

Machine Learning Classification of Augmented Raman Serum Spectra for Chronic Heart Failure Detection

Yulia A. Khristoforova*, Ekaterina O. Salnikova, and Irina A. Matveeva

Samara National Research University, 34 Moskovskoe shosse, Samara 443080, Russian Federation

*e-mail: khristoforova.yua@ssau.ru

Abstract. In this work, we propose a promising non-invasive approach based on the Raman spectroscopy of human serum for chronic heart failure (CHF) diagnosis. Due to the limited sample size, which limits the performance of machine learning classifiers, this study explores data augmentation techniques to improve the classification of low- and high-grade CHF spectra using k-nearest neighbor (kNN) and partial least squares discriminant analysis (PLS-DA) algorithms. Raman spectra of 151 patients with CHF of the different stages were acquired at 532 nm excitation from serum samples collected at Samara City Clinical Hospital. Two augmentation approaches were systematically evaluated: (1) combined approach based on the linear spectral transformations (wavenumber shifting $\pm 1-2 \text{ cm}^{-1}$, intensity stretching 0.9-1.1) and (2) Wasserstein Generative Adversarial Network (WGAN)-based synthetic spectrum generation, expanding training datasets 10-fold while preserving physiochemical realism. Augmentation based on linear spectral transformations yielded algorithm-specific results: kNN showed no significant receiver operating characteristic area under curve (ROC AUC) improvement (0.67 ± 0.11 original vs. $0.67-0.69$ augmented), while PLS-DA achieved statistically significant gains (0.71 ± 0.11 vs. $0.80-0.81$; $z = 2.3-2.6$, $p < 0.05$). WGAN augmentation proved superior across both methods, with k-NN reaching 0.74 ± 0.08 and PLS-DA achieving 0.83 ± 0.09 . These findings establish WGAN as an optimal augmentation strategy for Raman-based CHF classification, achieving clinically relevant performance (ROC AUC > 0.80) from limited cohorts while enabling biomarker identification for cardiovascular diagnostics.

Keywords: Raman spectroscopy; chronic heart failure; data augmentation; kNN; PLS-DA; binary classification; WGAN.

Paper #9555 received 18 Mar 2026; revised manuscript received 20 Apr 2026; accepted for publication 21 Apr 2026; published online 7 May 2026. [doi: 10.18287/JBPE26.12.020304](https://doi.org/10.18287/JBPE26.12.020304).

1 Introduction

Cardiovascular diseases (CVDs) are the leading cause of death globally, claiming 17.9 million lives each year [1]. Chronic heart failure (CHF) affects over 64 million people worldwide and is a key component of the CVD burden, which is projected to increase by 90% from 2025 to 2050 due to population aging [1]. In Russia, the prevalence of CHF in European regions is about 7-10% but these cases are often underdiagnosed due to their subtle symptoms [2]. The number of severe CHF cases

(HIIIB, HIIII) has grown more sharply, comprising about 38% of all CHF cases and contributes to higher hospitalization and mortality rates.

Early and accurate diagnosis of CHF is crucial, as it allows for timely intervention to improve patient outcomes and reduce hospitalizations. Current diagnostic methods for CHF, such as NT-proBNP/BNP biomarker assays, echocardiography, and clinical scoring systems (e.g., HFA-PEFF) [3], face significant limitations, including reducing accuracy for comorbidities, a lack of standardization across tests, and reliance on specialized

equipment, which limits accessibility primary care. The diagnostic value of brain natriuretic peptide (BNP)/N-terminal part of its prohormone (NT-proBNP) is valuable [3, 4], but it is limited by various confounding factors, such as levels rising in conditions other than CHF, such as chronic kidney disease, atrial fibrillation, pericardial disease, pulmonary embolism, and aging [3, 4]. Obesity can also suppress their levels, making individualized interpretation necessary in the presence of comorbidities. These limitations are particularly significant for early detection of diseases in patients who may have asymptomatic or mild symptoms [3, 5]. NT-proBNP levels in these patients may remain normal or only slightly elevated, leading to missed diagnoses and delayed intervention. Therefore, the development of new promising diagnostic techniques is essential to overcome these challenges.

Non-invasive techniques such as Raman spectroscopy are becoming increasingly important for detecting chemical changes in the biofluids of patients with different diseases [6–9]. These techniques offer the potential for label-free analysis of biofluids and tissue, which is not possible with current diagnostic methods that rely on invasive biomarkers. This highlights the need for more advanced spectral techniques, especially in resource-limited settings.

Raman spectroscopy is an optical, non-destructive analytical method based on inelastic scattering of photons by molecular bond vibrations and rotations. It produces a unique “chemical fingerprint” spectrum where each peak corresponds to specific vibrational or rotational modes of molecules within the analyte, enabling detailed characterization of complex biological samples like tissues [7] or fluids [8, 10, 11].

In biological samples, the Raman spectrum provides a composite contribution from biochemical components, including proteins, lipids, nucleic acids, and metabolites [12]. These components can be used to identify alterations in molecular composition associated with the progression of the disease [6]. Integration with machine learning algorithms [8, 11, 13] can help extract hidden spectral similarities and differences between different groups, for example between diseased and healthy groups or between different stages of diseases. This can be done by identifying subtle peak shifts and intensity variations that may be overlooked by traditional analysis methods.

Developing new diagnostic approaches requires statistically reliable data in large volumes. However, collecting real biomedical spectral data is time-consuming and resource-intensive, often taking years to acquire samples and perform spectral measurements under controlled conditions. As a result, many studies rely on small sample sizes, typically dozens [14, 15] rather than thousands. Small sample sizes can lead to overfitting, poor generalization, and inaccurate metric estimates, especially for high-dimensional spectral data [16]. In other cases, researchers acquire numerous spectra from a single sample [14], limiting generalizability to diverse patient populations.

Another approach is applying augmentation methods [17–20] to real data for generating synthetic data, which simulates variability in spectral profiles while preserving biochemical features, thereby enhancing model robustness without extended collection efforts. Common transformations for spectral data augmentation include frequency shifting, intensity stretching, noise injection, and generative models [17–19]. Data augmentation can help avoid overfitting on a limited number of real-world samples, allowing training on larger datasets that can produce more stable and reliable models [17–20]. For example, Zhao et al. [18] demonstrated skin cancer detection using Raman spectroscopy, where data augmentation strategies – including random noise addition, spectral shifting, linear combinations, and 1D-GAN-generated spectra – increased training data from 512 to 14608 spectra. The classification performance of PLS-DA, PC-LDA, and SVM was essentially independent of the number of 1D-GAN synthesized spectra, whereas 1D-CNN improved until reaching a plateau at around $n = 10000$ and LR showed no meaningful benefit even at the maximum augmentation size ($n=30000$). It should be noted, analysis of the extended dataset not only can elevate performance across convolutional neural network (CNN) and conventional machine learning algorithms (PLS-DA, SVM) by 2–4% but also enhanced robustness on noisy or shifted test spectra. Therefore, the results achieved in studies on clinical Raman applications emphasize the critical role of augmentation in stabilizing machine learning models.

In this study, we aim to investigate the impact of data augmentation methods on the performance of machine learning models trained on Raman spectroscopy data. We apply these models to serum spectra from patients with CHF, specifically to distinguish between initial clinical phase (HIIA) and advanced phases (HIIIB, HIIIC) of the disease in accordance with Strazhesko-Vasilenko classification. The use of augmentation techniques is necessary due to the limited size of our spectral dataset, which is a common challenge in pilot biomedical studies that involve Raman spectroscopy of biofluids.

2 Materials and Methods

2.1 Experimental Data

Study protocols were approved by the Ethics Committee of Samara State Medical University (protocol #268, September 11, 2023). All involved subjects agreed to participate in this study by informed consent. Human serum were taken from 151 patients with initial clinical phase (HIIA) and advanced phases (HIIIB, HIIIC) of CHF at the Cardiology Department of Samara City Clinical Hospital No. 1 (N.I. Pirogov). CHF diagnoses were confirmed via NT-proBNP testing results. Blood serum sampling was performed at the early stage of hospitalization, before treatment in the surgical department had been initiated. Detailed information on medications, comorbidities, and diet was not systematically recorded or balanced between groups.

Therefore, we acknowledge that differences in these confounding factors may influence the observed spectral patterns and should be considered a limitation of our work.

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 $10\ \mu\text{l}$ was placed on aluminum foil and allowed to air dry at room temperature until it was completely dry. The total set of experimental data includes 151 spectra, consisting of 69 spectra of patients with initial CHF stages and 82 with advanced stages.

2.2 Spectroscopic Measurements

Raman measurements were performed using an EnSpectr R532 ADF U300 spectrometer equipped with a 532 nm excitation laser and a $50\times$ LMPlan objective. Spectra were recorded in the $89\text{--}3892\ \text{cm}^{-1}$ range with a spectral resolution of $1.5\ \text{cm}^{-1}$. The diameter of the laser spot in focus was 5 microns. Human blood serum was analyzed using a laser power of 27 mW. Before recording the spectral characteristics of the blood serum sample, a preliminary recording of the ambient background signal was performed. After that, the background component was automatically subtracted from the registered serum spectrum using the built-in software algorithm. The Raman spectra were registered with 2 s exposure time and 10 accumulations to improve the signal-to-noise ratio. Each of the resulting spectra is a discrete set of 3502 parameters.

2.3 Augmentation Procedure

In this work, we used several approaches for augmentation of Raman spectra: (1) combined augmentation approach, incorporating wavenumber axis shifting and intensity axis stretching, and (2) synthetic spectrum generation via Wasserstein generative adversarial networks (WGAN) method [22].

A combined approach was used to generate two augmented sets of Raman spectra. For the first set, the original data was shifted by $\pm 1\ \text{cm}^{-1}$ and stretched by 0.9 arbitrary units (a.u.) in intensity. For the second set, the data was shifted by $\pm 2\ \text{cm}^{-1}$ and stretched by 1.1 a.u. in intensity.

The use of WGAN makes obtaining Raman data much more complicated than the combined approach based on simple linear transformations. WGAN generates synthetic Raman spectra using a minimax framework with (Wasserstein-1) distance. This augmentation was performed separately on the low and high stages CHF class subsets to preserve class-specific discriminative features. The generator takes random noise and label and produces artificial spectra from random noise, while the discriminator estimates the distance between real and synthetic spectral distributions. If the discriminator is able to successfully distinguish

between synthetic and real spectra, it will signal the generator to adjust its parameters in order to improve the realism of the generated spectra and minimize the loss function. This process of iterative refinement will continue until the synthetic spectra are indistinguishable from the real spectra, at which point, the discriminator will no longer be able to differentiate between the two [22]. Gradient penalty enforces the 1-Lipschitz constraint, ensuring training stability and realistic spectral peak morphology preservation. This approach effectively augments limited Raman datasets while maintaining biochemical feature integrity for downstream classification tasks.

The WGAN was implemented with a fully connected critic-generator structure applied separately to each class. For the first class (69 spectra) and the second class (82 spectra), both of length 744, the generator takes a 690-dimensional latent vector sampled from a standard normal distribution and maps it through two dense layers (512 and 1024 units) with ReLU activation, a dropout layer (rate 0.2), and a final dense layer (744 units) with tanh activation to produce synthetic spectra in the range $[-1, 1]$. The critic processes a spectrum of length 744 through three dense layers (1024, 512, 256 units) with ReLU activation, a dropout layer (rate 0.2), and a final scalar output. The model was trained with RMSprop (learning rate 0.001), updating the critic 5 times per generator step, after z-score normalization of each spectrum.

2.4 Analyzing and Preprocessing

The spectral range between 800 and $1800\ \text{cm}^{-1}$ was selected for further augmentation and analysis. This corresponds to a set of 744 discrete parameters, representing the Raman intensity at specific wavenumbers. The experimental dataset was randomly divided into training and test sets: 80% of the spectra were allocated to the training set, while the remaining 20% formed the test set. The test set was held out from all augmentation procedures and was not used at any stage of synthetic data generation. Augmentation using three approaches was applied exclusively to the training set, which initially comprised 120 spectra and was expanded to 1320 spectra. In this way, two independent models were constructed: one trained on the original real spectra only, and the second trained on the original spectra together with the augmented spectra. Both models were subsequently evaluated on the same held-out test set consisting exclusively of real spectra. This design allowed for a direct comparison of model performance under identical testing conditions and ensured that the test results reflected the effect of augmentation on generalization to unseen real data rather than any influence from synthetic samples.

Afterward, Raman spectra of both extended training and test sets were preprocessed by smoothing using the Savitzky-Golay filter (15 window was selected as the standard choice for spectroscopic applications, with a polynomial order of $p = 3$ to balance peak intensity

preservation against residual noise) and normalization to standard normal variation.

To classify Raman spectra of low and high CHF classes, k NN and PLS-DA binary model were built.

The k NN algorithm classifies Raman spectra by computing Euclidean distances in spectral feature space and assigning the class label of the majority among the k closest training spectra. In our model, optimal $k=6$ neighbors were determined through minimization of root mean square error (RMSE), balancing bias-variance trade-off while preserving spectral discrimination between serum samples of CHF patients with different stages patients with different stages.

PLS-DA constructs latent variables (LV) through partial least squares regression between spectral predictors and class indicators, maximizing covariance to achieve optimal class separation.

Additionally, to assess the validity of the WGAN-generated spectra, PLS-DA models were constructed and evaluated on three separate datasets: (1) the original real spectra ($n=120$), (2) the synthetically generated spectra alone ($n=1200$), and (3) the combined dataset of real and synthetic spectra ($n=1320$). The PLS-DA model trained on the original real spectra served as a reference to characterize the inherent discriminative ability of the experimental data. The model built on the generated spectra alone was used to evaluate whether the synthetic data preserved the same spectral patterns and class structure as the real data. Finally, the PLS-DA model fitted on the merged dataset (real and synthetic spectra) allowed us to investigate how the addition of augmented samples affected classification performance, stability, and latent-space structure.

ROC-AUC served as the primary metric for evaluating classification performance improvement, with its standard deviation assessing model stability. Comparative analysis was conducted between k NN and PLS-DA classification results obtained from both original and augmented datasets, employing the z -statistic for significance testing. The γ correlation coefficient between ROC curves of the compared datasets was calculated [23]. Following z -statistic computation, the null hypothesis H_0 (no significant stability improvement) was either accepted or rejected. A significance level of $\alpha=0.05$ was adopted, corresponding to a critical value $z_{crit}=1.96$; thus, H_0 was rejected when $|z|\geq 1.96$ and accepted when $|z|<1.96$.

3 Results and Discussion

Figure 1 shows the 532 nm stimulated serum Raman spectra of patients with low and high CHF stages and differenced spectrum. Biological samples, particularly human serum, are known to contain carotenoids that produce characteristic resonance Raman signals under 532 nm laser excitation. These carotenoids generate prominent Raman bands at 1150 cm^{-1} ($\nu(\text{C}-\text{C})$) and 1515 cm^{-1} ($\nu(\text{C}=\text{C})$) [24]. Additionally, both spectral profiles exhibit a distinct peak at 1004 cm^{-1} , attributable to carotenoid $\text{C}-\text{CH}_3$ rocking vibrations, which may overlap with the phenylalanine ring mode. Other

prominent peaks also are observed at 1445 cm^{-1} (CH_2/CH_3 scissoring & bending in proteins and lipids) and 1665 cm^{-1} (Amide I ($\text{C}=\text{O}$ stretching) in proteins). The differenced spectrum has prominent features at 1004 , 1150 , 1515 cm^{-1} that corresponds to the typical Raman bands in serum. Blood sampling at the early stage of hospitalization, before treatment reduces the likelihood that the spectra were influenced by in-hospital therapy. Nevertheless, information on medications taken prior to admission was not available, and the potential effects of comorbidities, diet, and other confounders.

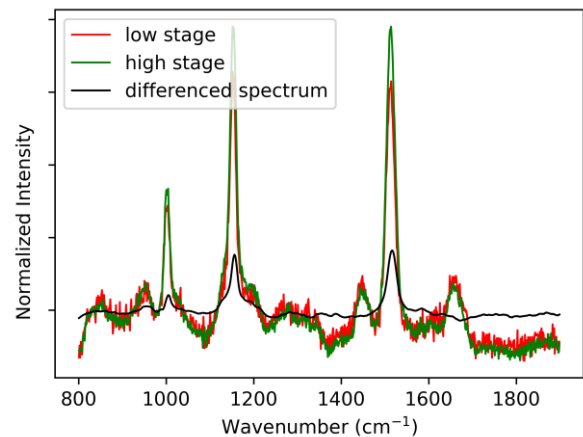


Fig. 1 Serum Raman spectra of patients with low and high CHF stages.

The original dataset was extended 10-fold, resulting in three sets of 1320 spectra including original and augmented spectra. Two sets were generated using a combined approach with different coefficients for linear transformations, and one set was generated using WGAN. Figure 2 presents original and augmented Raman spectra.

Fig. 2(a) shows an example of spectrum transformation through wavenumber shifting and stretching. Wavenumber axis shifting ($\pm 1-2\text{ cm}^{-1}$) simulates instrumental variations inherent to Raman spectroscopy, including laser wavelength drift, spectrometer calibration inaccuracies, and sample positioning effects. These perturbations represent realistic experimental uncertainty encountered across multiple measurements and instruments. The intensity axis stretching ($S_y=0.9-1.1$) takes into account physiological and measurement variability, especially biological factors. This includes fluctuations in serum analyte concentrations, variations in protein binding, and differences in carotenoid content between patients. Such transformations preserve intrinsic spectral shape and peak position relationships while introducing statistically plausible perturbations observed in real-world serum Raman datasets.

k NN classification showed that augmentation using combined approach through linear spectral transformations (wavenumber shifting and intensity stretching) did not produce statistically significant

improvements in ROC AUC compared to the original dataset. This lack of significant improvement persisted across both augmented datasets created with different linear transformation coefficients. The ROC AUC values for CHF stage classification using *k*NN were 0.67 ± 0.11 (original dataset), 0.69 ± 0.10 (augmented set with $\pm 1 \text{ cm}^{-1}$ shift and 0.9 stretch), and 0.67 ± 0.09 (augmented set with $\pm 2 \text{ cm}^{-1}$ shift and 1.1 stretch). Nevertheless, model performance was stable across datasets that differed by an order of magnitude in size (10-fold expansion), without any observed degradation in classification ability.

In contrast to *k*NN results, PLS-DA classification demonstrated statistically significant ROC AUC improvement following spectral augmentation using

linear transformations, according to *z*-value analysis. Specifically, the original dataset yielded an AUC of 0.71 ± 0.11 , while the two augmented datasets produced AUC values of 0.80 ± 0.10 and 0.81 ± 0.09 , respectively.

We can assume the observed difference between *k*NN and PLS-DA can be explained by their intrinsic working principles. *k*NN is a distance-based classifier that relies on local neighborhood structure in the raw feature space; linear spectral transformations (wavenumber shifts and intensity stretches) produce smooth, controlled perturbations that preserve the global topology and neighbor relationships, so the model shows only minor changes in ROC-AUC without statistically significant improvement.

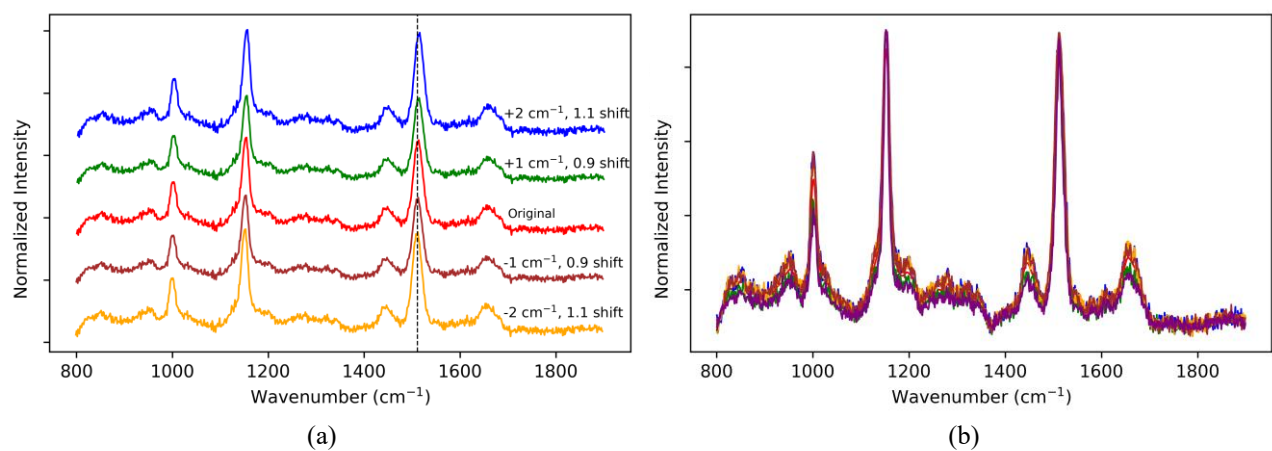


Fig. 2 Original and augmented serum Raman spectra using different approaches: (a) combined method based on linear transformations, (b) WGAN synthesized spectra.

Table 1 Comparison of the performance of classification models on the initial and augmented datasets (SD – standart deviation).

Classification method	Analyzed data	ROC AUC ± SD	z-value	Improving
<i>k</i> NN	Original set	0.67 ± 0.11	–	–
	Combined approach ($\pm 1 \text{ cm}^{-1}$ shift, 0.9 stretch)	0.69 ± 0.10	<1.96	No
	Combined approach ($\pm 2 \text{ cm}^{-1}$ shift, 1.1 stretch)	0.67 ± 0.09	<1.96	No
	WGAN	0.74 ± 0.08	2.38	Yes
PLS-DA	Original set	0.71 ± 0.11	–	–
	Combined approach ($\pm 1 \text{ cm}^{-1}$ shift, 0.9 stretch)	0.80 ± 0.10	2.30	Yes
	Combined approach ($\pm 2 \text{ cm}^{-1}$ shift, 1.1 stretch)	0.81 ± 0.09	2.56	Yes
	WGAN	0.83 ± 0.09	2.95	Yes

In contrast, PLS-DA is a covariance-based latent-variable method that explicitly seeks directions of maximum covariance between spectra and class labels; the enlarged, augmented training set provides richer sampling of the spectral manifold, enabling more robust latent-variable estimation and clearer class separation, which leads to a statistically significant gain in ROC-AUC.

Fig. 2(b) shows an example of serum Raman spectra generated using WGAN. Application of WGAN-generated synthetic spectra further enhanced classification performance across both algorithms: k NN and PLS-DA. For k NN, the augmented WGAN dataset yielded ROC AUC = 0.74 ± 0.08 , representing a significant improvement over the initial 0.67 ± 0.11 . This statistically significant enhancement ($z \approx 2.38$, $p < 0.05$) demonstrates WGAN's capacity to generate realistic spectral perturbations beyond simple linear transformations.

The performance of PLS-DA has been further improved, achieving an AUC of 0.83 ± 0.09 on the WGAN-expanded dataset. This represents a 17% relative improvement over the baseline (0.71 ± 0.11) and a 4% absolute gain over augmentation based on linear transformations. A z -test comparison has confirmed the high statistical significance of the difference between the results of PLS-DA with and without WGAN augmentation ($z \approx 2.6$, $p < 0.01$).

To evaluate the representativeness of the WGAN-generated spectra, PLS-DA models were built and assessed on three different datasets: (1) the original real spectra only, (2) the synthetic spectra generated by the WGAN, and (3) the combined dataset of real and synthetic spectra. The PLS-DA model trained on the original real spectra achieved an accuracy of 0.71 ± 0.11 , reflecting the baseline discriminative capability of the measured experimental data. The model fitted on the synthetic spectra alone reached 0.85 ± 0.08 , indicating that the generated spectra preserved sufficient class-separating information and spectral patterns to support a stable classification, even in the absence of real data. Finally, the PLS-DA model trained on the merged dataset (real and synthetic spectra) yielded an accuracy of 0.83 ± 0.09 , which is slightly higher than that of the real-only model while maintaining a comparable level of variability. These results suggest that the WGAN-generated spectra are consistent with the real data in terms of class structure and underlying spectral features.

WGAN augmentation proved effective due to its ability to synthesize physiochemically plausible Raman spectra that preserve peak morphology, baseline characteristics, and signal-to-noise relationships characteristic of real serum measurements. Reduced standard deviation across both techniques indicates improved model stability and resistance to overfitting, with dataset expansion directly contributing to variance reduction (standard deviation (SD) decreased from 0.11 to 0.08–0.09). This improved performance is crucial for clinical translation, where patient cohorts display

significant biological diversity, as larger training sets better capture the full range of physiological and experimental variations inherent in real-world Raman serum measurements.

These findings support the use of WGAN as an optimal augmentation strategy for Raman spectroscopic classification of CHF. This approach allows us to minimize the link between the limited number of experimental datasets available and the statistical power needed for reliable diagnosis, while preserving accuracy, which is crucial for discovering important spectral trends.

Additionally, it is interesting to observe the different levels of differentiation among the k NN and PLS-DA methods. The comparison between k NN and PLS-DA reveals that these algorithms have different sensitivities to the size of the dataset. PLS-DA shows a pronounced dependence on the dataset size, achieving better generalization on the 10-fold expanded dataset. This enhanced performance is likely due to PLS-DA's ability to exploit increased spectral variance from linear transformations, which simulates biological sample variability observed in real-world serum measurements. These transformations effectively capture physiological fluctuations and experimental inconsistencies, providing a more representative training distribution for PLS-DA. Covariance-based methods, such as PLS-DA, require robust latent variable construction, and this augmentation provides them with a more representative distribution for training, leading to better performance.

Results from the augmented datasets (10-fold expansion to $n = 1200$ training spectra) demonstrate reduced SD in ROC AUC compared to the original dataset ($n = 120$), indicating lower variance and enhanced model stability across both k NN and PLS-DA classifiers. For k NN, SD decreased from 0.11 (original) to 0.08 (WGAN-augmented), a 27% variance reduction; PLS-DA showed similar improvement from 0.11 to 0.09 (17% relative gain). This narrower spread reflects more consistent performance across cross-validation folds, minimizing overfitting risks inherent to small biomedical cohorts.

Notably, neither k NN nor PLS-DA exhibited performance degradation post-augmentation, maintaining or improving ROC-AUC despite 10-fold dataset expansion. Nevertheless, this stability may however stem from the limited variability introduced by the employed augmentation methods (linear transformations and WGAN), which simulate controlled spectral perturbations rather than the full spectrum of real-world biological heterogeneity. Actual serum samples often display unpredictable physiological variations – such as patient-specific protein profiles, medication effects, comorbidities (e.g., renal dysfunction, diabetes), diurnal hormone fluctuations, and dietary influences on carotenoid/lipid content – that exceed the controlled perturbations generated by current augmentation methods (linear transformations, WGAN).

Summary, results from the current cohort provisionally confirm the feasibility of using Raman spectroscopy to differentiate CHF stages, establishing

proof-of-concept for future clinical validation. Raman's molecular sensitivity enables detection of systemic metabolic shifts during CHF noninvasively. Advanced CHF stages (HIIB–III) were accompanied by significant elevation of N-terminal pro-B-type natriuretic peptide (NT-proBNP) and reduced glomerular filtration rate, confirming disease severity. However, these biochemical differences may reflect both direct CHF effects and comorbid conditions more prevalent in severe cases – for instance, type 2 diabetes mellitus occurred in 21% of early-stage (HIIA) vs. 53% of advanced-stage (HIIB–III) patients. Raman spectroscopy captures these composite metabolic shifts noninvasively, demonstrating feasibility for CHF staging. In accordance with our spectral data progressive CHF stages can characterize by reduced carotenoid content (1004, 1150, 1515 cm^{-1}) that can reflect antioxidant depletion and increased protein/lipid ratios (1445, 1665 cm^{-1}) indicating inflammation, validated by PLS-DA (ROC AUC = 0.83). Also, augmentation results of this work yielded positive outcomes, confirming classification performance's dependence on effective sample size, validating augmentation as a viable strategy when real sample accrual is constrained. Ideally, model robustness requires continuous expansion with authentic biological specimens to capture full clinical heterogeneity. However, given persistent resource limitations in pilot biomedical spectroscopy studies, continued augmentation research remains essential – exploring advanced techniques like variational autoencoders (VAE), diffusion models, and physics-informed GANs. Future validation should include a variety of classification tasks, such as multi-diagnosis cohorts (for example, CHF with comorbidities) and sub-class analyses (HIIA subtypes). Additionally, it is important to cross-validate the model against independent serum collections in order to ensure its generalization against complex physiological variability found in real-world clinical settings.

4 Conclusions

This study focuses early CHF detection in high-risk patients by differentiating early-stage (HIIA) from progressive stages (HIIB–III) using serum Raman analysis to identify initial biochemical changes, positioning it as a promising screening tool. We evaluated data augmentation strategies for enhancing Raman spectroscopic classification of CHF using limited human serum samples. Linear spectral transformations (wavenumber shifting and intensity stretching)

demonstrated algorithm-specific efficacy, providing statistically significant ROC AUC improvements for PLS-DA ($0.71 \rightarrow 0.80\text{--}0.81$, $p < 0.05$) but not k NN, highlighting fundamental differences in distance-based versus covariance-based classification sensitivities.

WGAN-based augmentation emerged as the optimal methodology, achieving higher performance across both algorithms: k NN improved from 0.67 ± 0.11 to 0.74 ± 0.08 (27% variance reduction), while PLS-DA reached 0.83 ± 0.09 (17% relative gain). The consistent standard error reduction not only underscores enhanced model stability critical for heterogeneous clinical cohorts but also demonstrates that increased case numbers through synthetic data expansion directly contribute to lower variance, enabling more reliable performance as patient cohort sizes grow.

Key findings demonstrate that WGAN effectively simulates physiochemically realistic spectral variance – encompassing biological heterogeneity, carotenoid content fluctuations, and instrumental effects – beyond the capabilities of linear methods. PLS-DA exhibited pronounced dataset size dependence, leveraging expanded feature spaces for optimal latent variable construction. These results establish WGAN augmentation as a good tool for Raman-based diagnostics, particularly for CHF, enabling higher CHF staging (ROC AUC > 0.80) from small patient cohorts.

Future work can explore an expanded class of augmentation methods, including variational autoencoders (VAE), diffusion models, and physics-informed GAN architectures, to further enhance spectral realism. Systematic investigation of classifier performance dependencies on varying dataset sizes ($n = 50\text{--}5000$), augmentation ratios, and sample heterogeneity will optimize augmentation protocols for clinical deployment. Hybrid augmentation strategies combining WGAN with machine learning classification approach and clinical metadata (NT-proBNP levels, ejection fraction) may yield personalized diagnostic models with enhanced prognostic capability.

Funding

This research was funded by Russian Science Foundation, grant number 25-75-00146; <https://rscf.ru/project/25-75-00146/>.

Disclosures

The authors declare that they have no conflicts of interest or competing financial interests related to this work.

References

1. Cardiovascular diseases (CVDs), WHO, 31 July 2025 (accessed 2 March 2026) [[https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))].
2. A. S. Galyavich, S. V. Nedogoda, G. P. Arutyunov, Yu. N. Belenkov, “About the classification of heart failure,” Russian Journal of Cardiology 28(9), 5584 (2023). [in Russian]
3. B. Bozkurt, A. J. S. Coats, H. Tsutsui, et al., “Universal definition and classification of heart failure: a report of the Heart Failure Society of America, Heart Failure Association of the European Society of Cardiology, Japanese Heart

- Failure Society and Writing Committee of the Universal Definition of Heart Failure: Endorsed by the Canadian Heart Failure Society, Heart Failure Association of India, Cardiac Society of Australia and New Zealand, and Chinese Heart Failure Association,” *European Journal of Heart Failure* 23(3), 352–380 (2021).
4. S. A. Hill, R. A. Booth, P. L. Santaguida, A. Don-Wauchope, J. A. Brown, M. Oremus, U. Ali, A. Bustamam, N. Sohel, R. McKelvie, C. Balion, and P. Raina, “Use of BNP and NT-proBNP for the diagnosis of heart failure in the emergency department: a systematic review of the evidence,” *Heart Failure Reviews* 19(4), 421–438 (2014).
 5. T. A. McDonagh, M. Metra, M. Adamo, et al., “2023 Focused Update of the 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: Developed by the task force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) With the special contribution of the Heart Failure Association (HFA) of the ESC,” *European Heart Journal* 44(37), 3627–3639 (2023).
 6. 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).
 7. L. A. Bratchenko, Y. A. Khristoforova, Y. G. Loginova, A. A. Moryatov, V. P. Zakharov, O. I. Kaganov, and I. A. Bratchenko, “Raman *in Vivo* Analysis of Melanoma Invasion Level,” *Journal of Biomedical Photonics & Engineering* 11(3), 030303 (2025).
 8. L. A. Bratchenko, Y. A. Khristoforova, I. A. Pimenova, E. N. Tupikova, M. A. Skuratova, G. A. Dvoynikov-Sechnoy, S. Wang, P. A. Lebedev, and I. A. Bratchenko, “SERS-based technique for accessible and rapid diagnosis of multiple myeloma in blood serum analysis,” *Light: Advanced Manufacturing*, 6(2), 284–294 (2025).
 9. H. Yilmaz, A. Ramoji, A. Winterfeld, H. Salehi, A. Ozkul, and J. Popp, “Applications of Raman Spectroscopy in Pandemic Virology: A Comprehensive Review,” *ACS Photonics* 13(6), 1568–1590 (2026).
 10. D. W. Shipp, F. Sinjab, and I. Notingher, “Raman spectroscopy: techniques and applications in the life sciences,” *Advances in Optics and Photonics* 9(2), 315–428 (2017).
 11. L. A. Bratchenko, Y. A. Khristoforova, I. A. Pimenova, M. S. Snegerev, V. I. Kupaev, P. A. Lebedev, Y. V. Kistenev, and I. A. Bratchenko, “Comparative Study Into the Effect of Detector Noises and Sensitivity on the Serum SERS Analysis: Example of Non-Communicable Diseases Discrimination,” *Journal of Biophotonics* 18(4), e202400475 (2025).
 12. J. De Gelder, K. De Gussem, P. Vandenabeele, and L. Moens, “Reference database of Raman spectra of biological molecules,” *Journal of Raman Spectroscopy* 38(9), 1133–1147 (2007).
 13. N. M. Ralbovsky, I. K. Lednev, “Towards development of a novel universal medical diagnostic method: Raman spectroscopy and machine learning,” *Chemical Society Reviews* 49(20), 7428–7453 (2020).
 14. S. Jiang, X. Sha, S. Qu, Z. Jiang, Y. Shen, Q. Wang, M. Kang, and G. Fang, “Label-Free SERS spectroscopy of Serum: Early screening and therapeutic evaluation for ovarian tumor patients,” *Microchemical Journal* 215, 114180 (2025).
 15. S. Ahmad, M. I. Majeed, H. Nawaz, M. R. Javed, N. Rashid, M. Abubakar, F. Batool, S. Bashir, M. Kashif, S. Ali, M. Tahira, S. Tabbasum, and I. Amin, “Characterization and prediction of viral loads of Hepatitis B serum samples by using surface-enhanced Raman spectroscopy (SERS),” *Photodiagnosis and Photodynamic Therapy* 35, 102386 (2021).
 16. C. Beleites, U. Neugebauer, T. Bocklitz, C. Krafft, and J. Popp, “Sample size planning for classification models,” *Analytica Chimica Acta* 760, 25–33 (2013).
 17. X. Zhang, H. Li, X. Tian, C. Chen, Y. Su, M. Li, J. Lv, C. Chen, and X. Lv, “Application of spectral small-sample data combined with a method of spectral data augmentation fusion (SDA-Fusion) in cancer diagnosis,” *Chemometrics and Intelligent Laboratory Systems* 231, 104681 (2022).
 18. J. Zhao, H. Lui, S. Kalia, T. K. Lee, and H. Zeng, “Improving skin cancer detection by Raman spectroscopy using convolutional neural networks and data augmentation,” *Frontiers in Oncology* 14, 1320220 (2024).
 19. M. Wu, S. Wang, S. Pan, A. C. Terentis, J. Strasswimmer, and X. Zhu, “Deep learning data augmentation for Raman spectroscopy cancer tissue classification,” *Scientific Reports* 11(1), 23842 (2021).
 20. S. Laitrakun, S. Deepaisarn, S. Gulyanon, C. Srisumarnk, N. Chiewnawintawat, A. Angkoonsawaengsuk, and N. Jaikaew, “Toward Practical Augmentation of Raman Spectra for Deep Learning Classification of Contamination in HDD,” *Journal of Information and Communication Convergence Engineering* 21(3), 208–215 (2023).
 21. D. Ma, L. Shang, J. Tang, Y. Bao, J. Fu, and J. Yin, “Classifying breast cancer tissue by Raman spectroscopy with one-dimensional convolutional neural network,” *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 256, 119732 (2021).
 22. M. Patil, M. M. Patil, and S. Agrawal, “WGAN for Data Augmentation,” in *GANs for Data Augmentation in Healthcare*, A. Solanki, M. Naved (Eds.), Springer International Publishing, 223–241 (2023).
 23. 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).
 24. J. Udensi, J. Loughman, E. Loskutova, and H. J. Byrne, “Raman Spectroscopy of Carotenoid Compounds for Clinical Applications - A Review,” *Molecules* 27(24), 9017 (2022).