Comparison of Machine Learning Methods for the Analysis of Serum Raman Spectra in the Detection of Chronic Heart Failure

Elena Sorokina[1], Yulia Khristoforova[1*], Ivan Bratchenko[1], Maria Skuratova[2], Petr Lebedev[3], and Valery Zakharov[1]

[1] Samara National Research University, 34 Moskovskoe shosse, Samara 443080, Russian Federation
[2] Samara City Clinical Hospital №1 named after N. I. Pirogov, 80 Polevaya str., Samara 443096, Russian Federation
[3] Samara State Medical University, 159 Tashkentskaya str., Samara 443095, Russian Federation
[*]e-mail: khristoforovajulia@gmail.com

Abstract. The high sensitivity of Raman spectroscopy to the chemical composition of substances at the molecular level makes it a valuable tool for the diagnosis of chronic heart failure (CHF) through analysis of blood serum. Raman spectroscopy offers a label-free, fast detection method, with highly specific and accurate results when combined with machine learning (ML) techniques. However, it is essential to carefully choose the appropriate ML algorithm for analyzing high-dimensional spectral data to achieve reliable and correct results that are primarily based on the true chemical features and differences of the studied samples, specimens or structures as not all algorithms may deliver the high performance. In this study, we compared four approaches: (1) multivariate curve resolution alternating least squares method in combination with logistic regression (MCR-LR), (2) with linear kernel support vector machine (MCR-SVM), (3) projection on the latent structures with discriminant analysis (PLS-DA), and (4) projection on the latent structures with linear kernel support vector machine (PLS-SVM). These methods were applied to 193 Raman spectra from patients with CHF and 78 from control cases. We found that PLS-DA and PLS-SVM demonstrated the best ROC AUCs, with average value of 0.950 (0.91−0.97, 0.95 CI) and 0.99 (0.94−1.00, 0.95 CI), while MCR-LR and MCR-SVM achieved only 0.50 (0.46−0.53, 0.95 CI) and 0.59 (0.54−0.65, 0.95 CI), respectively.

Keywords: Raman spectroscopy, spectral analysis, multivariate curve resolution alternating least squares method, projection on the latent structures with discriminant analysis, linear kernel support vector machine, chronic heart failure.

Paper #9195 received 4 Dec 2024; revised manuscript received 24 Dec 2024; accepted for publication 24 Dec 2024; published online 30 Dec 2024. doi: 10.18287/JBPE24.10.040319.

1 Introduction

Chronic heart failure (CHF) is a clinical syndrome that occurs with structural or functional abnormalities of heart that results in elevated intracardiac pressures and/or inadequate cardiac output [1]. The chronic heart failure (CHF) is diagnosed based on physical examination and a series of tests included echocardiography and electrocardiogram, and brain natriuretic peptide (BNP) or N-terminal pro-BNP (NT-proBNP) protein level, both of which tend to be elevated in people with heart failure [2]. The BNP and NT-proBNP protein level has been recommended as a biomarker for the diagnosis of CHF that allows one to predict of all stages of the CHF disease, but this diagnostic approach is time-consuming and expensive. Therefore, it is extremely important to develop rapid diagnostic techniques for CHF identification.

Today, Raman spectroscopy has become widely used in medical research for diagnostic purposes. In numerous studies, biological fluids [3, 4] and tissues [3, 5−7] are being analyzed using Raman spectroscopy to diagnose skin diseases [5], cardiovascular disorders [6, 8], cancer [7, 9], etc. The principle of the Raman spectroscopy is based on the inelastic scattering of the radiation following its interaction with the vibrational modes of molecules. The Raman spectrum registered as a result of this phenomenon consists of a set of characteristic Raman peaks that correspond to the specific molecular vibrational modes which makes it a unique "fingerprint" of a biological sample. Therefore, Raman spectrum can be analyzed to identify of the relative sample's composition and identify any changes that may occur due to disease.

In the analysis of Raman spectra, machine learning (ML) methods have been found to be useful for a variety of applications [10], including classification [5, 11], clustering [12], regression analysis [13, 14] of the experimental data. The ML applications involve creating prediction, regression and classification statistical models based on the analysis of high-dimensional spectral data, with reducing the dimensionality and identifying the significant features that can be used to interpret the physical meaning of the results. Then, ML models are validated using new spectra to ensure their performances.

The projection on latent structures (PLS) can be considered a reference classifier as it is widely employed for various biomedical applications. For example, in study by I. Behl et al. [15] Raman spectra of brush biopsies from the buccal mucosa/tongue were subjected to PLS-DA to differentiate the healthy samples from oral mucosal dysplastic lesions. S Zhu et al. [16] achieved high classification accuracies more than 85% by PLS-LDA diagnostic model based on serum SERS to diagnose pulmonary cryptococcal infection and to assess the degrees of infection. In study by Khristoforova et al. [17] PLS-DA was also used to analyze autofluorescence and Raman data of *in vivo* human skin lesions to identify amelanotic melanoma cases. However, the cited works [15−17] do not explain the choice of the currently widely used PLS-DA method. The question arises as to whether the performance will depend on the ML methods. Additionally, PLS does not essentially analyze spectra, focusing solely on the classification task. Therefore, we want to compare the PLS approach with the multivariate curve resolution alternating least squares (MCR-ALS) analysis, which is used to reconstruct the spectral profiles contributed to complex Raman spectra of biological substances. This allows for both the physical interpretation of the components that the original Raman spectra are composed of, and for classification in combination with differentiating algorithms.

In this study, we applied 4 combinations of ML algorithms: MCR-ALS analysis in combination with logistic regression and linear kernel support vector machine (SVM), and supervised PLS algorithm with discriminant analysis and linear kernel SVM to analyze Raman spectra of human serum of patients with CHF and control group. Therefore, the aim of this study is to select the efficient ML approach to determine the potential of Raman spectroscopy to identify CHF cases by analyzing the *in vitro* serum samples spectra.

2 Materials and Methods

2.1 Studied Samples

The study protocols were approved by the ethical committee of Samara State Medical University (protocol No 268, 11 September 2023). All involved subjects agreed to participate in this study by informed consent. In this study, the *in vitro* analysis of human serum was performed at Cardiology Department Samara City Clinical Hospital №1 named after N. I. Pirogov and Samara National Research University. The final diagnosis of CHF was made based on the results of an NT-proBNP test.

In this work, a total of 78 patients were included in the *in vitro* experimental study: 14 patients affected by CHF of 1 and 2 stages and 64 subjects of control group. The control group included subjects from the same Cardiology Department of Samara City Clinical Hospital №1 named after N. I. Pirogov, but with an unconfirmed diagnosis of CHF. The serum samples were collected from patients in the morning, after they had fasted. Each sample was placed in a labelled, sterile tube. The tubes were immediately frozen at −16 °C and then transferred to the laboratory for Raman analysis in the freezer. Immediately before the analysis, the samples were defrosted at room temperature. A sample of serum in a volume of 10 μl was placed on aluminum foil and allowed to air dry at room temperature until it was completely dry. In laboratory, for each sample 4−6 Raman spectra were registered from different surface points. A total of 271 Raman spectra were analyzed, including 193 spectra from CHF cases and 78 from control cases.

2.2 Experimental Setup

The experimental setup consisted of the EnSpectr R532 spectrometric system (Spektr-M, Chernogolovka, Russia)) and the ADF U300 microscope (ADF, China) with 50× LMPlan Objective. A laser with a central wavelength of 532 nm was used to acquire Raman spectra in the 89−3892 cm^{-1} region with a spectral resolution of 1.5 cm^{-1}. The laser spot size at the sample surface was approximately 5 μm. All Raman spectra were registered during 2 sec acquisition time with 10 accumulations.

2.3 Spectra Preprocessing

Statistical analysis and data processing were conducted using Python 3, including numpy, scipy, matplotlib, and scikit-learn libraries.

Before recording the Raman spectra of the dried serum samples, a background signal was recorded and then subtracted from the recorded spectrum of the serum using a built-in software algorithm. For the subsequent analysis, the spectral range of 800−1800 cm^{-1} was used, which corresponds to a discrete set of 744 parameters representing the Raman intensities, corresponding to specific values of the wavenumber.

The preprocessing algorithm includes 3 steps: baseline removal with asymmetric least squares [18, 19], smoothing by Savitsky-Golay filter (2 order of polynomial degree, 15 window width size, 0 polynomial order) [20, 21], and normalization by standard normal variation.

2.4 Machine Learning Methods

2.4.1 MCR-ALS

We used the MCR-ALS method [22] to decompose the Raman spectra, which represents a combination of contributions from Raman-active molecular components, into individual spectral components. In this context, each Raman spectrum will be just a linear combination of the spectra of "pure" spectral components:

$$X = C * S^T, \quad (1)$$

where X is a knowing matrix of serum Raman spectra, S is a matrix of derived "pure" spectral components (profiles), and C is a matrix of concentrations. Therefore, the application of MCR-ALS method for analysis of 532-nm stimulated Raman spectra of human serum allows us to recognize the spectra profiles of the most significant serum components and their contribution to the samples.

A principal component analysis (PCA) has been conducted to estimate the correct number of significant spectral components (S) into which the experimental spectral set can be resolved. When the spectra were decomposed into 8 components, the first 4 MCR components explained 97% variances of the initial Raman spectra and eigenvalues of 5, 6, 7, 8 components are strongly dropped. Additionally, we also divided the experimental set into different numbers of components to determine the optimal number. The serum spectra were divided into a different number of components to determine the optimal number. When the spectra were decomposed into 8 components, spectral profiles above the 5th order were noise-like. When the spectra were decomposed into 5 components, the concentration of 5th component was almost entirely contained within 4th component in accordance with exploratory factor analysis. Therefore, we can assume that the number of useful spectral profiles describing serum Raman spectra is 4.

To perform the decomposition of serum Raman spectra using MCR-ALS method we used MCR-ALS GUI 2.0 software [23].

2.4.2 PLS

PLS reduces the spectral dimension to a smaller number of new, orthogonal components, called latent variables (LVs). LVs are linear combinations of the original spectral data and are determined based on the variability in the spectral values. This variability correlates with the different classes of spectra, meaning that only the spectral characteristics that distinguish between the classes being studied are used. In PLS analysis, it is important to determine the correct number of LVs to avoid overfitting of the model [24]. In this study, the number of LVs was determined in accordance with several criteria. We estimated RMSE of the performed K-fold cross validation, and, in addition, we estimated, the shapes and scores of LVs to discard the LVs related to possible noise contribution and scatter of LV scores.

2.4.3 Classification Models

We analyzed various combinations of machine learning algorithms, including MCR with logistic regression (MCR-LR), MCR with linear kernel support vector machine (MCR-SVM), PLS with linear discriminant analysis (PLS-DA), and PLS with linear kernel SVM, to build classification models to distinguish CHF versus control cases. Both PLS and MCR methods allow one to reduce the dimensionality of the spectral data. Considering that data dimensionality reduction had already been performed using MCR, we decided not to combine the MCR and PLS-DA methods.

LR estimates the probability that an observation belongs to a specific class based on a linear threshold that determined by the linear discriminant function. To build the LR classification model, each registered Raman spectra of human serum is represented as a set of concentrations of statistically significant MCR components (MCR-LR).

SVM builds a hyperplane that optimally separates the cases of CHF and control groups. This algorithm assumes that the larger the distance between the hyperplane and observations from the same class, the lower the average error rate of the classifier. In this study, only a linear model was employed to construct a dividing hyperplane. Linear kernel SVM enables the extraction of the most discriminative features, based on the proximity of a predictor vector to a dividing hyperplane.

The PLS scores are traditionally combined with linear discriminant analysis, so we built PLS-DA classification model for analysis of experimental cases based on the preprocessed Raman spectra of serum. Additionally, we combined PLS with a linear kernel SVM, where the SVM algorithm was applied to the scores of the useful LVs.

The experimental set was split into a training set (80%, 220 spectra) and testing (20%, 51 spectra) set. To evaluate the reliability and the stability of the model throughout the training process, 10-fold cross-validation was used. 10-fold cross-validation approach involves dividing the training set into 10 equal sized parts (folds) with each part (fold) being used in turn as a validation part, while the other 9 folds are used to train the model.

Then, the performance results were averaged of the 10 results obtained. This allows one to ensure the stability of the model and to avoid overfitting. After training the model with 10-fold cross-validation on the training set, the final model performance was evaluated on a test set that had not been used in the training process. The diagram of the proposed analysis procedure is presented in Fig. 1.

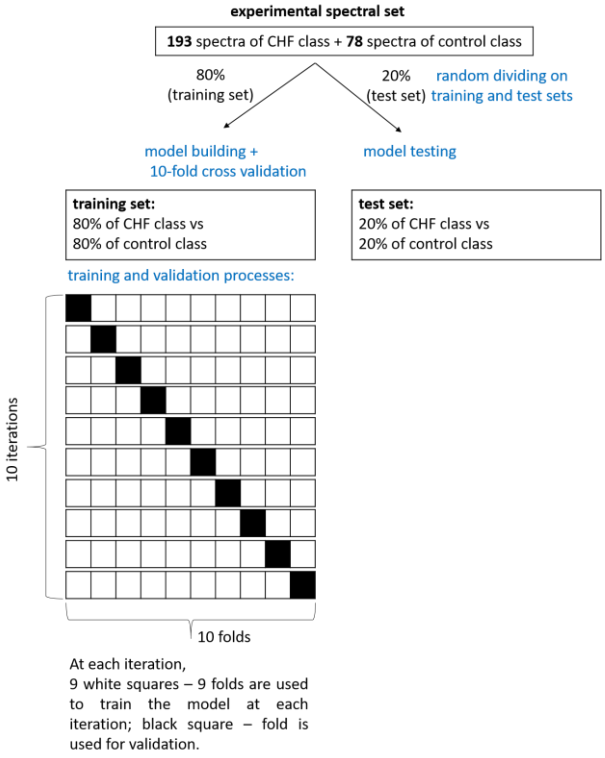


Fig. 1 Diagram of the model building procedure.

The performance of the ML algorithms for differentiating CHF cases from control serum was presented using a receiver operating characteristic (ROC) curve. For quantitative analysis, the area under the curve (AUC) was calculated. The statistical comparison between the ROC AUC values obtained using two different ML methods was conducted according to Hanley and McNeil [25].

3 Results and Discussion

3.1 Serum Raman Spectra

Fig. 2 shows the average serum Raman spectra of CHF and control cases with standard deviation. As is known, biological systems, in particular human serum, contain carotenoids, which give rise to a resonance Raman signal when excited by 532 nm laser [26]. This results in dominant carotenoid Raman bands at 1151 (ν(C-C)) and 1511 (ν(C=C)) cm^{-1}. Also, both spectra have prominent peak at 1004 cm^{-1} can be also assigned to carotenoids (C-CH_3 rocking vibration) and it can be overlapped with a phenylalanine band [26–28]. A deficiency of carotenoids in the blood can cause cell membranes to dry out, leading to cell rupture, fluid leakage, and disruption of cell control, ultimately causing cell death [27]. However, the exact effect of carotenoids, especially β-carotene, on cardiovascular disease remains unclear. For instance, in Ref. [29], it was found that carotenoids in blood serum have an inverse linear relationship with the prevalence of cardiovascular disease. Additionally, serum Raman spectra contains the band at 1445 cm^{-1} corresponded to δ(CH_2), δ(CH_3) in collagen and phospholipids and band at 1665 cm^{-1} corresponded to Amide I of collagen [26, 27]. The low intensity peaks are observed at extended Amide III band (1285, 1336 cm^{-1}).

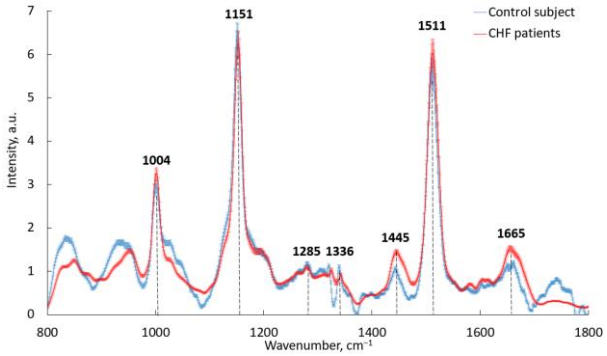


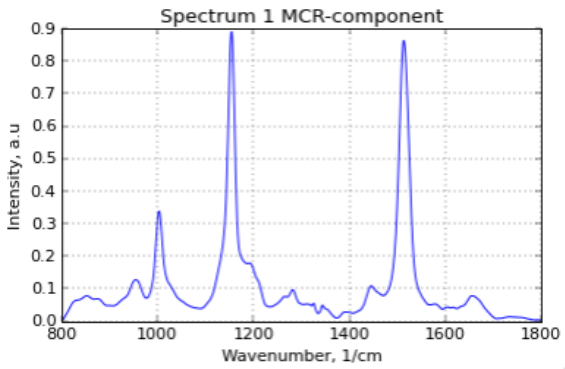
Fig. 2 The 532 nm stimulated Raman spectra of human serum. The shaded area represents the standard deviation.

3.2 MCR-ALS Analysis

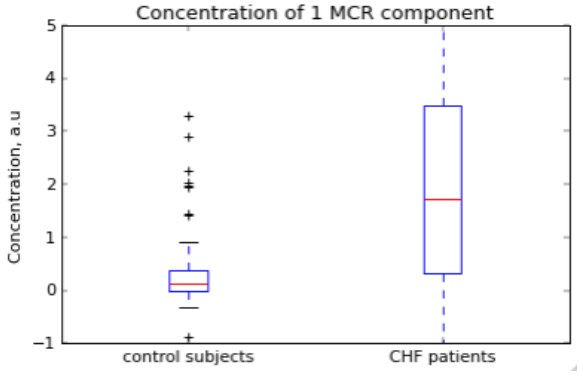
The progression of various diseases, including CHF, is accompanied by quantitative changes in the molecular components of biological tissues and fluids. The Raman spectra of serum of patients with CHF and control patients are expected to show differences in the intensities of the same set of spectral peaks. Therefore, we propose decomposing experimental set of CHF and control Raman spectra into the same spectral MCR components and analyzing the concentrations of these spectral MCR components as new features of the blood serum.

In accordance with PCA analysis 97% explained variance of the experimental data are described by the first 4 MCR components. The spectral profiles and concentrations of the obtained MCR components are presented in Fig. 3.

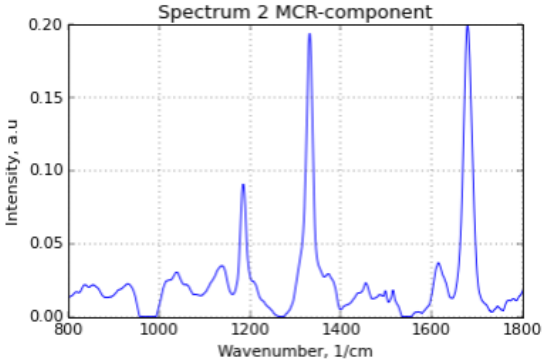
The spectral profile of the 1st MCR component shows the prominent 1004 cm^{-1}, 1151 cm^{-1} and 1511 cm^{-1} bands. This indicates that 1st MCR component mostly corresponds to a carotenoid in blood serum. In accordance with box-plot diagram (Fig. 3b), the concentration values of 1 MCR component is higher in Raman spectra of CHF patients in comparison with control group.



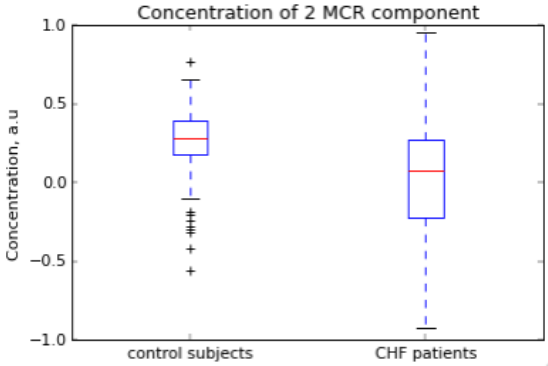
(a)



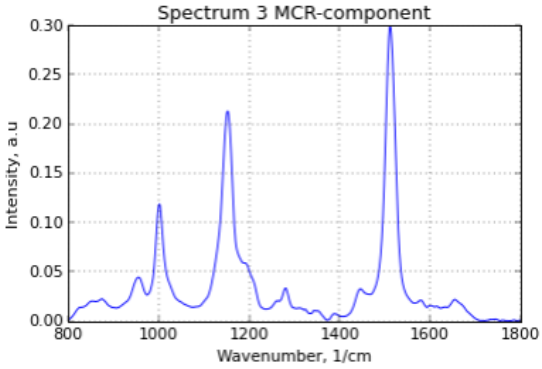
(b)



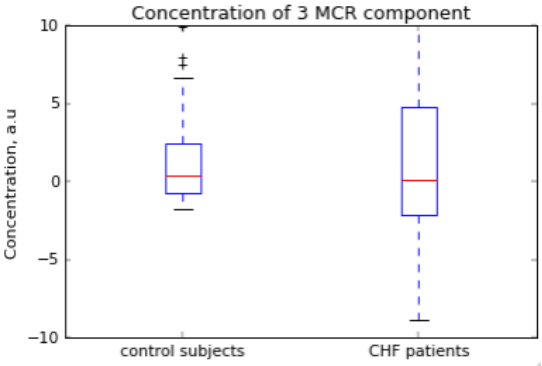
(c)



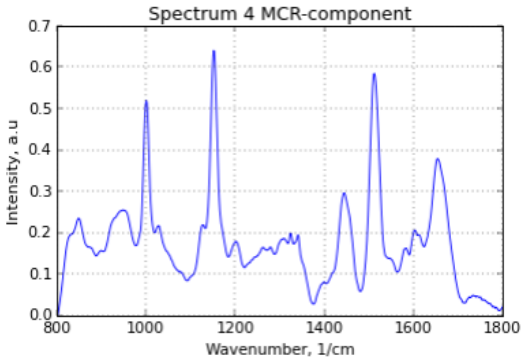
(d)



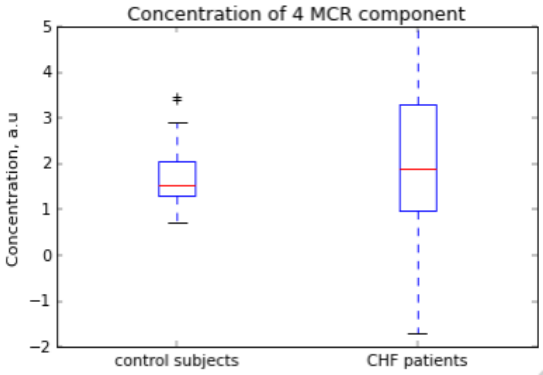
(e)



(f)



(g)



(h)

Fig. 3 MCR components extracted from the serum Raman spectra stimulated by the 532 nm laser wavelength. (a, c, e, g) The spectral profiles of the MCR components. (b, d, f, h) The relative concentrations of the MCR components.

The spectral profile of the 2nd MCR component are characterized by the 1336 cm^{-1} band associated with CH_3CH_2 in collagen, and Amide I band at 1665 cm^{-1}. Also, according to the box-plot diagram (Fig. 3d), the relative concentration values of 2nd MCR component has a shifted median and higher dispersion in the CHF group.

The spectral profile of the 3th MCR component has similar peaks to the 1st MCR components and also can be associated with protein and carotenoids contributions. The 4th MCR component represents the contribution of various proteins at 1445 cm^{-1}, 1665 cm^{-1} bands, carotenoids at 1004 cm^{-1}, 1151 cm^{-1} and 1511 cm^{-1} bands, phenylalanine at 1004 cm^{-1} band. The medians of the concentration relative values of 3th and 4th MCR components in opposite groups have close values while dispersion is higher in CHF group.

The Mann-Whitney *U*-test was used to evaluate the statistical significance of differences in relative concentration values between the CHF and control groups. It was achieved that 1st and 2nd MCR components had a statistically significant difference, with *p*-values of <0.001 and <0.0001 respectively. Therefore, to build classification models to differ CHF and control cases, we will use only the 1st and 2nd MCR components, basing the model on differences in carotenoid and protein levels.

3.3 Classification Models

To build the MCR-LR and MCR-SVM classification models, each Raman spectra of human serum is represented as a concentrations vector of 2 MCR components obtained using the MCR method. The values of 1st and 2nd MCR concentrations were used as predictor variables for 78 CHF and 271 control cases to differ them. The MCR-LR and MCR-SVM models were built to classify CHF and control cases and verified on the test set with 0.50 (0.46−0.53, 0.95 CI) and 0.59 (0.54−0.65, 0.95 CI) ROC AUC, respectively. The ROC curves of both models are presented in Fig. 4.

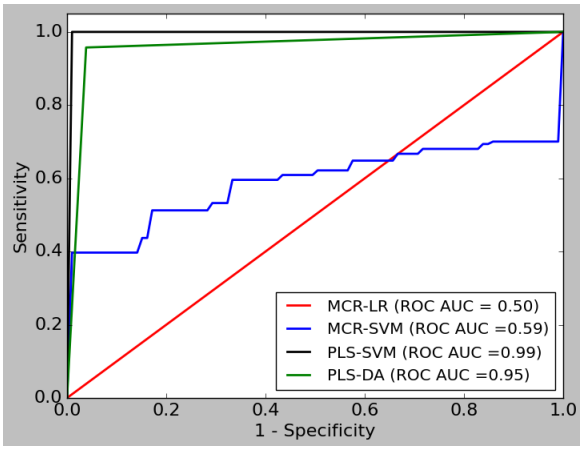
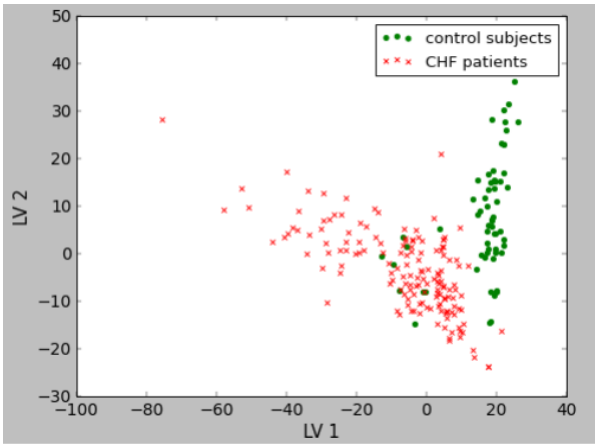


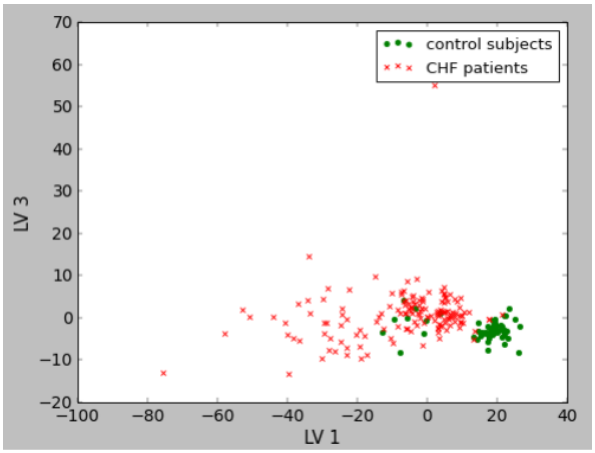
Fig. 4 ROC curves of the binary classification models.

We performed PLS analysis to enable more accurate differentiation of CHF and control cases analyzing preprocessed serum Raman spectra. To determine the optimal number of LVs for our classification model, firstly, we applied minimum of RMSE criterion. The RMSE value has decreased to 3 LVs for the training set and cross validation algorithm. In addition, scatter plots in Fig. 5 demonstrated that scores of 3 LVs allows us good separating the CHF and control cases on LV1−LV2 and LV1−LV3 plots. Therefore, analysis of estimation RMSE values and score distributions, has demonstrated that 3 PCs can be used for training a statistical PLS model in order to avoid overfitting.

The classification model based on analyses of preprocessed Raman spectra using PLS with DA achieves ROC AUC equaled to 0.95 (0.91−0.97, 0.95 CI). The corresponding ROC curve is presented in Fig. 4. The highest ROC AUC of 0.99 (0.94−1.00, 0.95 CI) was obtained when PLS was combined with the linear kernel SVM algorithm for classifying serum SERS spectra between the CHF and control groups.



(a)



(b)

Fig. 5 2D plots of the PLS scores obtained for training set: (a) LV1−LV2, (b) LV1−LV3.

In this study, comparing the ROC AUCs of MCR-LR, MCR-SVM, PLS-DA, and PLS-SVM models together, we revealed that the PLS-SVM method based on analysis of the preprocessed Raman spectra had the best performance with the largest ROC AUC (Fig. 4).

Nevertheless, it is need to estimate the statistical significance of obtained differences in ROC AUC values. The comparison of the MCR-LR and MCR-SVM models' performances equaled to 0.50 (0.46−0.53, 0.95 CI) and 0.59 (0.54−0.65, 0.95 CI) ROC AUCs, respectively, resulted in an insignificant statistical difference ($p = 0.24$). Also, ROC AUCs achieved for PLS-DA (0.95 (0.91−0.97, 0.95 CI)) and PLS-SVM (0.99 (0.94−1.00, 0.95 CI)) algorithms do not differ significantly from each other ($p = 0.09$), as they are statistically similar. However, when we compared the results obtained using different methods, we found a statistically significant improvement in the ROC AUC values for PLS compared to MCR ($p<0.001$ in all cases).

As we can see from the spectral profiles of the MCR components in Fig. 3, the MCR-ALS analysis does not allow us to reconstruct the spectral profile of individual vibrational molecular bonds. Instead, the spectral profile of each MCR component is a complex spectrum composed of a mixture of peaks corresponding to the contribution of several biomolecule groups. It is important to note that the individual MCR components are associated with real biomolecules, but their spectral profiles are not identical to those of pure substances. Therefore, a valuable application of MCR-ALS analysis is the physical interpretation of the differences and similarities between the classes being analyzed, while the extracted MCR components are not sufficient for optimal classification. The full Raman spectra analyzed using PLS and subsequently differentiating them using DA and SVM algorithms enables us to achieve a more accurate model compared to processing the spectra using MCR. PLS performs better at finding the input variables that have a closer relationship with the output variables.

Nevertheless, some studies have demonstrated successful results from the use of MCR [30] techniques for classification tasks. In our opinion, a comparison of modeling techniques for data prediction should be conducted with care, taking into account the type of the biochemical data involved.

4 Conclusions

In this study, we explore the performance of the MCR-LR, MCR-SVM, PLS-DA, and PLS-SVM machine learning algorithms in differentiating the serum Raman spectra of patients with CHF from control subjects. MCR enabled the identification of common patterns and differences in the MCR spectral components associated with the molecular components, however, it was not possible to effectively classify the CHF and control cases based on MCR components using LR and SVM. We found that PLS-DA and PLS-SVM demonstrated the best statistical significant ROC AUCs equaled to 0.95 (0.91−0.97, 0.95 CI) and 0.99 (0.94−1.00, 0.95 CI) in comparison with MCR-LR and MCR-SVM for classification of CHF and controls. Nevertheless, we MCR method allows one to extract the spectral components corresponding to different vibrational bonds which can be valuable for physical interpreting differences between the analyzed classes. Taken together, combination of Raman spectroscopy with MCR and PLS method has shown promising potential for rapid and accurate clinical diagnostics of CHF disease in the near future.

Disclosures

The authors declare no conflicts of interest.

References

1. T. A. McDonagh, M. Metra, M. Adamo, R. S. Gardner, A. Baumbach, M. Böhm, H. Burri, J. Butler, J. Čelutkienė, O. Chioncel, J. G. F. Cleland, A. J. S. Coats, M. G. Crespo-Leiro, D. Farmakis, M. Gilard, S. Heymans, A. W. Hoes, T. Jaarsma, E. A. Jankowska, M. Lainscak, C. S. P. Lam, A. R. Lyon, J. J. V. McMurray, A. Mebazaa, R. Mindham, C. Muneretto, M. Francesco Piepoli, S. Price, G. M. C. Rosano, F. Ruschitzka, A. Kathrine Skibelund, ESC Scientific Document Group, R. A. De Boer, P. Christian Schulze, M. Abdelhamid, V. Aboyans, S. Adamopoulos, S. D. Anker, E. Arbelo, R. Asteggiano, J. Bauersachs, A. Bayes-Genis, M. A. Borger, W. Budts, M. Cikes, K. Damman, V. Delgado, P. Dendale, P. Dilaveris, H. Drexel, J. Ezekowitz, V. Falk, L. Fauchier, G. Filippatos, A. Fraser, N. Frey, C. P. Gale, F. Gustafsson, J. Harris, B. Iung, S. Janssens, M. Jessup, A. Konradi, D. Kotecha, E. Lambrinou, P. Lancellotti, U. Landmesser, C. Leclercq, B. S. Lewis, F. Leyva, A. Linhart, M.-L. Løchen, L. H. Lund, D. Mancini, J. Masip, D. Milicic, C. Mueller, H. Nef, J.-C. Nielsen, L. Neubeck, M. Noutsias, S. E. Petersen, A. Sonia Petronio, P. Ponikowski, E. Prescott, A. Rakisheva, D. J. Richter, E. Schlyakhto, P. Seferovic, M. Senni, M. Sitges, M. Sousa-Uva, C. G. Tocchetti, R. M. Touyz, C. Tschoepe, J. Waltenberger, M. Adamo, A. Baumbach, M. Böhm, H. Burri, J. Čelutkienė, O. Chioncel, J. G. F. Cleland, A. J. S. Coats, M. G. Crespo-Leiro, D. Farmakis, R. S. Gardner, M. Gilard, S. Heymans, A. W. Hoes, T. Jaarsma, E. A. Jankowska, M. Lainscak, C. S. P. Lam, A. R. Lyon, J. J. V. McMurray, A. Mebazaa, R. Mindham, C. Muneretto, M. F. Piepoli, S. Price, G. M. C. Rosano, F. Ruschitzka, and A. K. Skibelund, "2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure," European Heart Journal 42(36), 3599–3726 (2021).
2. M. Onodera, M. Nakamura, F. Tanaka, T. Takahashi, S. Makita, T. Ishisone, Y. Ishibashi, K. Itai, T. Onoda, M. Ohsawa, K. Tanno, K. Sakata, S. Omama, K. Ogasawara, A. Ogawa, T. Kuribayashi, K. Sakamaki, and A. Okayama, "Plasma B-Type Natriuretic Peptide Is Useful for Cardiovascular Risk Assessment in Community-Based Diabetes Subjects: Comparison With Albuminuria," International Heart Journal 53(3), 176–181 (2012).
3. Y. Khristoforova, L. Bratchenko, and I. Bratchenko, "Raman-Based Techniques in Medical Applications for Diagnostic Tasks: A Review," International Journal of Molecular Sciences 24(21), 15605 (2023).

4. S. Abalde-Cela, L. Wu, A. Teixeira, K. Oliveira, and L. Diéguez, "Multiplexing Liquid Biopsy with Surface-Enhanced Raman Scattering Spectroscopy," Advanced Optical Materials 9(15), 2001171 (2021).
5. I. A. Bratchenko, L. A. Bratchenko, Y. A. Khristoforova, A. A. Moryatov, S. V. Kozlov, and V. P. Zakharov, "Classification of skin cancer using convolutional neural networks analysis of Raman spectra," Computer Methods and Programs in Biomedicine 219, 106755 (2022).
6. Y. A. Khristoforova, L. A. Bratchenko, M. A. Skuratova, E. A. Lebedeva, P. A. Lebedev, and I. A. Bratchenko, "Raman spectroscopy in chronic heart failure diagnosis based on human skin analysis," Journal of Biophotonics 16(7), e202300016 (2023).
7. L. Huang, H. Sun, L. Sun, K. Shi, Y. Chen, X. Ren, Y. Ge, D. Jiang, X. Liu, W. Knoll, Q. Zhang, and Y. Wang, "Rapid, label-free histopathological diagnosis of liver cancer based on Raman spectroscopy and deep learning," Nature Communications 14(1), 48 (2023).
8. H. Yang, C. Zhao, R. Li, C. Shen, X. Cai, L. Sun, C. Luo, and Y. Yin, "Noninvasive and prospective diagnosis of coronary heart disease with urine using surface-enhanced Raman spectroscopy," Analyst 143(10), 2235–2242 (2018).
9. H. Leng, C. Chen, R. Si, C. Chen, H. Qu, and X. Lv, "Accurate screening of early-stage lung cancer based on improved ResNeXt model combined with serum Raman spectroscopy," Journal of Raman Spectroscopy 53(7), 1302–1311 (2022).
10. N. M. Ralbovsky, I. K. Lednev, "Raman spectroscopy and chemometrics: A potential universal method for diagnosing cancer," Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 219, 463–487 (2019).
11. I. Matveeva, I. Bratchenko, Y. Khristoforova, L. Bratchenko, A. Moryatov, S. Kozlov, O. Kaganov, and V. Zakharov, "Multivariate Curve Resolution Alternating Least Squares Analysis of In Vivo Skin Raman Spectra," Sensors 22(24), 9588 (2022).
12. A. Ospanov, I. Romanishkin, T. Savelieva, A. Kosyrkova, S. Shugai, S. Goryaynov, G. Pavlova, I. Pronin, and V. Loschenov, "Optical Differentiation of Brain Tumors Based on Raman Spectroscopy and Cluster Analysis Methods," International Journal of Molecular Sciences 24(19), 14432 (2023).
13. S. Shafaq, M. Irfan Majeed, H. Nawaz, N. Rashid, M. Akram, N. Yaqoob, A. Tariq, S. Shakeel, A. Ul Haq, M. Saleem, M. Zaman Nawaz, and R. Zaki Abdul Bari, "Quantitative analysis of solid dosage forms of Losartan potassium by Raman spectroscopy," Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 272, 120996 (2022).
14. K. Bērziņš, S. D. L. Harrison, C. Leong, S. J. Fraser-Miller, M. J. Harper, A. Diana, R. S. Gibson, L. A. Houghton, and K. C. Gordon, "Qualitative and quantitative vibrational spectroscopic analysis of macronutrients in breast milk," Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 246, 118982 (2021).
15. I. Behl, G. Calado, A. Vishwakarma, D. Traynor, S. Flint, S. Galvin, C. M. Healy, M. L. Pimentel, A. Malkin, H. J. Byrne, and F. M. Lyng, "Classification of cytological samples from oral potentially malignant lesions through Raman spectroscopy: A pilot study," Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 266, 120437 (2022).
16. S. Zhu, Y. Li, H. Gao, G. Hou, X. Cui, S. Chen, and C. Ding, "Identification and assessment of pulmonary Cryptococcus neoformans infection by blood serum surface-enhanced Raman spectroscopy," Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 260, 119978 (2021).
17. I. Bratchenko, Y. Khristoforova, L. Bratchenko, A. Moryatov, S. Kozlov, E. Borisova, and V. Zakharov, "Optical Biopsy of Amelanotic Melanoma with Raman and Autofluorescence Spectra Stimulated by 785 nm Laser Excitation," Journal of Biomedical Photonics & Engineering 7(2), 020308 (2021).
18. P. H. C. Eilers, H. F. M. Boelens, "Baseline correction with asymmetric least squares smoothing," Leiden University Medical Centre Report 1(1), 5 (2005).
19. S. He, W. Zhang, L. Liu, Y. Huang, J. He, W. Xie, P. Wu, and C. Du, "Baseline correction for Raman spectra using an improved asymmetric least squares method," Analytical Methods 6(12), 4402–4407 (2014).
20. A. Savitzky, M. J. E. Golay, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures," Analytical Chemistry 36(8), 1627–1639 (1964).
21. I. Pence, A. Mahadevan-Jansen, "Clinical instrumentation and applications of Raman spectroscopy," Chemical Society Reviews 45(7), 1958–1979 (2016).
22. J. Felten, H. Hall, J. Jaumot, R. Tauler, A. De Juan, and A. Gorzsás, "Vibrational spectroscopic image analysis of biological material using multivariate curve resolution–alternating least squares (MCR-ALS)," Nature Protocols 10(2), 217–240 (2015).
23. J. Jaumot, A. De Juan, and R. Tauler, "MCR-ALS GUI 2.0: New features and applications," Chemometrics and Intelligent Laboratory Systems 140, 1–12 (2015).
24. I. A. Bratchenko, L. A. Bratchenko, "Comment on "Combining derivative Raman with autofluorescence to improve the diagnosis performance of echinococcosis"," Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 252, 119514 (2021).
25. J. A. Hanley, B. J. McNeil, "A method of comparing the areas under receiver operating characteristic curves derived from the same cases," Radiology 148(3), 839–843 (1983).

26. K. Kralova, M. Kral, O. Vrtelka, and V. Setnicka, "Comparative study of Raman spectroscopy techniques in blood plasma-based clinical diagnostics: A demonstration on Alzheimer's disease," Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 304, 123392 (2024).
27. M. Saleem, M. Bilal, S. Anwar, A. Rehman, and M. Ahmed, "Optical diagnosis of dengue virus infection in human blood serum using Raman spectroscopy," Laser Physics Letters 10(3), 035602 (2013).
28. H. Horiue, M. Sasaki, Y. Yoshikawa, M. Toyofuku, and S. Shigeto, "Raman spectroscopic signatures of carotenoids and polyenes enable label-free visualization of microbial distributions within pink biofilms," Scientific Reports 10(1), 7704 (2020).
29. M. Wang, R. Tang, R. Zhou, Y. Qian, and D. Di, "The protective effect of serum carotenoids on cardiovascular disease: a cross-sectional study from the general US adult population," Frontiers in Nutrition 10, 1154239 (2023).
30. I. A. Matveeva, A. I. Komlev, O. I. Kaganov, A. A. Moryatov, and V. P. Zakharov, "Multidimensional Analysis of Dermoscopic Images and Spectral Information for the Diagnosis of Skin Tumors," Journal of Biomedical Photonics & Engineering 10(1), 010307 (2024).