Supporting Information

Machine Learning-driven SERS platform for precise detection and analysis of vascular calcification

Wei Li[a, b], Zhilian You[b], Dawei Cao[c], Naifeng Liu[a]*

[a]Department of Cardiology, Zhongda Hospital, School of Medicine, Southeast University, Nanjing, 210009, China

[b]Department of Cardiology, Affiliated Hospital of Yangzhou University, Yangzhou University, Yangzhou, 225001, China

[c]School of Information Engineering, Yangzhou Polytechnic Institute, Yangzhou, 225002, China

*Corresponding authors: liunf@seu.edu.cn

1. Materials and Methods

1.1. Materials

Hexadecyltrimethylammonium chloride (CTAC), chloroauric acid ($HAuCl_4$), cetyltriethylammnonium bromide (CTAB), trisodium citrate dihydrate, $NaBH_4$, NaOH, NaOL, silver nitrate, hydroquinone solution all purchased from Sinopharm Chemical Reagent Suzhou Co. All reagents were analytical grade and were used as received without further purification.

1.2. Preparation of GNBPs

GNBPs used in the experiment was prepared via a seed-mediated method (Huang et al., 2014). Firstly, CTAC (66 mM, 4 mL) was mixed with $HAuCl_4$ (1 mM, 1 mL) followed by adding HNO_3 solution (10 mM, 1 mL). After 15 min of vigorous agitation at 800 r/min, $NaBH_4$-NaOH mixed solution (v:v=1:1, 1 mL) and trisodium citrate dihydrate solution (1 M, 100 μL) were added. Thus, the seed solution was prepared after 1 h aging in an 80 °C water bath. Subsequently, deionized water (19 mL), CTAB (0.91 g) and NaOL (0.038 g) were mixed in a breaker under vigorous stirring at 800 r/min, $HAuCl_4$ (0.5 mM, 6.25 mL) and silver nitrate (2 mM, 1.5 mL) were added sequentially during vigorous stirring at 800 r/min. Then, HCl solution (0.5 M, 0.5 mL), hydroquinone solution (0.1 M, 10 mL) and the seed solution (150 μL) were added sequentially after the color of the solution turns from yellow to colorless. The reaction should last for 24 h. After two consecutive centrifugations and washing, the reaction solution was added to deionized water to obtain a successfully prepared GNBPs solution.

1.3. GNBPs substrate fabrication

PDMS substrate was prepared in the cleaned Petri dishes. Prepolymer gel (15 mL) and curing agent (2 mL) were mixed in the Petri dish and placed in the vacuum box for 1 h. Following, the PDMS substrates were peeled off and cut into small piece for further use. The GNBPs substrate was fabricated via the liquid–liquid interface assembly strategy. Briefly, n-hexane (6 mL) and GNBPs solution (6 mL) were added into the breaker followed by adding ethanol (4 mL) dropwise to act as the inducer of self-assembly. After 2 min, a dense metallic luster nanofilm was formed and the prepared PDMS was applied to transfer the nanofilm; thus, the GNBPs substrate was prepared.

1.4. Establishment of VC rat model

All animal experiments were conducted in accordance with the "Guidelines for the Care and Use of Laboratory Animals" of Yangzhou University. These experiments were approved by the Experimental Animal Welfare and Ethics Committee of Yangzhou University (Approval No. 202304001). Fifty rats were divided into healthy group (HL) and vascular calcification (VC) group , with twenty-five rats in each group. Rats in VC group were intraperitoneally injected with vitamin D3 (3×10^5IU/kg) once a day for 30 days, while rats in HL group were intraperitoneally injected with the same volume of olive oil once a day for 30 days. After the construction of the model, the rats in each group were anesthetized with Aferdine (10ml/kg) and scanned by CT to evaluate the calcification of aorta. After 30 days of first injection, all rats were euthanized with sodium pentobarbital(150 mg/kg), and the aortic arch-thoracic aorta-

abdominal aorta were carefully isolated the rats for alizarin red staining to evaluate aortic calcification.Some of the thoracic aorta of rats were taken and embedded in paraffin to make tissue sections for HE staining and Von Kossa staining to assess the calcium deposits in the thoracic aorta.

1.5. SERS measurement and characterization

Transmission electron microscopy (TEM) observation was obtained by using a FEI Tecnai G2 F30 S-twin field-emission transmission electron microscopy. The scanning electron microscopy (SEM) image was obtained via Hitachi S-4800II field emission scanning electron microscope. UV-Vis absorption spectra were recorded with a Shimadzu UV3600 spectroscopy. Raman measurements were conducted with a Thermo Fisher DXR Smart Raman spectrometer. A laser with the wavelength of 785 nm was applied as the excitation source with the laser power and collection time of 80 mW and 10 s, respectively.

1.6. Multivariate analysis

To enhance the signal-to-noise ratio of the raw SERS spectra, eliminate background interference, and correct for instrument effects, all raw SERS spectral data were preprocessed sequentially using Origin software. The preprocessing steps included Raman shift truncation, Savitzky-Golay smoothing, airPLS baseline correction, and normalization. The Raman shift range from 600 to 1800 cm^{-1} was truncated in the SERS spectra. Savitzky-Golay smoothing algorithm was employed to smooth the spectral data, with a smoothing window size of 12 data points and a polynomial order of 3. Adaptive Iteratively Reweighted Penalized Least Squares (airPLS) was used for

baseline correction, with a smoothing parameter λ set to 10 and the number of iterations set to 30. The spectral data was then normalized to a common scale (0-1). We performed an average value analysis on the normalized SERS spectra of the VC ret group (n=25) and the HL rat group (n=25), resulting in their representative spectra. The representative spectrum of HL rat serum was subtracted from the representative spectrum of VC rat serum to obtain the differential SERS spectrum between the two. The differences between VC and HL rats were preliminarily analyzed through the characteristic peaks of the differential SERS spectrum using Origin software. To further distinguish between the SERS spectra, a diagnostic and identification model for VC was established using PCA-LDA method. For PCA, the Scree Plot, Score Plot, Loading Plot, and Box Plot were analyzed by Origin software. The Scree Plot showed the percentage of variance explained by each principal component and the cumulative variance explained, indicating how many principal components were needed to capture a significant amount of the total variance in the SERS dataset. After PCA, the information content of different numbers of PCs was used as input for LDA. When PCs was 13, accounting for 92.7% of the total variance, the model has the highest accuracy. The accuracy, sensitive, specificity, and AUC of the model were analyzed by MATlab software. Meanwhile, the model was also compared with commonly used models such as SVM, KNN, DT, LR, NB, ANN, PCA-ANN, PCA-KNN, and PCA-SVM. In this study, the input layer of the ANN model contained 1,200 nodes, corresponding to the Raman spectral features in the 600 ~ 1800 cm^{-1} range. The model was designed with three hidden layers, containing 512, 256, and 128 nodes,

respectively. Each layer employed ReLU activation functions to effectively process nonlinear data and extract complex features. To prevent overfitting and improve the model's generalization ability, L2 regularization was applied with a strength (λ) of 0.001. Additionally, the training iterations (epochs) were set to 200 to ensure optimal performance. The output layer contained a single node, using a sigmoid activation function for the binary classification task of predicting vascular calcification.

2. Results and discussion

2.1. Characterization of GNBPs substrate

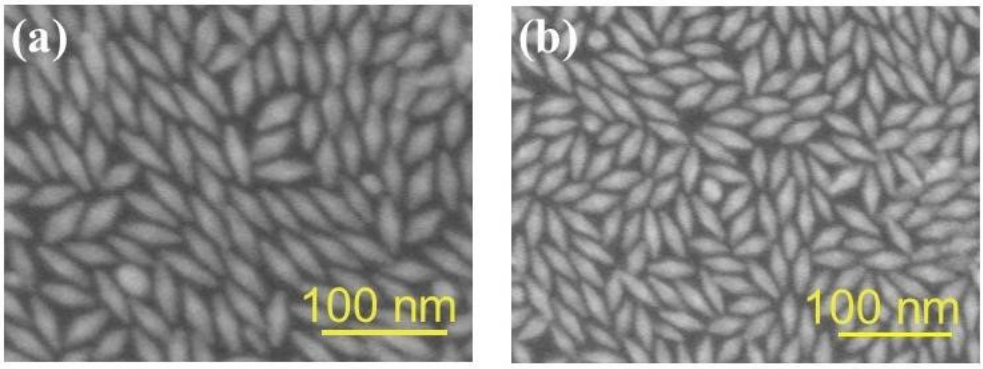

Fig. S1 Optical images of the substrate before and after adding samples. (a) SEM image of GNBPs substrate. (b) SEM image of GNBPs Substrates after adding the sample.

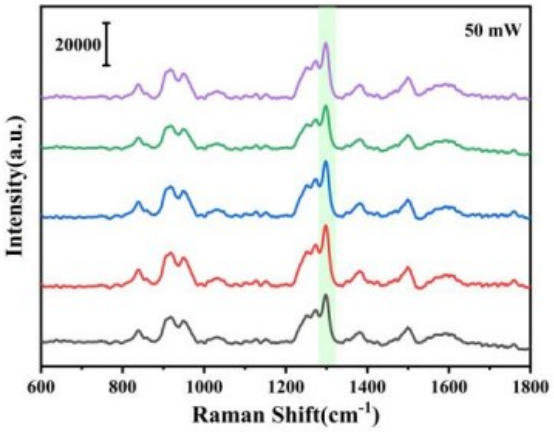

Fig. S2 Repeatability of laser power at 50 mW.

2.2. Comparison of SERS spectra between VC and HL rat serum

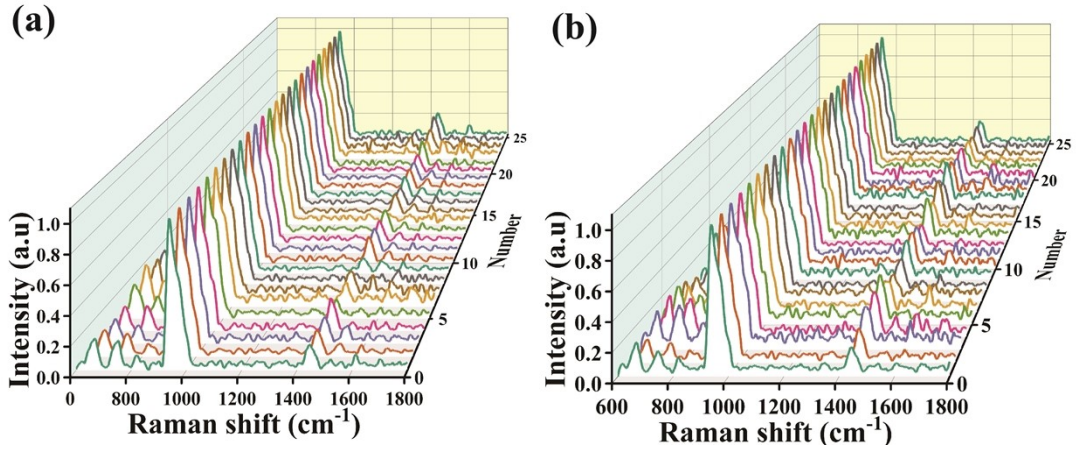

Fig. S3 (a) SERS Spectra of Pre-treated HL Rats (n=25). (b) SERS Spectra of Pre-treated vc Rats (n=25).

2.3. Multivariate analysis

Table S1 SERS peak positions and vibrational mode assignments based on the Reference

Raman shift (cm^{-1})	Assignment	Reference
612	Cholesterol	1
663	C-S stretching mode of cystine	2
757	Tryptophan, Phosphatidylethanolamine	3
816	Proline, hydroxyproline, tyrosine, $\nu_2 PO_2^-$ stretch of nucleic acids	4
846	Monosaccharides (α-glucose), (C-O-C) skeletal mode	5
891	Saccharide band	1
952	ν_s (CH3) of proteins (α-helix)	6
979	C-C stretching β-sheet (proteins)	7
1000	Phenylalanine, Bound & free NADH	8
1033	Phenylalanine mode, C-H in-plane phenylalanine (proteins) ν (CO) ν (CC) ν (CCO) (polysaccharides, pectin)	5, 7, 9
1094	Lipid, ν (C-N)	10
1121	The strong C-O band of ribose	11
1139	Palmitic acid, Fatty acid	1

1157	Carotenoids (absent in normal tissue)	12
1204	Amide III & CH_2 wagging vibrations from glycine backbone & proline side chains, Collagen, Tyrosine, phenylalanine (IgG)	6, 11, 13, 14,
1314	CH_3CH_2 twisting mode of collagen	15
1357	Tryptophan	4
1450	CH_2 bending, C-H deformation bands, Bending modes of methyl groups	16, 17
1532	Carotenoid (absent in normal tissues)	12
1620	ν (C==C), porphyrin	18
1655	Amide I (proteins), C==O stretching (lipids)	7
1708	One of absorption positions for the C==O stretching vibrations of cortisone	19
1746	C==O stretch (lipid)	20

Fig. S4 showed the difference in spectral intensity between VC rat and HL rat at at 663, 757, 846, 979, 1000, and 1450 cm^{-1}. In each figure, the data points for the HL rat group (gray) and the VC rat group (red) represented the Raman intensity values of different samples. The plot indicated that the spectral intensity of VC rat was generally higher than that of HL rat at all key characteristic peaks, especially at 757, 846, 1000, and 1450 cm^{-1}, where the differences were particularly pronounced. These significant differences suggested that these characteristic peaks played a key role in distinguishing between vascular calcification and healthy states. The plot also marked the 25%-75% quartile range and the median line, further highlighting the differences between the two groups. The spectral intensity distribution of VC rat was broader and generally higher, reflecting the changes in chemical composition under pathological conditions.

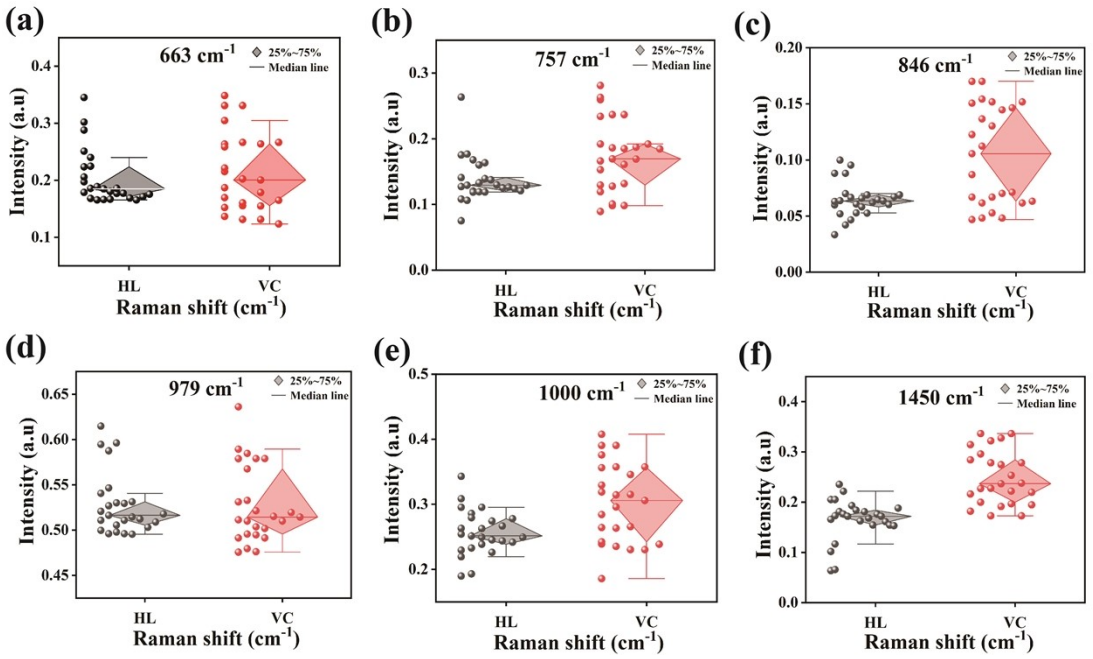

Fig. S4 The scatter plots of key characteristic peaks in the loading plots of PC1, and PC2 distinguished VC rats and HL rats. (a) 663 cm^{-1}, (b) 757 cm^{-1}, (c) 846 cm^{-1} , (d) 979 cm^{-1} , (e) 1000 cm^{-1} , and (e) 1450 cm^{-1}.

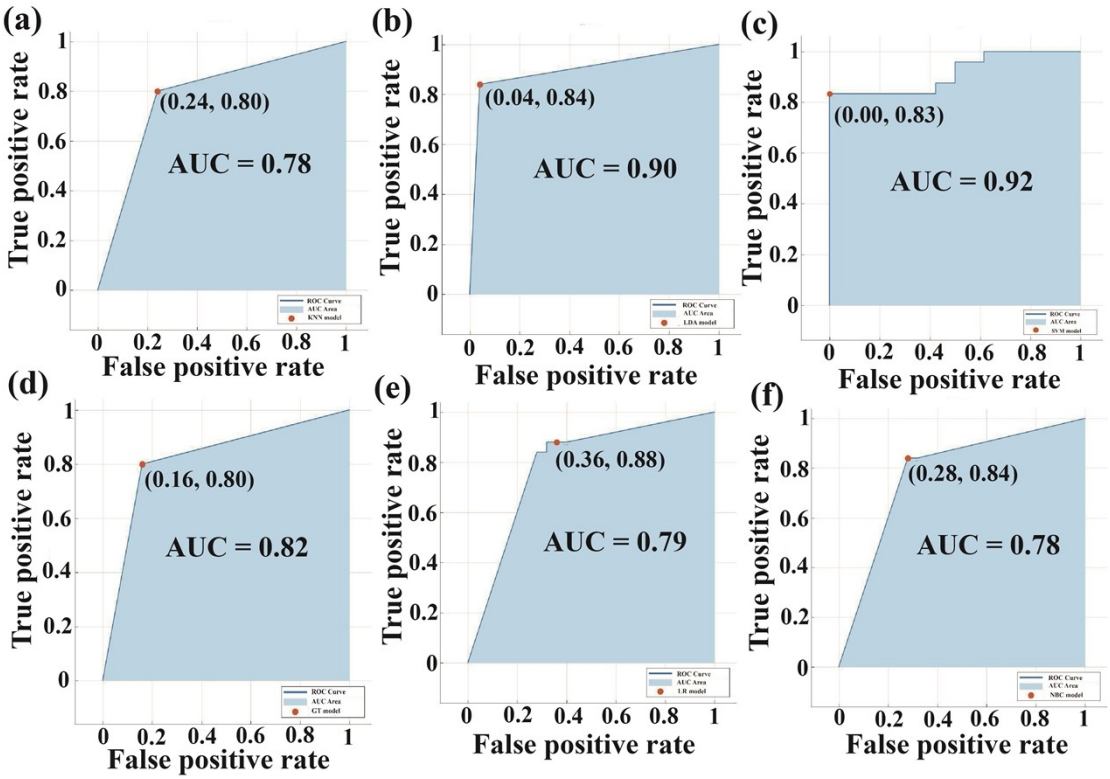

Fig. S5. The AUC values of (a) KNN, (b) LDA, (c) SVM, (d) DT, (e) LR, (f) NB.

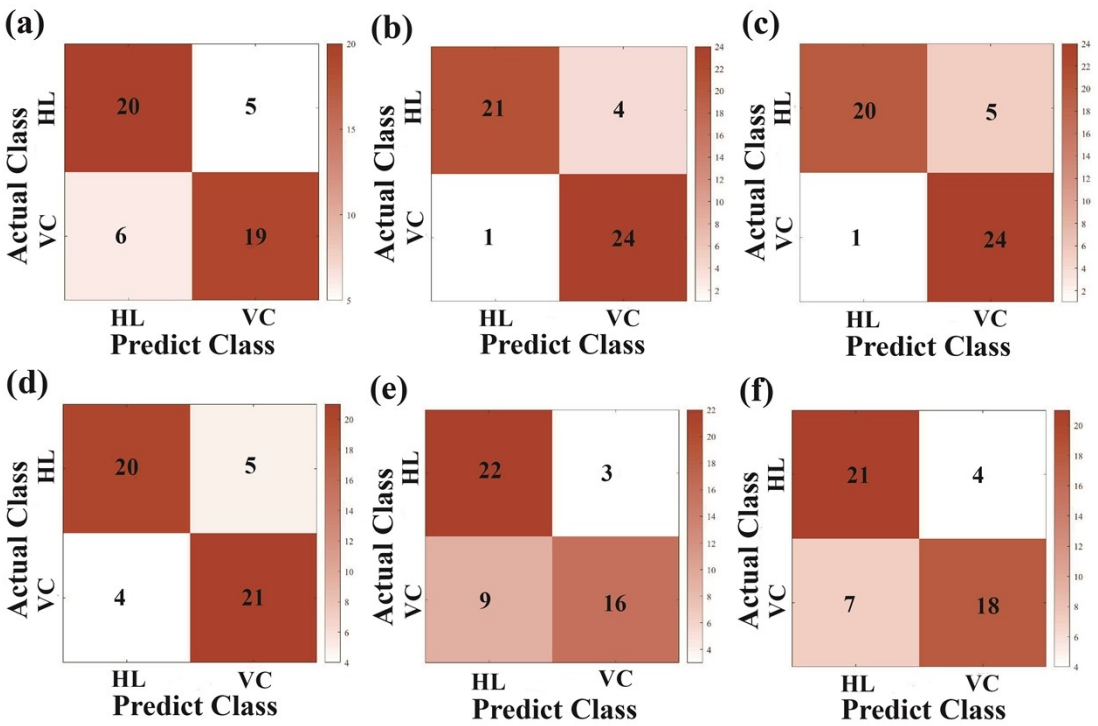

Fig. S6. The confusion matrices of (a) KNN, (b) LDA, (c) SVM, (d) DT, (e) LR, (f) NB.

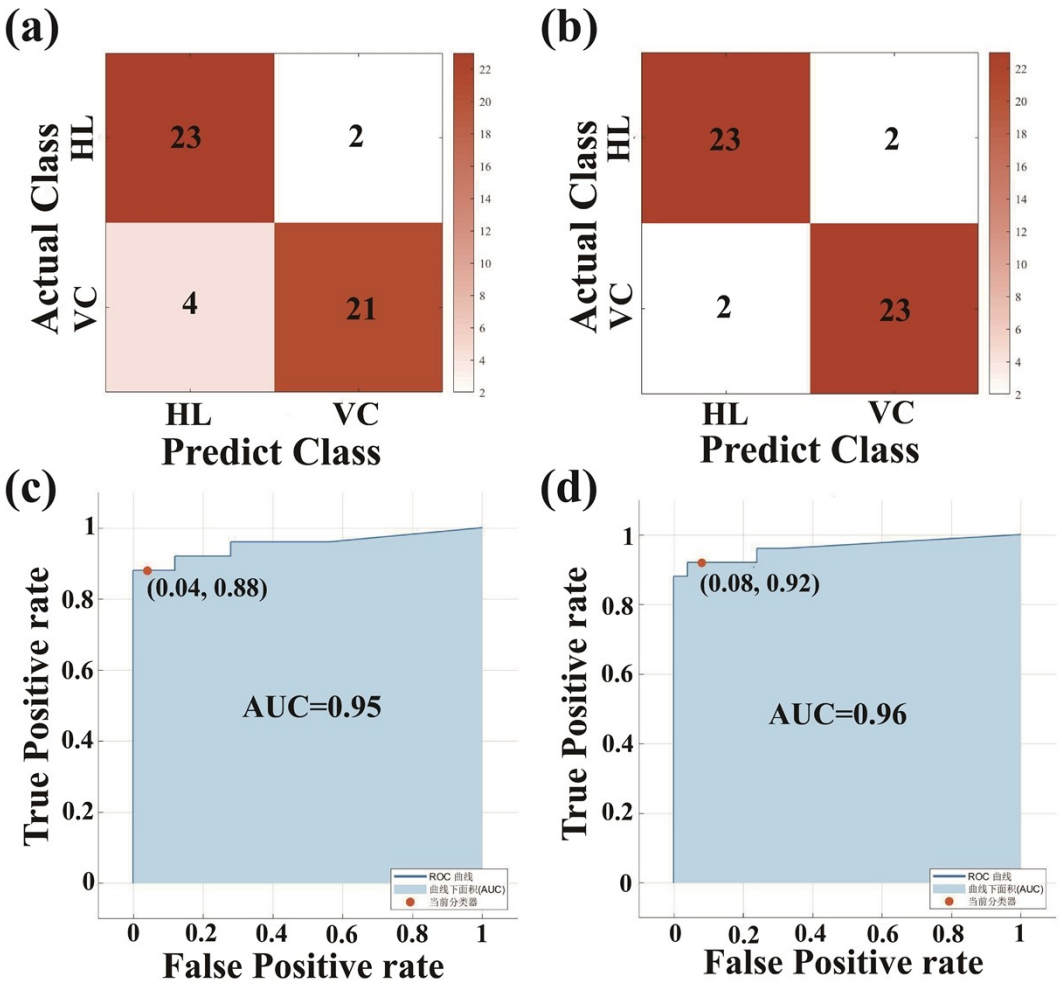

Fig. S7 The confusion matrices of (a) ANN, (b) PCA-ANN. The AUC values of (c) ANN, (d) PCA-ANN.

Reference

1. C. Krafft, L. Neudert, T. Simat and R. Salzer, *Spectrochim. Acta A*, 2005, **61**, 1529-1535.

2. N. Stone, C. Kendell, N. Shepherd, P. Crow and H. Barr, *J. Raman Spectrosc.*, 2002, **33**, 564-573.

3. N. Stone, C. Kendall, J. Smith, P. Crow and H. Barr, *Faraday Discuss.*, 2004, **126**, 141-157.

4. W. T. Cheng, M. T. Liu, H. N. Liu and S. Y. Lin, *Microsc. Res. Tech.*, 2005, **68**, 75-79.

5. G. Shetty, C. Kedall, N. Shepherd, N. Stone and H. Barr, *Br. J. Cancer*, 2006, **94**, 1460-1464.

6. R. J. Lakshimi, V. B. Kartha, C. M. Krishna, J. G. R. Solomon, G. Ullas and P. UmaDevi, *Radiat. Res.*, 2002, **157**, 175-182.

7. I. Notingher, C. Green and C. Dyer, *J. R. Soc. Interface*, 2004, **1**, 79-90.

8. R. Malini, K. Venkatakrishma and J. Kurien, *Biopolymers*, 2006, **81**, 179-193.

9. J. W. Chan, D. S. Taylor, T. Zwerdling, S. T. Lane, K. Ihara and T. Huser, *Biophys. J.*, 2006, **90**, 648-656.

10. S. M. Ronen, A. Stier and H. Degani, *FEBS Lett.*, 1990, **266**, 147-149.

11. R. K. Dukor, *Biomed. Appl.*, 2002, **5**, 3335-3359.

12. A. Mahadevan-Jansen and R. Richards-Kortum, *IEEE EMBS Chicago*, Oct 30-Nov 2, **1997**.

13. G. I. Dovbeshko, N. Y. Gridina, E. B. Kruglova and O. P. Pashchuk, *Talanta*, 2000, **53**, 233-246.

14. P. G. L. Andrus and R. D. Strickland, *Biospectroscopy*, 1998, **4**, 37-46.

15. D. P. Lau, Z. Huang, H. Lui, C. S. Man, K. Berean, M. D. Morrison and H. Zeng, *Lasers Surg. Med*., 2003, **32**, 210-214.

16. S. Sigurdsson, P. A. Philipsen, L. K. Hansen, L. Laesen, M. Gniadecka and H. C. Wulf, *IEEE Trans. Biomed. Eng.*, 2004, **51**, 10.

17. Y. Y. Tan, A. G. Shen, J. W. Zhang, N. Wu, L. Feng, Q. F. Wu, Y. Ye and J. M. Hu, Machine Learning & Cybernetics, Nov 2-5, **2003**.

18. E. Kateinen, M. Elomaa, U.-M. Laakkonen, E. Sippola, P. Niemela, J. Suhonen and K. Jarninen, *J. Forensic Sci.*, 2007, **52**, 88-92.

19. R. A. Shaw and H. H. Mantsch, *J. Mol. Struct.*, 1999, **13**, 480-481.

20. N. Stone, C. Kendall, J. Smith, P. Crow and H. Barr, *Faraday Discuss*., 2004, **126**, 141-157.