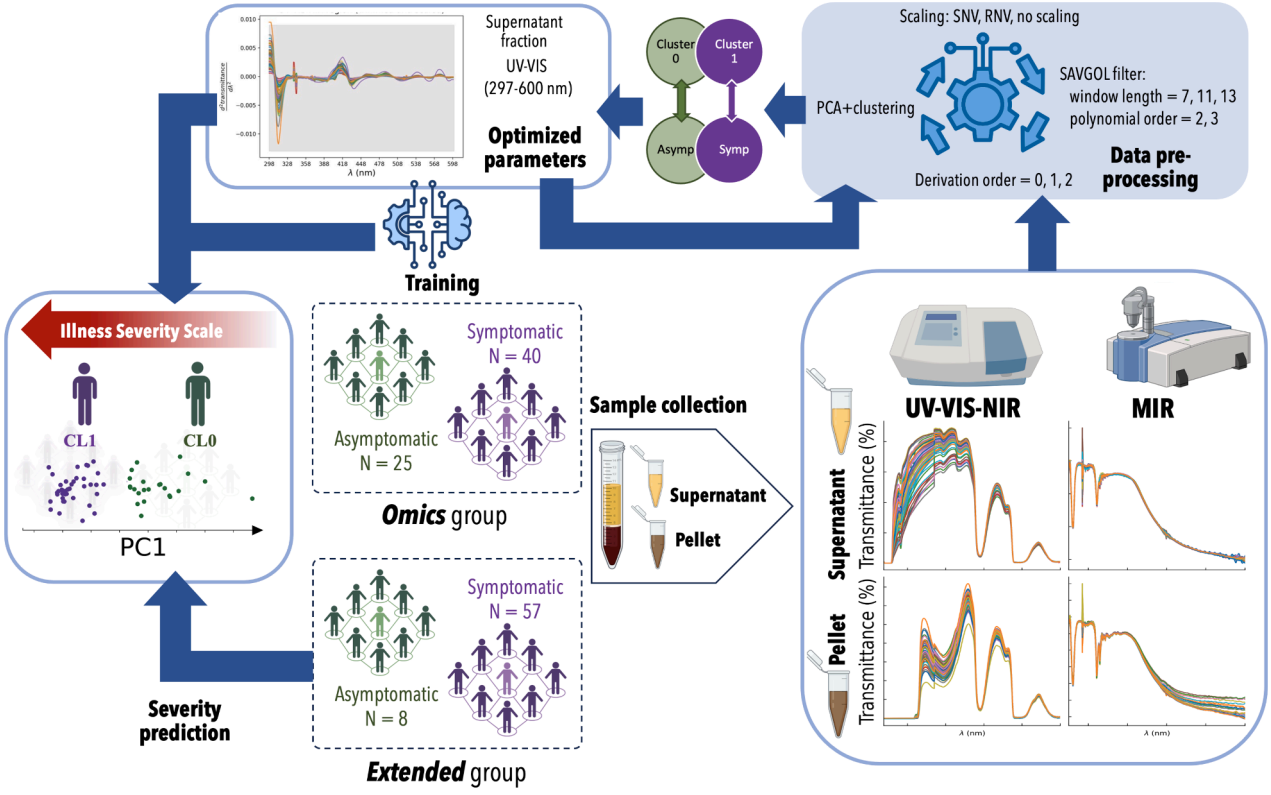


Fig. 1. Schematic of the spectroscopic workflow and optimization framework for diagnosing and assessing PCC severity. The figure outlines the process from patient cohort division, with 65 patients from previous proteomics and lipidomics studies (***Omics group***) and 65 new patients (***Extended group***), to the collection and processing of blood samples into supernatant and pellet fractions. UV–VIS–NIR–MIR spectroscopy is applied to both fractions, and various spectral preprocessing techniques are used, including trimming, scatter correction, and second-order derivation, with a focus on the optimized spectral range of 297–600 nm. The optimization process, driven by the highest Jaccard score, ensures the best alignment between clinical classification (symptomatic *vs.* asymptomatic) and spectral clustering. The figure also details the clustering of patients, visualized as two clusters (*CL0* and *CL1*), and the iterative optimization cycle that selects the spectral parameters most correlated with clinical severity. Finally, the diagnostic model's validation across the extended cohort is highlighted.

nm) at 0.964 cm^{-1} intervals (3738 wavelengths). Each spectrum was measured twice, so a total of 520 spectra were obtained in each region. Two aliquots of about 2 mL from each patient were taken for this analysis. Before the measurement, each aliquot was heated to 36 °C for 10 min. A total amount of 350 μL was employed from each plasma supernatant sample (300 μL and 50 μL to measure in the UV–VIS-NIR and the MIR region, respectively). Transmittance spectra of pellet samples were also measured following the same protocol but diluting in water to 1:1 in that case to facilitate the cuvette cleaning.

2.4. Spectra preprocessing

The Spectra Manager™ II software (Jasco Inc., Hachioji, Tokyo, Japan) was first used to read all the spectra files and for baseline correction. Further data pretreatment and analysis were performed employing own developed codes in Python, based on the following packages: Pandas [42], Numpy [43], Scipy [44], Scikit-learn [45], Matplotlib [46] and Nippy [47]. The pretreatment of the spectra consisted first in the calculation of the average of the two measurements obtained for each sample and the removal of outliers according to the Tukey test (patients S_123, S_103, S_25, S_9 and P_94), remaining a total of 126 spectra for the supernatant fraction (_S) and 129 for the pellet fraction (_P).

The preprocessing of the spectra was performed making use of the Nippy package, involving the following steps: trimming, scatter correction, Savitzky-Golay filter, and second order derivation. Trimming and scaling the spectra was optimized according to the analysis performance (favoring the relation between spectra and clinical preclassification, see Section 2.5.), eventually selecting the 297–600 nm region, which corresponds to the UV–VIS region. The scaling method selected was Robust Normal Variate (RNV), a variation of Standard Normal Variate (SNV) less sensitive to outliers. This scaling procedure involves normalization on a per-spectrum basis, removing the mean and scaling according to the interquartile range (IQR). These type of scaling methods are widely used in infrared spectroscopy [48]. Their objective is to minimize scattering associated to changes in particle-size distribution, that may cause additive or multiplicative effects, the alteration of the spectral profile, and may contribute to the variability of the raw data set [49]. A Savitzky-Golay filter [50] with a polynomial order of 3 and a filter window of 7 wavelength channels, was applied for the purpose of smoothing the data. Finally, the second derivative of the preprocessed spectra was calculated in order to highlight the differences between samples. The parameters of the different pretreatment steps (the trimming range, the scaling method, the window filter, the polynomial order and the derivative order) were optimized to favor the relation between the spectra classification of patients and the clinical data.

2.5. Data mining, selection of optimal spectral region and preprocessing parameters

The analysis of spectral data aimed to identify regions significantly associated with clinical outcomes. Given the high dimensionality of the spectral data relative to the sample size, principal component analysis (PCA) was first used for dimensionality reduction and pattern identification on the preprocessed spectra. This was followed by a clustering analysis, identifying two clusters that were later compared to clinical classification of patients. This approach enabled the identification of the spectral region that best correlated with clinical classifications, ensuring robust and meaningful associations between spectral features and patient outcomes. The explained variance for each component as well as the wavelength channel contributions to each principal component were determined. Then the samples were plotted in the space defined by the first two eigenvectors, and an unsupervised classification was performed on these coordinates. The K-means algorithm was used for this aim due to its superior performance (higher average silhouette score) compared to other tested methods such as Gaussian Mixture or DBSCAN models. The same silhouette score criterion was also employed to choose the optimal number of clusters, which was found to be two for all the proposed spectral regions and preprocessing parameters considered.

The complete analysis (trimming or selection of the spectral range, scatter correction scaling, Savitzky Golay filtering, derivation, PCA and clustering) was repeated for different spectral regions and preprocessing parameters. Spectral regions were selected first by considering the different transmittance peaks and then manually modifying their limits narrowing and widening the intervals, uprising new proposed spectral ranges. The preprocessing parameters considered included: SNV, RNV and no-scaling for scatter correction; polynomial order of 2 and 3 and window length of 7, 13, and 15 for the Savitzky Golay filter; and order of 0, 1 and 2 for spectra derivation. Each spectral region considered was submitted to all combinations of the considered preprocessing parameters, and subsequently subjected to PCA and clustering analysis. The selection of the spectral region and preprocessing parameters considered for the analysis were optimized by comparing the clusters classification (*CL0* and *CL1*) with a first general classification based on clinical data using the category *Ext* + *90d* + *9m* (asymptomatic and symptomatic) by means of the similarity Jaccard score, J (Eq. (1)). This coefficient is a function of the elements of the contingency table for clinical data classification and cluster assignation (in our case), where n_{Ai}, n_{Si} denote the number of patients assigned to the *symptomatic* or *asymptomatic* clinical groups and identified in cluster *i*.

$$\text{contingency table}\,(\text{clinical data} \times \text{clusters}) = \begin{pmatrix} n_{A0} & n_{S0} \\ n_{A1} & n_{S1} \end{pmatrix}$$

$$J = \frac{1}{2}\left[\frac{n_{A0}}{n_{A0}+n_{S0}+n_{A1}} + \frac{n_{S1}}{n_{S1}+n_{S0}+n_{A1}}\right] \quad (1)$$

The region and parameters that showed the best performance in terms of clustering quality and its correlation with clinical classifications were selected. Specifically, the 297–600 nm range was identified as optimal because it provided the highest Jaccard similarity score when comparing cluster classifications (*CL0* and *CL1*) with clinical classifications (*asymptomatic* and *symptomatic*). This region was thus selected for further analysis.

2.6. Statistical analysis

Clusters identification was followed by their characterization and relation to the clinical data available for each patient, aiming to identify the clinical information that best correlated with cluster separation. Clinical data were compared between clusters both individually (each clinical characteristic isolated) and by clinical categories (*ext*, *90d*, *9m*, *prev medication*, *COVID medication*, *prev diseases* and *covid status*, Table S1), in order to identify and evaluate significancy.

For the quantification of possible discrepancies between clusters different statistical tests were used. The non-parametric Mann-Whitney U [51] test was employed to compare pairs of distributions (one for each cluster) of number of clinical variables presented by each patient (two binomial distributions). The χ^2 test of independence and Barnard's test were employed to analyze the dependency of the number of patients that did and did not present a given clinical variable and belonged to each cluster, a 2 × 2 (cluster × clinical variable) contingency table. The χ^2 test is only valid for large sample sizes (all expected values in the contingency table ≥ 5), while Barnard's test is an exact test meant for lower number of samples, converging to the χ^2 test in the limit of large samples. For this reason, Barnard's test is always applied and χ^2 is only employed if the validation criterion is satisfied. The p-value < 0.05 criterion was used for statistical significance. Violin plots, color maps and bar diagrams were built for data visualization and clusters comparison. The Jaccard score was employed again to compare the clusters classification and clinical classifications using different criteria, in order to find the clinical criterion that best associates with clusters separation.

2.7. Comparative analysis with lipidomics and proteomics studies on the same patients and validation with an extended group

The results obtained from the previous analysis for the ***Omics group*** were compared with those from the two previous lipidomics [8] and proteomics studies [29] using the Jaccard score and contingency tables. Then, the spectral data from the 65 additional patients (the ***Extended group***) were employed to validate the model's predictive ability independently of the initial dataset. By introducing this second group, we aimed to test the robustness and accuracy of the model across a broader PCC patient population, thus enhancing its potential clinical applicability.

These data were transformed into the PCA space, subjected to cluster prediction, then the severity score was determined as the PC1 coordinate and finally the results were compared with clinical data to ensure the robustness and accuracy of our findings.

3. Results and discussion

3.1. Spectral characterization of UV–VIS–NIR and MIR regions in PCC blood samples from patients analyzed in proteomics and lipidomics studies

A comprehensive spectral analysis was conducted on both the supernatant and pellet fractions of blood samples obtained from 62 patients (3 patients in the ***Omics group*** were outliers). These samples, previously analyzed in lipidomics [28] and proteomics [29] studies, were subjected to detailed examination across the UV–VIS–NIR (190–2500 nm) and MIR (2500–25000 nm) spectral regions (Fig. 2). The UV–VIS–NIR region is particularly sensitive to electronic transitions in biomolecules, making it ideal for detecting changes in electronic states, while the MIR region offers detailed information on vibrational modes, enabling a thorough characterization of molecular structures and interactions.

Studying both supernatant and pellet fractions is useful for a comprehensive understanding of blood's molecular composition and

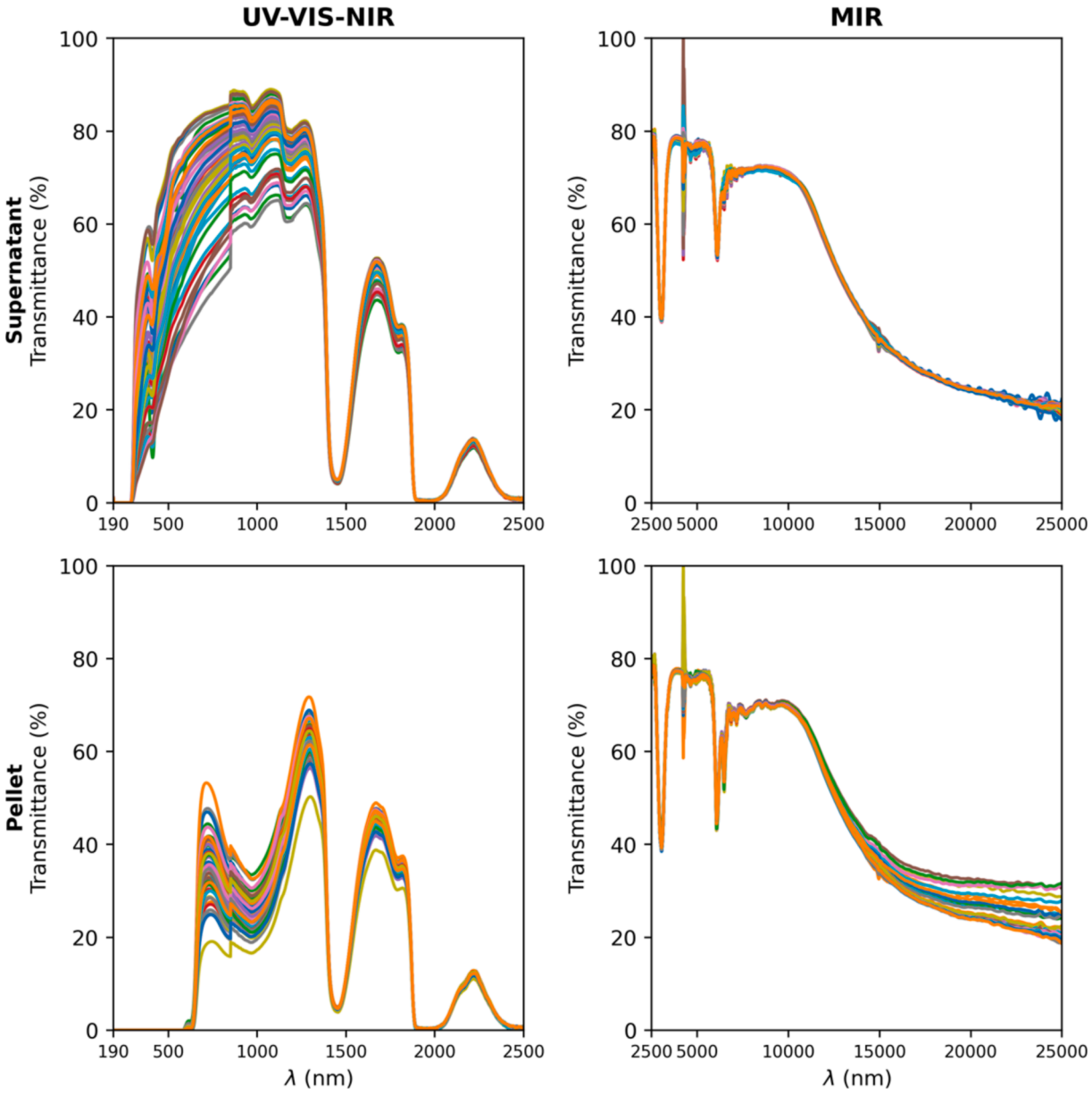


Fig. 2. Transmittance spectra of plasma samples from 62 patients of the Omics group, common to our previous lipidomics [28] and proteomics [29] studies in the UV–VIS–NIR and the MIR region using the supernatant (top) and pellet (down) fraction of the blood.

interactions, as they contain complementary information. The supernatant fraction, primarily containing plasma, provides insights into soluble proteins, lipids, and other circulating biomolecules. In contrast, the pellet fraction, composed of cellular components and other particulate matter, reveals information about the intracellular environment and cell-associated molecules. By analyzing the UV–VIS–NIR and MIR spectra of both fractions, distinctive patterns and absorptive features can be identified, which may indicate specific biochemical changes or disease states. Note the discontinuity at the end of the UV–VIS–NIR spectrum (2500 nm) where transmittance drops to 0 %, and the start of the MIR spectrum (2500 nm) where transmittance is 100 %. This discrepancy is due to the use of different equipment: in the NIR region, a quartz cuvette is used, which is suitable for UV but absorbs completely beyond ~2500 nm, causing the 0 % transmittance. In the MIR region, the sample is measured without a cuvette using an ATR accessory made of non-absorbing materials, resulting in 100 % transmittance. Additionally, a discontinuity is found in the spectra at 850 nm due to a grating change. Since the scaling procedure involves a second derivative, this could have negative effects, but since it is out of the selected region, no specific considerations are required.

3.1.1. Supernatant fraction analysis

The UV–VIS–NIR spectra of the supernatant fraction (Fig. 2) revealed significant spectral variability, particularly in the 297–1500 nm region, likely corresponding to electronic transitions in various biomolecules. In the UV region, a sharp decrease in transmittance is observed starting around 300 nm, corresponding to increased absorbance primarily attributed to aromatic amino acids such as tryptophan (~280 nm), tyrosine (~275 nm), and phenylalanine (~257 nm). This region also captures absorption from peptide bonds (below 240 nm) and nucleic acid bases (~260 nm) [52,53]. The variations in this spectral range likely reflect differences in protein concentrations or conformational changes among PCC patients, consistent with previous studies on serum protein analysis in this condition [54].

The VIS region (400–700 nm) exhibits a series of broad, overlapping bands with varying transmittance levels. Variability in this region may arise from differences in oxygenation status, hemoglobin concentration, or the presence of hemoglobin breakdown products such as bilirubin [55]. Significant absorption maxima within the 400–430 nm range – the Soret band – are characteristic of all hemoproteins. Oxyhemoglobin exhibits two closely positioned maxima at 540 and 578 nm, whereas deoxyhemoglobin shows a single absorption peak at 555 nm. These variations could be linked to altered respiratory function or hemolytic processes associated with PCC. Typical features of blood absorption spectra include bands at 350 nm, 412 nm, 576 nm, and 620 nm due to the absorption of iron porphyrin, which is also part of hemoglobin, along with a protein absorption band at 280 nm that overlaps with the luminescence band for plasma proteins (λ_{max} = 330 nm) [56]. The 400–500 nm range, on the other hand, could reflect the presence of various carotenoids [57].

In the NIR region (700–2300 nm), broad and overlapping bands are observed. This region is sensitive to the presence of water, lipids, and proteins, with broad and complex spectral features attributed to overlapping absorption bands of these molecules. Variability in this region may indicate changes in the hydration state, lipid composition, or protein secondary structure, reflecting metabolic alterations and inflammatory responses in PCC patients. Strong absorption bands at ~1450 nm and ~1940 nm, seen as transmittance minima, are characteristic of water [58]. Hemoglobin and its derivatives also contribute to the NIR spectrum, with distinct features observed between 700–1100 nm. Oxygenated and deoxygenated hemoglobin show different absorption patterns in this range, allowing for their differentiation [59,60]. Additionally, other specific absorption bands are centered at 2178 nm for albumin, 2240 nm for globulins, and 2200 nm for urea [61]. Lipid signatures are observed as C—H stretching overtones at ~1200 nm and ~1700 nm [62] and ~2300 nm [63], while weak absorption features around 1408 and 2326 nm may be associated with glucose [64].

The MIR region, typically spanning from 2500 to 25000 nm, provides rich molecular information in biological samples. This region is characterized by fundamental vibrational modes of molecules, offering high specificity for various biochemical components [56,65,66]. The most prominent are the amide bands: Amide A (~3030 nm) and Amide B (~3225 nm), which arise from N-H stretching vibrations; Amide I (~6060 nm), primarily due to C═O stretching; Amide II (~6450 nm), resulting from N—H bending and C—N stretching; and Amide III (~7690 nm), associated with C—N stretching and N—H bending [56]. Although there is some variability in absorption intensities, these differences are relatively subtle, suggesting that the overall biochemical composition of plasma remains similar across the cohort. Beyond 8000 nm, the MIR region shows a significant and consistent decrease in transmittance across all patients, continuing down to 25000 nm. This steady decline indicates increased absorption by various molecular bonds, such as C—H, O—H, and N—H, reflecting the complex biochemical composition of the plasma. The uniformity of this decrease among patients suggests common biochemical features within the PCC cohort. Key absorption features in this range include the stretching vibrations of aliphatic and aromatic hydrocarbons (C—H stretching) around 8500–9500 nm, and the O—H stretching bands of water and alcohols near 9600–11,500 nm. The fingerprint region (approximately 12500–15500 nm) also provides highly specific molecular information due to the presence of multiple bending and stretching vibrations of functional groups like C—O, C—N, and P═O.

3.1.2. Pellet fraction analysis

Following the examination of the plasma supernatant, the focus shifted to the spectral characteristics of the pellet fraction, which offered crucial insights into the cellular components of blood in PCC patients. This analysis highlighted the distinct biochemical environments between the pellet and the supernatant.

In the UV–VIS–NIR spectra (190–2500 nm), the pellet fraction shows dramatically lower transmittance compared to the supernatant across the entire range, with the difference being most pronounced from 190 nm to about 1300 nm (Fig. 2). Starting from 190 nm, the pellet exhibits very low transmittance (high absorbance) of around 0 %, while the supernatant shows much higher transmittance of 40–60 % in the same region. This stark contrast in the UV region (190–400 nm) likely reflects the high concentration of nucleic acids, aromatic amino acids, and other UV-absorbing cellular components in the pellet. The difference remains significant in the visible region (400–700 nm), where the pellet continues to show lower transmittance. This is likely due to the presence of hemoglobin and other cellular pigments in the pellet fraction, primarily from erythrocytes.

The pellet spectra also displayed a prominent dip in transmittance around 950–1100 nm, which was less evident in the supernatant. This feature could be also attributed to the strong absorption of hemoglobin in the NIR region, further emphasizing the cellular nature of the pellet fraction. Notably, in this region, both the supernatant and pellet exhibit significant variability between patients, reflecting individual differences in hemoglobin levels and other cellular components. Beyond 1500 nm, the spectra of the supernatant and pellet fractions become more similar, suggesting that the major differences in transmittance and variability are concentrated in the lower wavelengths. This similarity in the higher wavelength region indicates that the overall biochemical composition of plasma and cellular components shares common features in this part of the spectrum, despite the distinct characteristics observed at lower wavelengths.

The examination of the MIR region (2500–25000 nm) reveals that the spectra for both the supernatant and pellet fractions are quite similar and consistent across patients up to approximately 15000 nm. Important features, including a dip in transmittance around 3000 nm and another wide dip near 6247 nm, and a gradual decrease in transmittance from about 8000 nm onwards, are almost identical in both fractions.

However, notable differences emerge between supernatant and pellet fractions, particularly in the 15000–25000 nm range. The higher transmittance observed in the pellet fraction suggests lower absorption by these molecular species, which could be attributed to several factors. The intact cellular structures in the pellet might restrict certain low-frequency vibrations, reducing absorption. Additionally, the cellular environment could alter the vibrational behavior of large biomolecules compared to their state in solution, as found in the supernatant. While cellular components are more concentrated in the pellet, their structural arrangement might lead to less efficient absorption in this specific spectral range.

Notably, the pellet fraction displays greater variability between samples in the 15000–25000 nm range. This increased inter-patient variability could reflect differences in cellular composition, membrane integrity, or alterations in large biomolecular structures among patients with post-COVID condition. The diversity in cellular responses to the viral infection and subsequent recovery processes might manifest as spectral variations in this region. In contrast, the supernatant fraction shows lower transmittance (higher absorption) and more consistency between samples in this range. This could indicate that the molecular species responsible for absorption in this region, possibly including smaller biomolecules or metabolites, are more uniformly distributed in the cell-free fraction.

These spectral differences between supernatant and pellet fractions provide valuable insights into the biochemical alterations associated with the post-COVID condition. The variability observed in different regions of the spectra suggests that both plasma and cellular components exhibit sensitivity to individual metabolic changes, though in different ways. In some spectral regions, the supernatant shows greater variability, indicating sensitivity to metabolic alterations, while in other regions, the pellet exhibits more pronounced changes, reflecting uniform cellular alterations across patients. This highlights the complex nature of biochemical responses in PCC patients, with both plasma and cellular fractions contributing distinct and complementary information.

3.2. *Identification of spectral regions correlating with patient symptoms through comprehensive analysis and unsupervised clustering*

To identify spectral regions most closely associated with patient symptoms, a comprehensive analysis was conducted across the entire UV–VIS–NIR and MIR spectrum of both supernatant and pellet fractions. The spectral data underwent preprocessing, including Robust Normal Variate (RNV) scaling and Savitzky–Golay filtering, followed by the calculation of the second-order derivative. These steps significantly enhanced spectral features, accentuating peaks and troughs to reveal subtle differences between samples, particularly in regions associated with protein and metabolite absorptions.

A systematic approach was then employed (see Methods), applying PCA and unsupervised clustering algorithms. Specifically, a K-means clustering model was used to identify two distinct clusters in the space formed by the two first eigenvectors from the PCA. The resulting clusters were evaluated against patient symptom data and compared with the clinical pre-classification using the Jaccard score to determine the optimal spectral range and preprocessing parameters for analysis.

This comprehensive scanning and clustering process revealed that the region between 297 and 600 nm from the supernatant spectra showed the strongest correlation with patient symptoms (Fig. S1). This finding suggests that the electronic transitions of specific biomolecules within this spectral window may be particularly relevant to the underlying biochemical changes in post-COVID condition.

Clustering in the MIR region was less effective, yielding lower Jaccard similarity scores compared to the UV–VIS region. For instance, the full MIR range achieved a Jaccard score of 0.375, while a specific subregion (4100–4400 nm) achieved a slightly higher score of 0.381. These results indicate that, although the MIR region contains relevant biochemical information, its ability to differentiate between symptomatic and asymptomatic PCC patients is limited. Representative scatter plots illustrating clustering performance for these MIR regions are provided in the Supporting Information (Fig. S2). This analysis reinforces the diagnostic superiority of the UV–VIS region, particularly 297–600 nm, in aligning spectral data with clinical classifications and symptom severity.

Further analysis of this optimal spectral range provided additional insights. PCA applied to the preprocessed data within the 297–600 nm region showed that the first two principal components accounted for 85.57 % and 5.47 % of the variance, respectively. The transmittance values at each wavelength channel served as features, resulting in a large set of corelated data points that presented an overall transmittance pattern. Fig. 3a illustrates the transmittance spectra colored by their PC1 value, facilitating the identification of the proportionality between PC1 and transmittance in the 297–335 nm region. This spectral range had the highest contribution to PC1, making it essential for distinguishing between patients. This indicates that variations in transmittance within this wavelength range are the most critical for differentiating patient spectra. For PC2, the 390–460 nm region was the most significant contributor, characterized by changes in the curvature of local transmittance minima within that range (Fig. 3b). This means that differences in the spectral data within these wavelengths are particularly important for the second component's variability.

The scatter plot in Fig. 3c shows the distribution of samples in the space defined by the first two principal components. It reveals that most of the variation in the data is captured along the PC1 axis, which aligns with PC1 having the higher explained variance. The plot also shows the two clusters identified by the K-means algorithm: cluster 0 (*CL0*, green) and cluster 1 (*CL1*, purple). The clear separation along the PC1 axis indicates that this component plays a significant role in differentiating the clusters. The components of these clusters (23 patients in *CL0* and 39 in *CL1*) were compared with those obtained from a clinical pre-classification using the Jaccard score. This was done to determine the optimal spectral range for analysis, specifically the spectral range that maximizes the Jaccard score between the clusters obtained from the spectroscopy data and those obtained from the clinical information. As anticipated, given the variance distribution, the clusters are separated by a PC1 threshold around 0.0025, *i.e.* patients from cluster 1 have a PC1 value <0.0025 and patients from cluster 0 have a PC1 value >0.0025.

After identifying clusters based solely on spectral data and selecting the wavelength window that maximizes similarity with those derived from clinical information, a detailed characterization was conducted by analyzing the data of patients within each group. This analysis aimed to explore the relationship between the spectral clusters and the clinical manifestation of symptoms.

Qualitative and quantitative discrepancies between clusters are summarized in Fig. 4. The matrix representation of clinical characteristics reveals that *CL0* predominantly consists of asymptomatic patients, whereas *CL1* is mainly composed of symptomatic patients. This distinction is evident as patients in cluster 0 show fewer clinical symptoms across various categories compared to those in cluster 1. The upper subfigure in Fig. 4 presents a bar diagram quantifying this observation, showing the percentage of patients from each cluster presenting different clinical characteristics. The results indicate that the symptomatic cluster (*CL1*) has a higher proportion of patients with various clinical symptoms compared to the asymptomatic cluster (*CL0*). This consistent pattern across multiple clinical characteristics underscores the robustness of the clustering in the spectra in differentiating between asymptomatic and symptomatic patients. This separation indicates that the PC1 value is inversely correlated with the severity of symptoms, as patients with lower PC1 values exhibit a higher prevalence of symptoms. Specifically, a lower PC1 value corresponds to an increased number of clinical characteristics associated with PCC, suggesting that PC1 can be used as a quantitative measure of symptom severity.

A quantitative analysis was performed to assess the independence of the number of patients presenting each characteristic within the

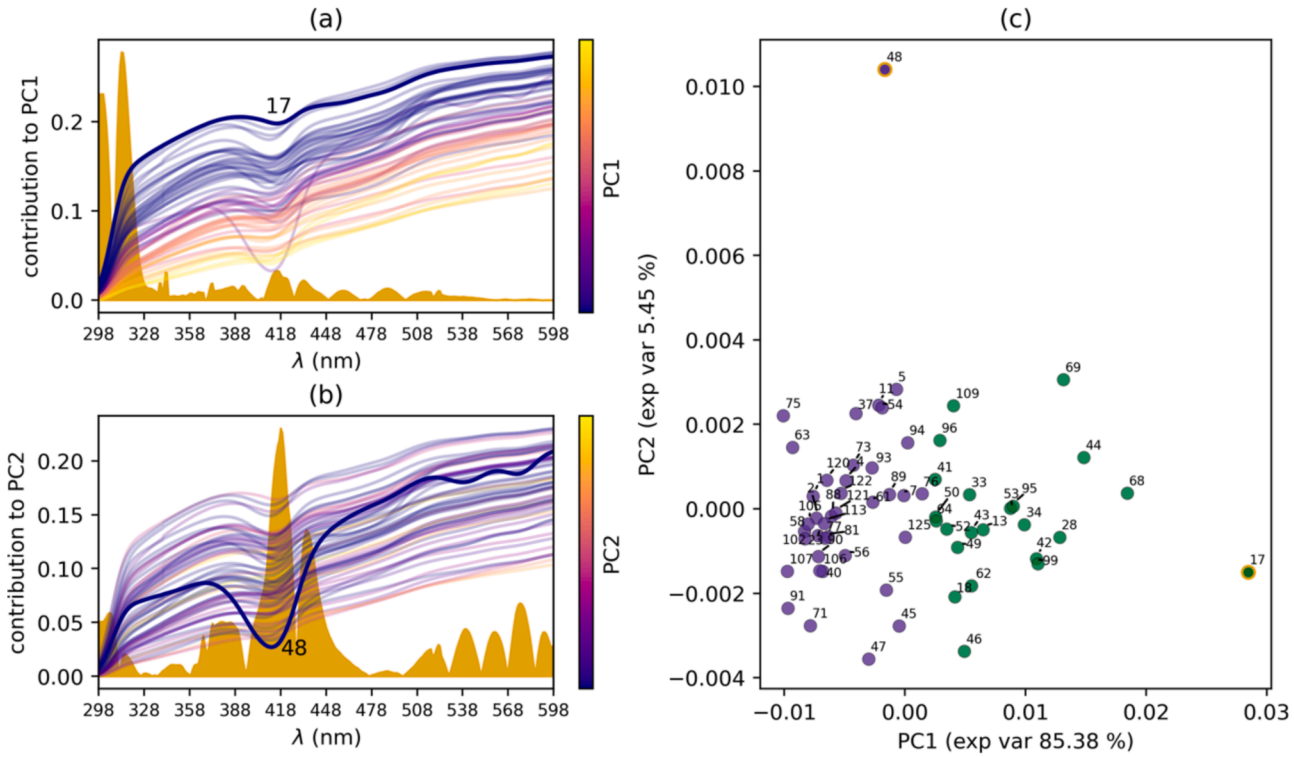


Fig. 3. Analysis of wavelength contributions to principal components PC1 and PC2, with a focus on differentiating spectral profiles of PCC patients. (a) Absolute values of each wavelength's contribution to PC1 are shown in yellow as a background overlay, helping to identify key wavelengths that influence this component. The original spectra are color-coded according to their PC1 values, with darker shades indicating higher PC1 contributions. Patient 17, with the highest PC1 value, is highlighted with a thicker line to underscore its distinct spectral signature within this component. (b) Similar analysis for PC2, with absolute values of each wavelength's contribution displayed in yellow as a background overlay. The original spectra are color-coded by PC2 values, emphasizing the relative impact of each wavelength on this component. Patient 48, which exhibits the highest PC2 value, is highlighted with a thicker line to indicate its unique profile in relation to PC2. (c) Scatter plot of samples in the PC1-PC2 space, where clusters CL0 and CL1 are represented in green and purple, respectively. Patients 17 and 48, exhibiting the highest PC1 and PC2 values, are outlined with a yellow border to highlight their distinct clustering positions and to emphasize the separation achieved between asymptomatic and symptomatic patients. The plot indicates that most variability is captured along the PC1 axis, reflecting symptom severity.

identified clusters, using Barnard's and $\chi 2$ tests where applicable. Out of the evaluated characteristics, 26 were found to be statistically significant (p-value < 0.05) for distinguishing between clusters using Barnard's test. Among these, 19 characteristics were also significant according to both Barnard's and $\chi 2$ tests. This statistical significance is visually represented in Fig. 4, with red asterisks indicating $\chi 2$ test significance and black asterisks indicating Barnard's test significance. Notably, the lowest p-values were associated with *Conventional oxygen therapy* and *Hospitalization* (Table S2). These characteristics pertain to the acute phase of infection rather than PCC symptoms or those present at the time of blood extraction. Other significant characteristics were categorized into groups related to symptoms on the extraction day (*ext*), symptoms observed 90 days (*90d*) and 9 months (*9m*) post-diagnosis, as well as medications administered during COVID infection (*Covid medication*). Interestingly, all patients classified in *CL0* (identified as primarily asymptomatic) reported no symptoms at the time of blood extraction, 90 days, and 9 months post-diagnosis, except for five misclassified patients (codes 13, 17, 46, 64 and 99) according to the spectral data.

To further explore the relationship between illness severity and spectral data, Fig. 5 presents violin plots of the PC1 coordinates of patients grouped by their hospitalization status, conventional oxygen therapy requirement, and ICU admission during COVID infection. The characteristics *Hospitalization* and *Conventional oxygen therapy* were selected due to their statistical significance in differentiating clusters, as shown in Table S2. Additionally, *Admission to ICU or similar* was included to further delineate patient subgroups. The violin plots clearly demonstrate that greater illness severity is associated with lower PC1 values, indicating that patients who required more intensive treatment during the acute phase of COVID-19 exhibit distinct spectral signatures.

Grouping clinical data into categories can provide additional insights not identified by individual analysis. Seven categories were analyzed: *Ext* (symptoms at the time of extraction), *90d* (symptoms 90 days post-diagnosis), *9m* (symptoms 9 months post-diagnosis), *previous medication*, *previous diseases*, *COVID medication*, and *COVID status* (Table S1). For each patient, the sum of symptoms or clinical characteristics belonging to each clinical group was calculated, resulting in seven distributions for each cluster. To evaluate the significance of the differences between clusters, the Mann-Whitney *U* test was used. Statistical significance in this context would indicate that the two populations belong to different distributions, or that they differ in their median values if equal variance is satisfied.

The p-values obtained for each clinical group (Fig. 6) suggest that six out of seven clinical groups have two significantly different distributions for each cluster. The *previous diseases* group, with a p-value of 0.025, is within the range of significance but presents the highest p-value among the significant groups. The *previous medication* group is the only one not showing significant differences between clusters, with a p-value of 0.065. However, medication administered during the COVID infection presented two clearly differentiated distributions for the two clusters, achieving a p-value of <0.005. This result may be attributed to the positive correlation between illness severity and the intensity of medication treatment. Symptoms on the extraction day showed clear significance, as did symptoms during the COVID infection, 90 days after diagnosis, and 9 months after diagnosis. These results suggest that blood plasma compounds, such as proteins and lipids, are altered during the

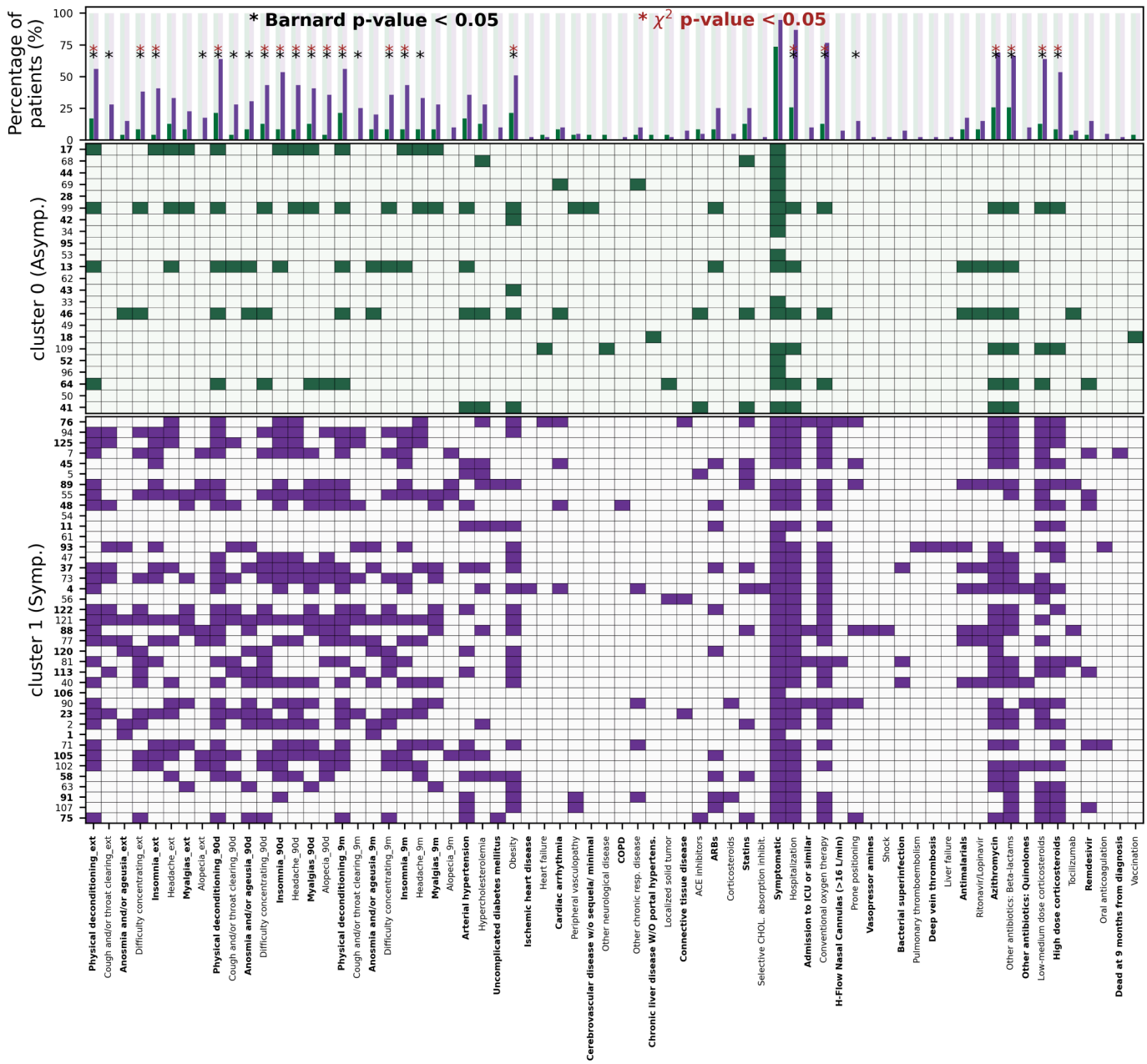


Fig. 4. *Top*: Bar diagram showing the percentage of patients from each cluster presenting various clinical characteristics. Green bars represent *CL0* (asymptomatic) and purple bars represent *CL1* (symptomatic). The asterisks (*) indicate statistical significance: black asterisks for Barnard's p-value < 0.05 and red asterisks for $\chi 2$ p-value < 0.05. *Bottom*: Matrix representation illustrating the severity scale based on UV–VIS data and its correlation with clinical characteristics. The data is derived from 62 patients previously analyzed in proteomics and lipidomics studies. The y-axis represents patient codes, ordered by PC1 values whereas the x-axis lists different clinical characteristics (details in Table S2). Each element in the matrix is color-coded to indicate the presence of a symptom: green for *CL0* and purple for *CL1*. Absence of color denotes the absence of the clinical characteristic. This visualization highlights the trend of symptom presence in each cluster, with *CL1* showing a higher prevalence of symptoms compared to *CL0*. The ordering by PC1 values emphasizes the correlation between PC1 and the severity of symptoms, with lower PC1 values indicating higher symptom prevalence.

COVID infection in a way that persists even 9 months after the disease onset. Furthermore, the medians of the distributions for cluster 0 are 0 for all clinical categories, and lower than those for cluster 1 across all clinical groups except for *prev medication*, reinforcing the identification of clusters *CL0* and CL1 with asymptomatic and symptomatic patients, respectively.

To complete the cluster characterization, an additional analysis using the Jaccard and ϕ similarity scores was performed to identify the clinical category best associated with cluster separation. Different clinical categories from Table S1 were selected, along with an *oxygen* group (separating patients who received conventional oxygen therapy during the COVID infection from those who did not) and a *significant* group (separating patients who presented more than one clinical characteristic from those selected as significant in Fig. 4 or Table S3 from those who did not). Clinical groups are ordered by ascending Jaccard similarity score (Fig. 7). The *hosp*, *oxygen*, *ext* + *90d* + *9m*, and *significant* groups present the highest similarity scores (both Jaccard and ϕ), indicating the strongest association with cluster separation from spectral data. The *ext* + *90d* + *9m* criterion, used to optimize the spectral range for analysis, was expected to have a high Jaccard score. Conventional oxygen therapy and hospitalization during the COVID infection, identified as the two most significant clinical characteristics for differentiating clusters using

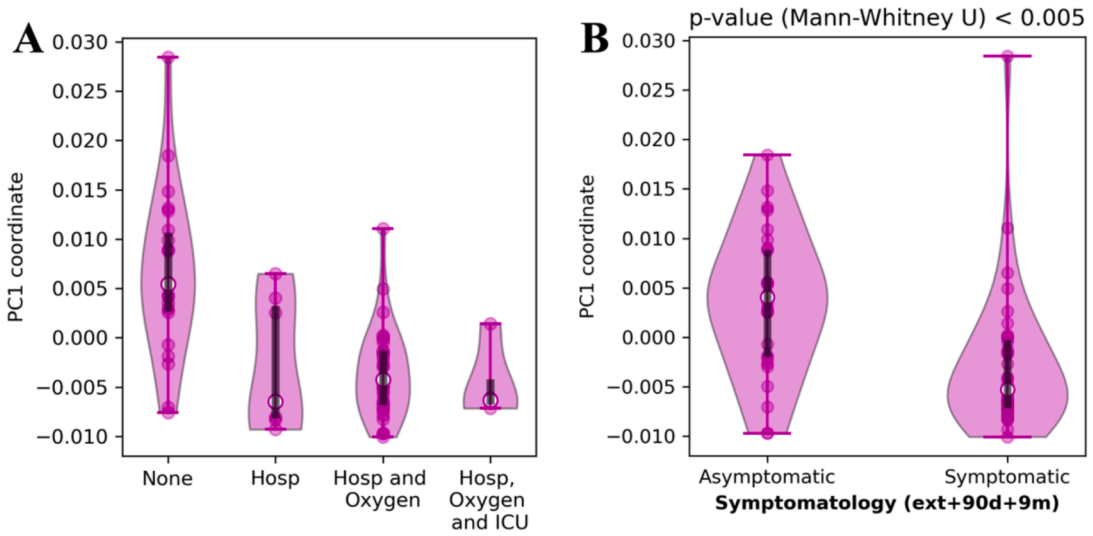


Fig. 5. A. Violin plots showing the distribution of PC1 values for patients grouped by hospitalization status, ICU admission, and receipt of conventional oxygen therapy during COVID infection. The categories include: *none* (no hospitalization or oxygen therapy), *Hosp* (hospitalized), *Hosp and Oxygen* (hospitalized and received oxygen therapy), and *Hosp, Oxygen and ICU* (hospitalized, received oxygen therapy, and admitted to ICU). **B.** Violin plots showing the distribution of PC1 values for patients grouped by symptomatology according to the presence of symptoms at the time of extraction, 90 days post-diagnosis, and 9 months post-diagnosis (*Ext + 90d + 9m*). White dots represent the mean values, and black vertical lines indicate the interquartile range for each group. The p-value for the Mann-Whitney *U* test is shown, indicating the statistical significance of differences in distribution between asymptomatic and symptomatic patients.

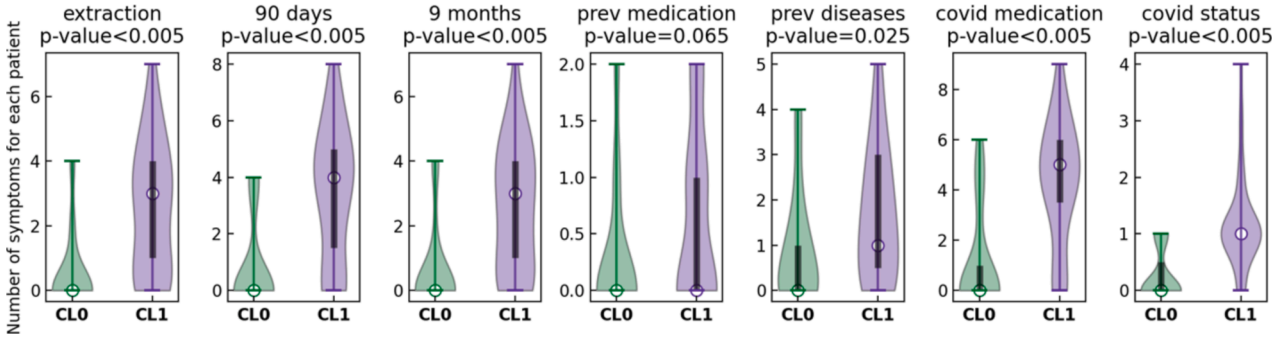


Fig. 6. Violin plots showing the distribution of the number of symptoms for each patient, divided by clinical categories: symptoms at extraction, 90 days post-diagnosis, 9 months post-diagnosis, previous medication, previous diseases, medication during COVID infection, and COVID status. Median values are indicated by white dots, and interquartile ranges are represented by black vertical lines. The p-values for the Mann-Whitney *U* test are provided for each group, highlighting the statistical significance of differences in distribution between cluster 0 (*CL0*) and cluster 1 (*CL1*).

Barnard's test (Table S3), also rank highly in cluster association according to the Jaccard score. Notably, these high Jaccard scores are achieved with single characteristics rather than groups of characteristics, as seen in most other cases. The *significant* group (using a threshold of more than one symptom) also presents a high Jaccard score.

While the Jaccard and ϕ scores generally align, some discrepancies exist. In this analysis, the Jaccard score is preferred because it does not penalize imbalances between non-diagonal elements in the contingency table. Since we aim to measure the degree of association between the two classifications ignoring these imbalances is more suitable. Therefore, hospitalization during the COVID infection emerges as the clinical criterion most strongly associated with cluster separation, even surpassing the combination of individually significant symptoms or the *ext + 90d + 9m* criterion used to optimize the spectral region and scaling method.

3.3. Comparative analysis between spectral, proteomic and lipidomic studies

To compare the spectral analysis with previous lipidomic and proteomic studies conducted on the same cohort of patients, Jaccard similarity scores were calculated between clusters obtained from spectral data, symptoms (*ext + 90d + 9m*, Table S1), hospitalization status, and proteomic and lipidomic classifications (Table 1). The similarity scores for these different combinations are summarized in Fig. 8. Additionally, for combinations involving cluster separation, contingency tables are included to illustrate the comparison of cluster assignments between spectral data, proteomics, lipidomics, symptomatology (*ext + 90d + 9m*), and hospitalization during COVID infection.

The Jaccard scores provide a valuable insight into the concordance between different methods of classifying patient data. While the scores are not exceptionally high, they do show a reasonable degree of similarity, suggesting that UV–VIS spectroscopy captures significant biochemical information related to patient symptoms and clinical outcomes. Specifically, the similarity scores between spectral data and clinical symptom-based classification (0.667) and hospitalization status (0.681) are promising, indicating that UV–VIS spectroscopy can effectively reflect clinical severity and patient status. The score of 0.599 between spectral data and proteomic classifications and 0.643 with lipidomic classifications indicate that while VIS spectroscopy is less aligned with these more detailed molecular analyses, it still captures

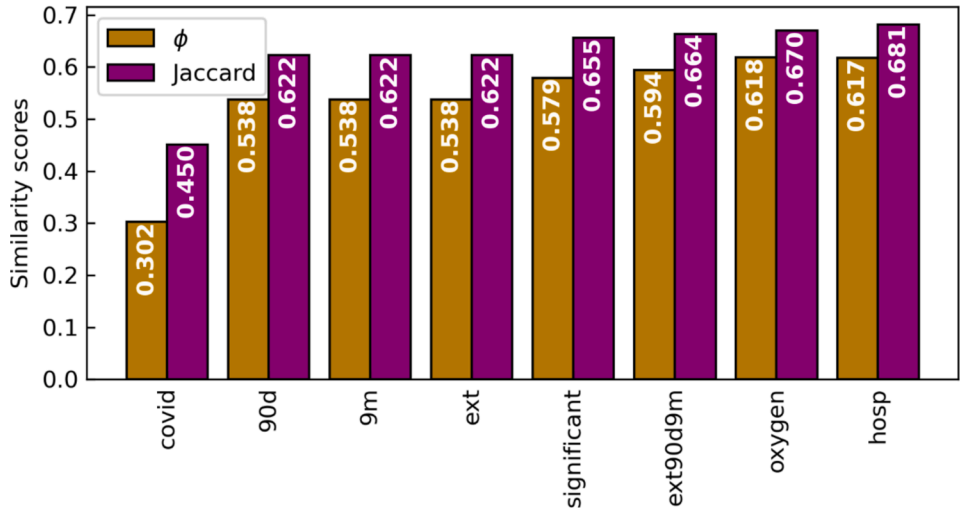


Fig. 7. Jaccard and ϕ similarity scores for the comparison between cluster classifications and various clinical characteristic criteria. The clinical categories analyzed include COVID status (*COVID*), symptoms 90 days post-diagnosis (*90d*), symptoms 9 months post-diagnosis (*9m*), symptoms at extraction (*ext*), significant symptoms (*significant*), combined symptoms at extraction, 90 days, and 9 months (*ext* + *90d* + *9m*), oxygen therapy requirement (*oxygen*), and hospitalization status (*hosp*). Higher scores indicate a stronger association between the clinical characteristics and the clusters from spectra.

Table 1
Jaccard similarity coefficients for the comparisons between clusters obtained from spectra, clusters obtained from the proteomics [29] and lipidomics studies [28], classification of patients by Hospitalization and Ext + 90d + 9m.

Jaccard score	Proteomics [29]	Lipidomics [28]	Symptoms	Hospitalization
Spectra	0.599	0.643	0.667	0.681
Proteomics		0.767	0.792	0.816
Lipidomics			0.967	0.762

relevant biochemical changes. The slightly lower scores may be attributed to the broader scope of UV–VIS spectroscopy, which might not detect the same specific molecular details as proteomics and lipidomics but does provide a holistic view of the biochemical state. The extremely high similarity between proteomics and lipidomics (0.767) aligns with expectations, given that both techniques provide comprehensive molecular profiles. This high degree of agreement supports the validity of using these methods as benchmarks for evaluating the effectiveness of UV–VIS spectroscopy.

3.4. Assessing the predictive ability of the spectroscopic analysis with an expanded patient cohort

The PCA analysis performed on the spectral data of the ***Omics group*** shows that the projection of the samples on the two first eigenvectors (PC1 and PC2) can effectively classify patients into symptomatic and asymptomatic. Moreover, it provides a quantitative measurement of their severity. Once this analysis is completed, the spectra of the additional 64 patients from the ***Extended group*** (Fig. S3) can be scaled and transformed to the same PCA space. Using the clustering model obtained from the ***Omics group***, they can also be classified into the previously defined clusters. Among the new 64 patients of the ***Extended group*** (1 patient was an outlier), 23 were assigned to the asymptomatic cluster, while the remaining 41 were predicted to belong to the symptomatic cluster. However, only 13, 9, and 15 patients from the ***Extended group*** were actually asymptomatic during blood extraction, 90 days after diagnosis, and 9 months after diagnosis, respectively. This discrepancy suggests that while the clusters from spectra provide a good general indication of patient status, there may be some misclassification or the presence of subclinical symptoms not captured by the initial symptom reports.

Fig. S4 presents a matrix representation illustrating the severity scale of the 64 patients based on UV–VIS data and its correlation with clinical characteristics, classified according to the developed predictive algorithm. The Figure highlights the trend of symptom presence in each cluster, showing that the symptomatic cluster (*CL1*) consistently has a higher number of clinical symptoms. This underscores the robustness of the UV–VIS spectroscopic analysis in correlating with clinical manifestations. However, some patients predicted to be asymptomatic from their PC1 value still exhibited symptoms at various follow-up points, suggesting potential subclinical presentations or variations not fully captured by the spectral data alone. This indicates that while the model is generally effective, there is room for improvement in refining its

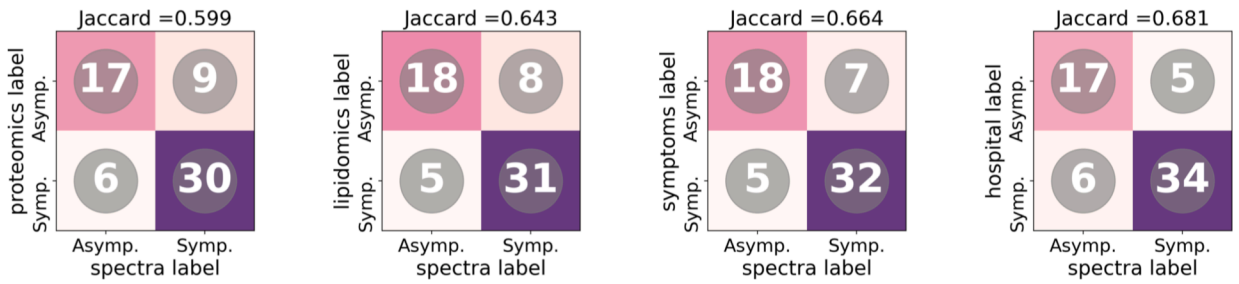


Fig. 8. Jaccard similarity scores between the groups obtained from spectra, lipidomics, proteomics analysis, patient symptomatology, and patient hospitalization during COVID-19 infection. Only patients from the Omics group were used for this comparison. Contingency tables illustrate the comparison of cluster assignments between spectral data and proteomics [29], lipidomics [28], symptomatology (ext + 90d + 9m), and hospitalization during COVID infection.

predictive capabilities to better account for subtle clinical variations.

Violin plots of the distributions of the number of symptoms for each clinical group are shown in Fig. 9. These plots reveal that for all clinical groups, the median of symptoms in the symptomatic group was higher than that in the asymptomatic group. This difference was statistically significant for all clinical groups except for those assessing symptoms at the time of extraction, 90 days, and 9 months post-diagnosis. In these specific cases, statistical significance was likely hindered by the smaller number of asymptomatic patients (13, 9, and 15, respectively) compared to the symptomatic patients (51, 55, and 49, respectively). Despite this, the higher median values in the symptomatic group suggest that the lack of significance may be due to the unbalanced sample sizes rather than an absence of a true effect. The other clinical groups, which had more balanced patient numbers, reinforced the robustness of these findings. For instance, groups considering previous diseases and COVID medication showed clear statistical significance, indicating that these factors are strongly associated with the patients' symptomatic status as determined by the clusters from spectra.

The analysis of the ***Extended group*** underscores the robustness of UV–VIS spectroscopy in distinguishing between symptomatic and asymptomatic patients. The predictive ability of the method to classify and quantify the severity of the PCC, demonstrated its potential as a diagnostic tool. The consistent finding that symptomatic patients have higher median symptom counts across various clinical categories suggests that UV–VIS spectroscopy is capturing meaningful biochemical signals related to disease severity and symptom burden. However, the variability in certain categories and the occasional misclassification of asymptomatic patients highlight areas for further refinement. Incorporating additional spectral regions, refining preprocessing techniques, and primarily increasing the sample size could improve the model's accuracy in detecting subtle clinical variations.

4. Conclusions

This study investigated the potential of UV–VIS–NIR–MIR spectroscopy for the rapid diagnosis and severity assessment of post-COVID condition (PCC). By analyzing blood samples from 130 patients, including 65 previously studied in lipidomic and proteomic analyses, spectral regions strongly correlated with patient symptomatology and clinical outcomes were identified.

The findings reveal that the 297–600 nm range of the supernatant fraction provides the most significant information for differentiating between asymptomatic and symptomatic PCC patients. Unsupervised clustering of spectral data in this range effectively separated patients into two distinct groups, which aligned well with reported symptom severity and clinical characteristics such as hospitalization status and oxygen therapy requirement during acute COVID-19 infection.

The clusters from spectra showed significant associations with various clinical parameters, including persistent symptoms at different time points post-infection. Notably, the clustering based on spectral data correlated strongly with hospitalization status during acute infection, suggesting that the method captures not only current symptomatology but also reflects the severity of the initial COVID-19 episode.

Comparative analysis with proteomic and lipidomic data from the same patient cohort revealed moderate to strong concordance, supporting the validity of spectroscopic analysis in capturing relevant biochemical changes associated with PCC. While UV–VIS spectroscopy may not provide the same level of molecular detail as these omics approaches, it offers a holistic view of the biochemical state that correlates well with clinical outcomes and symptom burden.

The robustness of our approach was further assessed through the analysis of an expanded patient cohort, demonstrating the potential of this method for widespread clinical application. Additionally, this spectroscopic method presents notable advantages in terms of computational efficiency and practical applicability. Unlike proteomics and lipidomics, which require extensive profiling of individual biomolecules, spectroscopic analysis involves simpler data processing, resulting in faster data acquisition and lower computational demands, making it suitable for rapid, cost-effective implementation. The accessibility of spectroscopy also makes it highly adaptable for use in both portable and point-of-care (POC) devices, enabling real-time assessments in clinical or primary care settings. Such POC applications would allow healthcare providers to monitor PCC severity without relying on complex infrastructure, thus expanding accessibility and facilitating timely, targeted interventions for individuals affected by long-term COVID-19. In addition to its computational efficiency and practical applicability, the proposed spectroscopic method demonstrates a high level of environmental sustainability. A greenness assessment using the MoGAPI [67] framework yielded a score of 90 [68], indicating the environmentally friendly nature of the approach. This further supports its suitability for widespread clinical application, as the method not only offers rapid and cost-effective implementation but also aligns with green chemistry principles.

Nonetheless, certain limitations should be acknowledged. The sample size, while adequate for initial findings, could be expanded in future studies to enhance the model's generalizability. Additionally, the possibility of misclassification due to overlapping spectral features or the complexity of PCC symptoms suggests the need for further refinement. Future work should consider expanding the spectral range, integrating additional machine learning algorithms to improve model accuracy and robustness, and integrating this approach into clinical practice to fully realize its potential in managing PCC.

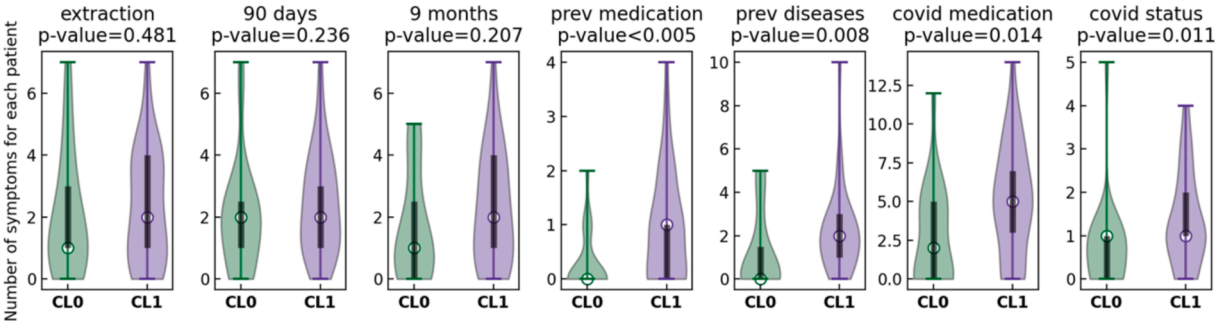


Fig. 9. Violin plots of the distributions of number of symptoms for each patient in the ***Extended group*** divided by clinical categories. Patients in clusters 0 (*CL0*) and 1 (*CL1*) are from the ***Extended group*** (64 new patients) and were classified according to the projection of the spectra on the first eigenvector obtained from the analysis of the ***Omics group***. Median values and interquartile range are shown as white dots and black vertical lines, respectively. The Mann-Whitney *U* test p-value is shown for each category of clinical variables.

5. Declaration of AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT4o from OpenAI and Claude from Anthropic in order to improve language and readability. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

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.

Acknowledgments

This work was supported by the Spanish Agencia Estatal de Investigación (AEI) and the ERDF (PID2019-111327GB-I00, PDC2022-133402-I00, PID2022-141534OB-I00 and CNS2023-144353), by Xunta de Galicia (ED431C 2021/21 and Centro de investigación do Sistema universitario de Galicia accreditation 2023-2027, ED431G 2023/03) and the European Union (European Regional Development Fund – ERDF). P. A.-R. thanks Xunta de Galicia for her predoctoral contract (ED481A-2024-073). We would like to thank nurse Paula García Fandiño for her efficient and crucial work in gathering samples for this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.saa.2024.125474.

Data availability

Data will be made available on request.

References

[1] R.S.A. Razak, A. Ismail, A.F. Abdul Aziz, L.S. Suddin, A. Azzeri, N.I. Sha'ari, Post-COVID syndrome prevalence: a systematic review and meta-analysis, BMC Public Health 24 (1) (2024) 1–19, https://doi.org/10.1186/s12889-024-19264-5.

[2] A. Sugiyama, T. Takafuta, T. Sato, Y. Kitahara, Y. Yoshinaga, K. Abe, C. Chanroth, A.G. Ataa, Z. Phyo, A. Kurisu, K. Ko, T. Akita, E. Kishita, M. Kuwabara, J. Tanaka, Natural course of post-COVID symptoms in adults and children, Sci. Rep. 14 (1) (2024) 1–9, https://doi.org/10.1038/s41598-024-54397-y.

[3] M. Parotto, M. Gyöngyösi, K. Howe, S.N. Myatra, O. Ranzani, M. Shankar-Hari, M. S. Herridge, Post-acute sequelae of COVID-19: understanding and addressing the burden of multisystem manifestations, Lancet Respir. Med. (2023) 739–754, https://doi.org/10.1016/S2213-2600(23)00239-4.

[4] E.W. Ely, L.M. Brown, H.V. Fineberg, Long covid defined, N. Engl. J. Med. 391 (18) (2024) 1746–1753, https://doi.org/10.1056/nejmsb2408466.

[5] J.V.A. Franco, L.I. Garegnani, M.-I. Metzendorf, K. Heldt, R. Mumm, C. Scheidt-Nave, Post-Covid-19 conditions in adults: systematic review and meta-analysis of health outcomes in controlled studies, BMJ Med. 3 (1) (2024) e000723.

[6] J.B. Soriano, S. Murthy, J.C. Marshall, P. Relan, J.V. Diaz, A clinical case definition of post-COVID-19 condition by a delphi consensus, Lancet Infect. Dis. 22 (4) (2022) e102–e107, https://doi.org/10.1016/S1473-3099(21)00703-9.

[7] C. Fernández-de-las-Peñas, D. Palacios-Ceña, V. Gómez-Mayordomo, L. L. Florencio, M.L. Cuadrado, G. Plaza-Manzano, M. Navarro-Santana, Prevalence of post-COVID-19 symptoms in hospitalized and non-hospitalized COVID-19 survivors: a systematic review and meta-analysis, Eur. J. Intern. Med. 92 (2021) 55–70, https://doi.org/10.1016/j.ejim.2021.06.009.

[8] D.M. Altmann, E.M. Whettlock, S. Liu, D.J. Arachchillage, R.J. Boyton, The immunology of long COVID, Nat. Rev. Immunol. 23 (2023) 618–634, https://doi.org/10.1038/s41577-023-00904-7.

[9] C. Fernandez-de-las-Peñas, K.I. Notarte, R. Macasaet, J.V. Velasco, J.A. Catahay, A. T. Ver, W. Chung, J.A. Valera-Calero, M. Navarro-Santana, Persistence of post-COVID symptoms in the general population two years after SARS-CoV-2 infection: a systematic review and meta-analysis, J. Infect. 8 (2) (2024) P77–P88, https://doi.org/10.1016/j.jinf.2023.12.004.

[10] H.E. Davis, L. McCorkell, J.M. Vogel, E.J. Topol, Long COVID: major findings, mechanisms and recommendations, Nat. Rev. Microbiol. 21 (2023) 133–146, https://doi.org/10.1038/s41579-022-00846-2.

[11] S. Yu, X. Li, Z. Xin, L. Sun, J. Shi, Proteomic insights into SARS-CoV-2 infection mechanisms, diagnosis, therapies and prognostic monitoring methods, Front. Immunol. 13 (2022) 923387, https://doi.org/10.3389/fimmu.2022.923387.

[12] C.B. Messner, V. Demichev, D. Wendisch, L. Michalick, M. White, A. Freiwald, K. Textoris-Taube, S.I. Vernardis, A.S. Egger, M. Kreidl, D. Ludwig, C. Kilian, F. Agostini, A. Zelezniak, C. Thibeault, M. Pfeiffer, S. Hippenstiel, A. Hocke, C. von Kalle, A. Campbell, C. Hayward, D.J. Porteous, R.E. Marioni, C. Langenberg, K. S. Lilley, W.M. Kuebler, M. Mülleder, C. Drosten, N. Suttorp, M. Witzenrath, F. Kurth, L.E. Sander, M. Ralser, Ultra-high-throughput clinical proteomics reveals classifiers of COVID-19 infection, 11-24.e4, Cell Syst. 11 (1) (2020), https://doi.org/10.1016/j.cels.2020.05.012.

[13] H. Babačić, W. Christ, J.E. Araújo, G. Mermelekas, N. Sharma, J. Tynell, M. García, R. Varnaite, H. Asgeirsson, H. Glans, J. Lehtiö, S. Gredmark-Russ, J. Klingström, M. Pernemalm, Comprehensive proteomics and meta-analysis of COVID-19 host response, Nat. Commun. 14 (1) (2023) 1–18, https://doi.org/10.1038/s41467-023-41159-z.

[14] L. Goracci, E. Petito, A. Di Veroli, E. Falcinelli, C. Bencivenga, E. Giglio, C. Becattini, E. De Robertis, G. Vaudo, P. Gresele, A platelet lipidomics signature in patients with COVID-19, Platelets 34 (1) (2023) 2200847, https://doi.org/10.1080/09537104.2023.2200847.

[15] M. Costanzo, M. Caterino, Targeted lipidomics data of COVID-19 patients, Data Br. 48 (2023) 109089, https://doi.org/10.1016/j.dib.2023.109089.

[16] A. Viode, K.K. Smolen, P. van Zalm, D. Stevenson, M. Jha, K. Parker, O. Levy, J. A. Steen, H. Steen, Longitudinal plasma proteomic analysis of 1117 hospitalized patients with COVID-19 identifies features associated with severity and outcomes, Sci. Adv. 10 (21) (2024), https://doi.org/10.1126/sciadv.adl5762.

[17] E. Duijvelaar, J. Gisby, J.E. Peters, H.J. Bogaard, J. Aman, Longitudinal plasma proteomics reveals biomarkers of alveolar-capillary barrier disruption in critically Ill COVID-19 patients, Nat. Commun. 15 (1) (2024) 1–16, https://doi.org/10.1038/s41467-024-44986-w.

[18] B. Chatterjee, S.S. Thakur, Valuable contributions and lessons learned from proteomics and metabolomics studies of COVID-19, J. Proteome Res. 23 (10) (2024) 4171–4187, https://doi.org/10.1021/acs.jproteome.4c00340.

[19] M. Alexovič, T. Bondarchuk, C. Uličná, J. Sabo, Blood proteomics of COVID-19 infection: an update, Clin. Chim. Acta. 562 (2024) 119881, https://doi.org/10.1016/j.cca.2024.119881.

[20] C. Yang, C.P. Shannon, S.J. Tebbutt, Unravelling long COVID: insights from proteomics and considerations for comprehensive understanding, eBioMedicine 101 (2024) 105023, https://doi.org/10.1016/j.ebiom.2024.105023.

[21] C. Iosef, M.J. Knauer, M. Nicholson, L.R. Van Nynatten, G. Cepinskas, S. Draghici, V.K.M. Han, D.D. Fraser, Plasma proteome of long-COVID patients indicates HIF-mediated vasculo-proliferative disease with impact on brain and heart function, J. Transl. Med. 21 (1) (2023) 1–21, https://doi.org/10.1186/s12967-023-04149-9.

[22] M.A. Patel, M.J. Knauer, M. Nicholson, M. Daley, L.R. Van Nynatten, G. Cepinskas, D.D. Fraser, Organ and cell-specific biomarkers of long-COVID identified with targeted proteomics and machine learning, Mol. Med. 29 (1) (2023) 1–15, https://doi.org/10.1186/s10020-023-00610-z.

[23] M.A. Patel, M.J. Knauer, M. Nicholson, M. Daley, L.R. Van Nynatten, C. Martin, E. K. Patterson, G. Cepinskas, S.L. Seney, V. Dobretzberger, M. Miholits, B. Webb, D. D. Fraser, Elevated vascular transformation blood biomarkers in long-COVID indicate angiogenesis as a key pathophysiological mechanism, Mol. Med. 28 (1) (2022) 1–12, https://doi.org/10.1186/s10020-022-00548-8.

[24] C. Cervia-Hasler, S.C. Brüningk, T. Hoch, B. Fan, G. Muzio, R.C. Thompson, L. Ceglarek, R. Meledin, P. Westermann, M. Emmenegger, P. Taeschler, Y. Zurbuchen, M. Pons, D. Menges, T. Ballouz, S. Cervia-Hasler, S. Adamo, M. Merad, A.W. Charney, M. Puhan, P. Brodin, J. Nilsson, A. Aguzzi, M.E. Raeber, C.B. Messner, N.D. Beckmann, K. Borgwardt, O. Boyman, Persistent complement dysregulation with signs of thromboinflammation in active long covid, Science 383 (2024) 273. https://www.science.org/doi/10.1126/science.adg7942.

[25] J.J. Kovarik, A. Bileck, G. Hagn, S.M. Meier-Menches, T. Frey, A. Kaempf, M. Hollenstein, T. Shoumariyeh, L. Skos, B. Reiter, M.C. Gerner, A. Spannbauer, E. Hasimbegovic, D. Schmidl, G. Garhöfer, M. Gyöngyösi, K.G. Schmetterer, C. Gerner, A multi-omics based anti-inflammatory immune signature characterizes long COVID-19 syndrome, iScience 26 (1) (2023) 105717, https://doi.org/10.1016/j.isci.2022.105717.

[26] Y. López-Hernández, J. Monárrez-Espino, D.A.G. López, J. Zheng, J.C. Borrego, C. Torres-Calzada, J.P. Elizalde-Díaz, R. Mandal, M. Berjanskii, E. Martínez-Martínez, J.A. López, D.S. Wishart, The plasma metabolome of long COVID patients two years after infection, Sci. Rep. 13 (1) (2023) 1–14, https://doi.org/10.1038/s41598-023-39049-x.

[27] X. Gu, S. Wang, W. Zhang, C. Li, L. Guo, Z. Wang, H. Li, H. Zhang, Y. Zhou, W. Liang, H. Li, Y. Liu, Y. Wang, L. Huang, T. Dong, D. Zhang, C.C.L. Wong, B. Cao, Probing long COVID through a proteomic lens: a comprehensive two-year longitudinal cohort study of hospitalised survivors, eBioMedicine 98 (2023) 104851, https://doi.org/10.1016/j.ebiom.2023.104851.

[28] P.F. Garrido, L.S. Castillo-Peinado, F. Priego-Capote, I. Barrio, Á. Piñeiro, M. J. Domínguez-Santalla, E. Rodríguez-Ruiz, R. Garcia-Fandino, Lipidomics signature in post-COVID patient sera and its influence on the prolonged inflammatory response, J. Infect. Public Health 17 (4) (2024) 588–600, https://doi.org/10.1016/J.JIPH.2024.01.017.

[29] A. Seco-González, P. Antelo-Riveiro, S.B. Bravo, P.F. Garrido, M.J. Domínguez-Santalla, E. Rodríguez-Ruiz, Á. Piñeiro, R. Garcia-Fandino, Proteomic analysis of post-COVID condition: insights from plasma and pellet blood fractions, J. Infect. Public Health 17 (12) (2024) 102571, https://doi.org/10.1016/J.JIPH.2024.102571.

[30] R.K. Sahu, S. Mordechai, Spectroscopic techniques in medicine: the future of diagnostics, Appl. Spectrosc. Rev. 51 (6) (2016) 484–499, https://doi.org/10.1080/05704928.2016.1157809.
[31] Y. Khristoforova, L. Bratchenko, I. Bratchenko, Raman-based techniques in medical applications for diagnostic tasks: a review, Int. J. Mol. Sci. 24 (21) (2023), https://doi.org/10.3390/IJMS242115605.
[32] H.J. Byrne, F. Bonnier, J. McIntyre, D.R. Parachalil, Quantitative analysis of human blood serum using vibrational spectroscopy, Clin. Spectrosc. 2 (2020) 100004, https://doi.org/10.1016/j.clispe.2020.100004.
[33] L. Rodriguez-Saona, D.P. Aykas, K.R. Borba, A. Urtubia, Miniaturization of optical sensors and their potential for high-throughput screening of foods, Curr. Opin. Food Sci. 31 (2020) 136–150, https://doi.org/10.1016/j.cofs.2020.04.008.
[34] Y. Qi, D. Hu, Y. Jiang, Z. Wu, M. Zheng, E.X. Chen, Y. Liang, M.A. Sadi, K. Zhang, Y. P. Chen, Recent progresses in machine learning assisted raman spectroscopy, Adv. Opt. Mater. 11 (14) (2023) 2203104, https://doi.org/10.1002/adom.202203104.
[35] S. Hu, H. Li, C. Chen, C. Chen, D. Zhao, B. Dong, X. Lv, K. Zhang, Y. Xie, Raman spectroscopy combined with machine learning algorithms to detect adulterated suichang native honey, Sci. Rep. 12 (1) (2022) 1–7, https://doi.org/10.1038/s41598-022-07222-3.
[36] M.I. Rumaling, F.P. Chee, A. Bade, N.H. Hasbi, S. Daim, F. Juhim, M. Duinong, R. Rasmidi, Methods of optical spectroscopy in detection of virus in infected samples: a review, Heliyon 8 (9) (2022) e10472.
[37] A. Bedair, K. Okasha, F.R. Mansour, Spectroscopic methods for COVID-19 detection and early diagnosis, Virol. J. 19 (152) (2022) 1–13, https://doi.org/10.1186/s12985-022-01867-2.
[38] M.E. Raypah, A.N. Faris, M.M. Azlan, N.Y. Yusof, F.H. Suhailin, R.H. Shueb, I. Ismail, F.H. Mustafa, Near-infrared spectroscopy as a potential COVID-19 early detection method: a review and future perspective, Sensors 22 (12) (2022) 4391, https://doi.org/10.3390/s22124391.
[39] B.F.O. Coelho, S.L.P. Nunes, C.A. de França, D. dos S. Costa, R.F. do Carmo, R. M. Prates, E.F.S. Filho, R.P. Ramos, On the Feasibility of Vis–NIR Spectroscopy and Machine Learning for Real Time SARS-CoV-2 Detection, Spectrochim. Acta - Part A Mol. Biomol. Spectrosc. 308 (2024) 123735, https://doi.org/10.1016/j.saa.2023.123735.
[40] IDIS. (s.f.). Unidad de Epigenómica – Instituto de Investigación Sanitaria de Santiago de Compostela https://www.idisantiago.es/plataformas-de-apoyo-comun/biobanco/ (Accessed Nov 17, 2024).
[41] Declaration of Helsinki, Recommendations guiding medical doctors in biomedical research involving human subjects, Med. J. Aust. 1 (7) (1976) 206–207, https://doi.org/10.5694/j.1326-5377.1976.tb140527.x.
[42] J. Reback, W. McKinney, Jbrockmendel, J. Van den Bossche, T. Augspurger, P. Cloud, Gfyoung, Sinhrks, A. Klein, M. Roeschke, S. Hawkins, J. Tratner, C. She, W. Ayd, Petersen, T.; Garcia, M.; Schendel, J.; Hayden, A.; MomIsBestFriend; Jancauskas, V.; Battiston, P.; Seabold, S.; chris-b1; h-vetinari; Hoyer, S.; Overmeire, W.; alimcmaster1; Dong, K.; Whelan, C.; Mehyar, M. Pandas-Dev/Pandas: Pandas 1.0.3. Zenodo 2020. Available from: doi: 10.5281/zenodo.3715232.
[43] C.R. Harris, K.J. Millman, S.J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N.J. Smith, R. Kern, M. Picus, S. Hoyer, M.H. van Kerkwijk, M. Brett, A. Haldane, J.F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, T.E. Oliphant, Array programming with NumPy, Nature 585 (2020) 357–362, https://doi.org/10.1038/s41586-020-2649-2.
[44] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S.J. van der Walt, M. Brett, J. Wilson, K.J. Millman, N. Mayorov, A.R.J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E.W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E.A. Quintero, C.R. Harris, A.M. Archibald, A.H. Ribeiro, F. Pedregosa, P. van Mulbregt, A. Vijaykumar, A. Bardelli, Pietro, A. Rothberg, A. Hilboll, A. Kloeckner, A. Scopatz, A. Lee, A. Rokem, C.N. Woods, C. Fulton, C. Masson, C. Häggström, C. Fitzgerald, D.A. Nicholson, D.R. Hagen, D. V. Pasechnik, E. Olivetti, E. Martin, E. Wieser, F. Silva, F. Lenders, F. Wilhelm, G. Young, G.A. Price, G.L. Ingold, G.E. Allen, G.R. Lee, H. Audren, I. Probst, J. P. Dietrich, J. Silterra, J.T. Webber, J. Slavič, J. Nothman, J. Buchner, J. Kulick, J. L. Schönberger, J.V. de Miranda Cardoso, J. Reimer, J. Harrington, J.L. C. Rodríguez, J. Nunez-Iglesias, J. Kuczynski, K. Tritz, M. Thoma, M. Newville, M. Kümmerer, M. Bolingbroke, M. Tartre, M. Pak, N.J. Smith, N. Nowaczyk, N. Shebanov, O. Pavlyk, P.A. Brodtkorb, P. Lee, R.T. McGibbon, R. Feldbauer, S. Lewis, S. Tygier, S. Sievert, S. Vigna, S. Peterson, S. More, T. Pudlik, T. Oshima, T.J. Pingel, T.P. Robitaille, T. Spura, T.R. Jones, T. Cera, T. Leslie, T. Zito, T. Krauss, U. Upadhyay, Y.O. Halchenko, Y. Vázquez-Baeza, SciPy 1.0: fundamental algorithms for scientific computing in python, Nat. Methods 17 (3) (2020) 261–272, https://doi.org/10.1038/s41592-019-0686-2.
[45] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, Scikit-learn: machine learning in python, J. Mach. Learn. Res. 12 (2011) 2825–2830.
[46] J.D. Hunter, Matplotlib: a 2D graphics environment, Comput. Sci. Eng. 9 (3) (2007) 90–95, https://doi.org/10.1109/MCSE.2007.55.
[47] J. Torniainen, I.O. Afara, M. Prakash, J.K. Sarin, L. Stenroth, J. Töyräs, Open-source python module for automated preprocessing of near infrared spectroscopic data, Anal. Chim. Acta 1108 (2020) 1–9, https://doi.org/10.1016/j.aca.2020.02.030.
[48] T. Fearn, C. Riccioli, A. Garrido-Varo, J.E. Guerrero-Ginel, On the geometry of SNV and MSC, Chemom. Intell. Lab. Syst. 96 (1) (2009) 22–26, https://doi.org/10.1016/j.chemolab.2008.11.006.
[49] C. Pasquini, Near infrared spectroscopy: a mature analytical technique with new perspectives – a review, Anal. Chim. Acta 10 (26) (2018) 8–36, https://doi.org/10.1016/j.aca.2018.04.004.
[50] V. Corradi, B.I. Sejdiu, H. Mesa-Galloso, H. Abdizadeh, S.Y. Noskov, S.J. Marrink, D.P. Tieleman, Emerging diversity in lipid-protein interactions, Chem. Rev. 119 (9) (2019) 5775–5848, https://doi.org/10.1021/acs.chemrev.8b00451.
[51] H.B. Mann, D.R. Whitney, On a test of whether one of two random variables is stochastically larger than the other, Ann. Math. Stat. 18 (1) (1947) 50–60, https://doi.org/10.1214/aoms/1177730491.
[52] F.-X. Schmid, Biological macromolecules: uv-visible spectrophotometry, in: Encyclopedia of Life Sciences, John Wiley & Sons Ltd, 2001, https://doi.org/10.1038/npg.els.0003142.
[53] S. Soltani, A. Ojaghi, F.E. Robles, Deep UV dispersion and absorption spectroscopy of biomolecules, Biomed. Opt. Express 10 (2) (2019) 487, https://doi.org/10.1364/boe.10.000487.
[54] A. Talla, S.V. Vasaikar, G.L. Szeto, M.P. Lemos, J.L. Czartoski, H. MacMillan, Z. Moodie, K.W. Cohen, L.B. Fleming, Z. Thomson, L. Okada, L.A. Becker, E. M. Coffey, S.C. De Rosa, E.W. Newell, P.J. Skene, X. Li, T.F. Bumol, M. Juliana McElrath, T.R. Torgerson, Persistent serum protein signatures define an inflammatory subcategory of long COVID, Nat. Commun. 14 (1) (2023) 1–16, https://doi.org/10.1038/s41467-023-38682-4.
[55] V.V. Volkov, J. McMaster, J. Aizenberg, C.C. Perry, Mapping blood biochemistry by Raman spectroscopy at the cellular level, Chem. Sci. 13 (1) (2022) 133–140, https://doi.org/10.1039/d1sc05764b.
[56] G.A. Zalesskaya, N.P. Mit'Kovskaya, O.A. Galai, A.V. Kuchinskii, O.V. Laskina, Change in the absorption spectra of blood exposed to a low-frequency magnetic field, J. Appl. Spectrosc. 74 (2) (2007) 223–229, https://doi.org/10.1007/s10812-007-0035-6.
[57] K.J. Scott, Detection and measurement of carotenoids by UV/VIS spectrophotometry, F2.2.1-F2.2.10, Curr. Protoc. Food Anal. Chem. 00 (1) (2001), https://doi.org/10.1002/0471142913.faf0202s00.
[58] J. Wang, R. Xu, S. Yang, Estimation of plant water content by spectral absorption features centered at 1,450 Nm and 1,940 Nm regions, Environ. Monit. Assess. 157 (1–4) (2009) 459–469, https://doi.org/10.1007/s10661-008-0548-3.
[59] B.L. Horecker, The absorption spectra of hemoglobin and its derivatives in the visible and near infra-red regions, J. Biol. Chem. 148 (1) (1943) 173–183, https://doi.org/10.1016/s0021-9258(18)72329-6.
[60] D.M. Mancini, L. Bolinger, H. Li, K. Kendrick, B. Chance, J.R. Wilson, Validation of near-infrared spectroscopy in humans, J. Appl. Physiol. 77 (6) (1994) 2740–2747, https://doi.org/10.1152/jappl.1994.77.6.2740.
[61] J.W. Hall, A. Pollard, Near-infrared spectroscopic determination of serum total proteins, albumin, globulins, and urea, Clin. Biochem. 26 (6) (1993) 483–490, https://doi.org/10.1016/0009-9120(93)80013-K.
[62] K. Jansen, M. Wu, A.F.W. Van der Steen, G. Van Soest, Photoacoustic imaging of human coronary atherosclerosis in two spectral bands, Photoacoustics 2 (1) (2014) 12–20, https://doi.org/10.1016/j.pacs.2013.11.003.
[63] G. Ripoll, S. Failla, B. Panea, J.F. Hocquette, S. Dunner, J.L. Olleta, M. Christensen, P. Ertbjerg, I. Richardson, M. Contò, P. Albertí, C. Sañudo, J.L. Williams, Near-infrared reflectance spectroscopy for predicting the phospholipid fraction and the total fatty acid composition of freeze-dried beef, Sensors 21 (12) (2021) 4230, https://doi.org/10.3390/s21124230.
[64] J. Yadav, A. Rani, V. Singh, B.M. Murari, Prospects and limitations of non-invasive blood glucose monitoring using near-infrared spectroscopy, Biomed. Signal Process. Control 18 (2015) 214–227, https://doi.org/10.1016/j.bspc.2015.01.005.
[65] K.Z. Liu, M.H. Shi, H.H. Mantsch, Molecular and chemical characterization of blood cells by infrared spectroscopy: a new optical tool in hematology, Blood Cells, Mol. Dis. 35 (3) (2005) 404–412, https://doi.org/10.1016/j.bcmd.2005.06.009.
[66] C. Delrue, R. Speeckaert, M. Oyaert, T. Kerre, S. Rottey, R. Coopman, W. Huvenne, S. De Bruyne, M.M. Speeckaert, Infrared spectroscopy: a new frontier in hematological disease diagnosis, Int. J. Mol. Sci. 24 (23) (2023) 17007, https://doi.org/10.3390/ijms242317007.
[67] F.R. Mansour, J. Płotka-Wasylka, M. Locatelli, Modified GAPI (MoGAPI) tool and software for the assessment of method greenness: case studies and applications, Analytica 5 (3) (2024) 451–457, https://doi.org/10.3390/analytica5030030.
[68] https://fotouhmansour.github.io/MoGAPI/ (Accessed Nov 17, 2024).