

University of Nebraska - Lincoln

## DigitalCommons@University of Nebraska - Lincoln

---

Faculty Publications from the Department of  
Electrical and Computer Engineering

Electrical & Computer Engineering, Department  
of

---

9-18-2022

### SERS spectroscopy with machine learning to analyze human plasma derived sEVs for coronary artery disease diagnosis and prognosis

Xi Huang

*University of Nebraska-Lincoln*, [huangxi.1989.u.s@gmail.com](mailto:huangxi.1989.u.s@gmail.com)

Bo Liu

*University of Nebraska Medical Center*

Shenghan Guo

*The State University of New Jersey, Arizona State University*

Weihong Guo

*The State University of New Jersey*

Ke Liao

*University of Nebraska Medical Center*

*See next page for additional authors*

Follow this and additional works at: <https://digitalcommons.unl.edu/electricalengineeringfacpub>



Part of the [Computer Engineering Commons](#), and the [Electrical and Computer Engineering Commons](#)

---

Huang, Xi; Liu, Bo; Guo, Shenghan; Guo, Weihong; Liao, Ke; Hu, Guoku; Shi, Wen; Kuss, Mitchell; Duryee, Michael J.; Anderson, Daniel R.; Lu, Yongfeng; and Duan, Bin, "SERS spectroscopy with machine learning to analyze human plasma derived sEVs for coronary artery disease diagnosis and prognosis" (2022).

*Faculty Publications from the Department of Electrical and Computer Engineering*. 692.

<https://digitalcommons.unl.edu/electricalengineeringfacpub/692>

This Article is brought to you for free and open access by the Electrical & Computer Engineering, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications from the Department of Electrical and Computer Engineering by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.





---

**Authors**

Xi Huang, Bo Liu, Shenghan Guo, Weihong Guo, Ke Liao, Guoku Hu, Wen Shi, Mitchell Kuss, Michael J. Duryee, Daniel R. Anderson, Yongfeng Lu, and Bin Duan

## RESEARCH ARTICLE

# SERS spectroscopy with machine learning to analyze human plasma derived sEVs for coronary artery disease diagnosis and prognosis

Xi Huang<sup>1</sup>  | Bo Liu<sup>2,3</sup>  | Shenghan Guo<sup>4,5</sup> | Weihong Guo<sup>4</sup> | Ke Liao<sup>6</sup> | Guoku Hu<sup>6</sup> | Wen Shi<sup>2,3</sup> | Mitchell Kuss<sup>2,3</sup> | Michael J. Duryee<sup>7</sup> | Daniel R. Anderson<sup>3</sup> | Yongfeng Lu<sup>1</sup>  | Bin Duan<sup>2,8,9</sup> 

<sup>1</sup>Department of Electrical and Computer Engineering, University of Nebraska Lincoln, Lincoln, Nebraska, USA

<sup>2</sup>Mary & Dick Holland Regenerative Medicine Program, University of Nebraska Medical Center, Omaha, Nebraska, USA

<sup>3</sup>Division of Cardiovascular Medicine, Department of Internal Medicine, University of Nebraska Medical Center, Omaha, Nebraska, USA

<sup>4</sup>Department of Industrial and Systems Engineering, Rutgers, The State University of New Jersey, Piscataway, New Jersey, USA

<sup>5</sup>School of Manufacturing Systems and Networks, Arizona State University, Mesa, Arizona, USA

<sup>6</sup>Department of Pharmacology and Experimental Neuroscience, University of Nebraska Medical Center, Omaha, Nebraska, USA

<sup>7</sup>Division of Rheumatology, Department of Internal Medicine, University of Nebraska Medical Center, Omaha, Nebraska, USA

<sup>8</sup>Department of Surgery, College of Medicine, University of Nebraska Medical Center, Omaha, Nebraska, USA

<sup>9</sup>Department of Mechanical and Materials Engineering, University of Nebraska-Lincoln, Lincoln, Nebraska, USA

## Correspondence

Daniel R. Anderson, Division of Cardiovascular Medicine, Department of Internal Medicine, University of Nebraska Medical Center, Omaha, NE, USA.

Email: [danderso@unmc.edu](mailto:danderso@unmc.edu)

Yongfeng Lu, Department of Electrical and Computer Engineering, University of Nebraska Lincoln, Lincoln, NE, USA.

Email: [ylu2@unl.edu](mailto:ylu2@unl.edu)

Bin Duan, Mary & Dick Holland Regenerative Medicine Program, University of Nebraska Medical Center, Omaha, NE, USA.

Email: [bin.duan@unmc.edu](mailto:bin.duan@unmc.edu)

## Funding information

University of Nebraska Collaboration Initiative funding; UNL and UNMC Sciences, Engineering, and Medicine Initiative funding

## Abstract

Coronary artery disease (CAD) is one of the major cardiovascular diseases and represents the leading causes of global mortality. Developing new diagnostic and therapeutic approaches for CAD treatment are critically needed, especially for an early accurate CAD detection and further timely intervention. In this study, we successfully isolated human plasma small extracellular vesicles (sEVs) from four stages of CAD patients, that is, healthy control, stable plaque, non-ST-elevation myocardial infarction, and ST-elevation myocardial infarction. Surface-enhanced Raman scattering (SERS) measurement in conjunction with five machine learning approaches, including Quadratic Discriminant Analysis, Support Vector Machine (SVM), K-Nearest Neighbor, Artificial Neural network, were then applied for the classification and prediction of the sEV samples. Among these five approaches, the overall accuracy of SVM shows the best prediction results on both early CAD detection (86.4%) and overall prediction (92.3%). SVM also possesses the highest sensitivity (97.69%) and specificity (95.7%). Thus, our study demonstrates a promising strategy for noninvasive, safe, and high accurate diagnosis for CAD early detection.

Xi Huang and Bo Liu contributed equally to this work.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Bioengineering & Translational Medicine* published by Wiley Periodicals LLC on behalf of American Institute of Chemical Engineers.

## KEYWORDS

coronary artery disease, diagnostics, machine learning, small extracellular vesicles, spectrogram

## 1 | INTRODUCTION

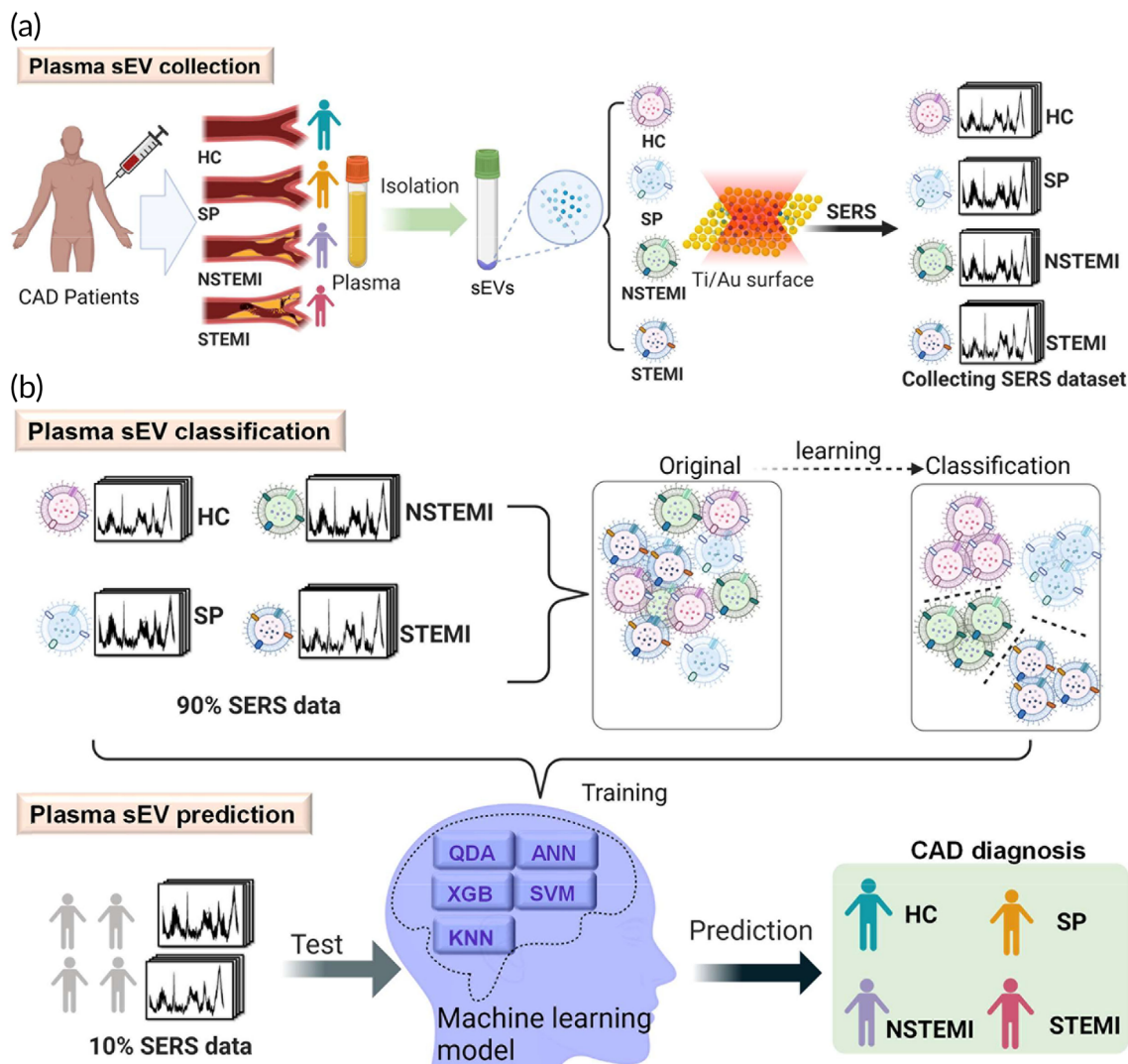
Coronary artery disease (CAD) is represented by the accumulation of atheromatous plaques within the walls of the arteries that supply blood to the heart.<sup>1</sup> CAD is one of the major cardiovascular diseases and remains the leading causes of death worldwide.<sup>2–4</sup> CAD is responsible for about 7 million deaths worldwide.<sup>5,6</sup> Based on the degree of stenosis and plaque characteristics, CAD patients can present clinically with cardiac symptoms and can be divided into different categories, that is, patients with stable nonobstructive plaques (SP), non-ST-elevation myocardial infarction (NSTEMI) and ST-elevation myocardial infarction (STEMI). This reflects the continuum of CAD and with increased severity there are decreased lumen areas, greater plaque burden, more plaque rupture all of which are associated with a greater risk of mortality.<sup>1,7</sup> Thus, there is still a huge need to develop new diagnostic and therapeutic approaches for CAD treatment, especially an early accurate CAD detection and timely intervention, which is expected to avert many late CAD events and deaths.

In the past few decades, various preventive and therapeutic strategies have substantially improved the prognosis of patients suffering from CAD. Many advanced techniques have been developed, reported, and clinically applied to the diagnostic and prognostic workup of CAD, such as electrocardiogram,<sup>8,9</sup> echocardiography,<sup>10</sup> intravascular imaging,<sup>11,12</sup> coronary angiography.<sup>13,14</sup> Although these diagnostic methods have revolutionized the management of CAD patients, the prevalence of adverse cardiac events remains high. Imaging approaches, like coronary angiography, are invasive and should not be used to diagnose the early-stage CAD. Some chemical agents used in cardiac stress testing, such as radiocontrast media may have potential side effects for CAD patients. As well, these stress tests are designed to detect significant CAD which represents a luminal loss of more than 50%. Novel biomarkers can predict and differentiate between CAD types in both the early and late stages, and may reduce unnecessary invasive coronary angiography and thus enhance predictive value.<sup>15</sup> However, the detection sensitivity of these biomarkers has been lower due to the use of large amounts of capturing and labeled detecting antibodies, which may increase the risk of false positivity due to nonspecific binding with nontarget analytes, especially for early CAD stage.<sup>16,17</sup> In addition, patients may not have any symptoms in the early CAD stage reflective of nonobstructive CAD stage (SP stage). Therefore, the prompt, economical, and accurate low-risk diagnosis and prognosis for CAD in asymptomatic patients are crucial to allow timely prevention and therapeutic treatments to improve patients' quality of life prior to the event. Enhancing the early diagnosis of CAD and utilization of therapeutic approaches to prevent progression is crucial for the management and prevention of CAD.

Small Extracellular vesicles (sEVs) are small lipid-bilayer enveloped assemblies with sizes ranging from 20 nm to several micrometers.<sup>18,19</sup>

sEVs are secreted by all cells in both normal and diseased tissues, and can be further categorized based on their biogenesis, size, and biophysical properties, such as exosomes, apoptotic bodies, microvesicles, ectosomes, and other vesicles.<sup>20–23</sup> sEVs are found in most biological fluids and contain a wide variety of cargo, such as proteins, lipids, nucleic acids and metabolites.<sup>21,24,25</sup> These cargoes are representative of their cellular origin and reflective of the pathological condition of the origin tissue and cells, which may serve as noninvasive diagnostic biomarkers in biological fluids.<sup>26–28</sup> Numerous studies have reported the value of exploiting sEV in diagnostic and therapeutic applications in the central nervous,<sup>29,30</sup> cancer,<sup>31,32</sup> and visceral organ diseases.<sup>33,34</sup> The cargoes of these sEVs reflect the molecular content and pathology of their original cells.<sup>23</sup> Therefore, sEVs isolated from liquid and tissue biopsies can serve as potential biomarkers to follow disease progression.<sup>35–39</sup> Different diseases are expected to alter either the sEV contents or the sorting and packaging process<sup>37,39</sup> and these alterations are expected to be detectable and be useful in diagnosis for evaluating disease activity and/or the response to therapy.<sup>40</sup> Plasma as an important component of blood, has the characteristics of representing systemic disease pathology.<sup>41</sup> In the cardiovascular system, sEVs are associated with endothelial cells, cardiac myocytes, vascular cells, progenitor and stem cells, and play an essential role in the development, injury and disease of the cardiovascular system.<sup>28,42</sup> Thus, cardiovascular-derived plasma sEVs have great potential as potential diagnostic biomarkers for CAD screening.

Surface-enhanced Raman scattering (SERS) is a commonly used sensing technique in which inelastic light scattering by molecules is greatly enhanced when the molecules are absorbed onto corrugated metal surfaces (usually Au).<sup>43,44</sup> The label-free, nondestructive and noninvasive characteristics of SERS enable its biomedical application to the diagnosis of diseases, such as neurodegenerative disorders,<sup>45</sup> cancer,<sup>46,47</sup> or diabetes.<sup>48</sup> This innovative technique has also been used to diagnose lung cancer by combining with exosomes by pattern analysis of SERS data.<sup>49,50</sup> Thus, SERS has the potential to differentiate sEVs based on their different membrane lipid/protein contents along with other various functional groups. However, the Raman signatures of sEVs are expected to be highly complex due to the overlapping. The common solution is to analyze the entire Raman spectra as “fingerprint” input by leveraging the power of machine learning (ML). The sEVs collected from patients with different stages of CAD have “impact” on the entire Raman spectra (spectral shapes), although the changes are usually small and very difficult to be detected. By ML algorithms and large training data set, it is possible to detect common “patterns” from the hundreds of Raman spectra with training data and then the algorithm can perform prediction in blind tests. ML has been extensively applied in analyzing spectroscopic signals for complex bio-samples and has achieved satisfying results. Thus, by using ML assisted analysis on a high-dimensional SERS database, valuable



**FIGURE 1** Schematic illustration of ML-assisted sEV analysis for CAD diagnosis. (a) Isolation of sEVs from the human plasma of patients with different CAD stages, including HC, SP, NSTEMI and STEMI stages, and collection of spectroscopic data of plasma-derived sEVs by SERS. (b) Overview of ML-assisted plasma sEV classification and CAD diagnosis using sEVs SERS signal patterns. This illustration is created by [BioRender.com](https://www.biorender.com) with an authorized license

information is expected to be extracted for accurate estimation and practical prediction of known CAD stages which when validated can be used clinically as a diagnostic tool.

In this study, we demonstrated a noninvasive, label-free SERS technique to diagnose CAD by assessing and monitor progression of the disease. To the best of our knowledge, direct, label-free in vitro characterization of sEVs for the early CAD stage diagnosis through SERS at a biomolecular level has not previously been demonstrated. A Raman spectral library of plasma-derived sEVs from patients with various degrees of CAD, including SP, NSTEMI, and STEMI CAD stages, was developed. Plasma from patients without CAD was used as healthy control (HC) group. We hypothesized that SERS data combining with ML algorithms can accurately classify the sEVs from different CAD stages and be used to predict potential risks in CAD patients. Firstly, we isolated and characterized sEVs from human plasma samples with various degrees of CAD and collected their SERS signals

(Figure 1a). Then, the SERS spectra of plasma-derived sEVs were obtained by Raman microscopy and analyzed using five supervised ML models (Figure 1b), including Quadratic Discriminant Analysis (QDA), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Artificial Neural network (ANN), and XGBoost (XGB). Then, 90% of SERS data set of sEVs were used to train the models. The supervised models classified the sEVs data into four clusters (one HC group and three stages of CAD). The remaining 10% spectral data were used to predict their CAD stages through the models. The five methods used in this study are most common algorithms in ML for classification and prediction. Each method represents a typical type of algorithms in ML theory. All methods have predictive errors and statistical noises in the data, especially for large data set or data set with sampling limitations. Therefore, it is important to understand the performance difference among these methods. Thus, we compared the diagnosis performances of the five models and demonstrated robust classifications

and high accurate diagnosis in plasma-derived sEVs for early-stage CAD detection and CAD progression monitoring through the introduction of ML algorithms.

## 2 | RESULTS AND DISCUSSION

### 2.1 | Plasma-derived sEV isolation and characterizations

Plasma-derived sEVs were isolated from patient plasma using a standard procedure of ultracentrifugation as shown in Figure 2a. Ten samples from each group were used to evaluate plasma-derived sEV features. The AFM images showed that the vesicle morphology of the most vesicles was flattened sphere-like with nanoscale sizes (Figure 2b), which was consistent with previous results.<sup>51</sup> The vesicles typically consisted of membrane vesicles of 50–200 nm in diameter according to the results of nanoparticle tracking analysis (NTA) (Figure 2c,d). The sEVs in the HC group had relatively smaller average size (Figure 2c). Additionally, the average size of vesicles in NSTEMI group was the largest among the four groups with significant difference (Figure 2c). The ranges of sEV concentration had wide distribution, and there was no significant difference among different groups (Figure 2d). The markers of sEVs, the tetraspanin (CD63) and endosomal pathway protein (Alix) were detected by Western blotting (Figure 2e). Both Alix and CD63 were expressed in sEVs and Calnexin (negative marker) was not detected. These results confirmed that the recovered vesicles were sEVs. Taken together, the size and content of typical sEV protein markers indicated that sEVs were successfully isolated from the human plasma samples from different stages of CAD patients.

### 2.2 | Averaged Raman spectra for four stages of CAD and QDA classification

To provide an overview of the Raman spectra of four CAD stages, all spectra after standard normal variate processing were simply averaged and shown in Figure 3a. In the fingerprint region, spectra showed Raman peaks appeared to originate from lipids and proteins which are the major contributors to sEV surface. For example, the vibrations contributed to symmetric ring breathing of tryptophan appeared as a peak at  $755\text{ cm}^{-1}$ . A peak at  $830\text{ cm}^{-1}$  was observed corresponding to C—O—O vibration typical of phospholipids. Other peaks such as  $856\text{ cm}^{-1}$  (glycogen),  $879\text{ cm}^{-1}$  (C—C stretch proline ring),  $1005\text{ cm}^{-1}$  (phenylalanine),  $1126\text{ cm}^{-1}$  (C—C vibrations in lipid),  $1244\text{ cm}^{-1}$  (amide III),  $1450\text{ cm}^{-1}$  ( $\text{CH}_2$  bending vibration of proteins/lipids), and  $1668\text{ cm}^{-1}$  (amide I) were also observed.

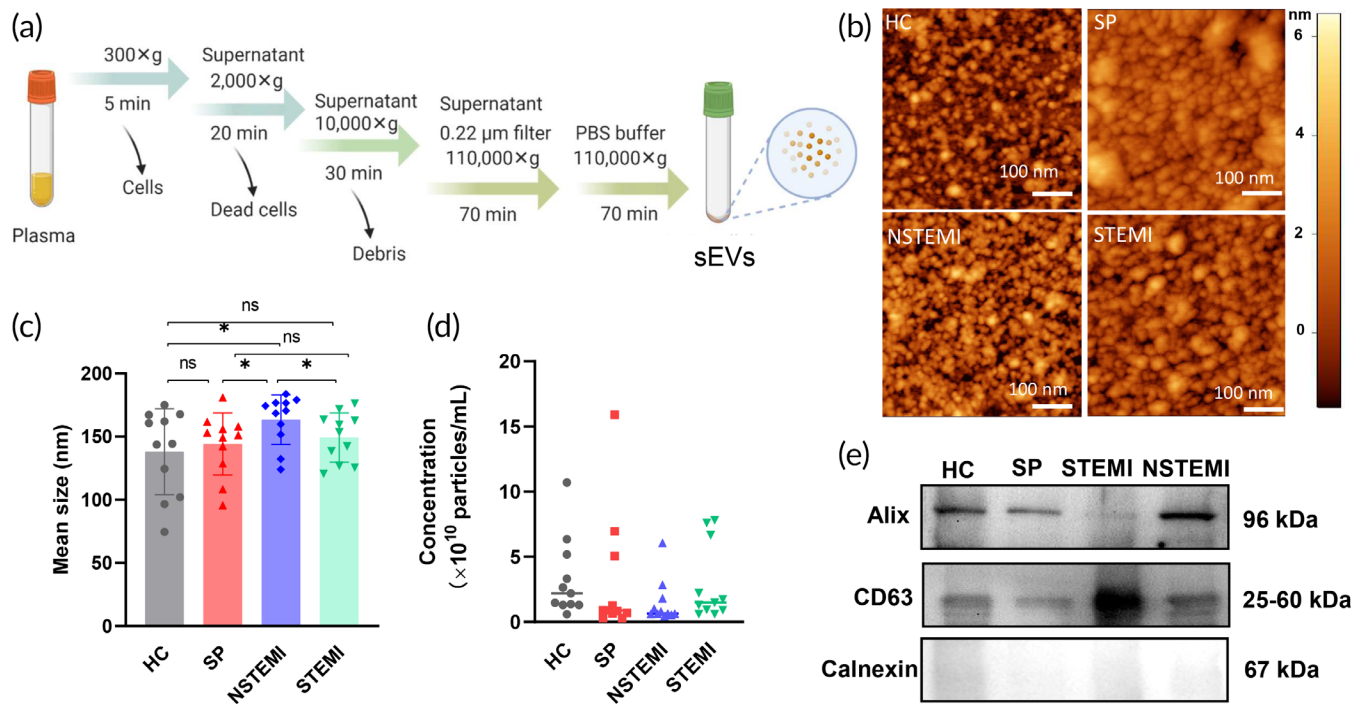
When we closely look at each individual spectrum as shown in Figure S1, the variations in spectra were significantly notable not only across the four CAD stages but even within the same CAD stage. The large variations are expected, because in spectroscopic measurement, the obtained data is semiquantitative, which means that analyzing

only based on a single pathology-related Raman spectroscopic peak is unlikely to be reliable and not suitable for diagnosis or classification of the disease status. The solution for better understanding of the data is to apply ML methods to extract the diagnosis or classification information based on the entire spectral pattern.

The first method we applied is discrimination analysis, which has been widely used to classify and predict Raman spectra of various biological and biomedical samples.<sup>52</sup> In this study, we chose QDA method since the classification (decision) boundaries of QDA can be learned quadratically and flexibly compared to another often used method, linear discrimination analysis (LDA). The PCA plot colored by QDA classification results is shown in Figure 3b, where each colored dot represents a spectrum. It clearly shows the capabilities to separate different sEV subpopulations based on the SERS spectra; however, overlaps are still observed. The overall classification accuracy performed by cross-validation is 80.26% with a sensitivity of 96.28% and a specificity of 74.37%. The sensitivity and specificity were calculated by counting HC group as negative and other three CAD stages together as positive. The heat map (Figure 3c) summarizes the prediction possibility (range 0–1) by QDA for the HC group and three stages of CAD. Red color indicates the highest possibility (1), while blue color indicates the lowest possibility (0). By supervised QDA, the heat map shows most predictions fall in the correct stages, but mispredictions and swings between two stages (yellow color,  $\sim 0.5$ ) are still observed. To draw each receiver operating characteristic (ROC) curve, we counted the representation stage as a positive response, and the other three stages as the negative response for the calculation of  $\text{Sensitivity}_{\text{acc}}$  and  $\text{Specificity}_{\text{acc}}$ . As an example, in the ROC curve of HC, the  $\text{Sensitivity}_{\text{acc}}$  and  $\text{Specificity}_{\text{acc}}$  are calculated as results of HC response against other three stages (SP, NSTEMI, and STEMI) responses. Area under the curve (AUC) of ROC is the indicator of the goodness of fit for the model, and a value of 1 indicates a perfect fit and a value near 0.5 indicates that the model cannot discriminate among the stages. The AUC results by QDA for each stage range from 0.9366 to 0.9824, indicating the QDA method shows good fitting results (Figure 3d). Furthermore, due to possible overfitting by the cross-validation method, we also verified the QDA model by randomly splitting data into training and testing sets. In each round, 90% of data (total 697 spectra) was used as a training set, and the remaining 10% of data (total 78 spectra) was used as a testing set. Only the training set was used to train the model, and the testing set was used for blind prediction. A total of 50 rounds were performed to evaluate the loss and accuracy change of the testing set. As shown in Figure 3e, we found the averaged overall accuracy of QDA is  $79.9\% \pm 3.9\%$ , with a minimum of 71.8% at round 10th and a maximum of 89.7% at round 12th, which is consistent with the cross-validation results.

### 2.3 | KNN, ANN, SVM, and XGB classifications and predictions

Besides QDA, we also implemented four other ML algorithms, including KNN, ANN, SVM, and XGB, to classify and predict CAD stages

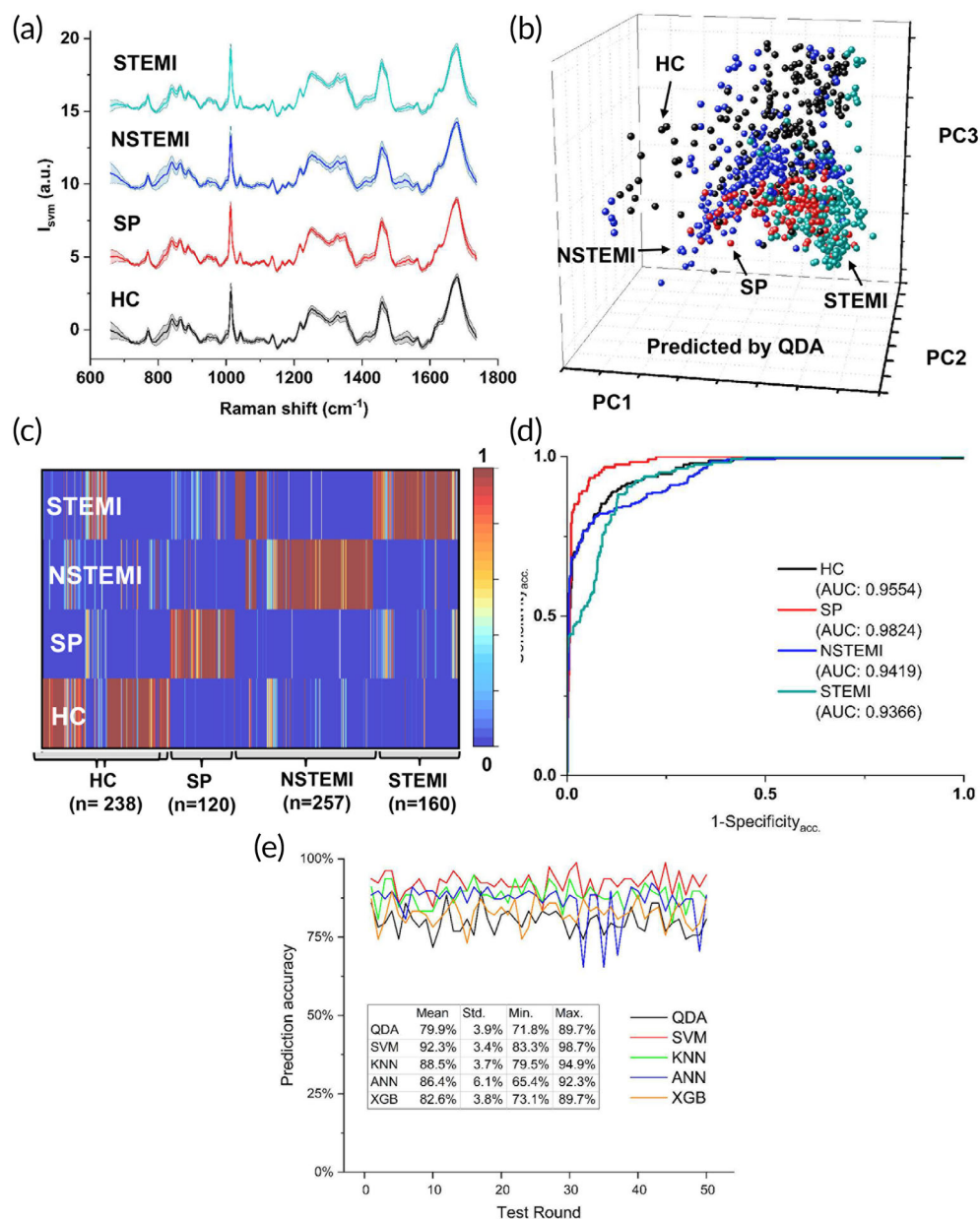


**FIGURE 2** Isolation and characterizations of sEV from human plasma. (a) Scheme of isolation procedure of sEVs from human plasma. (b) AFM images of sEVs. (c, d) The size and concentration of sEVs by NTA test ( $n = 11$ ,  $*p < 0.05$ ). (e) Western blot analysis of biomarker proteins on sEVs

and compare the classification performances. Each method has its own advantages and weaknesses.<sup>53</sup> QDA method allows nonlinear fitting through data and the relationship between data and data interpretation could be established practically. KNN method depends on the nearby adjacent samples rather than the algorithm of discriminating the class domain for classification, which may be more suitable for dense overlapped data sets among multiple classes. ANN has more tolerance to noise and missing data, and is good at handling high-dimensional data sets, but may have difficulty with data interpretation and algorithm structure understanding. SVM is also durable to noise and has advantage in handling a small set of data and overfitting issues. XGB is fast, but ideally requires nonoverlapped data set and an inappropriate training set may lead to distorting the decisions.

Same as QDA, the raw Raman data was preprocessed and performed by PCA initially. Fifteen PCs were chosen (Figure S2) as the new input variables to the models. Then to compare the classification accuracies from different methods, the model parameters also need to be estimated properly since those parameters will significantly affect the efficiency of the classification. Thus, obtaining the best possible set of parameters is critical, both from the computing cost and computing accuracy. Among these four classification methods, all of them have parameters that need to be determined. For ANN, the input is 15 PCs, and the output is the 4-classification group. To avoid overfitting or underfitting, an ANN structure with two hidden layers was chosen. The first hidden layer was set to have 10 neurons, and the second hidden layer was set to have 6 neurons. For XGB, 5000 decision trees were chosen. For SVM, the radial basis function kernel was chosen. In KNN, the best accuracy was obtained when  $K = 1$  after

experimenting  $K = 1-30$ . For all methods, 90% of the full data set was randomly split for training while the remaining 10% was used for testing. The process was also repeated for 50 rounds. Figure 3e shows the prediction accuracies for all the five classification methods used in this study. Results demonstrated that the SVM provided the highest and most robust prediction accuracy of 92.3% with a SD of 3.4% among all five methods. XGB provides a low prediction accuracy of 82.6%, but still better than previously used QDA method (79.9%). Among the 50 round tests of ANN, although the averaged prediction accuracy of ANN is 86.4%, there were 3 rounds of prediction accuracy below 70%, indicating ANN may lack stability due to a randomly chosen training set. The averaged confusion matrix from the 50 round tests is also shown in Table S1. For early CAD prediction (SP stage), QDA, SVM, and ANN had the prediction accuracy of 88.7%, 85.9%, and 86.4%, respectively, much higher than the other two methods, which indicates these ML algorithms may work better for early CAD detection. For the most severe stage (STEMI) prediction, all five algorithms have similar performance. QDA, SVM, KNN, ANN, and XGB had the prediction accuracy of 84.7%, 88.5%, 86.8%, 89.1%, and 84.2%, respectively. For HC group prediction, SVM and KNN have higher prediction accuracies of 95.7% and 91.6% over other three algorithms. Through Table S2, we found SVM has the overall best performance over other four algorithms for prediction of all stages. Table 1 shows the averaged confusion matrix of sensitivity and specificity from the 50 round tests. SVM method also had the highest sensitivity (97.2%) and specificity (95.8%) among all the methods used in this study, since major classification errors occurred within the disease groups (SP to be misclassified as NSTEMI and NSTEMI to be



**FIGURE 3** Raman spectra and QDA analysis. (a) Averaged Raman spectra for four stages of CAD (HC, SP, NSTEMI, STEMI); (b) The PCA plot colored by QDA classification results of the total 775 Raman spectra from four stages of CAD. (c) Heat map of the prediction possibility (range 0–1) by QDA for four stages of CAD. (d) ROC curves for each representation stage and their AUC values. (e) Prediction accuracy of test sets performed in QDA, SVM, KNN, ANN, and XGB in 50 test rounds by randomly splitting 90% of data into a training set and 10% into a test set

misclassified as SP). QDA shows the lowest specificity (74.8%) among the five methods.

## 2.4 | SVM as the best ML method in this study

The PCA plot colored by SVM classification results is shown in Figure 4a, where each colored dot represents a spectrum. The heat map (Figure 4b) summarizes the prediction possibility (range 0–1) by SVM for the HC group and three stages of CAD. Red color indicates the highest possibility (1) while the blue color indicates the lowest possibility (0). The heat map shows almost all predictions fall in the correct stages with a few swings between two stages (yellow color,  $\sim 0.5$ ). The AUC results by SVM for each stage range from 0.9888 to 0.9967, indicating the SVM method shows excellent fitting results in

Figure 4c, Figure 4d,e shows the decision boundaries of SVM projected in the 2D PCA plot, PC1 vs PC2, and PC2 vs PC3, respectively.

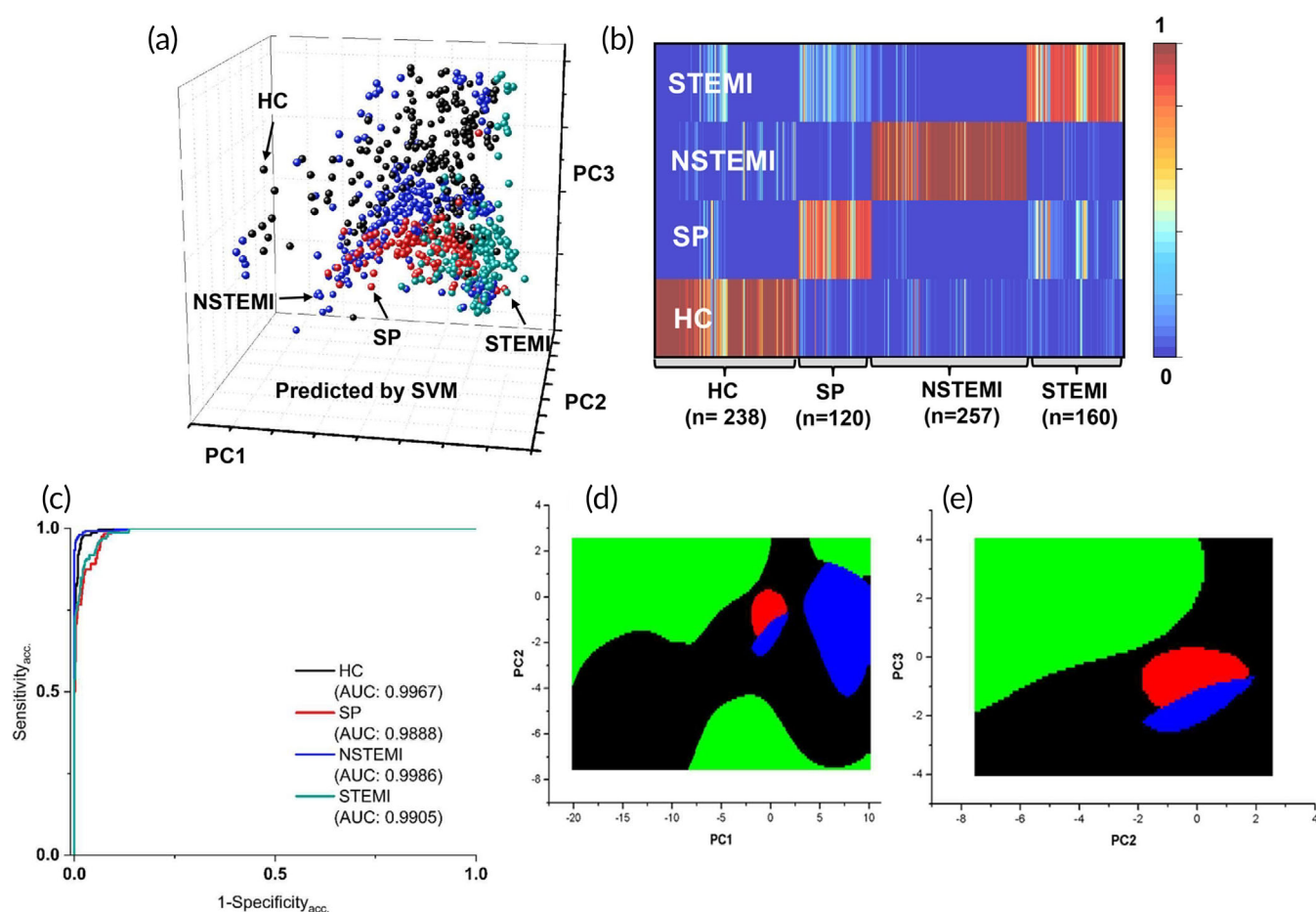
## 2.5 | Challenges and future directions

sEVs serve as a mediator of intercellular communication between cells, and can be used as a noninvasive indicator of disease,<sup>54–56</sup> which is the strategy applied in this study to diagnose CAD status via detecting and analyzing SERS signals from sEVs. However, although as a promising clinical approach, the application of our study still faces several challenges. The first limitation is related to the major Raman signals that are mostly derived from sEV surface molecules (e.g., membrane proteins, lipids).<sup>49,57,58</sup> Thus, SERS signals may be influenced by various membrane molecules.<sup>59</sup> In order to identify



**TABLE 1** Averaged confusion matrix of sensitivity and specificity from the 50 round tests

Model	Sample comparison		Predicted class		Sensitivity	Specificity
			Positive	Negative		
QDA	SP + NSTEMI + STEMI vs HC	Positive	51.8	1.9	96.5%	74.8%
		Negative	6.0	17.8		
SVM	SP + NSTEMI + STEMI vs HC	Positive	52.2	1.5	97.2%	95.8%
		Negative	1.0	22.8		
KNN	SP + NSTEMI + STEMI vs HC	Positive	50.5	3.2	94.0%	91.6%
		Negative	2.0	21.8		
ANN	SP + NSTEMI + STEMI vs HC	Positive	51.0	2.7	95.0%	88.7%
		Negative	2.7	21.1		
XGB	SP + NSTEMI + STEMI vs HC	Positive	49.0	4.7	91.2%	82.4%
		Negative	4.2	19.6		



**FIGURE 4** SVM analysis. (a) A PCA plot colored by SVM classification results of the total 775 Raman spectra from four stages of CAD. (b) Heat map of the prediction possibility (range 0–1) by SVM for four stages of CAD. (c) ROC curves for each representation stage and their AUC values. (d) Decision boundaries of SVM projected in 2D PCA plot (PC1 vs PC2). (e) Decision boundaries of SVM projected in the 2D PCA plot (PC2 vs PC3). (Black: HC; Red: SP; Green: STEMI; Blue: NSTEMI)

protein signatures existing in both surface and interior of sEV, we will analyze sEV cargoes using other advanced technologies such as proteomics.<sup>60</sup> The second challenge is the patient number involved in this study. One solution is to expand the obtained Raman library;

however, another issue may come out that the measurements between different batches of patient samples obtained from national wide may vary largely due to uncertain uniformity among different SERS methods and system errors coming from different Raman

instruments.<sup>44</sup> Thus, it has to optimize the calibration condition before each test to improve the precision of detection and analysis. The third challenge is that the isolated plasma sEVs may contain a variety of sEVs derived from other organs and cells that are not related to the CAD.<sup>34,61</sup> Therefore, our future study will focus on expanding the sample number and types of sEV samples to improve the robustness and reliability of our approach, such as detection of sEVs from heart tissues with different CAD stages.

### 3 | CONCLUSIONS

In summary, we successfully isolated sEVs from the human plasma samples from four stages of CAD patients, that is, HC, SP, NSTEMI, and STEMI. SERS measurements in conjunction with five ML algorithms were then applied for the classification and prediction of the sEV samples. The overall accuracy was 79.9%, 92.3%, 88.5%, 86.4%, and 82.6% for QDA, SVM, KNN, ANN, and XGB, respectively. Among these five approaches, SVM shows the best prediction results on both early CAD detection (86.4%) and overall prediction (92.3%). SVM also possesses the highest sensitivity (97.69%) and specificity (95.7%). Thus, our study demonstrates a promising strategy for noninvasive, safe, and high accurate diagnosis for early CAD detection.

## 4 | MATERIALS AND METHODS

### 4.1 | sEV isolation and characterizations

Deidentified human plasma samples were donated by the Nebraska Cardiovascular Biobank and Registry (IRB approved protocol 133-14-EP). The human plasma was obtained from patients at the time of cardiac catheterization and immediately stored at  $-80^{\circ}\text{C}$  for further use. Samples were obtained from patients who presented with chest pain and a positive stress test, and those who presented with a NSTEMI or a STEMI. Patients who had nonobstructive CAD (i.e.,  $<50\%$  lesions) at catheterization were defined as stable CAD patients (SP). The bank ID, Gender, Age, Race of each patient is recorded in Table S2. sEVs were isolated from patient plasma with three CAD stages, and the number of patients is 15 (SP stage), 15 (NSTEMI stage), and 17 (STEMI stage), respectively. The control samples (HC stage) were from 17 patients that had no signs of CAD. Among all sEV samples, 13 HC samples, 6 SP samples, 13 NSTEMI samples, and 8 STEMI samples were chosen for SERS test, and other samples not used in SERS test were applied for NTA. The human plasma was first centrifuged at  $300 \times g$  for 5 min,  $1000 \times g$  for 20 min and then at  $10,000 \times g$  for 30 min sequentially. The supernatant was then filtered with a  $0.22 \mu\text{m}$  filter, and ultracentrifuged (Sorval X + 80 Ultracentrifuge, Thermo Fisher) at  $100,000 g$  for 70 min. Subsequently, the sEV pellet was washed with PBS and ultracentrifuged at  $100,000 \times g$  for 70 min. The collected sEVs were reconstituted in PBS buffer and then preserved at  $-80^{\circ}\text{C}$ . For SERS analysis, parts of exosomes were resuspended in phosphate buffer (PB) (0.1 M,

$\text{pH} = 7.4$ ). The size and concentration of the final sEVs were examined by NTA using a NanoSight (NS300) measurement. The exosome morphology was evaluated as previously described by using atomic force microscopy (AFM, Bruker).<sup>62</sup>

### 4.2 | Western blot

The markers (CD63 and Alix) of the isolated sEVs were correspondingly detected by Western blotting as reported previously.<sup>62,63</sup> The sEVs were lysed using the Mammalian Cell Lysis kit (Sigma-Aldrich) and quantified using Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific, 23227). The samples were preheated at  $60^{\circ}\text{C}$  for 15 min. Proteins were electrophoresed in an SDS-polyacrylamide gel followed by transferring to PVDF membranes. The membranes were blocked with 5% BSA in TBS-Tween 20 and were then probed with antibodies specific for CD63 (1:1000, ab216130, Abcam), Alix antibody (1:1000, Proteintech), and calnexin (1:1000, Proteintech, negative biomarker for sEVs) overnight at  $4^{\circ}\text{C}$ . After three washes in TBS-tween 20, membranes were incubated with the secondary antibody (Thermo Scientific) for 1 h and washed again. For visualization, blots were exposed to SuperSignal West Dura Extended Duration substrate and measured by the FluorChem R system (ProteinSimple).

### 4.3 | SERS measurements

The sEV samples were immediately measured within 12 hours after taking from  $-80^{\circ}\text{C}$ . Gold-coated glass slides (Ti/Au 40 nm/100 nm, Deposition Research Lab Inc.) were used for better suppression of fluorescence background and surface plasmonic enhancement for SERS. Five microliters of each sEV samples was dropped onto the gold slide and then measured immediately before drying at room temperature. Raman spectra were recorded using a commercial micro-Raman microscope (Renishaw InVia Reflection) with 633-nm diode laser excitation. The laser power was set to 10 mW. The laser beam was focused on the  $5 \mu\text{l}$  droplet by using a  $50\times$  microscope objective with a numerical aperture of 0.75 (Leica n PLAN EPI  $50\times/0.75$ ). The laser spot is estimated to be  $1 \mu\text{m}$  in diameter. In the experiment, due to the sample availability, the sample size was experimentally boosted by automated measuring at different spots of the sEV droplet in  $4 \times 5$  grid. Each Raman spectrum obtained for following analyzes was recorded with an exposure time of 1 s and accumulated by 10 times. The Raman data set is shown in Figure S3. Since the measurement was automated and conducted on the sEV droplets by mapping program, spectra with high background levels (high PBS peak and high background level) were obtained and then should be removed from the data set (Figure S4) to minimize the buffer influence on the following machine learning analysis. Thus, the total number of Raman spectra obtained is 775, including 238 from 13 HC samples, 120 from 6 SP samples, 257 from 13 NSTEMI samples, and 160 from 8 STEMI samples.

#### 4.4 | Data preprocessing and classification methods

Data preprocessing of the raw Raman spectra included baseline correction and normalization. Baseline correction was performed by Vancouver Raman algorithm with five-point boxcar smoothing and a five-order polynomial fit. After baseline correction, the spectra were normalized using the standard normal variate technique, which can remove multiplicative error and preserve each preprocessed spectrum having same contribution to the following classification analysis.

#### 4.5 | ML and classification methods

Classification analysis was adopted to identify the CAD stage based on SERS measurements. Raman spectra were processed to reduce dimensionality, and then input to classifiers for CAD stage prediction. Cross-validation (CV) was adopted for robust model training and validation. The diagnostic capability of the classification models, characterized by sensitivity and specificity, was analyzed with Receiver Operating Characteristics (ROC).<sup>64</sup>

- *Dimensionality reduction.* Since the preprocessed spectra were of high dimensionality, classification analysis on these data directly can be computationally expensive. Principal component analysis (PCA),<sup>65</sup> as a widely used method for dimensionality reduction and feature extraction, was adopted to extract crucial information, that is, *features*, from the spectra data. For a  $n \times p$  data matrix,

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p], \mathbf{x}_i = [x_{1i}, x_{2i}, \dots, x_{ni}]^T, i = 1, \dots, p, \quad (1)$$

a row vector in  $\mathbf{X}$  belongs to  $\mathbf{R}^p$ . PCA uses singular value decomposition (SVD)<sup>66</sup> to extract  $l$  principal components (PCs) from  $\mathbf{X}$ , with  $l < p$ . Each PC consists of element

$$t_{kj} = [x_{j1}, x_{j2}, \dots, x_{jp}] \cdot \mathbf{w}_k, j = 1, 2, \dots, n, k = 1, 2, \dots, l, \quad (2)$$

where  $\mathbf{w}_k$  is the *weights* extracted by SVD that map each row of with  $\mathbf{X}$  to a PC. The extracted PCs are ranked based on the percentage of data variance they explain. For example, the 1st PC explains the highest percentage of data variance among all the PCs; the 2nd PC explains the 2nd largest percentage of data variance, so on and so forth. These PCs are then the features to be used in subsequent classification analysis. In this study, R software was used for PCA implementation.

- *Feature selection.* The PCA reduces the high dimensionality of the Raman spectra (1008 independent variables from  $648 \text{ cm}^{-1}$  to  $1747 \text{ cm}^{-1}$ ) to a few PCs. To determine the optimal number of PCs for following machine learning models, *Quadratic Discriminant Analysis (QDA)* was performed with leave-one-out cross-validation

method.<sup>67</sup> QDA is a classic type of binary classifier, which assumes normally distributed classes with unequal covariance. The two classes respectively follow  $F_0 = N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$  and  $F_1 = N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ . For a feature vector  $\mathbf{t}_j = [t_{j1}, t_{j2}, \dots, t_{jl}]$ , QDA predicts the *likelihood ratio* of the two classes,

$$L_j = \frac{\sqrt{2\pi|\boldsymbol{\Sigma}_1|}^{-1} \exp\left(-\frac{1}{2}(\mathbf{t}_j - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}_1^{-1} (\mathbf{t}_j - \boldsymbol{\mu}_1)\right)}{\sqrt{2\pi|\boldsymbol{\Sigma}_0|}^{-1} \exp\left(-\frac{1}{2}(\mathbf{t}_j - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_0^{-1} (\mathbf{t}_j - \boldsymbol{\mu}_0)\right)}, j = 1, 2, \dots, n \quad (3)$$

Given a threshold of discrimination,  $T$ , if  $L_j < T$ , the  $\mathbf{t}_j$  is assigned class 0. Otherwise, it is assigned class 1. As shown in Figure S4, we compared the QDA classification accuracy for varying number of top PCs, and 15 PCs were chosen (accounts for almost 91.2% of variations in the data) for a good balance between high accuracy and small number of features. Thus, these 15 PCs are used in all following ML models.

- *Classification analysis.* Four other supervised learning classification methods, SVM, KNN, ANN and XGB were also performed by R software on these 15 PCs which served as input variables. The 90% SERS data set of sEVs were used to train the models. The remaining 10% spectral data were used to predict the CAD stages through the models.
- *Analysis of diagnostic capability.* To gain insights into the diagnostic capability of the proposed method, we adopted *receiver operating characteristic (ROC)* curves to analyze the classifiers' sensitivity and specificity. The conventional ROC is a graphic representation of the diagnostic capability for a binary classification model. It visualizes the model's *true positive rate* (TPR) against the *false positive rate* (FPR) as the discriminant threshold varies.

TPR: rate of correctly predicting class 1 (TPR = sensitivity);

FPR: rate of falsely predicting class 1 (FPR = 1 – specificity);

$$TPR = \frac{TP}{TP + FN}, FPR = \frac{FP}{FP + TN} \quad (4)$$

For a complete ROC curve, the area under curve (AUC) represents the capability of accurately predicting the positive cases, which is the larger the better. In our study, there are four classes. To enable the analysis with ROC curves, we counted the representation stage as positive response and other three stages as the negative response for calculation of  $\text{Sensitivity}_{\text{acc}}$  and  $\text{Specificity}_{\text{acc}}$  in ROC curves. However, in other conditions, Sensitivity and Specificity for the five ML models were calculated by counting HC group as negative and other three CAD stages as positive.

#### 4.6 | Statistical analysis

The mean size of small extracellular vesicles is expressed as means  $\pm$  SD. The statistical differences among multiple groups were analyzed

by ANOVA. The  $p$  values are shown in the figures as  $*p < 0.05$ , which are considered to be statistically significant.

## AUTHOR CONTRIBUTIONS

**Xi Huang:** Data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); validation (equal); writing – original draft (equal). **Bo Liu:** Data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); validation (equal); writing – original draft (equal). **Shenghan Guo:** Data curation (supporting); formal analysis (equal); software (equal). **Weihong Guo:** Conceptualization (equal); data curation (supporting); formal analysis (equal); software (equal). **Ke Liao:** Data curation (supporting). **Guoku Hu:** Data curation (supporting). **Wen Shi:** Data curation (supporting); methodology (supporting). **Mitchell Kuss:** Data curation (supporting); methodology (supporting). **Michael Duryee:** Resources (equal); writing – original draft (supporting). **Daniel Anderson:** Conceptualization (equal); funding acquisition (equal); supervision (equal); writing – review and editing (equal). **Yongfeng Lu:** Conceptualization (equal); funding acquisition (equal); supervision (equal); writing – review and editing (equal). **Bin Duan:** Conceptualization (equal); funding acquisition (equal); project administration (lead); supervision (equal); writing – review and editing (equal).

## ACKNOWLEDGMENTS

This work has been supported by University of Nebraska Collaboration Initiative funding (B.D.) and UNL and UNMC Sciences, Engineering, and Medicine Initiative funding (G.K.H., D.A., Y.F.L., and B.D.).

## PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1002/btm2.10420>.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## ORCID

Xi Huang  <https://orcid.org/0000-0002-4477-3748>

Bo Liu  <https://orcid.org/0000-0003-2455-2021>

Yongfeng Lu  <https://orcid.org/0000-0002-5942-1999>

Bin Duan  <https://orcid.org/0000-0002-5647-3793>

## REFERENCES

- Malakar AK, Choudhury D, Halder B, Paul P, Uddin A, Chakraborty S. A review on coronary artery disease, its risk factors, and therapeutics. *J Cell Physiol.* 2019;234(10):16812-16823.
- Khera AV, Kathiresan S. Genetics of coronary artery disease: discovery, biology and clinical translation. *Nat Rev Genet.* 2017;18(6):331-344.
- Musunuru K, Kathiresan S. Genetics of common, complex coronary artery disease. *Cell.* 2019;177(1):132-145.
- Sarnak MJ, Amann K, Bangalore S, et al. Chronic kidney disease and coronary artery disease: JACC state-of-the-art review. *J Am Coll Cardiol.* 2019;74(14):1823-1838.
- Mensah GA, Roth GA, Fuster V. The global burden of cardiovascular diseases and risk factors: 2020 and beyond. *J Am Coll Cardiol.* 2019;74(20):2529-2532.
- Sanchis-Gomar F, Perez-Quilis C, Leischik R, Lucia A. Epidemiology of coronary heart disease and acute coronary syndrome. *Ann Transl Med.* 2016;4(13):256.
- Shroff GR, Li S, Herzog CA. Trends in discharge claims for acute myocardial infarction among patients on dialysis. *J Am Soc Nephrol.* 2017;28(5):1379-1383.
- Deng M, Tang M, Wang C, et al. Cardiodynamicsgram as a new diagnostic tool in coronary artery disease patients with nondiagnostic electrocardiograms. *Am J Cardiol.* 2017;119(5):698-704.
- Lih OS, Jahmunah V, San TR, et al. Comprehensive electrocardiographic diagnosis based on deep learning. *Artif Intell Med.* 2020;103:101789.
- Pastore MC, Mandoli GE, Contorni F, et al. Speckle tracking echocardiography: early predictor of diagnosis and prognosis in coronary artery disease. *Biomed Res Int.* 2021;2021:1-11.
- Mintz GS, Guagliumi G. Intravascular imaging in coronary artery disease. *Lancet.* 2017;390(10096):793-809.
- Smilowitz NR, Mohanane D, Razzouk L, Weisz G, Slater JN. Impact and trends of intravascular imaging in diagnostic coronary angiography and percutaneous coronary intervention in inpatients in the United States. *Catheter Cardiovasc Interv.* 2018;92(6):E410-E415.
- Collet C, Capodanno D, Onuma Y, et al. Left main coronary artery disease: pathophysiology, diagnosis, and treatment. *Nat Rev Cardiol.* 2018;15(6):321-331.
- Raman SV, Zareba KM. Coronary artery disease testing: past, present, and future. *JACC Cardiovasc Imaging.* 2017;10(11):1359-1360.
- Fan Y, Li Y, Chen Y, et al. Comprehensive metabolomic characterization of coronary artery diseases. *J Am Coll Cardiol.* 2016;68(12):1281-1293.
- Lim WY, Thevarajah TM, Goh BT, Khor SM. Paper microfluidic device for early diagnosis and prognosis of acute myocardial infarction via quantitative multiplex cardiac biomarker detection. *Biosens Bioelectron.* 2019;128:176-185.
- Parizadeh SM, Ferns GA, Ghandehari M, et al. The diagnostic and prognostic value of circulating microRNAs in coronary artery disease: a novel approach to disease diagnosis of stable CAD and acute coronary syndrome. *J Cell Physiol.* 2018;233(9):6418-6424.
- Rojalin T, Koster HJ, Liu J, et al. Hybrid nanoplasmonic porous biomaterial scaffold for liquid biopsy diagnostics using extracellular vesicles. *ACS Sens.* 2020;5(9):2820-2833.
- Ullah M, Kodam SP, Mu Q, Akbar A. Microbubbles versus extracellular vesicles as therapeutic cargo for targeting drug delivery. *ACS Nano.* 2021;15:3612-3620.
- Shao H, Im H, Castro CM, Breakefield X, Weissleder R, Lee H. New technologies for analysis of extracellular vesicles. *Chem Rev.* 2018;118(4):1917-1950.
- Skotland T, Sagini K, Sandvig K, Llorente A. An emerging focus on lipids in extracellular vesicles. *Adv Drug Deliv Rev.* 2020;159:308-321.
- van Eijndhoven MAJ, Baglio SR, Pegtel DM. Packaging RNA drugs into extracellular vesicles. *Nat Biomed Eng.* 2020;4(1):6-8.
- Wang C, Li Z, Liu Y, Yuan L. Exosomes in atherosclerosis: performers, bystanders, biomarkers, and therapeutic targets. *Theranostics.* 2021;11(8):3996-4010.
- Panagopoulou MS, Wark AW, Birch DJS, Gregory CD. Phenotypic analysis of extracellular vesicles: a review on the applications of fluorescence. *J Extracell Vesic.* 2020;9(1):1710020.
- Zhang X, Zhang H, Gu J, et al. Engineered extracellular vesicles for cancer therapy. *Adv Mater.* 2021;33:e2005709.
- Castellani C, Burrello J, Fedrigo M, et al. Circulating extracellular vesicles as non-invasive biomarker of rejection in heart transplant. *J Heart Lung Transplant.* 2020;39(10):1136-1148.
- Liu B, Kong Y, Shi W, et al. Exosomes derived from differentiated human ADMSC with the Schwann cell phenotype modulate peripheral nerve-related cellular functions. *Bioact Mater.* 2022;14:61-75.
- Zamani P, Fereydouni N, Butler AE, Navashenaq JG, Sahebkar A. The therapeutic and diagnostic role of exosomes in cardiovascular diseases. *Trends Cardiovasc Med.* 2019;29(6):313-323.

29. Betzer O, Perets N, Angel A, et al. In vivo neuroimaging of exosomes using gold nanoparticles. *ACS Nano*. 2017;11(11):10883-10893.
30. Kanninen KM, Bister N, Koistinaho J, Malm T. Exosomes as new diagnostic tools in CNS diseases. *Biochim Biophys Acta*. 2016;1862(3):403-410.
31. Saari H, Lazaro-Ibanez E, Viitala T, et al. Microvesicle- and exosome-mediated drug delivery enhances the cytotoxicity of paclitaxel in autologous prostate cancer cells. *J Control Release*. 2015;220:727-737.
32. Soung YH, Ford S, Zhang V, Chung J. Exosomes in cancer diagnostics. *Cancers (Basel)*. 2017;9(1):8.
33. Sato K, Meng F, Glaser S, Alpini G. Exosomes in liver pathology. *J Hepatol*. 2016;65(1):213-221.
34. Waldenstrom A, Ronquist G. Role of exosomes in myocardial remodeling. *Circ Res*. 2014;114(2):315-324.
35. Bracci L, Lozupone F, Parolini I. The role of exosomes in colorectal cancer disease progression and response to therapy. *Cytokine Growth Factor Rev*. 2020;51:84-91.
36. Chang L, Ni J, Zhu Y, et al. Liquid biopsy in ovarian cancer: recent advances in circulating extracellular vesicle detection for early diagnosis and monitoring progression. *Theranostics*. 2019;9(14):4130-4140.
37. Garcia-Manrique P, Gutierrez G, Blanco-Lopez MC. Fully artificial exosomes: towards new theranostic biomaterials. *Trends Biotechnol*. 2018;36(1):10-14.
38. Goetzl EJ, Abner EL, Jicha GA, Kapogiannis D, Schwartz JB. Declining levels of functionally specialized synaptic proteins in plasma neuronal exosomes with progression of Alzheimer's disease. *FASEB J*. 2018;32(2):888-893.
39. Kalluri R, LeBleu VS. The biology, function, and biomedical applications of exosomes. *Science*. 2020;367(6478):eaau6977.
40. Ludwig S, Floros T, Theodoraki MN, et al. Suppression of lymphocyte functions by plasma exosomes correlates with disease activity in patients with head and neck cancer. *Clin Cancer Res*. 2017;23(16):4843-4854.
41. Ibsen SD, Wright J, Lewis JM, et al. Rapid isolation and detection of exosomes and associated biomarkers from plasma. *ACS Nano*. 2017;11(7):6641-6651.
42. Barile L, Moccetti T, Marban E, Vassalli G. Roles of exosomes in cardioprotection. *Eur Heart J*. 2017;38(18):1372-1379.
43. Langer J, Jimenez de Aberasturi D, Aizpurua J, et al. Present and future of surface-enhanced Raman scattering. *ACS Nano*. 2020;14(1):28-117.
44. Wang Z, Zong S, Wu L, Zhu D, Cui Y. SERS-activated platforms for immunoassay: probes, encoding methods, and applications. *Chem Rev*. 2017;117(12):7910-7963.
45. Gao X, Zheng P, Kasani S, et al. Paper-based surface-enhanced Raman scattering lateral flow strip for detection of neuron-specific enolase in blood plasma. *Anal Chem*. 2017;89(18):10104-10110.
46. Guerrini L, Pazos-Perez N, Garcia-Rico E, Alvarez-Puebla R. Cancer characterization and diagnosis with SERS-encoded particles. *Cancer Nanotechnol*. 2017;8(1):5.
47. Zhang J, Dong Y, Zhu W, et al. Ultrasensitive detection of circulating tumor DNA of lung cancer via an enzymatically amplified SERS-based frequency shift assay. *ACS Appl Mater Interfaces*. 2019;11(20):18145-18152.
48. Han HW, Yan XL, Dong RX, Ban G, Li K. Analysis of serum from type II diabetes mellitus and diabetic complication using surface-enhanced Raman spectra (SERS). *Appl Phys B*. 2008;94(4):667-672.
49. Shin H, Oh S, Hong S, et al. Early-stage lung cancer diagnosis by deep learning-based spectroscopic analysis of circulating exosomes. *ACS Nano*. 2020;14:5435-5444.
50. Park J, Hwang M, Choi B, et al. Exosome classification by pattern analysis of surface-enhanced Raman spectroscopy data for lung cancer diagnosis. *Anal Chem*. 2017;89(12):6695-6701.
51. Chopra N, Dutt Arya B, Jain N, et al. Biophysical characterization and drug delivery potential of exosomes from human Wharton's jelly-derived mesenchymal stem cells. *ACS Omega*. 2019;4(8):13143-13152.
52. Carmicheal J, Hayashi C, Huang X, et al. Label-free characterization of exosome via surface enhanced Raman spectroscopy for the early detection of pancreatic cancer. *Nanomedicine*. 2019;16:88-96.
53. Amornsamankul S, Pimpunchat B, Triampo W, Charoenpong J, Nuttavut N. A comparison of machine learning algorithms and their applications. *Int J Simul: Syst Sci Technol*. 2019;20(4):1-17.
54. Izadpanah M, Seddigh A, Ebrahimi Barough S, Fazeli SAS, Ai J. Potential of extracellular vesicles in neurodegenerative diseases: diagnostic and therapeutic indications. *J Mol Neurosci*. 2018;66(2):172-179.
55. Shah R, Patel T, Freedman JE. Circulating extracellular vesicles in human disease. *N Engl J Med*. 2018;379(10):958-966.
56. Théry C. Diagnosis by extracellular vesicles. *Nature*. 2015;523:161-162.
57. Chalapathi D, Padmanabhan S, Manjithaya R, Narayana C. Surface-enhanced Raman spectroscopy as a tool for distinguishing extracellular vesicles under autophagic conditions: a marker for disease diagnostics. *J Phys Chem B*. 2020;124(48):10952-10960.
58. Penders J, Nagelkerke A, Cunnane EM, et al. Single particle automated Raman trapping analysis of breast cancer cell-derived extracellular vesicles as cancer biomarkers. *ACS Nano*. 2021;15:18192-18205.
59. Shin H, Seo D, Choi Y. Extracellular vesicle identification using label-free surface-enhanced Raman spectroscopy: detection and signal analysis strategies. *Molecules*. 2020;25(21):5209.
60. Barberis E, Vanella VV, Falasca M, et al. Circulating exosomes are strongly involved in SARS-CoV-2 infection. *Front Mol Biosci*. 2021;8:632290.
61. Smolarz M, Pietrowska M, Matysiak N, Mielanczyk L, Widlak P. Proteome profiling of exosomes purified from a small amount of human serum: the problem of co-purified serum components. *Proteomes*. 2019;7(2):18.
62. Hu G, Liao K, Niu F, et al. Astrocyte EV-induced lincRNA-Cox2 regulates microglial phagocytosis: implications for morphine-mediated neurodegeneration. *Mol Ther Nucl Acids*. 2018;13:450-463.
63. Dagur RS, Liao K, Sil S, et al. Neuronal-derived extracellular vesicles are enriched in the brain and serum of HIV-1 transgenic rats. *J Extracell Vesic*. 2020;9(1):1703249.
64. Kumar R, Indrayan A. Receiver operating characteristic (ROC) curve for medical researchers. *Indian Pediatr*. 2011;48:277-287.
65. Daffertshofer A, Lamothe CJ, Meijer OG, Beek PJ. PCA in studying coordination and variability: a tutorial. *Clin Biomech (Bristol, Avon)*. 2004;19(4):415-428.
66. Klema V, Laub A. The singular value decomposition: its computation and some applications. *IEEE Trans Autom Contr*. 1980;25:164-176.
67. Browne MW. Cross-validation methods. *J Math Psychol*. 2000;44(1):108-132.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Huang X, Liu B, Guo S, et al. SERS spectroscopy with machine learning to analyze human plasma derived sEVs for coronary artery disease diagnosis and prognosis. *Bioeng Transl Med*. 2022;e10420. doi:[10.1002/btm2.10420](https://doi.org/10.1002/btm2.10420)