

Development of a Method for the Intelligent Analysis of Raman Spectra of Blood Serum

Irina A. Pimenova* and Irina A. Matveeva

Department of Laser and Biotechnical Systems, Samara National Research University, 34 Moskovskoe shosse, Samara 443086, Russian Federation

*e-mail: pimenova.0312@list.ru

Abstract. This paper presents a method for intelligent analysis of Raman spectra of blood serum for chronic heart failure (CHF) diagnosis using machine learning approaches. Raman spectra obtained from blood serum samples of 229 subjects (189 patients with CHF and 40 healthy volunteers) were analyzed using the multivariate curve resolution-alternating least squares (MCR-ALS) method. Eight spectral components were extracted and used as features for classification. Logistic regression, support vector machine, random forest, gradient boosting, and ensemble stacking methods were evaluated. Using only spectral features, the stacking model achieved an area under the receiver operating characteristic curve (ROC AUC) of 0.72 ± 0.04 with a validation accuracy of 0.83 ± 0.03 . The inclusion of clinical parameters (systolic and diastolic blood pressure, final systolic size, total bilirubin, and hematocrit) significantly improved classification performance, increasing the ROC AUC to 1.00 ± 0.00 with a validation accuracy of 0.95 ± 0.01 . Statistical analysis showed no significant differences between the stacking ensemble and individual classifiers ($p > 0.05$), indicating that classification performance was primarily determined by the informativeness of the extracted spectral and clinical features. The obtained results demonstrate the potential of combining Raman spectroscopy with machine learning methods for CHF screening and early diagnosis, particularly in settings where access to advanced imaging techniques is limited.

Keywords: alternating least squares; multidimensional curve resolution method; Raman spectroscopy; blood serum; chronic heart failure; machine learning; stacking.

Paper #9382 received 20 Oct 2025; revised manuscript received 10 Sep 2026; accepted for publication 11 Sep 2026; published online 30 Sep 2026. [doi: 10.18287/JBPE26.12.030306](https://doi.org/10.18287/JBPE26.12.030306).

1 Introduction

Cardiovascular diseases occupy a leading place among the causes of death and disability worldwide, which underlines the importance of their diagnosis for healthcare systems [1]. Chronic heart failure (CHF) is one of the most widespread cardiovascular syndromes associated with progressive impairment of cardiac function, reduced quality of life, and increased mortality risk [2]. Early diagnosis of CHF remains a challenging task, especially at the initial stages of disease development, when structural and functional changes may not yet be sufficiently pronounced. Conventional

diagnostic approaches, including echocardiography and biochemical blood analysis, often require expensive equipment, specialized personnel, and considerable examination time, which limits their widespread use in screening programs and preventive medical examinations [3].

Biochemical blood analysis methods are used to quantify the concentration of blood components, enzymes, and hormones, as well as to identify deviations from normal values [4]. However, the prognostic significance of blood analysis can be increased when considering a complex of changes in the biochemical composition of blood rather than isolated biomarkers [5].

In this regard, optical spectroscopic methods capable of providing information about the molecular composition of biological samples attract significant interest [6].

Raman spectroscopy makes it possible to analyze the features of biological objects at the molecular level [7]. In recent years, Raman spectroscopy has opened new perspectives for blood analysis [8]. This method allows obtaining spectral information about the molecular composition of complex biological samples, including blood serum, making it a powerful tool for detecting biochemical changes associated with cardiovascular diseases [9]. Due to its high sensitivity, noninvasiveness, and minimal sample preparation requirements, Raman spectroscopy is becoming an attractive method for clinical diagnostics and biomedical screening applications.

However, Raman spectra of biological samples are characterized by high dimensionality, overlapping spectral bands, noise sensitivity, and substantial inter-patient variability [10]. These factors complicate the interpretation of spectral data and require the use of advanced computational approaches for efficient analysis. In this context, machine learning and intelligent decision-support systems have become increasingly important tools for Raman spectra processing [11].

Various machine learning algorithms have been applied to the analysis of Raman spectra in biomedical research. Classical statistical and machine learning methods, including Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Logistic Regression, Support Vector Machines (SVM) have demonstrated effectiveness in classification tasks involving biological samples [12–14]. The development of information technology has also led to attempts to use artificial neural networks and deep learning methods for spectral data analysis [15]. In particular, convolutional neural networks have shown promising results in automated Raman spectrum recognition and feature extraction [16]. These approaches enable identification of complex nonlinear relationships within spectral data and improve diagnostic accuracy for various pathological conditions [17].

Despite significant progress, the problem of extracting informative and interpretable features from Raman spectra remains unresolved. In many studies, dimensionality reduction is performed using PCA or related statistical methods [18–19]. Although such approaches effectively reduce computational complexity, the resulting components are often difficult to interpret from a biochemical perspective [20]. Therefore, methods capable of extracting physically interpretable spectral components are of particular interest for biomedical applications [21].

One promising approach is multivariate curve resolution using alternating least squares (MCR-ALS), which allows decomposition of spectral data into concentration profiles and spectra of underlying biochemical components [22]. Unlike conventional statistical decomposition techniques, MCR-ALS provides improved interpretability of extracted features

and may facilitate identification of molecular changes associated with disease development [23].

Another important trend is the development of ensemble machine learning methods that combine predictions of several individual classifiers [24]. Ensemble approaches can improve the robustness and generalization ability of classification models, especially when analyzing complex biomedical datasets. In particular, stacking methods combine the advantages of multiple algorithms by constructing a meta-model based on predictions of basic classifiers [25].

The novelty of the proposed approach lies in the combined use of MCR-ALS decomposition, ensemble machine learning methods, and integration of Raman spectral data with clinical patient parameters for chronic heart failure diagnosis. Unlike conventional approaches based solely on spectral preprocessing and standard dimensionality reduction techniques, the proposed framework extracts interpretable biochemical components from Raman spectra and combines them with clinically significant features. In addition, several machine learning algorithms, including logistic regression, support vector machines, random forest, gradient boosting, and stacking ensembles, are comparatively evaluated for CHF classification.

The use of Raman spectroscopy in combination with intelligent data analysis systems is a promising area of research and can increase the effectiveness of population screening and preventive medical examinations. Such approaches can improve the accessibility of diagnostics, contribute to the early detection of cardiovascular diseases, and potentially prevent disease progression [26].

The aim of this study was to develop an intelligent method for the analysis of blood serum Raman spectra using the MCR-ALS method and machine learning algorithms, taking into account additional clinical data of patients with chronic heart failure.

2 Methods and Materials

In this work, the *in vitro* Raman spectra of the blood serum of 229 patients are analyzed, among which 189 are patients with chronic heart failure, 40 are control group (volunteers without heart failure). Blood serum samples were taken from the patient in the morning on an empty stomach and placed in sterile tubes, followed by freezing at a temperature of -16°C . Immediately before spectral analysis, the samples were thawed at room temperature.

The recording of Raman spectra was carried out using surface-enhanced Raman spectroscopy (SERS) technology, described in detail in Ref. [27]. Each 1.5 ml serum sample was applied to an aluminum foil with a layer of silver structures and dried for 30 min. The spectral characteristics of blood serum were analyzed using an experimental stand consisting of a spectrometric system (EnSpectr R785, Spektr-M, Chernogolovka, Russia) and a microscope (ADF U300, ADF, China). The spectra were excited in the near-infrared range using a laser module with a central wavelength of 785 nm. An LMPlan lens with a $50\times$ magnification was used to focus the radiation on the

sample and collect the scattered radiation. The diameter of the laser spot in focus was 5 μm . Human serum was analyzed at a laser power of 10 mW. The spectra were recorded four times for each sample with an exposure time of 4 s. Further, the spectra were subjected to the preprocessing described in Ref. [28].

In this work, the method of resolving multidimensional curves using the alternating least squares (MCR-ALS) method is used to analyze the spectra, which has already shown its effectiveness in isolating the biochemical components of the skin [29]. The main idea of MCR is to decompose the matrix of Raman spectra into two matrices, one of which represents the concentration profiles for each component, and the second represents the matrix of the spectra of these components [30].

In the framework of the MCR-ALS analysis, the term “components” refers to latent spectral profiles extracted from the original Raman spectra during matrix decomposition [31].

In this approach, the concentrations of each individual component isolated in the studied samples serve as features for subsequent classification. The following machine learning algorithms were applied to the profiles of the obtained component concentrations: Logistic Regression, SVM, Random Forest, and Gradient Boosting.

Logistic Regression is a form of multiple regression that aims to analyze the relationship between several independent variables and a single binary dependent variable [32–34]. Random forest is based on the construction of an ensemble of decision trees and is one of the most robust machine learning algorithms [35–37]. The principle of the support vector machine method is to find the optimal hyperplane that maximally separates classes in a multidimensional feature space [38–40]. Gradient boosting is a powerful ensemble learning method in which the sequential construction of decision trees aims to minimize the errors of previous models [41–43].

To improve the quality of the final model and combine the advantages of various methods, the stacking method was applied, in which the basic models formed meta-features, which were then fed to the input of the final model, the logistic regression model. This approach makes it possible to exploit the strengths of individual algorithms, improving classification generalization performance.

In addition to concentration profiles, the clinical data of patients were used as additional features for classification: systolic blood pressure, diastolic blood pressure, final systolic size, total bilirubin, hematocrit.

A 5-fold cross-validation was performed to verify the stability of the classification models. All machine learning algorithms and ROC curve analyses were implemented in Python.

3 Results and Discussions

Using MCR-ALS, eight components have been identified. This amount was determined during

preliminary studies, which examined the range from 3 to 12 components in order to optimally identify substances. Several datasets with different numbers of components were analyzed using various machine learning methods. The ROC curves obtained for logistic regression based on five data sets are shown in Fig. 1. Based on the analysis, it is concluded that the choice of eight components is optimal for achieving the objectives of the study. The primary objective was to retain as much information as possible, which would improve the accuracy of the analysis of the Raman spectra of blood serum without losing key data.

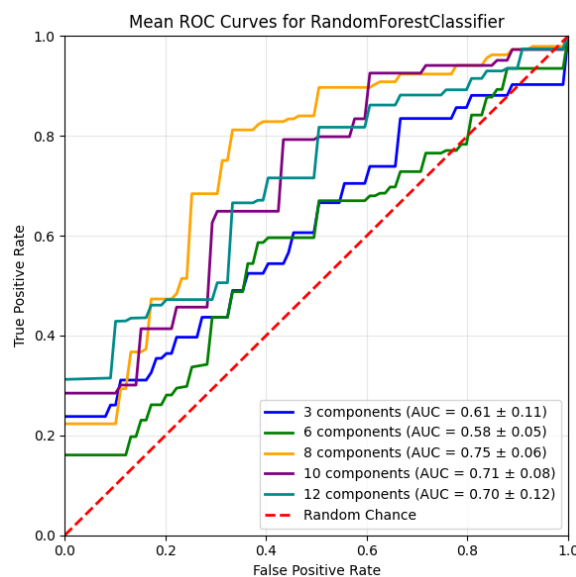


Fig. 1 ROC-curves for logistic regression for different numbers of components.

The main peaks for each component are shown in Table 1. The spectra of the components are shown in Fig. 2. For example, Component 1 was mainly associated with proteins and lipids due to the presence of peaks near 723 cm^{-1} , 1097 cm^{-1} , and 1329 cm^{-1} , corresponding to amide groups, C–O vibrations, and hydrocarbon chain deformations. Component 2 was related predominantly to protein structures and aromatic amino acids because of the characteristic Amide I band near 1652 cm^{-1} and aromatic ring vibrations at 1576 cm^{-1} . Component 3 was associated with metabolites and carbohydrate-related vibrations, while Components 4 and 5 reflected contributions from aromatic compounds, proteins, and lipid-associated molecular structures. Components 6–8 mainly corresponded to carbohydrates, amino acid residues, and low-frequency molecular vibrations observed in biological serum samples [44–45].

Therefore, the obtained concentration profiles reflect variations in the biochemical composition of blood serum associated with chronic heart failure.

As a result, the area under the ROC curve for the case of taking into account only spectral features (component concentrations) and using the basic logistic regression model was 0.72 ± 0.04 (validation accuracy: 0.83 ± 0.03), the support vector machine achieved an

ROC AUC of 0.66 ± 0.08 (validation accuracy: 0.83 ± 0.03), random forest 0.70 ± 0.07 (validation accuracy: 0.85 ± 0.01), gradient boosting 0.62 ± 0.05 (validation accuracy: 0.81 ± 0.02). For the stacking model, the area under the ROC curve was 0.72 ± 0.04 (validation accuracy: 0.83 ± 0.03). The addition of clinical patient data to the feature space has significantly improved the learning performance of the model. The ROC AUC increased to 1.00 ± 0.00 , while the validation accuracy was 0.95 ± 0.01 .

Table 1 Characteristic peaks and assumed molecular bonds of the components.

Component	Peak (cm ⁻¹)	Possible Assignment
Component 1	723.12	Amide group or methylene groups in lipids
	1329.03	C–H deformation in hydrocarbon chains (possibly Amide III)
	1096.99	C–O or S–O stretching in carbohydrates or carboxylic acids
Component 2	1652.38	C=O stretching in amide groups or carbohydrates (Amide I)
	1576.09	Aromatic ring stretching (e.g., phenyl groups)
	639.33	Amino acids associated with aromatic ring vibrations
Component 3	1393.58	Vibrations typical for creatinine or other metabolites
	1021.23	Carbohydrate or amino acid vibrations
	631.35	Amino acid residues (e.g., phenylalanine)
Component 4	1577.62	Aromatic vibrations related to phenyl groups
	1212.39	Vibrations in carbohydrates or amide groups
	1313.32	C–H deformation or aromatic ring vibrations
Component 5	624.35	Amino acids or carbohydrates
	1364.32	C–H or C–C deformations in hydrocarbon chains
	1604.25	Aromatic ring vibrations
Component 6	1053.51	C–O stretching in carbohydrates or metabolites
	677.01	Aromatic ring vibrations or possible methylene groups
	492.15	Low-frequency vibrations in organic molecules
Component 7	645.30	Aromatic rings or carbohydrates
	500.41	Methylene groups or low-frequency vibrations
	1142.56	C–O or methyl group stretching
Component 8	634.34	Amino acids or carbohydrates
	491.12	Vibrations in carbohydrates or amino acids
	1132.10	Presence of amide groups (Amide III region)

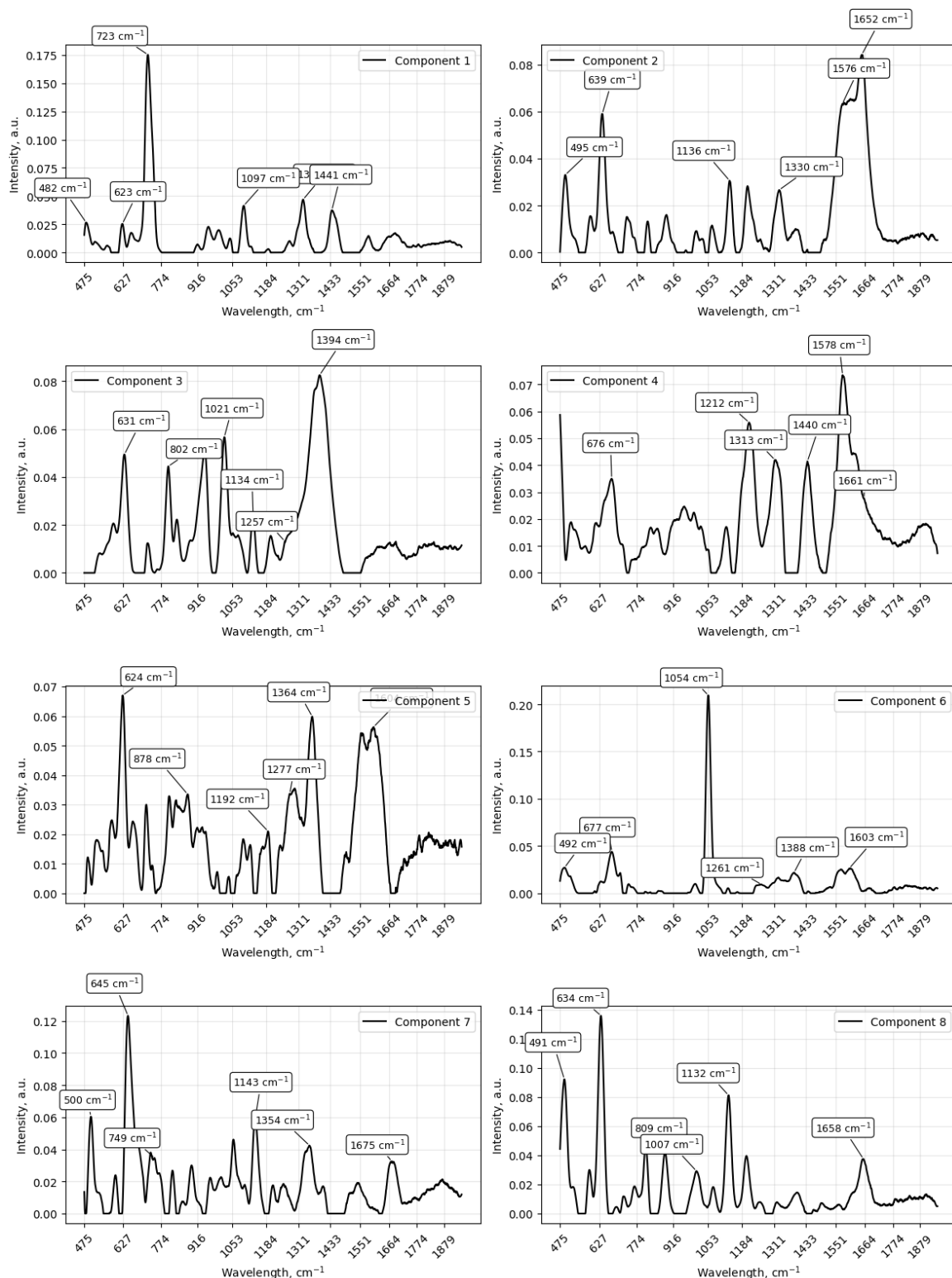


Fig. 2 Raman spectra obtained by *in vitro* MCR analysis of Raman scattering spectra of blood serum.

In the framework of this work, an ensemble classification model was implemented using the stacking method. The aim was to combine several basic models to improve the accuracy of the prediction of CHF based on spectral and clinical features. The ROC curves obtained as a result of the analysis are shown in Fig. 3.

To further validate the obtained results, a statistical significance analysis was conducted comparing the

baseline models with the stacked ensemble. The comparison was performed using the area under the ROC curve (AUC) obtained from 5-fold cross-validation. Since the performance values were calculated on the same data splits, paired statistical tests were applied, including the paired Student's *t*-test [46] and the nonparametric Wilcoxon signed-rank test [47].

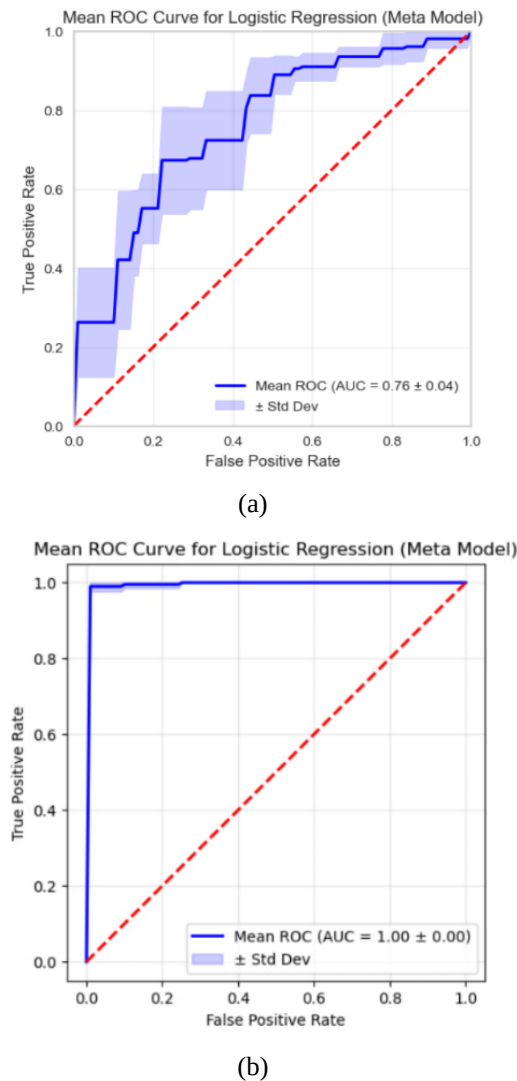


Fig. 3 ROC-curves for classification of healthy people and patients with chronic heart failure: (a) only component concentrations are taken as features; (b) component concentrations and clinical data are taken as features.

The results show that the differences between the stacked model and individual baseline classifiers are statistically insignificant ($p > 0.05$ in all cases). Specifically, the p -values for logistic regression were 0.3408 (t -test) and 0.3125 (Wilcoxon test), 0.2463 and 0.3125 for support vector machines, 0.8445 and 0.8125 for random forest, and 0.0722 and 0.1250 for gradient boosting, respectively. Although gradient boosting approached statistical significance ($p \approx 0.07$), it does not reach the generally accepted significance level of 0.05.

These results are consistent with the similar performance metrics reported above. The lack of statistically significant differences can be explained by several factors. First, the relatively small sample size (229 samples, of which only 40 were healthy controls) leads to increased variability in AUC estimates across different cross-validation folds, reducing the statistical power of the tests. Second, the baseline models demonstrate comparable performance levels and likely

capture similar patterns in the feature space formed by the eight MCR-ALS components. Under these conditions, the advantage of stacking is diminished because the metamodel does not receive sufficiently diverse or additional information.

Thus, the statistical analysis confirms that the stacking approach does not provide a statistically significant improvement over the individual classifiers in this study. This suggests that, within the considered framework, the overall classification performance is primarily determined by the informativeness of the extracted features rather than by the choice of a specific classification algorithm.

We conducted an additional experiment using only five clinical parameters (systolic and diastolic blood pressure, final systolic size, total bilirubin, and hematocrit). The obtained results demonstrated near-perfect classification performance (Fig. 4; Mean Accuracy: 0.96 ± 0.03 ; Mean AUC: 0.98 ± 0.03).

Such high performance can be explained by the strong association between these clinical variables and the diagnosis of chronic heart failure.

Among the clinical parameters used in this study, final systolic size represents a key echocardiographic indicator associated with structural cardiac remodeling and is directly related to the diagnosis of chronic heart failure. Other variables, such as blood pressure, bilirubin, and hematocrit, are not diagnostic criteria per se but reflect systemic and metabolic alterations accompanying the disease.

The combination of these features provides a strong discriminative signal, which may explain the high classification performance observed in this dataset.

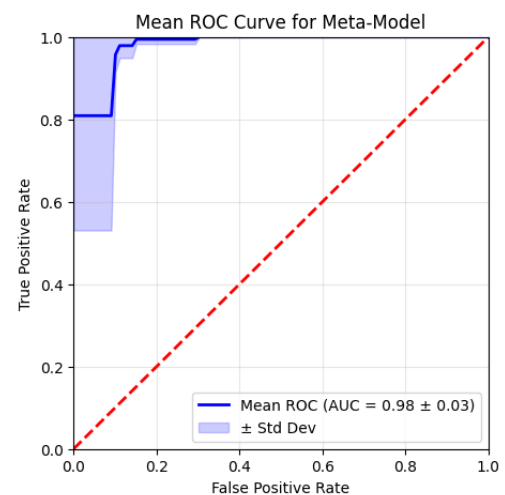


Fig. 4 ROC-curves for classification of healthy people and patients with chronic heart failure; clinical data are taken as features.

However, this result is likely to reflect dataset-specific characteristics and may lead to an overestimation of model performance, especially in the absence of external validation. In real-world clinical settings, variability of patient populations and overlapping clinical features typically result in lower separability. Despite the high

discriminative power of clinical parameters, it should be emphasized that these features are typically obtained using established diagnostic procedures, such as echocardiography, which may not always be available in all clinical settings.

In this context, Raman spectroscopy may provide important practical advantages. In particular, it can be considered as a potential tool for screening in settings where access to advanced imaging techniques is limited, or as a preliminary diagnostic approach prior to comprehensive clinical evaluation.

Furthermore, Raman spectroscopy reflects molecular-level changes in blood serum and therefore may be sensitive to early metabolic alterations that precede pronounced structural cardiac remodeling. This suggests its potential application for early-stage detection, where conventional clinical parameters may not yet exhibit strong diagnostic signals.

Thus, the value of Raman spectroscopy lies not in replacing established clinical diagnostics, but in complementing them, particularly in screening scenarios and early disease detection.

Machine learning methods have been applied to the analysis of biological fluids for the diagnosis and stratification of CHF in a number of recent studies. Thus, Sorokina et al. [48] conducted a systematic comparison of methods for intelligent analysis of Raman spectra of blood serum from patients with CHF and healthy volunteers using combinations of MCR and PLS with different classifiers (Logistic Regression, SVM). They showed that PLS-based approaches provide higher AUC values (up to 0.99), while MCR in combination with single classifiers demonstrated moderate performance (AUC 0.59–0.70), while leaving the possibility of biochemical interpretation of serum components. Other studies, such as De Bakker et al. [49] and Gao et al. [50] used proteomic and biomarker panels, applying ML algorithms to predict outcomes and classify patients with CHF, achieving a c index of 0.80–0.85 or an accuracy of approximately 85%. The approach by Woolley et al. [51] demonstrated the potential of clustering biomarker profiles to distinguish pathophysiological subtypes of HFpEF, but without direct application to automated diagnostics.

In contrast to these studies, this work uses SERS to directly assess the molecular composition of blood serum. The use of the MCR-ALS method allowed us to identify eight informative components with biochemical interpretation (amino acids, carbohydrates, lipids, amide groups). The resulting concentration profiles, combined with patient clinical data, were used to build ensemble machine learning models, including stacking base classifiers (Logistic Regression, SVM, Random Forest, and Gradient Boosting). This approach achieved an AUC of 1.0 ± 0.0 , which exceeds the MCR results reported by Sorokina et al. [48] and is comparable to the best results

from PLS methods, while maintaining the interpretability of features.

4 Conclusion

In this study, a method for intelligent analysis of Raman spectra of blood serum for chronic heart failure diagnosis was developed using MCR-ALS decomposition and machine learning approaches. Raman spectra obtained from 229 serum samples were decomposed into eight latent spectral components corresponding to major biochemical groups, including proteins, lipids, carbohydrates, amino acids, and metabolites. The concentration profiles of these components were used as features for classification.

Several machine learning algorithms, including logistic regression, support vector machine, random forest, gradient boosting, and stacking ensemble methods, were evaluated. Using only spectral features, the best classification performance corresponded to the stacking model and logistic regression model with an ROC AUC of 0.72 ± 0.04 and validation accuracy of 0.83 ± 0.03 . The integration of clinical parameters significantly improved classification results, increasing the ROC AUC to 1.00 ± 0.00 .

Statistical analysis demonstrated that differences between the stacking ensemble and individual classifiers were not statistically significant ($p > 0.05$), indicating that classification performance was primarily determined by the informativeness of the extracted spectral and clinical features rather than by the choice of a particular classification algorithm.

The obtained results demonstrate that Raman spectroscopy combined with machine learning methods can be considered a promising tool for blood serum analysis and chronic heart failure screening. In particular, the proposed approach may be useful in situations where rapid, minimally invasive, and cost-effective preliminary diagnostics are required. Furthermore, Raman spectroscopy may provide additional information about molecular-level biochemical changes associated with disease development.

Future work will focus on expanding the dataset, performing external validation on independent patient cohorts, and investigating additional biochemical and clinical markers to improve the robustness and clinical applicability of the proposed diagnostic approach.

Acknowledgment

This work was supported by the Russian Science Foundation project № 25-75-00146. <https://rscf.ru/project/25-75-00146/>.

Disclosures

All authors declare that there is no conflict of interest in this paper.

References

1. “Cardiovascular diseases (CVDs),” World Health Organization (WHO), 31 July 2025 (accessed 29 December 2025). [https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-%28cvds%29?utm_source].

2. V. Valente, R. Laborante, A. Abdin, P. M. Becher, M. Lainscak, M. Polovina, C. Gavina, and G. Savarese, "[The global epidemiology of heart failure: a comprehensive and contemporary review](#)," *European Journal of Heart Failure* xuag121 (2026).
3. P. A. Heidenreich, B. Bozkurt, D. Aguilar, L. A. Allen, J. J. Byun, M. M. Colvin, A. Deswal, M. H. Drazner, S. M. Dunlay, L. R. Evers, J. C. Fang, S. E. Fedson, G. C. Fonarow, S. S. Hayek, A. F. Hernandez, P. Khazanie, M. M. Kittleson, C. S. Lee, M. S. Link, C. A. Milano, L. C. Nnacheta, A. T. Sandhu, L. W. Stevenson, O. Vardeny, A. R. Vest, and C. W. Yancy, "[2022 AHA/ACC/HFSA guideline for the management of heart failure: a report of the american college of cardiology/american heart association joint committee on clinical practice guidelines](#)," *Circulation* 145(18), (2022).
4. D. Glick (Ed.), *Methods of Biochemical Analysis*, Hoboken, John Wiley & Sons, NJ, USA (2009).
5. C. G. Atkins, K. Buckley, M. W. Blades, and R. F. Turner, "[Raman spectroscopy of blood and blood components](#)," *Applied Spectroscopy* 71(5), 767–793 (2017).
6. H. J. Butler, L. Ashton, B. Bird, G. Cinque, K. Curtis, J. Dorney, K. Esmonde-White, N. J. Fullwood, B. Gardner, P. L. Martin-Hirsch, M. J. Walsh, M. R. McAinsh, N. Stone, and F. L. Martin, "[Using Raman spectroscopy to characterize biological materials](#)," *Nature Protocols* 11(4), 664–687 (2016).
7. N. Kuhar, S. Sil, T. Verma, and S. Umapathy, "[Challenges in application of Raman spectroscopy to biology and materials](#)," *RSC Advances* 8(46), 25888–25908 (2018).
8. J. Dybas, F. C. Alcicek, A. Wajda, M. Kaczmarek, A. Zimna, K. Bulat, A. Blat, T. Stepanenko, T. Mohaissen, E. Szczesny-Malysiak, D. Perez-Guaita, B. R. Wood, and K. M. Marzec, "[Trends in biomedical analysis of red blood cells – Raman spectroscopy against other spectroscopic, microscopic and classical techniques](#)," *TrAC Trends in Analytical Chemistry* 146, 116481 (2022).
9. 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).
10. J. Servet Lerdo De Tejada, J. Vališ, L. Hrubčík, D. Janstová, J. Tomeš, T. Uhlíková, M. Vazdar, J. Mareš, and P. Matějka, "[Mathematical and machine learning-assisted modelling of Raman spectroscopy for biomedical applications](#)," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 362, 128199 (2026).
11. S. Guo, J. Popp, and T. Bocklitz, "[Chemometric analysis in Raman spectroscopy from experimental design to machine learning-based modeling](#)," *Nature Protocols* 16(12), 5426–5459 (2021).
12. E. D. Omar, H. Mat, A. Z. Abd Karim, R. Sanaudi, F. Ibrahim, M. A. Omar, M. Z. H. Ismail, V. Jayaraj, and B. L. Goh, "[Comparative analysis of logistic regression, gradient boosted trees, SVM, and random forest algorithms for prediction of acute kidney injury requiring dialysis after cardiac surgery](#)," *International Journal of Nephrology and Renovascular Disease* 17, 197–204 (2024).
13. J.-M. Yin, Y. Li, J.-T. Xue, G.-W. Zong, Z.-Z. Fang, and L. Zou, "[Explainable machine learning-based prediction model for diabetic nephropathy](#)," *Journal of Diabetes Research* 2024(1), 8857453 (2024).
14. Md. S. Hossain, Md. A. Talukder, and Md. Z. Mahmud, "[Advancements in cardiovascular disease detection: leveraging data mining and machine learning](#)," *bioRxiv* 09, 584222 (2024).
15. J. Liu, M. Osadchy, L. Ashton, M. Foster, C. J. Solomon, and S. J. Gibson, "[Deep convolutional neural networks for Raman spectrum recognition: a unified solution](#)," *Analyst* 142(21), 4067–4074 (2017).
16. M. Kazemzadeh, C. L. Hisey, K. Zargar-Shoshtari, W. Xu, and N. G. R. Broderick, "[Deep convolutional neural networks as a unified solution for Raman spectroscopy-based classification in biomedical applications](#)," *Optics Communications* 510, 127977 (2022).
17. N. Blake, R. Gaifulina, L. D. Griffin, I. M. Bell, and G. M. H. Thomas, "[Machine Learning of Raman Spectroscopy Data for Classifying Cancers: A Review of the Recent Literature](#)," *Diagnostics* 12(6), 1491 (2022).
18. Z. Zhang, A. Castelló, "[Principal components analysis in clinical studies](#)," *Annals of Translational Medicine* 5(17), 351–351 (2017).
19. L. A. Bratchenko, I. A. Bratchenko, "[Avoiding overestimation and the ‘black box’ problem in biofluids multivariate analysis by Raman spectroscopy: interpretation and transparency with the SP-LIME algorithm](#)," *Journal of Raman Spectroscopy* 56(4), 353–364 (2025).
20. Y. Liu, Y. Wu, J. Wang, J. Qi, C. Zhou, and Y. Xue, "[Recent Advances in Raman Spectral Classification with Machine Learning](#)," *Sensors* 26(1), 341 (2026).
21. R. S. Senger, D. Scherr, "[Resolving complex phenotypes with Raman spectroscopy and chemometrics](#)," *Current Opinion in Biotechnology* 66, 277–282 (2020).
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. M. Ando, H. Hamaguchi, "[Molecular component distribution imaging of living cells by multivariate curve resolution analysis of space-resolved Raman spectra](#)," *Journal of Biomedical Optics* 19(1), 011016 (2013).
24. O. Sagi, L. Rokach, "[Ensemble learning: A survey](#)," *WIREs Data Mining and Knowledge Discovery* 8(4), e1249 (2018).

25. M. A. Pramudito, Y. N. Fuadah, Y. S. Kim, and K. M. Lim, “[Stacking Ensemble Machine Learning for Cardiac Safety Assessment Using hiPSC-CM MEA Data](#),” *Annals of Biomedical Engineering* 54(5), 1334–1344 (2026).
26. A. Chaichi, A. Prasad, and M. R. Gartia, “[Raman Spectroscopy and Microscopy Applications in Cardiovascular Diseases: From Molecules to Organs](#),” *Biosensors* 8(4), 107 (2018).
27. S. Z. Al-Sammarraie, L. A. Bratchenko, E. N. Typikova, V. P. Zakharov, I. A. Bratchenko, and P. A. Lebedev, “[Silver nanoparticles-based substrate for blood serum analysis under 785 nm laser excitation](#),” *Journal of Biomedical Photonics & Engineering* 8(1), 010301 (2022).
28. 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).
29. I. A. Matveeva, A. I. Komlev, O. I. Kaganov, A. A. Moryatov, V. P. Zakharov, “[Multidimensional analysis of dermoscopic images and spectral information for the diagnosis of skin tumors](#),” *Journal of Biomedical Photonics & Engineering* 10(1), (2024).
30. 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).
31. H. Noothalapati, K. Iwasaki, and T. Yamamoto, “[Biological and Medical Applications of Multivariate Curve Resolution Assisted Raman Spectroscopy](#),” *Analytical Sciences* 33(1), 15–22 (2017).
32. T. G. Nick, K. M. Campbell, “[Logistic Regression](#),” *Topics in Biostatistics* 404, 273–301 (2007).
33. S. Domínguez-Almendros, N. Benítez-Parejo, and A. R. Gonzalez-Ramirez, “[Logistic regression models](#),” *Allergologia et Immunopathologia* 39(5), 295–305 (2011).
34. E. Y. Boateng, D. A. Abaye, “[A review of the logistic regression model with emphasis on medical research](#),” *Journal of Data Analysis and Information Processing* 07(04), 190–207 (2019).
35. L. Breiman, “[Random forests](#),” *Machine Learning* 45(1), 5–32 (2001).
36. B. Talekar, “[A detailed review on decision tree and random forest](#),” *Bioscience Biotechnology Research Communications* 13(14), 245–248 (2020).
37. H. A. Salman, A. Kalakech, and A. Steiti, “[Random forest algorithm overview](#),” *Babylonian Journal of Machine Learning* 2024, 69–79 (2024).
38. C. Cortes, V. Vapnik, “[Support-vector networks](#),” *Machine Learning* 20(3), 273–297 (1995).
39. J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, “[A comprehensive survey on support vector machine classification: Applications, challenges and trends](#),” *Neurocomputing* 408, 189–215 (2020).
40. S. Winters-Hilt, A. Yelundur, C. McChesney, and M. Landry, “[Support vector machine implementations for classification & clustering](#),” *BMC Bioinformatics* 7(S2), S4 (2006).
41. J. H. Friedman, “[Stochastic gradient boosting](#),” *Computational Statistics & Data Analysis* 38(4), 367–378 (2002).
42. A. Natekin, A. Knoll, “[Gradient boosting machines, a tutorial](#),” *Frontiers in Neurorobotics* 7, (2013).
43. H. Belyadi, A. Haghighat, “[Supervised learning](#),” *Machine Learning Guide for Oil and Gas Using Python* (Elsevier) 169–295 (2021).
44. X. Yue, H. Li, J. Tang, J. Liu, and J. Jiao, “[Rapid and label-free screening of echinococcosis serum profiles through surface-enhanced Raman spectroscopy](#),” *Analytical and Bioanalytical Chemistry* 412(2), 279–288 (2020).
45. J. Wang, D. Lin, J. Lin, Y. Yu, Z. Huang, Y. Chen, J. Lin, S. Feng, B. Li, N. Liu, and R. Chen, “[Label-free detection of serum proteins using surface-enhanced Raman spectroscopy for colorectal cancer screening](#),” *Journal of Biomedical Optics* 19(8), 087003 (2014).
46. T. K. Kim, “[T test as a parametric statistic](#),” *Korean Journal of Anesthesiology* 68(6), 540 (2015).
47. F. Wilcoxon, “[Individual comparisons by ranking methods](#),” *Biometrics Bulletin* 1(6), 80 (1945).
48. E. Sorokina, Y. Khristoforova, I. Bratchenko, M. Skuratova, P. Lebedev, and V. Zakharov, “[Comparison of Machine Learning Methods for the Analysis of Serum Raman Spectra in the Detection of Chronic Heart Failure](#),” *Journal of Biomedical Photonics & Engineering* 10(4), 040319 (2024).
49. Y. Gao, X. Bai, J. Lu, L. Zhang, X. Yan, X. Huang, H. Dai, Y. Wang, L. Hou, S. Wang, A. Tian, and J. Li, “[Prognostic value of multiple circulating biomarkers for 2-year death in acute heart failure with preserved ejection fraction](#),” *Frontiers in Cardiovascular Medicine* 8, 779282 (2021).
50. M. De Bakker, T. B. Petersen, A. J. Rueten-Budde, K. M. Akkerhuis, V. A. Umans, J. J. Brugts, T. Germans, M. J. T. Reinders, P. D. Katsikis, P. J. Van Der Spek, R. Ostroff, R. She, D. Lanfear, F. W. Asselbergs, E. Boersma, D. Rizopoulos, and I. Kardys, “[Machine learning-based biomarker profile derived from 4210 serially measured proteins predicts clinical outcome of patients with heart failure](#),” *European Heart Journal - Digital Health* 4(6), 444–454 (2023).
51. R. J. Woolley, D. Ceelen, W. Ouwerkerk, J. Tromp, S. M. Figarska, S. D. Anker, K. Dickstein, G. Filippatos, F. Zannad, M. Metra, L. Ng, N. Samani, D. J. Van Veldhuisen, C. Lang, C. S. Lam, and A. A. Voors, “[Machine learning based on biomarker profiles identifies distinct subgroups of heart failure with preserved ejection fraction](#),” *European Journal of Heart Failure* 23(6), 983–991 (2021).