



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

Assessment of blood serum stability with Raman spectroscopy and explanatory AI

Verónica Mieites^{a,b}^{*}, María Gabriela Fernández-Manteca^{a,b}, Inés Santiuste Torcida^b, Fidel Madrazo Toca^b, María José Marín Vidalled^b, Olga M. Conde^{a,b,c}

^a University of Cantabria, Photonics Engineering Group, Plaza de la Ciencia S/N, Santander, 39005, Cantabria, Spain

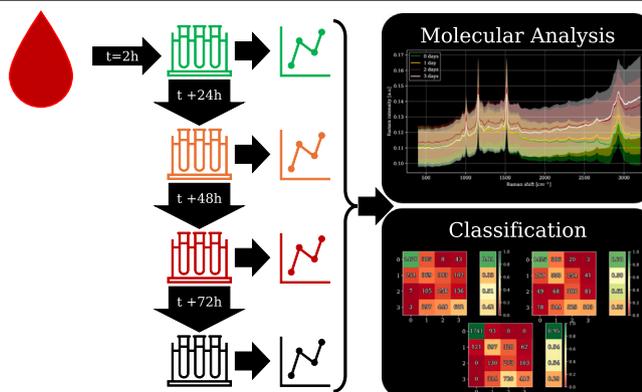
^b Valdecilla Health Research Institute (IDIVAL), Calle Cardenal Herrera Oría S/N, Santander, 39011, Cantabria, Spain

^c Biengineering, Biomaterials and Nanomedicine Research Network (CIBER-BBN), Calle de Melchor Fernández Almagro 3, Madrid, 28029, Spain

HIGHLIGHTS

- Raman analysis reveals stable serum analytes after ambient temperature storage.
- Training with the first 20 PCs yields high accuracy in degradation classification.
- Explainability of Raman data points to auto-fluorescence as a marker of serum degradation.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:

Raman spectroscopy
Serum degradation
Quality control
Storage delay
Spectral analysis
Explainable AI (XAI)

ABSTRACT

This study explores the potential of conventional Raman spectroscopy and commonly used spectral analysis pipelines for rapid and straightforward assessment of degradation in serum samples resulting from storage delays. Serum samples from 18 volunteers were processed within 2 h of extraction, which later on were analyzed via Raman spectroscopy over 4 days, while the corresponding serum vials were kept at room temperature. The resulting spectra were processed, including silicon normalization and a newly proposed outlier detection ensemble method. Next, baseline correction was performed, and spectral unmixing along with Principal Component Analysis (PCA) were applied. Several classification models (KNN, RF, and SVM) were trained and evaluated on three distinct balanced datasets: one including all data, one excluding low signal-to-noise ratio (SNR) data, and one excluding low-SNR data with baseline correction. Feature importance, assessed through random permutations, was used for explainability.

Spectral unmixing and PCA indicated limited spectral changes directly attributable to analyte degradation, with inter- and intra-sample variability dominating. Classification results showed that while removing the baseline led to inconclusive results, models trained on datasets retaining the baseline effectively identified non-degraded samples. These findings suggest that while conventional Raman spectroscopy may not be optimally sensitive to subtle analyte variations in serum stored at room temperature, the auto-fluorescence background holds promise as a potential biomarker for monitoring serum storage quality.

* Corresponding author at: University of Cantabria, Photonics Engineering Group, Plaza de la Ciencia S/N, Santander, 39005, Cantabria, Spain.
E-mail address: mieitesv@unican.es (V. Mieites).

<https://doi.org/10.1016/j.saa.2025.126297>

Received 25 October 2024; Received in revised form 23 April 2025; Accepted 24 April 2025

Available online 9 May 2025

1386-1425/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

1. Introduction

Sample handling is a crucial factor in all research areas since the way samples are treated, stored and maintained can heavily influence any derived result. Defining well-established quality control protocols is essential to ensure truthful and reproducible results. Biobanks are institutions designed to certify good sample quality. Their primary role is to assure proper sample conditions, both before and during storage, so that their materials can be used in any line of research with consistent results [1]. Variations in data provided by biobanks will be due to donor conditions and not to the introduction of artifacts during manipulation and storage [2]. However, samples that arrive at research facilities originate from various sources, such as hospitals, universities, or other research institutions, creating challenges in confirming their validity [3]. This issue is particularly critical with common samples, specifically with blood specimens, due to their wide availability and their sensitivity to changes in pre-analytical conditions [4]. In the last twenty years, great efforts have been made to find biomarkers that can serve as quality standards, to establish different quality control methods for fluid samples and to define sample conditions that suit different purposes. In the context of biomarker quantification, stability can be defined as how much an analyte deviates from initial concentrations over time [5]. In whole blood, serum, and plasma, many biomarkers have been extensively studied to assess their responses. Most studies focus on the need for refrigeration (+4 °C) or freezing (−20 °, −70 °C) to preserve analyte activity. Among others, different methods such as electrophoresis, chromatography, ELISA, mass spectrometry, and gas spectrometry, have been used to evaluate the stability of over 255 metabolites in serum [6–10], plasma [11], and both simultaneously [12–16]. Researchers have also examined the impact of extended precentrifugation time and of changes in precentrifugation storage temperature on serum and plasma properties [17–21]. Still, the diverse range of methods and analytes studied makes it challenging to establish a universal standard for evaluating serum stability (Tables 1 and 2).

Most of the mentioned methods demand some degree of sample preparation, a relatively substantial sample size, high costs, or the need for specific labeling, making them impractical for routine implementation within a laboratory's regular workflow, as testing every sample would be excessively time-consuming. Additionally, all of them search for specific analyte changes, therefore those methods are not capable of detecting global changes in a biological sample that would indicate that it has undergone a process that makes it no longer usable.

Utilizing inelastic scattering, Raman spectroscopy offers a molecular-specific method for molecular identification and chemical bond analysis [22]. It provides a “spectral signature” reflecting the biochemical composition of a sample [23]. Advantages of Raman spectroscopy include minimal sample preparation, low sample volume requirements, and its non-destructive nature, making it suitable for analyzing solid and liquid samples. This technique has been used to study blood components since the 1970s, with ongoing innovations in instrumentation and applications [24,25].

The multiple modalities of Raman Spectroscopy have been used before to detect chemical changes through traditional chemometry. For example, Magdas et al. applied a variation of linear discriminant analysis to distinguish between wine types using Fourier-Transform Raman [26,27]. Similarly, Ortiz et al. applied principal component analysis for the discrimination of real and counterfeit pharmaceutical powders [28]. Açıkgöz and Hamamci also applied a least squares method discriminant analysis to determine specific analytes in blood samples [29]. Neural-networks-based methods have also been used in conjunction with Raman spectroscopy, specifically in health-related experiments, to classify sub-types of breast cancer [30], to diagnose lupus [31] or to analyze cortical bone [32]. Nonetheless, there has been a recent uprising of explainability-based methods (explainable AI, XAI) in multiple fields [33–38] in order to not only provide chemometric or classification results but also to try to understand their biological

basis. Specifically in Raman, XAI was used to predict carbon-to-oxide ratios [39] and to diagnose thyroid [40] and breast [41] cancer, in a more informed way than using conventional chemometrics or classifiers only.

This work proposes the use of Raman spectroscopy as a potentially more rapid and straightforward alternative to existing serum analysis techniques for assessing changes in serum samples due to storage delays. The aim is to investigate the feasibility of using Raman spectroscopy, coupled with common analysis pipelines, to detect serum degradation. Through the analysis of Raman spectra from serum samples left at ambient temperature for multiple days, we studied the evolution of the main Raman bands corresponding to serum components using both supervised (spectral unmixing) and unsupervised (principal component analysis) techniques, combined with classification, and explainable AI (XAI) methods. The latter were incorporated as a diagnostic tool to investigate the shortcomings of the proposed analysis pipelines for the inspection of serum samples. This evaluation will highlight both the potential and the challenges associated with the use of Raman spectroscopy for this task.

2. Materials and methods

This section details the materials and methods employed to obtain the primary findings of this study.

2.1. Measurement protocol and system parameters

Serum samples were obtained from 38 volunteers, aged between 40 and 69 years, residing in the northern region of Spain. The samples have been acquired in the context of a cohort, Cohorte Cantabria, designed to provide information on morbidity related to lifestyle and socioeconomic aspects [44]. Blood samples were centrifuged to separate serum from the cellular fraction within 2 h after extraction, according to the faster processing time advised by the SPREC nomenclature [45]. A droplet of serum of 10 μ L was deposited on a microscopy slide and left to dry for 15 min inside a custom-made water evaporator (see Fig. S1 in the supplementary material), which minimizes the occurrence of cracks in the surface of the droplets and speeds up the drying process. Aluminum slides are used for the measurements instead of conventional fused silica slides to avoid the intrinsic Raman signal of fused silica, which would interfere with the molecular fingerprint of the samples [24,46–48]. The aluminum slides used in this work also help reduce measurement noise without the increased cost of using gold, silver, or SERS substrates [49,50]. The tube containing the remaining serum of each individual was left at room temperature for 24 h and then a second 10 μ L droplet was measured with Raman following the same procedure. This method was repeated over four days, creating a spectral dataset of 38 volunteers at four temporal states of sample degradation, labeled according to the SPREC nomenclature [45]: 00 h (Control, SER-SST-A-B-N-B-A), 24 h (SER-SST-A-B-N-H-A), 48 h (SER-SST-A-B-N-J-A) and 72 h (SER-SST-A-B-N-M-A). Out of the 38 volunteers, only the samples from 18 could be measured on four consecutive days. Therefore, the molecular analysis and classification results are provided for those 18 volunteers to ensure that the classes are balanced. The data from the remaining 20 volunteers was used as a test set for the classification tasks.

Raman spectra were acquired using a Jasco NRS-4500 spectrometer with a 532 nm laser (18 mW power). A 15 \times 15 grid of measurements was collected per droplet, with 170 μ m spacing to prevent laser-induced sample damage, using a 50 \times objective (3 μ m spot size). Spectra were obtained using a 17 μ m circular aperture, a 1200 lines/mm diffraction grating (0.5 cm^{-1} resolution), a 1 s exposure time, and 15 accumulations per data point. These parameters were chosen to optimize the signal-to-noise ratio (SNR) and minimize sample heating.

Table 1

Compilation of studies evaluating the stability of various analytes in serum and plasma under different pre-analytical conditions. Part I.

Blood product	Temp.	Analytes	Conclusions	Ref.
Serum, whole	RT, 6 °C	Vitamin D	Stable	[42]
Serum, plasma	RT, 4 °C, 30 °C, -20 °C	Hormones	All hormones but ACTH stable at 4 °C, 13 hormones stable in anticoagulants, 8 stable at 30 °C, BNP and NT-BNP stable for <24 h at RT	[13]
Serum	22 °C, 4 °C, -20 °C	25 analytes	All analytes but carbon dioxide, aspartate and alanine aminotransferases, lactate dehydrogenase, alkaline phosphate and cholesterol stable at all temperatures and times	[7]
Serum, plasma	RT	24 analytes	All stable over 56 h, but unstable when the specimens have delayed centrifugation times	[14]
Serum, plasma	RT	Folate	Unstable	[15]
Serum, plasma	RT, 4 °C, freeze-thawed	Endogeneous and added RNA	Under pre-centrifugation delays, stable in plasma at 4 °C, unstable in serum at 4 °C. With never-frozen, freeze-thawed and thawed plasma and serum at RT, stable.	[18]
Serum	RT, refrigerated, frozen, freeze-thawed	14 analytes	All stable but insulin and gastrin.	[12]

Table 2

Studies evaluating the stability of serum and plasma. Part II.

Blood product	Temp.	Analytes	Conclusions	Ref.
Serum	RT	225 analytes	With pre-centrifugation delays all 255 are unstable. With extended serum storage, 14% of the 225 increased, and 7% decreased. Lipids were stable but amino acids and nucleobases were not.	[19]
Serum	RT, wet ice, dry ice, freeze-thawed	127 metabolites	All stable in freeze-thawed cycles, wet ice, and dry ice. At RT, most were unstable, including phenylalanine amino acids, glycine and arginine. Amino acids with low protein frequency (i.e., tryptophan) were stable.	[10]
Serum	RT, 4 °C	Glucose, urea, phosphorus, creatinine, uric acid, cholesterol, triglyceride, total protein, albumin and calcium	Glucose, phosphorus and creatinine unstable. The rest, stable over 72 h.	[8]
Serum	RT, refrigerated, frozen, freeze-thawed	Fatty acids	Stable	[43]

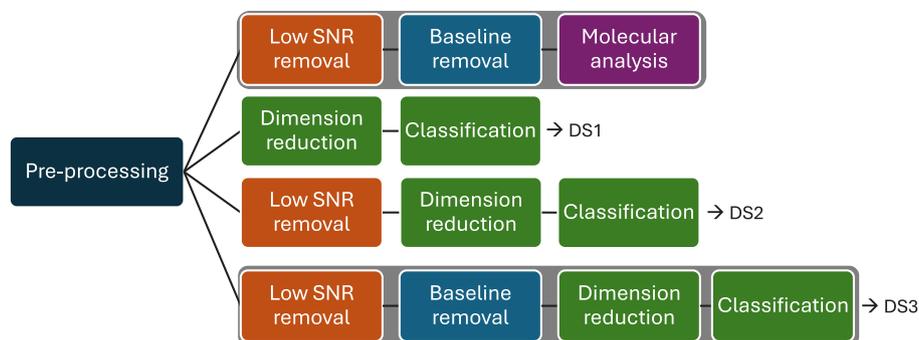


Fig. 1. Description of the analysis pipeline implemented in this work. The *pre-processing* block includes those steps that are common for all results, including outlier removal, data balancing, and silicon normalization, in that order. The shaded branches represent the common analysis pipelines implemented in Raman spectroscopy for molecular analysis (first row) and classification (last row). The details of each block in this figure are included in the following sections. DS1 (balanced), DS2 (balanced, high-SNR), and DS3 (balanced, high-SNR, no baseline) indicate the different datasets that will be used to evaluate the multiple combinations of spectral pre-processing for classification.

2.2. Spectral preprocessing

The analytical pipelines developed and utilized in this work for spectral analysis are illustrated in Fig. 1 and will be thoroughly described below.

Notably, we implemented a series of incremental classification steps to assess the impact of baseline removal and the presence of low signal data on the results. This approach aims to determine the necessity of additional processing for accurate detection of serum degradation.

2.2.1. Outlier detection and data balancing

To ensure data quality, outlier spectra were removed prior to analysis. Outliers can arise from sample imperfections (e.g., cracks, uneven surfaces), sample burning, out-of-focus measurements, or saturated spectra.

We propose an ensemble outlier detection strategy that incorporates five distinct methods: the boxplot method, which identifies outliers based on deviations exceeding 1.5 times the Interquartile Range (IQR) of Raman intensities [51]; Isolation Forest (iForest), an algorithm that isolates outliers by randomly partitioning data, with outliers requiring fewer partitions [52]; One-Class SVM (OC-SVM), which constructs a global boundary around the data, classifying anything outside as an outlier [53]; Local Outlier Factor (LOF), which calculates the local density deviation of a data point with respect to its neighbors (five neighbors), identifying points with significantly lower neighbor density as outliers [52]; and Angle-Based Outlier Detection (ABOD), which analyzes the variance in angles between data points to identify those with unusually high or low angle variances, indicating potential outliers [54].

Aside from the boxplot method, all others have one threshold-related hyperparameter used to establish what is classified as an outlier. In the case of the methods implemented in this article, this hyperparameter can be understood as the expected percentage of outliers in the data, which we established at 10% for the serum samples. As Raman spectra of uniform samples should be uniform, this intentionally high value implies that a significant number of data points could be misclassified as an outlier.

To mitigate the impact of potential false positives and method-specific biases, we implemented a consensus approach. A spectrum was identified as an outlier and removed only if at least four out of the five methods agreed on its classification. This consensus strategy enhanced the robustness of our outlier detection, ensuring that only spectra consistently identified as outliers were removed, while mitigating the weaknesses of each individual method. The raw, 18-patient dataset contained a total of 29,644 spectra, corresponding to 132 droplet measurements across all patients and four conditions. The consensus outlier detection method identified 2575 outliers, representing 8.6% of the total data.

Dried sample droplets often exhibit uneven surfaces due to water's surface tension, which prevents them from drying completely flat. This leads to inter- and intra-sample variability, characterized by low SNR samples spectra originating from out-of-focus regions. These effects are exacerbated by the reduced analyte concentration in the thinner areas of the sample. Consequently, the next step involved retaining only spectra with intensities exceeding the median intensity, effectively selecting well-focused regions with higher analyte concentrations. By utilizing the median, this process removed 50% of the outlier-free dataset, resulting in 13512 well-focused, high SNR spectra.

To mitigate the risk of biased models due to data imbalance [55, 56], the dataset was balanced by reducing the number of spectra per volunteer and per category to the lowest count observed within any single volunteer or category. This process resulted in a further 52% reduction of spectra, yielding a final dataset of 7056 spectra (from an initial 13512), distributed across 18 volunteers, with 392 spectra per volunteer, and 1764 spectra per category (0d, 1d, 2d, 3d). This balanced approach ensures that the model effectively learns class-specific features and minimizes the potential for overfitting.

2.2.2. Intensity normalization

To ensure accurate Raman measurements, the spectrometer is calibrated daily using a silicon sample after 1 h of thermal stabilization. Silicon, with its well-defined and strong Raman signal at approximately 521 cm^{-1} [57–59], serves as an ideal calibration standard. Specifically, the peak's position is used to calibrate the Raman shift and its intensity, recorded at each start-up, to normalize the Raman spectra of samples measured on the same day, thus accounting for daily variations on the instrument response.

2.2.3. Baseline correction

To remove the characteristic fluorescence of biological samples and focus only on the analytes, we compared multiple baseline removal methods (Table 3), including Modified Polynomial (ModPoly) [60], Asymmetric Least Squares (ALS) [61], Adaptive Iteratively Reweighted Penalized Least Squares (AIRPLS) [62], Statistics-Sensitive Nonlinear Iterative Peak-Clipping (SNIP), Morphological (Mor) [63,64] and Bubble-Fill [65]. Each of them is based on a different approach and dependent on different hyperparameters (Table 3 and Sec. 3 of the supplementary material) whose value yields different results.

2.3. Molecular analysis

Spectral unmixing was used to determine the composition of each spectrum by modeling it as a linear combination of twelve primary endmembers (see Table 4) [66–68]. These endmembers represent the most frequent spectral peaks observed in the Raman measurements. Each peak was modeled using a Gaussian profile, and the final spectrum was

Table 3

Baseline removals methods tested for this work. The processing time was measured for 13104 spectra without parallelizing the calculations. The results were achieved using a 10th Gen. Intel Core i7 Processor, with 64 Gb of RAM.

Method	Processing time (s)	Fundament	Hyperparameters
ModPoly	23	Polynomial fitting	Polynomial order
ALS	50	Whittaker-smoothing	Smoothing parameter, penalizing weighting factor
AIRPLS	23	Whittaker-smoothing	Smoothing parameter
SNIP	28	Local statistics	Window size
Mor	35	Morphological processing	Window size
BubbleFill	38	Morphological processing	Minimal bubble width

Table 4

Chemical bonds [72] associated with the twelve most frequent peaks detected across spectra in our dataset, sorted by increasing Raman shift (wavenumber).

Component	Raman shift [cm ⁻¹]
Phenylalanine	623.4
Tryptophan	878.4
Calcium	962.9
Phenylalanine	1005.4
Carotenoids	1157.4
Phenylalanine	1199.9
Phenylalanine	1212.9
Amide III	1286.9
Phospholipids	1447.9
Carotenoids	1515.9
Phenylalanine	1608.4
Amide I	1655.9

defined as the sum of these Gaussian peaks [69,70]. While the Voigt profile [71] is often more appropriate for Raman spectra, we opted for the Gaussian profile to significantly reduce computation times. This trade-off is considered reasonable as our subsequent analysis concentrates on peak regions, where the Gaussian and Voigt profiles show minimal divergence. For a comprehensive comparison of their fitting accuracy, please refer to Sec. 4E and Fig. S2 of the supplementary material.

Formally, each spectrum is described as follows. Let I be the detected Raman intensity at the wavenumber (Raman shift) k . Let a_i and b_i be the peak height and standard deviation, respectively, for a specific endmember i . Then, Eq. (1) was used to fit the spectra to any desired number of endmembers, by setting the a_i and b_i as free fitting parameters. In this work, we selected the number of endmembers as the twelve more frequently detected peaks across the dataset, since these exhibited the highest SNR.

$$I(k) = \sum_{i=1}^{i=12} a_i \cdot \exp\left(-\frac{(k - k_i)^2}{b_i^2}\right) \quad (1)$$

2.4. Dimensionality reduction

While Raman spectroscopy provides information on numerous molecular bonds, not all of them are relevant for every analysis. Chemometric analysis helps focus on specific peaks but often discards weaker signals. Peak ratios can also serve as biomarkers [32,73,74], though their selection typically relies on prior knowledge of chemical behavior. Alternatively, dimensionality reduction techniques have gained importance in the field due to the providing a low-dimensionality space (latent space) that is easier to work with, without having to discard a significant amount of data [75,76]. One of the most common dimensionality reduction techniques used in Raman spectroscopy is Principal Component Analysis (PCA) [77–79]. This method is based on finding the orthogonal directions that explain the greatest amount of variance in the data. Then, data can be mapped onto a space defined by said directions with the transformation given by Eq. (2) [80].

$$P = W^T X \quad (2)$$

In Eq. (2), X is the matrix containing the individual spectra, P the matrix that represents the data in the new space, i.e. the principal components (PCs), and W is the (weights) matrix containing the linear transformation that maximizes the variance explained by the consecutive P items. The dimensionality reduction property of PCA comes from the fact that the transformation is chosen to explain a high amount of the variance present in the data with fewer principal components than features in the original space.

Feature standardization is a necessary step to be performed before applying PCA [81]. Due to PCA being a variance-based method, if the dataset is comprised of features in different scales, those features centered around greater values will tend to have higher valued variances that will bias the calculations of PCA. Feature standardization re-scales the data to have an average of zero and a standard deviation of one so that PCA focuses only on the differences in the data and not on their scale.

To assess the separability of classes in the latent space, we employed the Fisher ratio (F_r), a metric that quantifies the separation between classes based on their means (μ_i), standard deviations (γ_i), and sample sizes (n_i), relative to the overall dataset mean (μ) [82], as indicated in Eq. (3). Each class represents the PCA-reduced spectra in this analysis at a specific time point (00 h, 24 h, 48 h, and 72 h). A higher F_r indicates greater inter-class distance and intra-class compactness, signifying well-separated and distinct classes.

$$F_r = \frac{\sum_{i=1}^C n_i (\mu_i - \mu)^2}{\sum_{i=1}^C n_i \gamma_i^2} \quad (3)$$

2.5. Classification

For this article, we wanted to test whether commonly used classification methods were enough to separate the samples by their deterioration state. Reducing the dimensionality before classifying ensures that all irrelevant data will be removed by the dimensionality reduction process. By doing so, noisy data will not hinder the classification process. For the classification, we used a K-Nearest Neighbors (KNN) classifier [83–85]. KNN classification is based on looking at the K closest neighbors of a given point to see which class they belong to. Then, a vote is cast under the principle that each sample should be in the same class than the majority of its neighbors [86]. Repeating this process for all data points divides the space according to the most voted classes.

In addition to KNN, we incorporated a Random Forest (RF) classifier [87,88]. This ensemble method functions by generating a multitude of decision trees, each trained on a random subset of the data and a random selection of features. For a given data point, each tree casts a vote for a particular class, and the Random Forest ultimately assigns the data point to the class that receives the majority of the votes. This aggregation of predictions from multiple trees typically leads to improved accuracy and reduced overfitting [87,88].

Finally, we used a Support Vector Machine (SVM) classifier. The principle of SVM lies in identifying a decision boundary, known as a hyperplane, that best segregates data points of distinct classes within a potentially high-dimensional feature space. The algorithm seeks to find

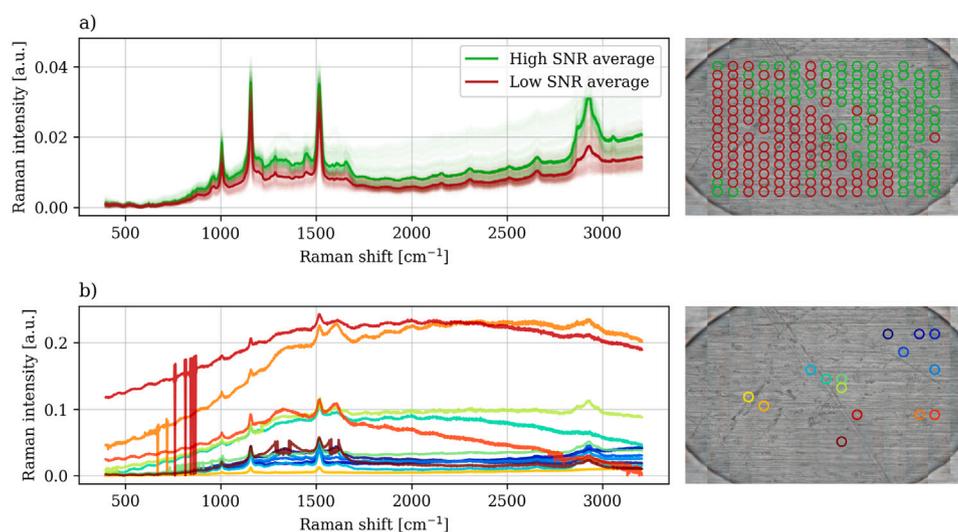


Fig. 2. Raman spectra measured on one of the serum droplets (a). High SNR data is depicted in green, and lower SNR data in red. The image of the droplet is also shown in the right column. In (b), the outliers measured for this droplet are shown, as well as their position in the droplet via color-coding.

the hyperplane that yields the largest margin, defined by the shortest distance between the hyperplane and the closest data points from each class [89–91].

The accuracy of N the predictions by the classifiers is calculated as

$$ACC = \sum_{n=1}^N \frac{1}{N} \begin{cases} 1, & \text{if } y_n = \hat{y}_n \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

where y is the expected class and \hat{y} the predicted class of a spectrum. The classifiers described in this article are also evaluated according to their per-class sensitivity, which indicates the accuracy of the classifier at correctly classifying data from one class. To further validate our results, we performed Leave-One-Out Cross-Validation (LOOCV), which consists of dividing the dataset into train and validation sets by keeping all of the data, in this case, of one individual (four classes) for validation and the rest for training. LOOCV is done as an alternative to randomly dividing the dataset into two groups. When the division is done randomly, some spectra of the same volunteer might end in the train group and some in the validation group (data leakage), which makes the classifier more accurate but not because it finds more features, but because it has trained on the same volunteers it is validating the accuracy on. Models trained on randomly divided data that keep the same individuals in both groups tend to give high accuracy values that are not generalizable when the classifier encounters a never-before-seen individual.

2.6. Explainability

Accuracy and sensitivity are good indicators of the performance of a model, but do not give any information about what parameters are helpful or relevant for the model to work. Although dimensionality reduction is used in an effort to remove less relevant information, it is possible that the dimension-reduced dataset still contains data that is not necessary to provide good classification results. In the context of this article, the aim is to know which principal components are the most important for improving the KNN classifier accuracy, but also what are the possible sources of classification error if the models fail. For this purpose, we opted for applying the random permutations method, which has recently been applied to multi-wavelength data [92]. This method focuses on randomly shuffling the data of each input feature (PC) and measuring the resulting decrease in classification accuracy. A significant drop in the accuracy after shuffling indicates the high importance of that feature. This process was repeated ten times for each feature (PC), fold, and model to ensure reliability.

3. Results and discussion

3.1. Preprocessing

3.1.1. Raw spectra

Fig. 2 depicts a serum droplet measurement. A 15×15 grid of measurement points was defined on the droplet (Fig. 2, right column), where Raman spectra were acquired. However, not all spectra were usable. Some (Fig. 2, b) exhibited detector saturation, or excessive auto-fluorescence, and were thus identified as outliers and removed according to the methodology detailed in Section 2.2.1.

Furthermore, the removal of low signal-to-noise ratio data further reduced the dataset size (Fig. 2, a), but significantly improved data quality, retaining only the most reliable spectra for analysis. This improvement is particularly evident in the fingerprint region of the spectrum (approximately $500\text{--}1800\text{ cm}^{-1}$). While the main spectral bands are visible in both average spectra, features such as those at 1450 cm^{-1} and 1655 cm^{-1} are much more pronounced in the high-SNR spectra.

3.1.2. Normalization and baseline correction

The silicon peak measurements taken for calibration are shown in Fig. 3. The samples have been calibrated with respect to the maximum intensity measured in silicon, as described in the methods section.

The performance of the six baseline correction methods is shown in Fig. 4. Of the tested methods, only ModPoly uses the overall spectrum instead of looking at its local features, which makes it so that there are spectral regions where the baseline is overestimated or underestimated (Fig. 4, a). Additionally, hyper-parameter tuning is required for all methods, but those that are Whittaker-based have the lowest interpretability since the user requires a deep understanding of the implementation in order to choose appropriate parameters, even if the baseline is properly estimated (Fig. 4, b and c). Morphological baseline estimation was not the slowest method (Table 3), but it also led to a “staircase” effect that would require further smoothing to be used (Fig. 4, e), which will also be dependent on additional hyper-parameters. Finally, even though both SNIP (Fig. 4, d) and BubbleFill (Fig. 4, f) use a morphological element to estimate the baseline, BubbleFill’s is defined so that it can be directly set as the width of the widest peak in the spectra, which in our case is the asymmetric O-H stretch at approximately 3400 cm^{-1} related to the water content [93]. On the other hand, setting SNIP’s window size is not straightforward, since it prefers the use of small windows to avoid spreading the statistics of the

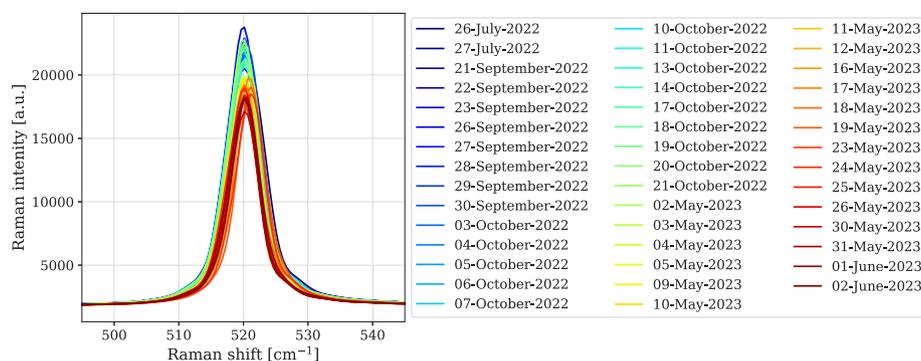


Fig. 3. Raw intensity measurements of the silicon calibration piece along the duration of the experiment.

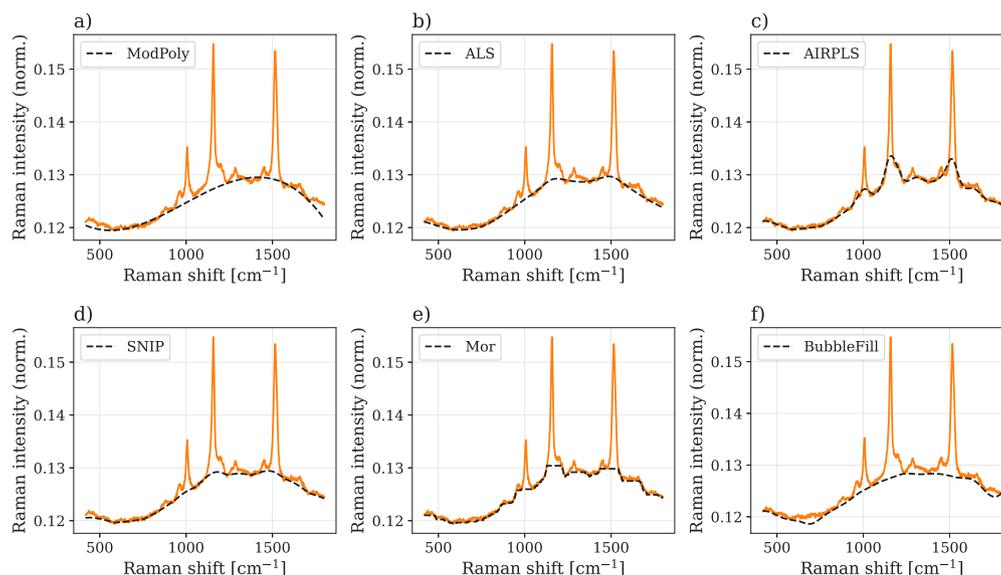


Fig. 4. Comparison of the different baseline methods (black dashed line) over an example spectra (orange solid line), including ModPoly (a), ALS (b), AIRPLS (c), SNIP (d), Mor (e) and BubbleFill (f).

peak's features. The previous evaluation led us to the use of BubbleFill with a minimum bubble size of 400 cm^{-1} as the smoother of choice for this work.

The dataset before and after baseline removal is depicted in Figs. 5 and 6, respectively. Removing the baseline severely reduces the standard deviation in general, visible only near prominent peaks. This indicates that the greater variation is found in the presence of analytes with strong spectral bands. Additionally, the peaks are more prominent and easier to see after removing the baseline, especially those related to the phenylalanine (around 1000 cm^{-1}) and carotenoids (near 1100 and 1500 cm^{-1}).

3.2. Molecular analysis

Peak detection was performed on the serum dataset after pre-processing. The twelve most frequently detected peaks were used to decompose the spectral measurements. On the serum, those peaks (Fig. 7, top) in our dataset are associated with the components detailed in Table 4. Most of the commonly detected peaks were those associated with vibrational modes of phenylalanine (5 bands), followed by those of carotenoids (2 bands). The rest of the peaks corresponded only to one component, i.e., calcium, amides I and III, tryptophan, and phospholipids.

Most bands exhibited a decrease in average and median intensity over time. However, there were exceptions. Phospholipids

(1447.9 cm^{-1}) and amide I (1655.9 cm^{-1}) showed a slight increase on the second day before decreasing. Phenylalanine (1608.4 cm^{-1}) also behaved differently, increasing on the third day but remaining otherwise stable. Interestingly, the consistent detection of phenylalanine and tryptophan aligns with the results found by Anton et al. who identified changes in these components using a targeted metabolomics approach [10]. Nevertheless, the distributions on different days showed substantial overlap, which is caused by inter- and intra-sample variability. Compared to this variability, the degradation-related changes were minor. Therefore, it is difficult to assess sample degradation using these peaks alone.

3.2.1. Datasets

Three different datasets were created for classification based on the intensity normalized, balanced data without outliers: (1) with all the spectra (i.e., balanced dataset), (2) with only high-SNR data (see Section 3.1.1, i.e., balanced, high-SNR) and (3) with high-SNR data and baseline removed (see Section 3.1.2, i.e., balanced, high-SNR, no baseline).

Within each dataset and sample degradation category (0–3 days), we calculated three key spectral statistics related to the spectral spread and symmetry: the average of the per-wavelength standard deviation (dispersion), the average of the per-wavelength kurtosis (shape), and the average of the per-wavelength skewness (asymmetry). The results are shown in Table 5.

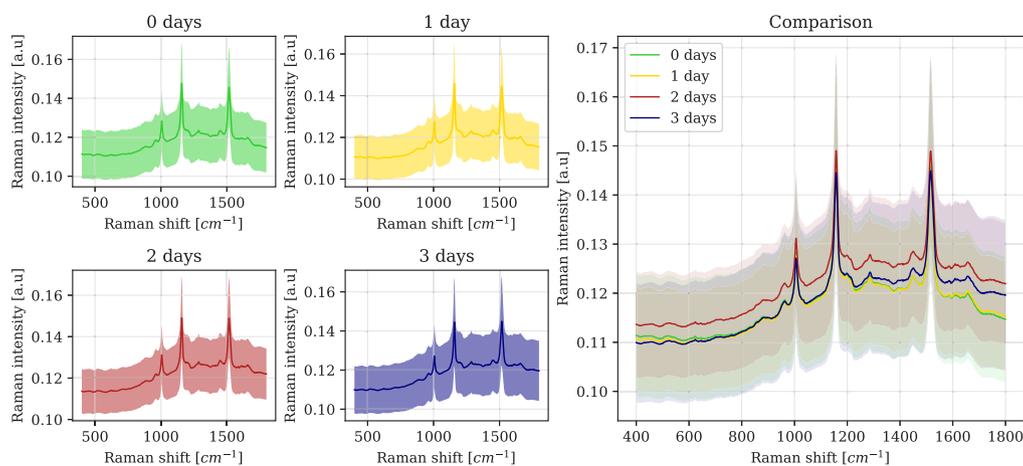


Fig. 5. Balanced dataset before baseline removal. The solid lines represent the average per day, and the shaded areas represent the standard deviation of the Raman spectra.

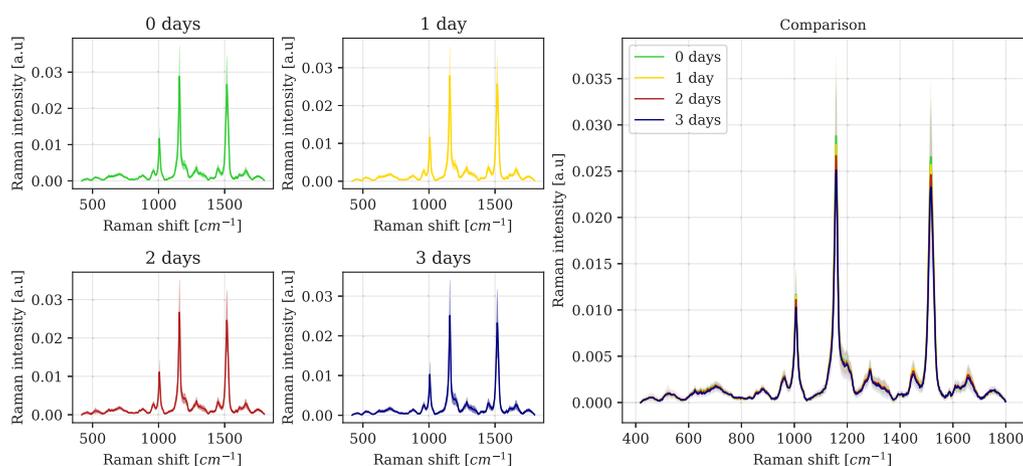


Fig. 6. Balanced dataset after baseline removal. The solid lines represent the average per day, and the shaded areas represent the standard deviation of the Raman spectra.

Table 5

Statistical descriptors (Standard (Std) deviation (Dev), Kurtosis, Skewness) for the three Raman spectra datasets.

Days	Balanced Std. Dev.	Kurtosis	Skewness
0	0,105	-1,146	0,245
1	0,095	-0,904	-0,025
2	0,102	-0,313	-0,127
3	0,119	-0,832	0,105
Days	Balanced and high-SNR Std. Dev.	Kurtosis	Skewness
0	0,074	-1,316	0,213
1	0,063	-0,821	-0,003
2	0,073	-0,116	0,339
3	0,097	-0,201	0,258
Days	Balanced, high-SNR, no baseline Std. Dev.	Kurtosis	Skewness
0	0,013	35,215	-1,670
1	0,012	5,119	0,561
2	0,014	73,194	-2,923
3	0,014	11,249	-1,103

The balanced dataset demonstrates consistent statistical properties across its four categories, with comparable spectral variability (standard deviations of 0.095–0.119). The negative kurtosis values (−1.146 to −0.313) indicate platykurtic distributions with flatter peaks and thinner tails, while the near-zero skewness (−0.127 to 0.245) suggests relatively symmetrical distributions within each category, with

minor tails towards higher or lower intensities in some cases. Overall, this suggests broad spectral features and symmetrical distributions with consistent intensity and variability across the balanced dataset's categories.

The balanced and high-SNR dataset exhibits lower spectral variability (standard deviations of 0.063 to 0.097) compared to the balanced

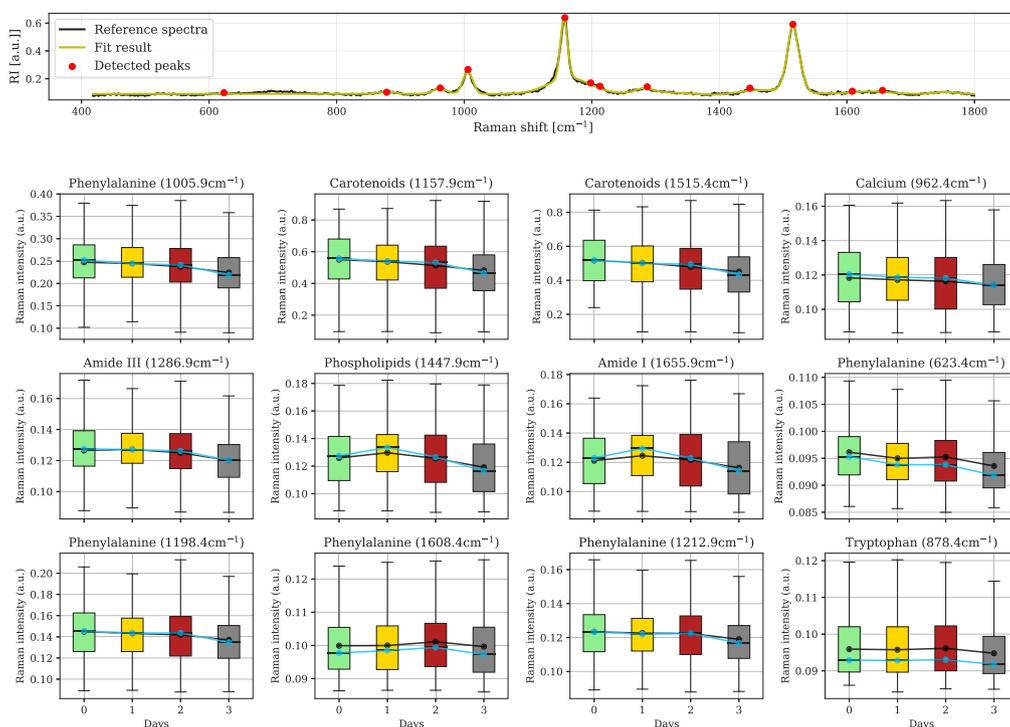


Fig. 7. On the top, one reference fit is shown, with the result in yellow, the reference spectra in black, and the position of the twelve peaks in red. On the bottom, the boxplots of the evolution across days of the twelve most frequently detected chemical bonds in serum are represented, sorted from most frequent to least frequent. The blue line represents the evolution of the median, and the black line, the evolution of the average.

dataset, suggesting that removing the out-of-focus samples further reduces spectral variations. Similar to the balanced dataset, the kurtosis values are negative (-1.316 to -0.116), indicating platykurtic distributions that may be even broader or flatter after focusing. The skewness values remain close to zero (0.003 to 0.339), suggesting largely symmetrical distributions with minor tails towards higher intensities in some categories.

The balanced, high-SNR, without baseline dataset shows the most significant statistical changes, with very low spectral variability (standard deviations of 0.012 to 0.014) due to baseline removal and focusing. Notably, the kurtosis has become highly positive (5.119 to 73.194), indicating leptokurtic distributions with sharp, prominent spectral peaks, varying significantly in sharpness and tail behavior across categories. The skewness values are more pronounced and both positive and negative (-2.923 to 1.670), suggesting asymmetrical spectral distributions.

The three datasets illustrate the impact of processing on Raman spectra's statistical characteristics. The balanced dataset provides a starting point with moderate variability, broad features, and symmetrical distributions. By removing the out-of-focus spectra in the balanced and high-SNR dataset, the variability is reduced while maintaining a generally platykurtic and symmetrical nature. The balanced, high-SNR, without baseline dataset shows the most significant changes; baseline removal combined with focusing isolates sharp, asymmetrical spectral peaks, evidenced by high positive kurtosis and more pronounced skewness. This progression demonstrates numerically how processing transforms spectral distributions from broad and symmetrical to sharp and potentially asymmetrical peaks, emphasizing different spectral information aspects in each class, consistent with the graphical representation of the dataset (Section 3.1.2). The balanced, high-SNR, no baseline dataset appears most informative for distinguishing categories based on the shape and asymmetry of their key Raman bands, as all the inter-class variation associated with different baseline levels is effectively removed.

3.2.2. Dimensionality reduction

Crucially, within each cross-validation fold, the validation group was standardized using parameters (mean and standard deviation) calculated solely from the training group. Similarly, principal component analysis was fitted exclusively to the training data within each fold, and the learned transformation was then applied to the corresponding validation data. Standardizing the validation data or performing PCA independently would constitute data leakage, where information from the unseen validation set contaminates the preprocessing, resulting in overly optimistic and unreliable performance metrics.

The first 20 principal components with PCA were kept according to a cumulative variance criterion (keeping more than 97.5% of the variance of the data), and were sorted by importance according to the Fisher ratio (Eq. (3)). The remaining PCs were not considered due to the low variance explained by each of them, which indicates that they represent mostly noise. The principal components that maximized separability by class differed between the three datasets (Fig. 8). PCA sorts the principal components in terms of variance, which means that the first PCs usually have very well-defined spectral features that turn into noise as the PC number increases. The fact that PC1 is Fisher-chosen only in the second dataset (Fig. 8, b) means that the PC that keeps the most variance is the one with the best separability only when the low-SNR data are removed, but the baseline is kept.

For the first dataset (Fig. 8, a), data collapses to zero on PC6 even though it was selected by the Fisher ratio as the one with the most separability between classes. A possible cause for this is that there are some spectra that have a value far from zero for PC6 at the sides of the main distribution, which means that its average will not be highly affected by them, while its standard deviation will. By the definition of the Fisher ratio, PC6 will increase in value due to these spectra while the separability stays mostly unchanged. For PC3 there is a wider spread of values but all categories have their bulk of data on the same position. However, we do see that the first two classes are virtually indistinguishable from each other, while the third and fourth days have slightly wider distributions of PC3.

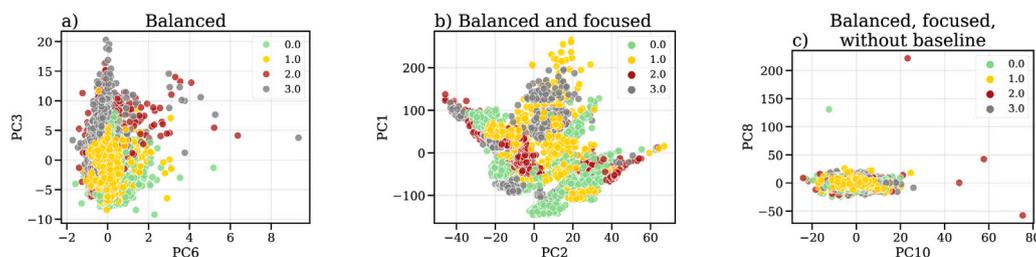


Fig. 8. Projection of the spectroscopic measurements over the latent space conformed by the two PCA principal components that maximize separability according to the Fisher ratio, for the intensity-normalized, balanced data without outliers (a), with additional removal of out-of-focus spectra (b), and with additional baseline removal (c). The labels 0, 1, 2, and 3 refer to the number of days the samples have been left at ambient temperature.

Table 6

Table showing the total training and evaluation times for three distinct datasets with the three classifiers. Results were generated using an 18-fold leave-one-patient-out cross-validation strategy, with feature performance included in the time, evaluated via 10-fold random permutations within each fold's train and validation splits. The models were trained using a 10th Gen. Intel Core i7 Processor, with 64 Gb of RAM.

	KNN	RF	SVM
Balanced	11.5 min	56.1 s	4.7 h
Balanced, high-SNR	3.2 min	42.6 s	27.6 min
Balanced, high-SNR, no baseline	3.4 min	29.56 s	1.3 h

By removing low-SNR spectra (Fig. 8, b), the data projection on the (PC2, PC1) plane has less overlap between classes, even if the individual labels do not collapse to a single cluster. Still, most of the clusters keep having some degree of overlap, indicating that the variance in the data is not due to differences between classes as much as due to variations within the same class (intra-sample variability).

Removing the baseline (Fig. 8, c) leads to a completely different projection map of the data. In this case, all distributions are mostly identical with similar average and standard deviation values. The reduced separability observed suggests that changes in the Raman spectra of serum samples left at ambient temperature are primarily driven by alterations in the auto-fluorescence baseline rather than analyte concentrations. While this study focused on analyte evolution through Raman spectroscopy, the observed impact of storage conditions on the auto-fluorescence baseline (Fig. 8, b) highlights its potential as an indicator of sample integrity. Further investigation of baseline changes over time are included in the following sections.

3.3. Classification

KNN, RF and SVM classification was done on the three generated datasets after applying PCA, according to Fig. 1: (DS1) the balanced dataset, (DS2) the balanced dataset with high-SNR data, and (DS3) the balanced dataset with high-SNR data and the baseline removed. At this point, it is worth noting that the balanced dataset contains spectra from 18 different donors. Each dataset requires a different classifier. To choose the number of neighbors for KNN, we maximized the average validation accuracy across the 18 folds of leave-one-out cross-validation (LOOCV), leading to 47, 5, and 23 neighbors for KNN for each dataset, respectively. The regularization parameter value for SVM was set at 100 with a radial basis function kernel, and 5 estimators were chosen for the RF. Both values were selected to avoid overfitting by evaluating their train and validation output. Table 6 presents the total training and feature performance time for each model on each dataset, revealing that SVM consistently took the longest, while RF was the fastest.

As expected, the training data exhibited strong performance, with confusion matrices (Fig. S3) showing predictions largely aligned with the true classes (centered close to the diagonal). While the validation accuracy displayed considerable variability in the cross-validation folds, ranging from below 10% to over 90%, the classification patterns observed on the confusion matrices (Fig. 9) were consistent across the KNN, RF, and SVM models for all three datasets. This consistency in behavior suggests that overfitting was not a significant issue in any model, as all reach similar conclusions. Particularly, all classifiers consistently

demonstrated higher sensitivity in identifying non-degraded samples across all datasets, but exhibited greater difficulty in distinguishing between non-degraded and 24-hour degraded serum, as well as between 48-hour and 72-hour degraded serum. This pattern, effectively illustrated by the per-class sensitivity visualization in Figs. 9 and 10, suggests a critical transition point in serum degradation between 24 and 48 h of storage at room temperature. A specific exception was observed in the confusion matrix of SVM classification of the balanced, high-SNR dataset (Fig. 9, h), where the more degraded the sample is, the harder it is for the classifier to identify it properly.

Across the three models (Fig. 10), RF achieved the highest sensitivity on the non-degraded samples of the balanced dataset, while SVM outperformed both KNN and RF on the remaining datasets for the same task. It must be noted that, on average, the total accuracy between the classifiers is comparable: 36%, 45.5%, and 39.8% for the first dataset; 54.5%, 54.5%, and 68.2% for the second dataset; and 34.5%, 34.0%, and 41.2%. Fig. 10 also highlights that the second-best detected class in nearly all cases corresponded to the longest degradation period, indicating better performance at identifying extreme conditions, while detecting finer distinctions between 24 and 48 h remains a challenge for these models. Interestingly, the dataset with the complete standard Raman spectroscopy preprocessing (baseline removal included) showed the poorest classification performance, implying that room temperature serum degradation might primarily involve alterations in the spectral baseline rather than specific metabolite shifts, aligning with the previously described overlapping analyte distributions found in the molecular analysis. Conversely, the least processed dataset still enabled the RF model to identify non-degraded samples in over 70% of cases correctly, suggesting that in time-sensitive scenarios, an RF model on minimally processed data could offer a rapid initial assessment of sample degradation status, warranting further evaluation with more comprehensive techniques.

Finally, we assessed the models' generalization ability using data from an external set of 20 volunteers. These individuals belonged to the same cohort as the primary dataset's 18 participants, but their spectra were excluded from the main analysis due to inconsistent measurements. Specifically, some volunteers had missing data points on certain days, such as when samples acquired before the weekend resulted in two skipped degradation time points. The resulting 20-volunteers dataset is not balanced: 1834, 1108, 506 and 1450 for the 0, 1, 2, and 3 days of sample degradation after outlier removal. Therefore, the average test accuracy will not be calculated and, instead, only the per-class sensitivity will be discussed here.

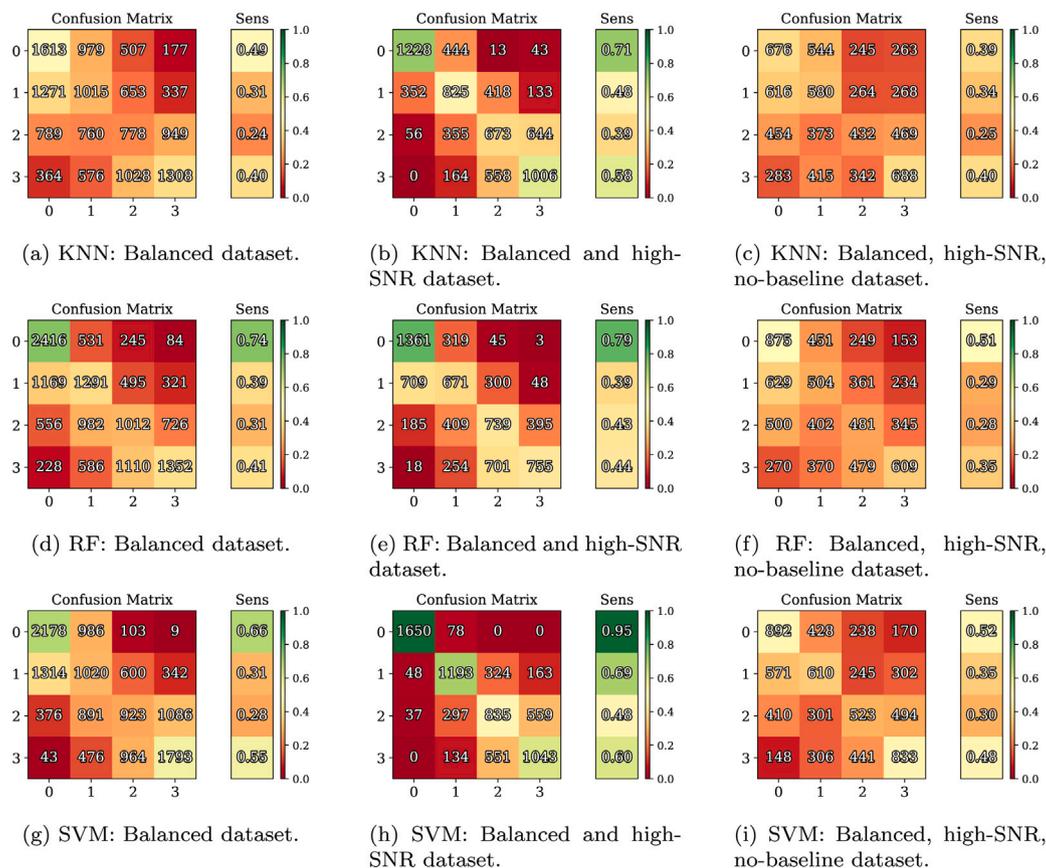


Fig. 9. Confusion matrices of the classifiers' validation output and class-wise sensitivity. The classes correspond to days of serum degradation at ambient temperature for the validation datasets, evaluated with KNN (a, b, c), RF (d, e, f), and SVM (g, h, i).

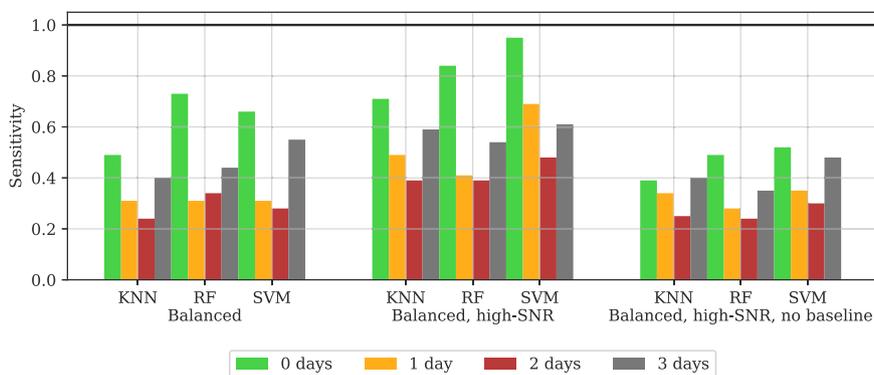


Fig. 10. Sensitivity of KNN, RF, and SVM models for predicting serum sample degradation at room temperature for the three validation datasets with different preprocessing steps (balanced, balanced with high SNR, and balanced with high SNR and no baseline) and degradation times (0, 1, 2, and 3 days). The bars represent the prediction sensitivity, and the horizontal black line represents the maximum sensitivity possible.

For this dataset (test dataset) of 20 individuals, we applied the preprocessing pipeline that yielded the best performance in our main analysis, which involved retaining the baselines and removing low-SNR data. Consistent with the cross-validation strategy used in the prior sections, the test data was standardized using the mean and standard deviation calculated from the original 18-patient training dataset and transformed according to the training-derived PCA transformation. This ensured that the standardization or dimension reduction processes were not influenced by the test set itself. The results are shown in Fig. 11.

These findings indicate that the models generalize well to the test set, achieving comparable or improved sensitivity in identifying non-degraded samples relative to the cross-validated validation sensitivity. This highlights their potential for robustly differentiating

between degraded and non-degraded serum samples at room temperature with Raman Spectroscopy if the baselines are kept, while accurately pinpointing the exact degradation stage remains a challenge.

3.3.1. Explainability

We evaluated feature importance through the average accuracy decrease according to the random permutations model for each classifier. The behavior for the three classifiers is depicted in Fig. 12. Although they varied in performance, the three classifiers showed good agreement on the importance of the principal components for the three datasets. For example, the classifiers agree that the most important principal component in the balanced dataset is PC3. In the balanced,

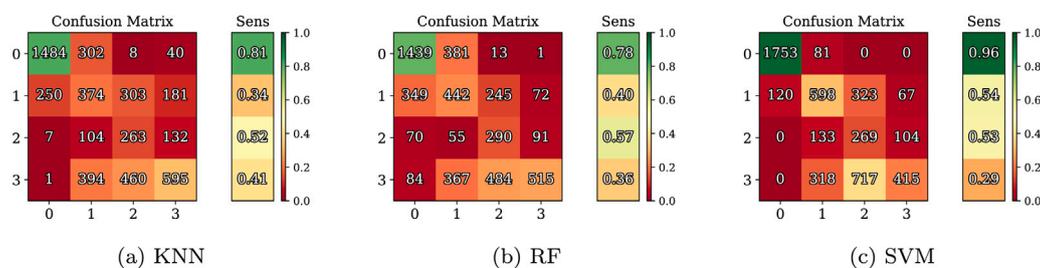


Fig. 11. Confusion matrices of the classifiers' test output and class-wise sensitivity. The classes correspond to days of serum degradation at ambient temperature for the validation dataset, evaluated with KNN (a), RF (b), and SVM (c).

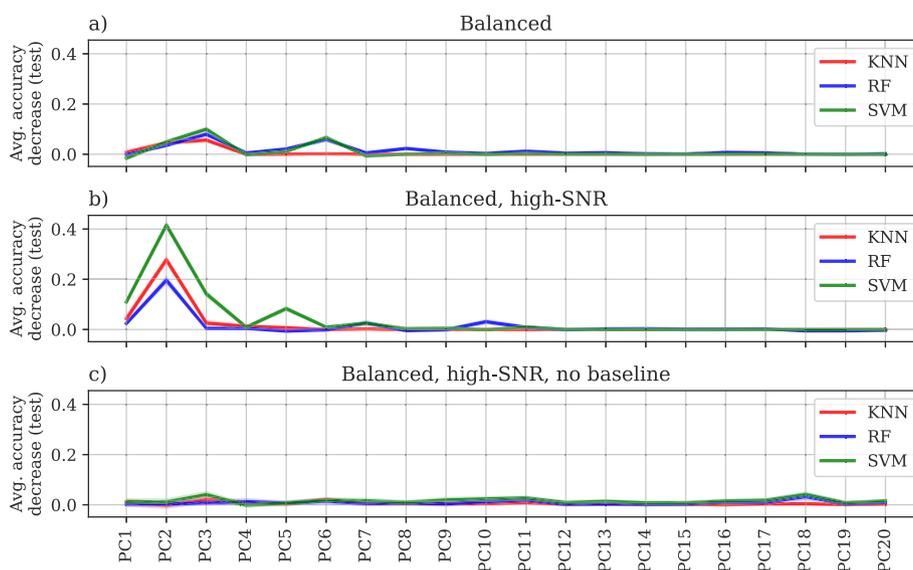


Fig. 12. Results of the accuracy decrease through feature permutation for the 20 studied principal components after 18-fold cross validation for the three classifiers. (a): balanced dataset; (b): balanced, high-SNR dataset; (c) balanced, high-SNR, no baseline dataset.

Table 7

Comparison of the two most important principal components (PC) for separability according to the Fisher ratio (FR), the K-nearest neighbors classifier (KNN), the random forest classifier (RF), and the support vector machine classifier (SVM). The components that appear more than once per dataset are highlighted in bold.

	FR	KNN	RF	SVM
Balanced	(PC6, PC3)	(PC3, PC2)	(PC3, PC6)	(PC3, PC6)
Balanced, high-SNR	(PC2, PC1)	(PC2, PC1)	(PC2, PC10)	(PC2, PC3)
Balanced, high-SNR, no baseline	(PC10, PC8)	(PC3, PC6)	(PC18, PC11)	(PC18, PC3)

high-SNR dataset, PC2 was deemed as the most relevant for classification by all models. Finally, for the balanced, high-SNR, no baseline dataset, which was the one with the worst accuracy, all principal components contributed approximately the same amount, consistent with the difficulty of the models to determine serum deterioration on this dataset accurately. The components with the best class separability in the PC space according to the Fisher ratio (Fig. 8) are the same as those detected with the random permutations method for each trained model (Table 7). This agreement between both metrics is significant when considering that the Fisher ratio uses the statistical properties of the data, whereas the random permutations method evaluates feature importance in terms of model performance.

To evaluate what the Fisher ratio and the random permutations are considering as important, i.e. to discuss the explainability of the results, we evaluated the weights assigned to each Raman wavenumber by each chosen PC according to the two most important validation set features.

The most important principal components detected for the balanced dataset are presented in Fig. 13. PC2 and PC3 show the presence of the most intense phenylalanine peak, along with those of carotenoids;

however, both of these PCs also include the spectral baseline. On the other hand, PC6 depicts a non-baselined version of the spectrum in which numerous spectral bands are highlighted, and whose values largely coincide with the most frequently detected peaks in the spectra employed for molecular analysis.

For the balanced, high-SNR dataset, four principal components were consistently selected by the various methods, as depicted in Fig. 14. As before, the effect of the baseline is evident in PC1, PC2, and PC3, exhibiting varying levels of prominence relative to the spectral peaks they include. Nevertheless, PC10, which was specifically selected by the RF model, displays a spectral profile resembling the derivative of the serum spectra, implying that a combination of the spectral baseline and the primary spectral peaks is optimal to provide effective classification outcomes.

Finally, for the balanced, high-SNR dataset with no baseline, there was no good agreement between the most important principal components, as shown in Fig. 15. Notably, the chosen principal components for this dataset reproduce the spectral shape worse than the other two. This occurs because the standardization that is applied prior to the

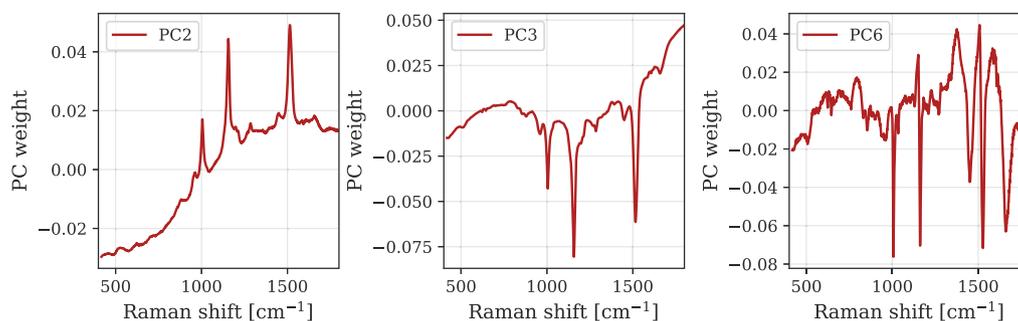


Fig. 13. Best PCs selected by the Fisher ratio and the three classifiers for the balanced dataset. There are only three PCs represented as all methods coincide in the most important ones (Table 7).

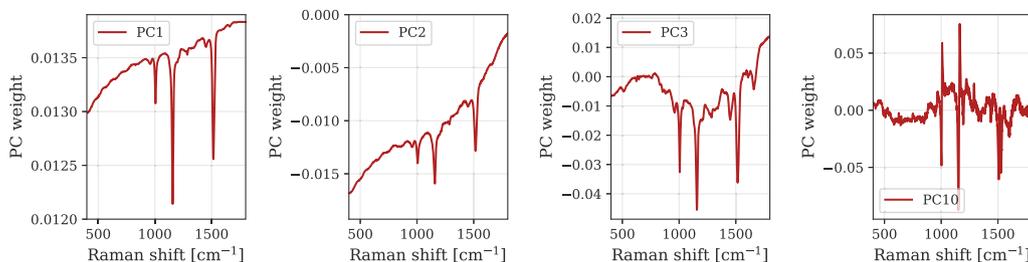


Fig. 14. Best PCs selected by the Fisher ratio and the three classifiers for the balanced, high-SNR dataset. There are only four PCs represented as selected by the multiple methods (Table 7).

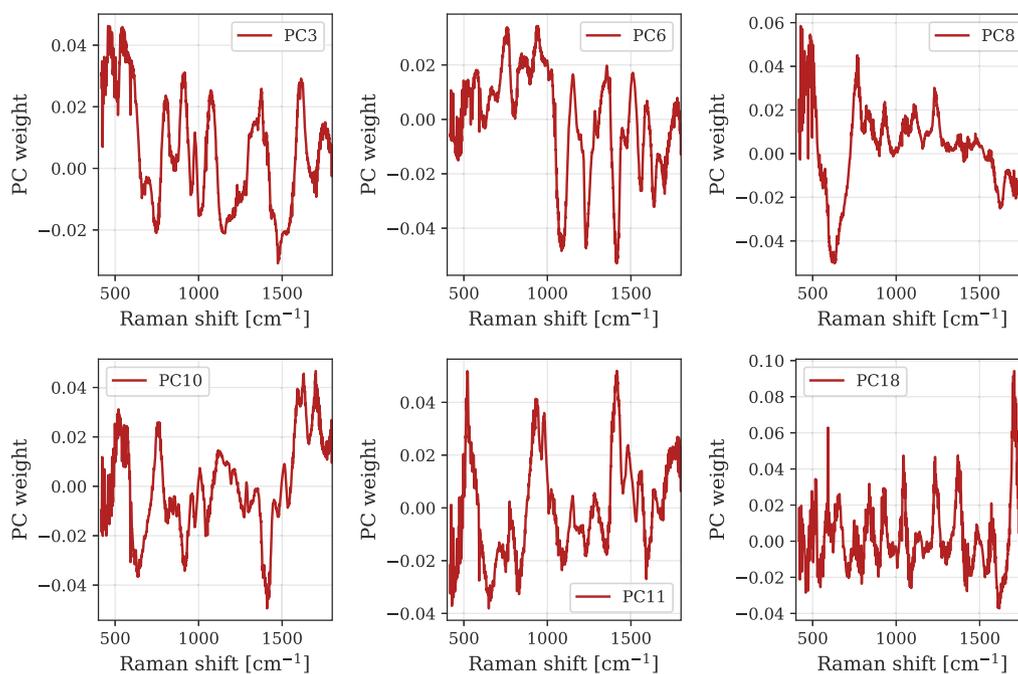


Fig. 15. Best PCs selected by the Fisher ratio and the three classifiers for the balanced, high-SNR, no baseline dataset. There are six PCs represented as selected by the multiple methods (Table 7).

dimension reduction process is no longer influenced by the baseline. Consequently, the way to interpret Fig. 15 is by focusing on the strongest peaks and valleys of each PC. Upon observation, we conclude that all PCs point to similar Raman shift values. However, there is no clear indication of some peaks being more important than others; rather, a mixture of peaks is given similar weights for all PCs. The only exception to this is PC8, which strongly highlights the phenylalanine peak at 624.4 cm^{-1} as a deep valley, while the rest of the spectrum has less intense peaks. Furthermore, combining these results with the low

variance explained by these components, and the observation that the classifiers fail at identifying even the non-degraded samples, suggests that most of the content of these PCs is very similar. This similarity arises because the spectra of this dataset are extremely similar between classes except on the main peaks (as shown in Fig. 6), ultimately hindering dimension reduction and classification.

The results presented in this section consistently indicate the baseline as the primary marker of serum sample degradation at room temperature. Anton et al. [10] have reported an evolution over time in

serum left at ambient temperature. Specifically, they observed changes in ratios related to lysophosphatidylcholine to total phosphatidylcholine. Notably, lysophosphatidylcholine is a type of lipid known to be detectable through fluorescence measurements [94]. Consequently, the Raman baseline could be related to this molecule. If so, it could act as a biomarker for serum degradation. Furthermore, the presence of phenylalanine and carotenoids on almost all principal components is also consistent. This aligns with the most commonly detected peaks and with findings in the literature, where results indicate their variation over time in serum samples left at room temperature [10].

4. Summary and conclusions

This work presented a thorough evaluation of Raman spectroscopy serum spectra when stored at ambient temperature for up to three days. The aim was to explore if conventional Raman spectroscopy analyzed with traditional pipelines can be used to detect serum deterioration so that research facilities could take advantage of the ease-of-use and non-invasive nature of Raman measurements. After preprocessing the measurements by removing outliers, low-SNR measurements, and spectral baselines, the resulting spectra were analyzed with spectral unmixing (supervised) to evaluate the evolution of the most common analytes in the molecular fingerprint. Even though, on average, some slight variations can be seen, most of the spectral variation is due to inter-sample and intra-sample differences, which does not appear to strongly correlate to the actual decay of the serum analytes.

Dimensionality reduction with PCA (unsupervised) was used to analyze each spectrum as a whole instead of looking at individual peaks. Overall, the PCA transformations generally resulted in overlapping data within the two main Fisher-ratio-detected components, particularly after baseline removal. Despite this overlap, considering the first 20 principal components to classify the data using KNN, RF, and SVM classifiers yielded promising results. Specifically, these classifiers sometimes achieved over 80% sensitivity in distinguishing degraded from non-degraded samples, on both the validation and test sets. These results are particularly encouraging when the baseline is retained in the dataset, which is further supported by the random permutations XAI method, as it identified the spectral baseline as a key feature included in the most important PCs. This suggests that the primary indicator of serum degradation detectable through conventional Raman spectroscopy resides in the spectral baseline, rather than in specific analyte changes.

While changes in conventional Raman spectral peaks of serum analytes were minimal after three days at room temperature, our classification results successfully distinguished degraded from non-degraded samples, primarily indicated by baseline shifts. This demonstrates the potential of Raman spectroscopy for detecting early signs of serum sample deterioration due to poor storage conditions. Although further investigation is needed to pinpoint the precise state of degradation and the origin of the baseline variations, our findings suggest that monitoring baseline changes offers a promising avenue for rapid serum quality assessment. Future research will focus on whole blood with pre-centrifugation delays to establish a detailed degradation timeline, ultimately aiming for non-invasive evaluation of blood quality using Raman spectroscopy.

Funding

The projects PREVAL21/07 (FUSIOMUSCLE), financed by the Health Research Institute Valdecilla (IDIVAL), project DTS22/00127 (hyPERfusioCAM) financed by ISCIII, and PT20/00067 and PT23/00058, financed by Plataforma ISCIII de Biobancos y Biomodelos, provided funds for this work.

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.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.saa.2025.126297>.

Data availability

The authors do not have permission to share data.

References

- [1] F. Betsou, Quality assurance and quality control in biobanking, in: *Biobanking of Human Biospecimens: Principles and Practice*, Springer International Publishing, Cham, 2017, pp. 23–49, http://dx.doi.org/10.1007/978-3-319-55120-3_2.
- [2] D. Simeon-Dubach, S.M. Zeisberger, S.P. Hoerstrup, Quality assurance in biobanking for pre-clinical research, *Transfus. Med. Hemotherapy* 43 (5) (2016) 353–358, <http://dx.doi.org/10.1159/000448254>.
- [3] A.-M. Simundic, N. Nikolac, I. Vukasovic, N. Vrkic, The prevalence of preanalytical errors in a Croatian ISO 15189 accredited laboratory, *Clin. Chem. Lab. Med.* 48 (7) (2010) 1009–1014, <http://dx.doi.org/10.1515/CCLM.2010.221>.
- [4] G. Lippi, G.C. Guidi, C. Mattiuzzi, M. Plebani, Preanalytical variability: the dark side of the moon in laboratory testing, *Clin. Chem. Lab. Med. (CCLM)* 44 (4) (2006) 358–365, <http://dx.doi.org/10.1515/CCLM.2006.073>.
- [5] R.G. Rioja, D.M. Espartosa, M. Segovia, M. Ibarz, M.A. Llopis, J.M. Bauça, I. Marzana, N. Barba, M. Ventura, I.G. del Pino, J.J. Puente, A. Caballero, C. Gómez, A.G. Álvarez, M.J. Alsina, V. Álvarez, Laboratory sample stability. Is it possible to define a consensus stability function? An example of five blood magnitudes, *Clin. Chem. Lab. Med. (CCLM)* 56 (11) (2018) 1806–1818, <http://dx.doi.org/10.1515/cclm-2017-1189>.
- [6] J.T. Anderson, A. Keys, Cholesterol in Serum and lipoprotein fractions: Its measurement and stability, *Clin. Chem.* 2 (3) (1956) 145–159, <http://dx.doi.org/10.1093/clinchem/2.3.145>.
- [7] J.G. Donnelly, S.J. Soldin, D.A. Nealon, J.M. Hicks, Stability of twenty-five analytes in human Serum at 22 C, 4 C, and -20 C, *Pediatr. Pathol. Lab. Med.* 15 (6) (1995) 869–874, <http://dx.doi.org/10.3109/15513819509027023>.
- [8] A. Marjani, Effect of storage time and temperature on Serum analytes, *Am. J. Appl. Sci.* 5 (8) (2008) 1047–1051, <http://dx.doi.org/10.3844/ajassp.2008.1047.1051>.
- [9] J. Giasson, M. Hernandez, Y. Chen, Stability of Serum carotene at various light and temperature conditions, *Arch. Pathol. Lab. Med.* 135 (12) (2011) 1529–1530, <http://dx.doi.org/10.5858/arpa.2011-0426-LE>.
- [10] G. Anton, R. Wilson, Z.-H. Yu, C. Prehn, S. Zukunft, J. Adamski, M. Heier, C. Meisinger, W. Römisch-Margl, R. Wang-Sattler, K. Hveem, B. Wolfenbuttel, A. Peters, G. Kastenmüller, M. Waldenberger, Pre-analytical sample quality: metabolite ratios as an intrinsic marker for prolonged room temperature exposure of Serum samples, *PLoS One* 10 (3) (2015) e0121495, <http://dx.doi.org/10.1371/journal.pone.0121495>.
- [11] N.E. Craft, E.D. Brown, J.C. Smith Jr., Effects of storage and handling conditions on concentrations of individual carotenoids, retinol, and tocopherol in plasma, *Clin. Chem.* 34 (1) (1988) 44–48, <http://dx.doi.org/10.1093/clinchem/34.1.44>.
- [12] N.P. Kubasik, M. Ricotta, T. Hunter, H.E. Sine, Effect of duration and temperature of storage on serum analyte stability: examination of 14 selected radioimmunoassay procedures, *Clin. Chem.* 28 (1) (1982) 164–165, <http://dx.doi.org/10.1093/clinchem/28.1.164>.
- [13] M.J. Evans, J.H. Livesey, M. Ellis, T.G. Yandle, Effect of anticoagulants and storage temperatures on stability of plasma and Serum hormones, *Clin. Biochem.* 34 (2) (2001) 107–112, [http://dx.doi.org/10.1016/S0009-9120\(01\)00196-5](http://dx.doi.org/10.1016/S0009-9120(01)00196-5).
- [14] B.L. Boyanton Jr., K.E. Blick, Stability studies of twenty-four analytes in human plasma and Serum, *Clin. Chem.* 48 (12) (2002) 2242–2247, <http://dx.doi.org/10.1093/clinchem/48.12.2242>.
- [15] R. Hannisdal, P.M. Ueland, S.J. Eussen, A. Svardal, S. Hustad, Analytical recovery of folate degradation products formed in human Serum and plasma at room temperature, *J. Nutr.* 139 (7) (2009) 1415–1418, <http://dx.doi.org/10.3945/jn.109.105635>.
- [16] S. Heiling, N. Knutti, F. Scherr, J. Geiger, J. Weikert, M. Rose, R. Jahns, U. Ceglarek, A. Scherag, M. Kiehntopf, Metabolite ratios as quality indicators for pre-analytical variation in Serum and EDTA plasma, *Metabolites* 11 (9) (2021) <http://dx.doi.org/10.3390/metabo11090638>.

- [17] T. Key, S. Oakes, G. Davey, J. Moore, L.M. Edmond, U.J. McLoone, D.I. Thurnham, Stability of vitamins A, C, and E, carotenoids, lipids, and testosterone in whole blood stored at 4 C for 6 and 24 hours before separation of serum and plasma, *Cancer Epidemiol. Biomarkers Prev.* 5 (10) (1996) 811–814.
- [18] N.B. Tsui, E.K. Ng, Y.D. Lo, Stability of endogenous and added RNA in blood specimens, Serum, and plasma, *Clin. Chem.* 48 (10) (2002) 1647–1653, <http://dx.doi.org/10.1093/clinchem/48.10.1647>.
- [19] B. Kamlage, S. Neuber, B. Bethan, S. González Maldonado, A. Wagner-Golbs, E. Peter, O. Schmitz, P. Schatz, Impact of prolonged blood incubation and extended Serum storage at room temperature on the human Serum metabolome, *Metabolites* 8 (1) (2018) <http://dx.doi.org/10.3390/metabo8010006>.
- [20] F. Betsou, E. Gunter, J. Clements, Y. DeSouza, K.A. Goddard, F. Guadagni, W. Yan, A. Skubitz, S. Somiari, T. Yeadon, R. Chuaqui, Identification of evidence-based biospecimen quality-control tools: A report of the international society for biological and environmental repositories (ISBER) biospecimen science working group, *J. Mol. Diagn.* 15 (1) (2013) 3–16, <http://dx.doi.org/10.1016/j.jmoldx.2012.06.008>.
- [21] J.-P. Trezzi, A. Bulla, C. Bellora, M. Rose, P. Lescuyer, M. Kiehnopf, K. Hiller, F. Betsou, LacaScore: a novel plasma sample quality control tool based on ascorbic acid and lactic acid levels, *Metabolomics* 12 (2016) 96, <http://dx.doi.org/10.1007/s11306-016-1038-1>.
- [22] P. Larkin, *Infrared and Raman Spectroscopy: Principles and Spectral Interpretation*, Elsevier, 2017, <http://dx.doi.org/10.1016/C2015-0-00806-1>.
- [23] M.J. Baker, S.R. Hussain, L. Lovergne, V. Untereiner, C. Hughes, R.A. Lukaszewski, G. Thiéfin, G.D. Sockalingum, Developing and understanding biofluid vibrational spectroscopy: a critical review, *Chem. Soc. Rev.* 45 (2016) 1803–1818, <http://dx.doi.org/10.1039/C5CS00585J>.
- [24] 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. McAnish, N. Stone, F.L. Martin, Using Raman spectroscopy to characterize biological materials, *Nat. Protoc.* 11 (4) (2016) 664–687, <http://dx.doi.org/10.1038/nprot.2016.036>.
- [25] C.G. Atkins, K. Buckley, M.W. Blades, R.F. Turner, Raman spectroscopy of blood and blood components, *Appl. Spectrosc.* 71 (5) (2017) 767–793, <http://dx.doi.org/10.1177/0003702816686593>, PMID: 28398071.
- [26] D.A. Magdas, F. Guyon, I. Feher, S.C. Pinzaru, Wine discrimination based on chemometric analysis of untargeted markers using FT-Raman spectroscopy, *Food Control* 85 (2018) 385–391, <http://dx.doi.org/10.1016/j.foodcont.2017.10.024>.
- [27] D.A. Magdas, S. Cinta Pinzaru, F. Guyon, I. Feher, B.I. Cozar, Application of SERS technique in white wines discrimination, *Food Control* 92 (2018) 30–36, <http://dx.doi.org/10.1016/j.foodcont.2018.04.043>.
- [28] R.S. Ortiz, K. de Cássia Mariotti, B. Fank, R.P. Limberger, M.J. Anzanello, P. Mayorga, Counterfeit cialis and viagra fingerprinting by ATR-FTIR spectroscopy with chemometry: Can the same pharmaceutical powder mixture be used to falsify two medicines? *Forensic Sci. Int.* 226 (1) (2013) 282–289, <http://dx.doi.org/10.1016/j.forsciint.2013.01.043>.
- [29] G. Açıkgöz, B. Hamamci, Determination of ethyl glucuronide (EtG) in blood samples using partial least squares discriminant analysis applied to surface-enhanced Raman spectroscopy, *Vib. Spectrosc.* 106 (2020) 103012, <http://dx.doi.org/10.1016/j.vibspec.2019.103012>.
- [30] J. Li, X. Wang, S. Min, J. Xia, J. Li, Raman spectroscopy combined with convolutional neural network for the sub-types classification of breast cancer and critical feature visualization, *Comput. Methods Programs Biomed.* 255 (2024) 108361, <http://dx.doi.org/10.1016/j.cmpb.2024.108361>.
- [31] C. Chang, H. Liu, C. Chen, L. Wu, X. Lv, X. Xie, C. Chen, Rapid diagnosis of systemic lupus erythematosus by Raman spectroscopy combined with spiking neural network, *Spectrochim. Acta Part A: Mol. Biomol. Spectrosc.* 310 (2024) 123904, <http://dx.doi.org/10.1016/j.saa.2024.123904>.
- [32] M. Unal, R. Ahmed, A. Mahadevan-Jansen, J.S. Nyman, Compositional assessment of bone by Raman spectroscopy, *Analyst* 146 (2021) 7464–7490, <http://dx.doi.org/10.1039/D1AN01560E>.
- [33] H. Robert Maier, F. Rosa Taghikhah, E. Nabavi, S. Razavi, H. Gupta, W. Wu, D.A. Radford, J. Huang, How much x is in XAI: Responsible use of “explainable” artificial intelligence in hydrology and water resources, *J. Hydrol. X* 25 (2024) 100185, <http://dx.doi.org/10.1016/j.hydroa.2024.100185>.
- [34] J. Ukwaththa, S. Herath, D. Meddage, A review of machine learning (ML) and explainable artificial intelligence (XAI) methods in additive manufacturing (3D printing), *Mater. Today Commun.* 41 (2024) 110294, <http://dx.doi.org/10.1016/j.mtcomm.2024.110294>.
- [35] A. Singhal, K.K. Agrawal, A. Quezada, A.R. Aguiñaga, S. Jiménez, S.P. Yadav, Explainable artificial intelligence (XAI) model for cancer image classification, *CMES - Comput. Model. Eng. Sci.* 141 (1) (2024) 401–441, <http://dx.doi.org/10.32604/cmescs.2024.051363>.
- [36] R. Gipiškis, C.-W. Tsai, O. Kurasova, Explainable AI (XAI) in image segmentation in medicine, industry, and beyond: A survey, *ICT Express* (2024) <http://dx.doi.org/10.1016/j.icte.2024.09.008>.
- [37] M. Miró-Nicolau, A. Jaume-i-Capó, G. Moyà-Alcover, Assessing fidelity in XAI post-hoc techniques: A comparative study with ground truth explanations datasets, *Artificial Intelligence* 335 (2024) 104179, <http://dx.doi.org/10.1016/j.artint.2024.104179>.
- [38] A.T. Karl, A randomized permutation whole-model test heuristic for self-validated ensemble models (SVEM), *Chemometr. Intell. Lab. Syst.* 249 (2024) 105122, <http://dx.doi.org/10.1016/j.chemolab.2024.105122>.
- [39] J. Yoo, Y. Cho, D.H. Kim, J. Kim, T.G. Lee, S.M. Lee, J. Choo, M.S. Jeong, Unraveling the role of Raman modes in evaluating the degree of reduction in graphene oxide via explainable artificial intelligence, *Nano Today* 57 (2024) 102366, <http://dx.doi.org/10.1016/j.nantod.2024.102366>.
- [40] L. Bellantuono, R. Tommasi, E. Pantaleo, M. Verri, N. Amoroso, P. Crucitti, M. Di Gioacchino, F. Longo, A. Monaco, A.M. Naciu, A. Palermo, C. Taffon, S. Tangaro, A. Crescenzi, A. Sodo, R. Bellotti, An explainable artificial intelligence analysis of Raman spectra for thyroid cancer diagnosis, *Sci. Rep.* 13 (1) (2023) 16590, <http://dx.doi.org/10.1038/s41598-023-43856-7>.
- [41] Y. Xie, X. Su, Y. Wen, C. Zheng, M. Li, Artificial intelligent label-free SERS profiling of serum exosomes for breast cancer diagnosis and postoperative assessment, *Nano Lett.* 22 (19) (2022) 7910–7918, <http://dx.doi.org/10.1021/acs.nanolett.2c02928>.
- [42] J.P.M. Wielders, F.A. Wijnberg, Preanalytical stability of 25(OH)-Vitamin D3 in human blood or serum at room temperature: Solid as a rock, *Clin. Chem.* 55 (8) (2009) 1584–1585, <http://dx.doi.org/10.1373/clinchem.2008.117366>.
- [43] J. Ren, E.L. Mozurkewich, A. Sen, A.M. Vahratian, T.G. Ferreri, A.N. Morse, Z. Djuric, Total Serum fatty acid analysis by GC-MS: Assay validation and Serum sample stability, *Curr. Pharm. Anal.* 9 (4) (2013) 331–339, <http://dx.doi.org/10.2174/1573412911309040002>.
- [44] M. Alonso-Peña, T. Dierssen, M.J. Marin, J. Alonso-Molero, I. Gómez-Acebo, I. Santiuste, J.V. Lazarus, P. Sanchez-Juan, G. Peralta, J. Crespo, M. Lopez-Hoyos, A. Peleteiro-Vigil, B.A. Lavin Gomez, O. Alvaro Melero, M.T. Arias-Loste, A. Batlle, J. Cabezas, J. Calvo Montes, J. Cayon de las Cuevas, L. Conde, L. Diego Gonzalez, C. Fariñas, S. Fernandez Luis, M. Fernandez Ortiz, S. Garcia Blanco, G. Garcia Lopez, M. Garcia Unzueta, J.C. Garrido Gomez, R. Gonzalez, P. Iruzueta, J. Martin Lazaro, L. Martin Ruiz, N. Martinez Magunacelaya, R. Martinez Santiago, J.M. Medina, M.J. Muruzabal Siges, A. Padilla, A. Peleteiro, L. Reyes-González, D. Ruiz, A. Santos-Laso, M.E. Sanz Piña, D. Sordo, S. Solorzano, R. Tejido, R. Wallman, M. Wunsch, Cantabria Cohort Collaborators, The Cantabria cohort, a protocol for a population-based cohort in northern Spain, *BMC Public Health* 23 (1) (2023) 2429, <http://dx.doi.org/10.1186/s12889-023-17318-8>.
- [45] F. Betsou, R. Bilbao, J. Case, R. Chuaqui, J.A. Clements, Y. De Souza, A. De Wilde, J. Geiger, W. Grizzle, F. Guadagni, E. Gunter, S. Heil, M. Kiehnopf, I. Koppandi, S. Lehmann, L. Linsen, J. Mackenzie-Dodds, R.A. Quesada, R. Tebbakha, T. Selander, K. Shea, M. Sobel, S. Somiari, D. Spyropoulos, M. Stone, G. Tybring, K. Valyi-Nagy, L. Wadhwa, Standard PREanalytical code version 3.0, *Biopreserv. Biobank.* 16 (1) (2018) 9–12, <http://dx.doi.org/10.1089/bio.2017.0109>.
- [46] J.V. Miller, E.G. Bartick, Forensic analysis of single fibers by Raman spectroscopy, *Appl. Spectrosc.* 55 (12) (2001) 1729–1732, <http://dx.doi.org/10.1366/0003702011954099>.
- [47] M.F. Escoriza, J.M. VanBriesen, S. Stewart, J. Maier, P.J. Treado, Raman spectroscopy and chemical imaging for quantification of filtered waterborne bacteria, *J. Microbiol. Meth.* 66 (1) (2006) 63–72, <http://dx.doi.org/10.1016/j.jmimet.2005.10.013>.
- [48] A. Shapiro, O.N. Gofrit, G. Pizov, J.K. Cohen, J. Maier, Raman molecular imaging: A novel spectroscopic technique for diagnosis of bladder cancer in urine specimens, *Eur. Urol.* 59 (1) (2011) 106–112, <http://dx.doi.org/10.1016/j.eururo.2010.10.027>.
- [49] M. Petersen, Z. Yu, X. Lu, Application of Raman spectroscopic methods in food safety: A review, *Biosensors* 11 (6) (2021) <http://dx.doi.org/10.3390/bios11060187>.
- [50] L. Cui, H.J. Butler, P.L. Martin-Hirsch, F.L. Martin, Aluminium foil as a potential substrate for ATR-FTIR, transflection FTIR or Raman spectrochemical analysis of biological specimens, *Anal. Methods* 8 (2016) 481–487, <http://dx.doi.org/10.1039/C5AY02638E>.
- [51] A. Siegel, *Statistics and Data Analysis: An Introduction*, John Wiley & Sons Australia, Limited, 1996.
- [52] Z. Cheng, C. Zou, J. Dong, Outlier detection using isolation forest and local outlier factor, in: Proceedings of the Conference on Research in Adaptive and Convergent Systems, RACS '19, Association for Computing Machinery, New York, NY, USA, 2019, pp. 161–168, <http://dx.doi.org/10.1145/3338840.3355641>.
- [53] M. Amer, M. Goldstein, S. Abdennadher, Enhancing one-class support vector machines for unsupervised anomaly detection, in: Proceedings of the ACM SIGKDD Workshop on Outlier Detection and Description, ODD '13, Association for Computing Machinery, New York, NY, USA, 2013, pp. 8–15, <http://dx.doi.org/10.1145/2500853.2500857>.
- [54] H.-P. Kriegel, M. Schubert, A. Zimek, Angle-based outlier detection in high-dimensional data, in: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '08, Association for Computing Machinery, New York, NY, USA, 2008, pp. 444–452, <http://dx.doi.org/10.1145/1401890.1401946>.
- [55] L. Zhou, Performance of corporate bankruptcy prediction models on imbalanced dataset: The effect of sampling methods, *Knowl.-Based Syst.* 41 (2013) 16–25, <http://dx.doi.org/10.1016/j.knosys.2012.12.007>.

- [56] P.H. Lee, Resampling methods improve the predictive power of modeling in class-imbalanced datasets, *Int. J. Environ. Res. Public Heal.* 11 (9) (2014) 9776–9789, <http://dx.doi.org/10.3390/ijerph110909776>.
- [57] I. De Wolf, Raman spectroscopy: About chips and stress, *Spectrosc. Eur.* 15 (2003).
- [58] P. Borowicz, M. Latek, W. Rzdokiewicz, A. Łaszcz, A. Czerwinski, J. Ratajczak, Deep-ultraviolet Raman investigation of silicon oxide: thin film on silicon substrate versus bulk material, *Adv. Nat. Sci.: Nanosci. Nanotechnol.* 3 (4) (2012) 045003, <http://dx.doi.org/10.1088/2043-6262/4/045003>.
- [59] W.-J. Lee, Y.-H. Chang, Growth without postannealing of monoclinic VO₂ thin film by atomic layer deposition using VCl₄ as precursor, *Coatings* 8 (12) (2018) <http://dx.doi.org/10.3390/coatings8120431>.
- [60] C.A. Lieber, A. Mahadevan-Jansen, Automated method for subtraction of fluorescence from biological Raman spectra, *Appl. Spectrosc.* 57 (11) (2003) 1363–1367, <http://dx.doi.org/10.1366/000370203322554518>.
- [61] P.H.C. Eilers, A perfect smoother, *Anal. Chem.* 75 (14) (2003) 3631–3636, <http://dx.doi.org/10.1021/ac034173t>.
- [62] Z.-M. Zhang, S. Chen, Y.-Z. Liang, Baseline correction using adaptive iteratively reweighted penalized least squares, *Analyst* 135 (2010) 1138–1146, <http://dx.doi.org/10.1039/B922045C>.
- [63] R. Perez-Pueyo, M.J. Soneira, S. Ruiz-Moreno, Morphology-based automated baseline removal for Raman spectra of artistic pigments, *Appl. Spectrosc.* 64 (6) (2010) 595–600, <http://dx.doi.org/10.1366/000370210791414281>.
- [64] C. Ryan, E. Clayton, W. Griffin, S. Sie, D. Cousins, SNIP, a statistics-sensitive background treatment for the quantitative analysis of PIXE spectra in geoscience applications, *Nucl. Instrum Methods Phys. Res. Sect. B: Beam Interactions Mater. Atoms* 34 (3) (1988) 396–402, [http://dx.doi.org/10.1016/0168-583X\(88\)90063-8](http://dx.doi.org/10.1016/0168-583X(88)90063-8).
- [65] G. Sheehy, F. Picot, F. Dallaire, K. Ember, T. Nguyen, K. Petrecca, D. Trudel, F. Leblond, Open-sourced Raman spectroscopy data processing package implementing a baseline removal algorithm validated from multiple datasets acquired in human tissue and biofluids, *J. Biomed. Opt.* 28 (2) (2023) 025002, <http://dx.doi.org/10.1117/1.JBO.28.2.025002>.
- [66] F. Kruse, A. Lefkoff, J. Boardman, K. Heidebrecht, A. Shapiro, P. Barloon, A. Goetz, The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data, *Remote Sens. Environ.* 44 (2) (1993) 145–163, [http://dx.doi.org/10.1016/0034-4257\(93\)90013-N](http://dx.doi.org/10.1016/0034-4257(93)90013-N), Airbone Imaging Spectrometry.
- [67] N. Keshava, J. Mustard, Spectral unmixing, *IEEE Signal Process. Mag.* 19 (1) (2002) 44–57, <http://dx.doi.org/10.1109/79.974727>.
- [68] L. Wang, C. Shi, C. Diao, W. Ji, D. Yin, A survey of methods incorporating spatial information in image classification and spectral unmixing, *Int. J. Remote Sens.* 37 (16) (2016) 3870–3910, <http://dx.doi.org/10.1080/01431161.2016.1204032>.
- [69] X. Yuan, R.A. Mayanovic, An empirical study on Raman peak fitting and its application to Raman quantitative research, *Appl. Spectrosc.* 71 (10) (2017) 2325–2338, <http://dx.doi.org/10.1007/978-1-4684-1342-7>.
- [70] C.-S. Choe, J. Lademann, M.E. Darvin, Gaussian-function-based deconvolution method to determine the penetration ability of petrolatum oil into in vivo human skin using confocal Raman microscopy, *Laser Phys.* 24 (10) (2014) 105601, <http://dx.doi.org/10.1088/1054-660X/24/10/105601>.
- [71] B.P. Asthana, W. Kiefer, Precise determination of frequency shifts by Raman difference spectroscopy using voigt profile for the Raman band, *Appl. Spectrosc.* 37 (4) (1983) 334–340.
- [72] Z. Movasaghi, S. Rehman, D.I.U. Rehman, Raman spectroscopy of biological tissues, *Appl. Spectrosc. Rev.* 42 (5) (2007) 493–541, <http://dx.doi.org/10.1080/05704920701551530>.
- [73] N. Shimodaira, A. Masui, Raman spectroscopic investigations of activated carbon materials, *J. Appl. Phys.* 92 (2) (2002) 902–909, <http://dx.doi.org/10.1063/1.1487434>.
- [74] N. Stone, P. Stavroulaki, C. Kendall, M. Birchall, H. Barr, Raman spectroscopy for early detection of laryngeal malignancy: Preliminary results, *Laryngoscope* 110 (10) (2000) 1756–1763, <http://dx.doi.org/10.1097/00005537-200010000-00037>.
- [75] W. Schumacher, S. Stöckel, P. Rösch, J. Popp, Improving chemometric results by optimizing the dimension reduction for Raman spectral data sets, *J. Raman Spectrosc.* 45 (10) (2014) 930–940, <http://dx.doi.org/10.1002/jrs.4568>.
- [76] O. Ryabchykov, S. Guo, T. Bocklitz, Analyzing Raman spectroscopic data, *Phys. Sci. Rev.* 4 (2) (2019) 20170043, <http://dx.doi.org/10.1515/psr-2017-0043>.
- [77] H. Shin, H. Jeong, J. Park, S. Hong, Y. Choi, Correlation between cancerous exosomes and protein markers based on surface-enhanced Raman spectroscopy (SERS) and principal component analysis (PCA), *ACS Sensors* 3 (12) (2018) 2637–2643, <http://dx.doi.org/10.1021/acssensors.8b01047>.
- [78] J.L.E. Campos, H. Miranda, C. Rabelo, E. Sandoz-Rosado, S. Pandey, J. Riikonen, A.G. Cano-Marquez, A. Jorio, Applications of Raman spectroscopy in graphene-related materials and the development of parameterized PCA for large-scale data analysis, *J. Raman Spectrosc.* 49 (1) (2018) 54–65, <http://dx.doi.org/10.1002/jrs.5225>.
- [79] S. Guo, P. Rösch, J. Popp, T. Bocklitz, Modified PCA and PLS: Towards a better classification in Raman spectroscopy-based biological applications, *J. Chemom.* 34 (4) (2020) e3202, <http://dx.doi.org/10.1002/cem.3202>, e3202 CEM-19-0167.R1.
- [80] F. Castells, P. Laguna, L. Sörnmo, A. Bollmann, J.M. Roig, Principal component analysis in ECG signal processing, *EURASIP J. Adv. Signal Process.* 2007 (1) (2007) 074580, <http://dx.doi.org/10.1155/2007/74580>.
- [81] K.M. Yeater, S.E. Duke, W.E. Riedell, Multivariate analysis: Greater insights into complex systems, *Agron. J.* 107 (2) (2015) 799–810, <http://dx.doi.org/10.2134/agronj14.0017>.
- [82] W. Zgallai, Biomedical signal processing and artificial intelligence in healthcare, in: *Developments in Biomedical Engineering and Bioelectronics*, Elsevier Science, 2020.
- [83] N.H. Othman, A.R.M. Radzol, K.Y. Lee, W. Mansor, Reduced featured k-NN classifier model optimal for classification of Dengue fever from Salivary Raman spectra, in: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, 2019, pp. 471–474, <http://dx.doi.org/10.1109/EMBC.2019.8856427>.
- [84] F.U. Ciloglu, A.M. Saridag, I.H. Kilic, M. Tokmakci, M. Kahraman, O. Aydin, Identification of methicillin-resistant staphylococcus aureus bacteria using surface-enhanced Raman spectroscopy and machine learning techniques, *Analyst* 145 (23) (2020) 7559–7570, <http://dx.doi.org/10.1039/d0an00476f>.
- [85] A.M. Laughney, V. Krishnaswamy, P.B. García-Allende, O.M. Conde, W.A. Wells, K.D. Paulsen, B.W. Pogue, Automated classification of breast pathology using local measures of broadband reflectance, *J. Biomed. Opt.* 15 (6) (2010) 066019, <http://dx.doi.org/10.1117/1.3516594>.
- [86] A. Mucherino, P.J. Papajorgji, P.M. Pardalos, K-nearest neighbor classification, in: *Data Mining in Agriculture*, Springer New York, New York, NY, 2009, pp. 83–106, http://dx.doi.org/10.1007/978-0-387-88615-2_4.
- [87] M. Pal, Random forest classifier for remote sensing classification, *Int. J. Remote Sens.* 26 (1) (2005) 217–222, <http://dx.doi.org/10.1080/01431160412331269698>.
- [88] A. Parmar, R. Katariya, V. Patel, A review on random forest: An ensemble classifier, in: J. Hemanth, X. Fernando, P. Lafata, Z. Baig (Eds.), *International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICD) 2018*, Springer International Publishing, Cham, 2019, pp. 758–763, http://dx.doi.org/10.1007/978-3-030-03146-6_86.
- [89] J. Cervantes, F. Garcia-Lamont, L. Rodriguez-Mazahua, A. Lopez, A comprehensive survey on support vector machine classification: Applications, challenges and trends, *Neurocomputing* 408 (2020) 189–215, <http://dx.doi.org/10.1016/j.neucom.2019.10.118>.
- [90] Y. Zhang, Support vector machine classification algorithm and its application, in: C. Liu, L. Wang, A. Yang (Eds.), *Information Computing and Applications*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 179–186, http://dx.doi.org/10.1007/978-3-642-34041-3_27.
- [91] R.G. Brereton, G.R. Lloyd, Support vector machines for classification and regression, *Analyst* 135 (2010) 230–267, <http://dx.doi.org/10.1039/B918972F>.
- [92] V. Mieites, G. Anichini, J. Qi, K. O'Neill, O.M. Conde, D.S. Elson, PoLambRimetry: a multispectral polarimetric atlas of lamb brain, *J. Biomed. Opt.* 29 (9) (2024) 096002, <http://dx.doi.org/10.1117/1.JBO.29.9.096002>.
- [93] I. Durickovic, Using Raman spectroscopy for characterization of aqueous media and quantification of species in aqueous solution, in: M.T. Stauffer (Ed.), *Applications of Molecular Spectroscopy To Current Research in the Chemical and Biological Sciences*, IntechOpen, Rijeka, 2016, <http://dx.doi.org/10.5772/64550>.
- [94] F. Jannah, S. Park, J.-M. Heo, N. Choi, J. Choo, J.-M. Kim, Rapid and sensitive detection of lysophosphatidylcholine using zwitterionic polydiacetylene vesicles and a microfluidic gradient sensor, *Sensors Actuators B: Chem.* 371 (2022) 132528, <http://dx.doi.org/10.1016/j.snb.2022.132528>.