BMC Medicine

RESEARCH

Open Access

Multi-cancer early detection based on serum surface-enhanced Raman spectroscopy with deep learning: a large-scale case–control study

Yuxiang Lin[1,2,3†], Qiyi Zhang[4,5†], Hanxi Chen[1,2,3†], Shuhang Liu[5], Kaiming Peng[6], Xiaojie Wang[2,7], Liyong Zhang[2,8], Jun Huang[9], Xiuqing Yan[10,11], Xueliang Lin[12], Uddin M. D. Hasan[13], Mahabub Sarwara[13], Fangmeng Fu[1,2,3*], Shangyuan Feng[5,12*] and Chuan Wang[1,2,3*]

Abstract

Background Early detection of cancer can help patients with more effective treatments and result in better prognosis. Unfortunately, established cancer screening technologies are limited for use, especially for multi-cancer early detection. In this study, we described a serum-based platform integrating surface-enhanced Raman spectroscopy (SERS) technology with resampling strategy, feature dimensionality enhancement, deep learning and interpretability analysis methods for sensitive and accurate pan-cancer screening.

Methods Totally, 1655 early-stage patients with breast cancer (BC, $n=569$), lung cancer (LC, $n=513$), thyroid cancer (TC, $n=220$), colorectal cancer (CC, $n=215$), gastric cancer (GC, $n=100$), esophageal cancer (EC, $n=38$), and 1896 healthy controls (HC) were enrolled. The serum SERS spectra were obtained from each participant. Data dimension enhancement was conducted by heatmap transformation and continuous wavelet transform (CWT). The dimensionalization SERS spectral data were subsequently analyzed by residual neural network (ResNet) as convolutional neural network (CNN) algorithm. Class activation mapping (CAM) method was performed to elucidate the potential biological significance of spectral data classification.

Results All participants were divided into a training set and a test set with a ratio of 7:3. The BorderlineSMOTE method was selected as the most appropriate resampling strategy and the deep neural network (DNN) model achieved desirable performance among all groups (accuracy rate: 93.15%, precision rate: 88:46%, recall rate: 85.68%, and F1-score: 86.98%), with the generated AUC values of 0.991 for HC, 0.995 for BC, 0.979 for LC, 0.996 for TC, 0.994 for CC, 0.982 for GC, and 0.941 for EC, respectively. Furthermore, the combination use of SERS spectra data and ResNet (form of heatmap) were also capable of effectively distinguishing different categories and making accurate

†Yuxiang Lin, Qiyi Zhang and Hanxi Chen contributed equally to this work.

*Correspondence:
Fangmeng Fu
ffm@fjmu.edu.cn
Shangyuan Feng
syfeng@fjnu.edu.cn
Chuan Wang
dr_chuanwang@fjmu.edu.cn
Full list of author information is available at the end of the article

predictions (accuracy rate: 94.75%, precision rate: 89.02, recall rate: 86.97, and F1-score: 87.88), with the AUC values of 0.996 for HC, 0.995 for BC, 0.988 for LC, 0.999 for TC, 0.993 for CC, 0.985 for GC, and 0.940 for EC, respectively. Additionally, strong wave number range of the spectral data was observed in the CAM analysis.

Conclusions Our study has offered a highly effective serum SERS-based approach for multi-cancer early detection, which might shed new light on cancer screening in clinical practice.

Keywords Multi-caner early detection, Serum, Raman spectroscopy, Deep neural network, Convolutional neural network

Background

Cancer is one major public health issue worldwide. For now, the global cancer burden is increasing rapidly, with approximately 19.3 million new cancer cases and 10.0 million cancer deaths estimated in 2020 [1]. It is well acknowledged that early cancer detection could help identify new cases with more effective treatment and reduce patient's economical burden [2]. Current cancer screening paradigms have been widely used in clinical practice, such as mammography for breast cancer (BC) [3], colonoscopy for colorectal cancer (CC) [4], and low-dose computed tomography (CT) for lung cancer (LC) [5]. Nevertheless, these screening methods are only designed for specific cancer type which are not valid for multi-cancer detection. In recent years, liquid biopsy-based methods, including cell-free DNA (cfDNA) [6], circulating tumor DNA (ctDNA) [7], proteins [8], metabolites [9], and cell-derived exosomes [10] have shed new light for simultaneous pan-cancer screening and increase the number of patients who are detected at earlier stages. However, assays based on these tests are of high cost and in demand of strict laboratory quality control. Therefore, the development and validation of a robust, low-cost, and easily repeated method for early cancer prediction are still necessary.

Raman is a molecular vibrational spectroscopic technique where a laser beam is directed onto the sample surface. A Raman spectrum can provide specific information of the tested sample, with changes in the qualitative and quantitative composition of a sample can lead to conversions of the Raman peak intensities, shapes, and locations [11]. Surface-enhanced Raman spectroscopy (SERS) is an enhancement method with nano-scaled metal substrates to increase the intensity of Raman scattering [12]. Numerous studies have demonstrated significant advances in SERS analysis based on human serum samples for detection of multiple cancer types, including breast [13–16], liver [17, 18], lung [19, 20], colon [21], and prostate cancer [22]. Despite of high sensitivity and selectivity, these studies only focus on limited cancer types with insufficient samples, which could not allow for validation and comprehensive large sets of data analysis. In addition, the reported SERS detection methods also included some patients with advanced stages, which might impact the sensitivity and accuracy of prediction. Meanwhile, the one-dimensional nature of early multi-cancer SERS spectral diagnostics poses a limitation on the application scope of deep learning algorithms. Hence, transforming spectral data into two-dimensional images holds some potential significance and applications. This approach aids researchers in visually analyzing the characteristics of different wavelengths associated with early cancer, while better aligning with the input requirements of convolutional neural networks. Moreover, leveraging interpretability analysis techniques can assist in understanding the decision-making behavior of neural networks in image classification tasks. Thus, a highly effective strategy with large-volume of clinical samples and efficient fitting algorithm is in critical need for multi-cancer early detection.

In this research, we assessed the performance and robustness of a SERS-based platform by integrating a large-scale dataset of 1655 early stage cancer patients and 1896 healthy controls. This proof-of-concept study provides a novel blood-based approach for sensitive and accurate pan-cancer screening.

Methods

Study design and population

This observational study was a retrospective analysis from Fujian Medical University Union Hospital between March 2021 and May 2023. Totally, 1655 patients diagnosed with breast cancer (BC, $n=569$), lung cancer (LC, $n=513$), thyroid cancer (TC, $n=220$), colorectal cancer (CC, $n=215$), gastric cancer (GC, $n=100$), esophageal cancer (EC, $n=38$), and 1896 healthy controls (HC) were enrolled. The inclusion criteria of eligible cancer patients were as follows: (I) diagnosed with early-stage cancer (stage 0, I, and II) with pathological confirmation; (II) no history of other malignant tumors; (III) treatment-naive before blood collection. Cancer staging was evaluated according to the 8th edition of the American Joint Committee on Cancer (AJCC) staging system. Healthy controls were frequency-matched to the cancer patients by gender or age (± 5 years) and randomly selected from persons who conducted routine health examinations.

The recruitment of this study was approved by the ethics committee of Fujian Medical University Union Hospital, China (No. 2021KY003, 2021KY004). Informed consent was obtained from each participant.

Sample preparation

Peripheral blood from each participant was collected with a serum collection tube (BD Biosciences). Serum was separated within 2 h after collection by a centrifuge at 2500 rpm (10 min, 4 ℃) and subsequently stored at − 80 ℃ until further analysis. In this study, the method of incubation with silver nanoparticles (AgNPs) as an enhanced substrate for SERS detection was adopted [23]. Specifically, 4.5 ml of 0.1 M NaOH solution was mixed with 5 ml of 0.06 M hydroxylamine hydrochloride solution thoroughly and rapidly poured into 90 ml of 1.1 mM AgNO3 solution. The reaction was stirred at room temperature until a colorless transparent solution changed to a milky gray silver colloid solution. The silver colloid was then centrifuged and the supernatant was discarded. Finally, 5 μl of serum was mixed with 5 μl of the silver colloid and dropped onto a grooved pure aluminum plate, followed by air-drying for SERS detection. The silver colloid prepared in our study exhibited an absorption peak at 425 nm with a half-width of 100 nm. As shown in Additional file 1: Figure S1, the average particle size of the prepared silver nanoparticles was approximately 40 nm observed by the transmission electron microscopy image. Throughout the entire process of silver colloid preparation, precautions were taken to prevent foreign particles like dust from entering the colloid solution to avoid colloidal sedimentation and inefficiency.

Raman measurement

The high-throughput SERS detection platform applied in this study was built in our laboratory which consists of an RMS1000 portable Raman spectrometer and a driver with stepper motors as the X–Y screw slide. Spectral signals were collected with the Raman spectrometer via a computer (Additional file 1: Figure S2). The measurement platform scanning spacing between the X–Y rows and columns was 16 mm, with accuracy of 25 um and a *Z*-axis height of 7.5 mm, of which the maximum measuring stroke is 200 mm*200 mm. The measurement time of single sample ranged from 5 s (single measurement) to 50 s (multiple re-measurements), and the maximum throughput of a single measurement was 100 samples. Serum samples from individuals were tested, and the collection range was between 200 cm^{-1} and 3000 cm^{-1} with a laser power of 20 mW. Four SERS signals from four different positions of each sample were collected with the integration time of 3000 ms. In order to remove abnormal samples caused by charge coupled device (CCD) saturation or cosmic rays, the outlier detection algorithm was then applied and the remaining spectra were averaged to obtain one SERS spectrum for each sample.

Data preprocessing

The spectral preprocessing procedure in this study includes the following steps: spectral cropping, baseline correction (including smoothing), normalization and outlier detection. Typically, Raman signals of biologically significant molecules are distributed within the wavenumber range of 400–1800 cm^{-1}, necessitating spectral cropping. Subsequently, background correction was performed using a multi-polynomial fitting algorithm (the Vancouver Raman algorithm from the BC Cancer Research Center) to reduce interference from fluorescence background [24]. Additionally, the Vancouver algorithm includes a curve smoothing function with adjustable sliding window size, which effectively filters noise and cosmic rays. Next, to eliminate the impact of spectral intensity changes between different spectra on data analysis and modeling, all Raman spectra involved in the same system need to undergo a data scaling process after subtracting the intrinsic fluorescence background from the original spectra, resulting in normalized Raman spectra with an integrated area of 1 to correct for spectral differences induced by systematic errors. Finally, to remove abnormal spectra caused by sample anomalies, PCA-DBSCAN algorithm is employed in this study to process spectral outliers.

Statistical analysis

All statistical analyses were performed by Python programming software (https://www.python.org, version 3.7). Imblearn library (version 0.0) was used for sample resampling strategy. PyTorch library (version 1.12.1) and torchcam library (version 0.3.2) were utilized for deep learning and image interpretability. The visualization of Raman spectral and feature peak was implemented by Origin software (version 2021, OriginLab). In order to eliminate the impact of imbalanced sample quantities on model performance, we conducted resampling on the training set with SMOTE, ADASYN, BorderlineSMOTE, SMOTEENN, and SMOTETomek by different power-law based modulation factors. Preliminary validation was carried out with a deep neural network (DNN) to identify optimal resampling method. Spectral data was dimensionally augmented by heatmap or continuous wavelet transform (CWT) transformation to better suit convolutional neural network (CNN) models. Transfer learning was applied with the ResNet18 network model for image classification. Finally, class activation map (CAM) analysis was employed to interpret the obtained images

and understand the behavior of the model in classifying serum SERS spectra.

Results

Participants' baseline characteristics

Totally, 3551 participants (including 1655 early stage cancer patients and 1896 healthy controls) were enrolled from Fujian Medical University Union Hospital between March 2021 and May 2023. For cancer group, there were 569 cases of breast cancer, 513 cases of lung cancer, 220 cases of thyroid cancer, 215 cases of colorectal cancer, 100 cases of gastric cancer, and 38 cases of esophageal cancer (Fig. 1).

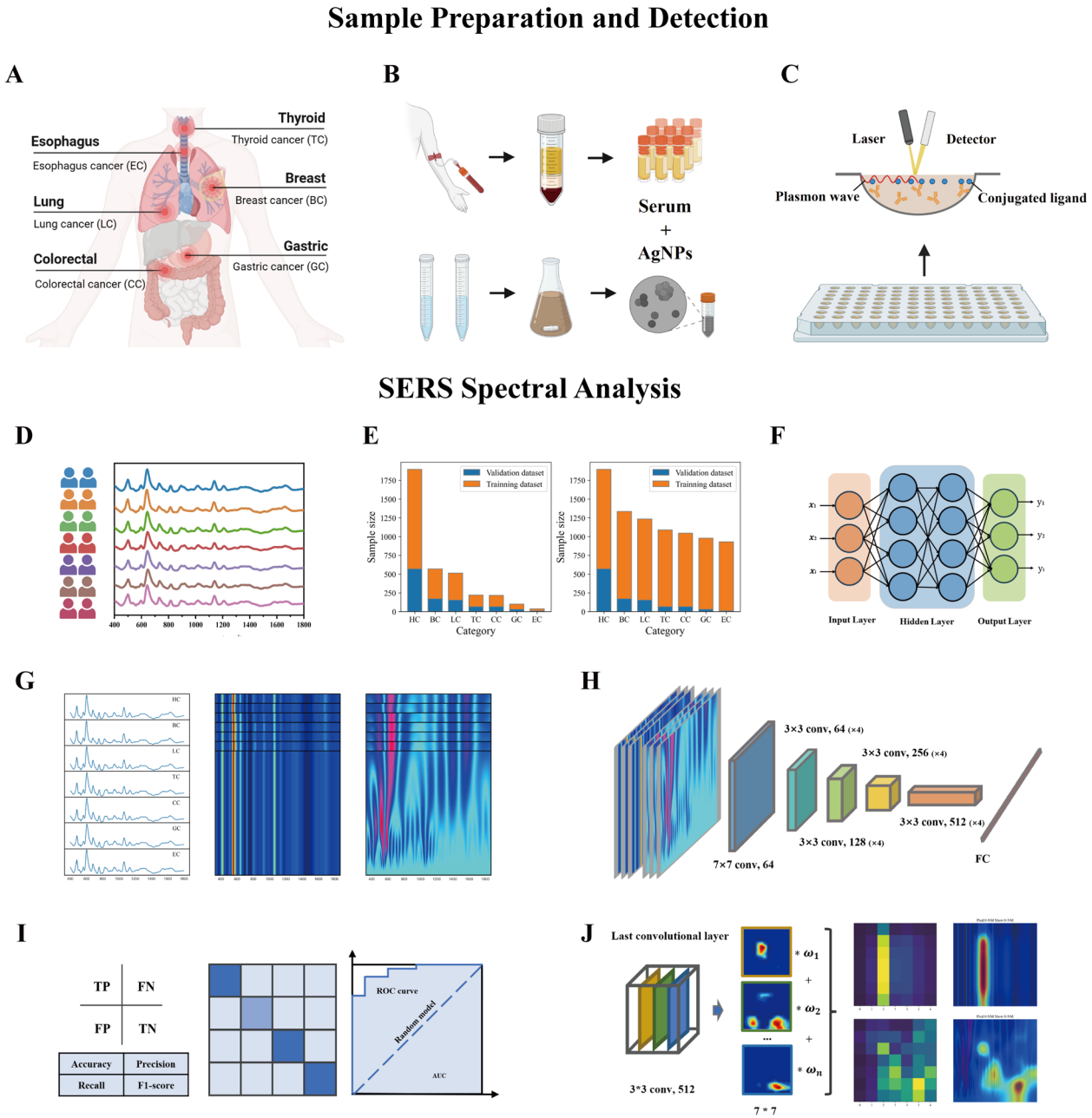

Fig. 1 Schematic diagram of serum SERS multi-cancer detection architecture. **A** Type of cancers enrolled in this study. **B** Serum collection and silver colloid preparation. **C** SERS spectrum acquisition. **D** Mean SERS spectra for each category of serum. **E** Resampling of the training set samples. **F** Deep neural network model. **G** Spectral data dimensionalization. **H** Convolutional neural network model. **I** Model evaluation. **J** Interpretability analysis of the convolutional neural network

SERS spectral analysis

In this study, the silver nanoparticles (AgNPs) were used as an enhanced substrate for SERS detection and a total of 3551 spectra were obtained. As shown in Fig. 2A, the SERS spectra of all serum were averaged by category and the spectral peaks of healthy controls, breast cancer, lung cancer, thyroid cancer, colorectal cancer, gastric cancer, and esophageal cancer patients were similar. In order to highlight the diversities among the groups, we obtained the difference spectra of cancer patients and healthy controls (Fig. 2B). It could be clearly observed that there are obvious differences in the characteristic peaks and intensities of each spectrum, indicating that there are distinct signals in SERS spectrum which can be used for cancer screening. It was also noted that there were multiple characteristic peaks in these seven types of serum samples (Fig. 3A), including 454, 494, 592, 609, 638, 708, 729, 788, 813, 854, 886, 922, 1012, 1134, 1208, 1286, 1580, and 1662 cm^{-1}. By peak analysis, significant differences among each category could be easily identified. These characteristic peaks could characterize specific components such as lipids, proteins, and nucleic acids. The common spectral peak attributions are listed in Table 1, which evidently supports the biological basis of differential signals in SERS spectra. Additionally, we further presented the intensity of spectral characteristic peaks by a heatmap (Fig. 3B) to visually demonstrate the differences and the most apparent characteristic peaks were displayed in a box and scatter diagram (Fig. 3C). As shown in the diagram, significant differences could be discovered in the mean value or intensity distribution range at specific peaks of 494, 638, 813, 922, 1012, 1134, and 1662 cm^{-1} for different types of serum samples. The presence of these characteristic peaks indicated that SERS spectra analysis could provide novel information and act as a powerful tool for the early screening of cancer.

Sample resampling

Considering the imbalanced distribution for each type of cancer sample in this study, a highly effective and robust resampling method is essential. Therefore, we firstly divided the sample into a training set and a test set with a ratio of 7:3. The resampling was then performed with the training set specifically. This partitioning could ensure that there were enough data for model training and evaluation without compromising the independence of test set. The resampling approaches consisted of oversampling methods (SMOTE, ADASYN, and BorderlineSMOTE) and the combination of oversampling and undersampling methods (SMOTEENN and SMOTETomek). In the resampling process, we referred to the power-law-based distribution strategy and introduced modulation factor γ which was described in our previous research [25]. The power-law distribution function could

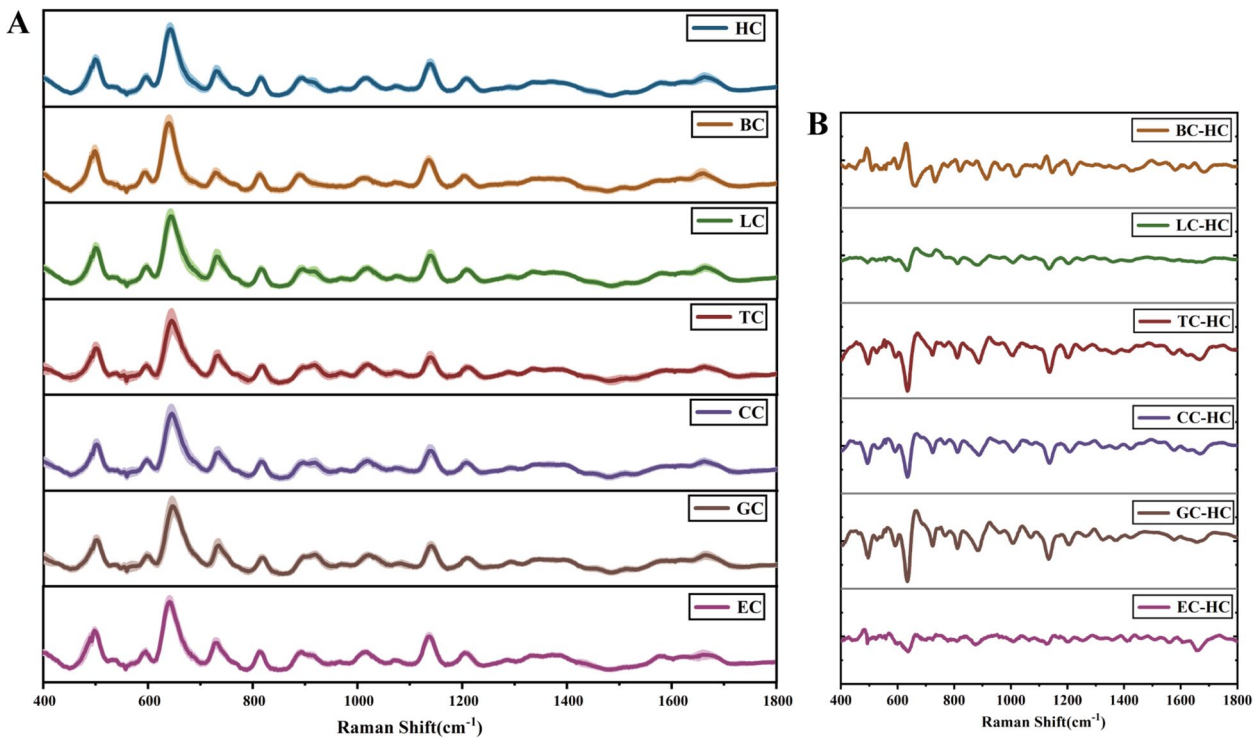

Fig. 2 SERS spectra and difference spectra of each category

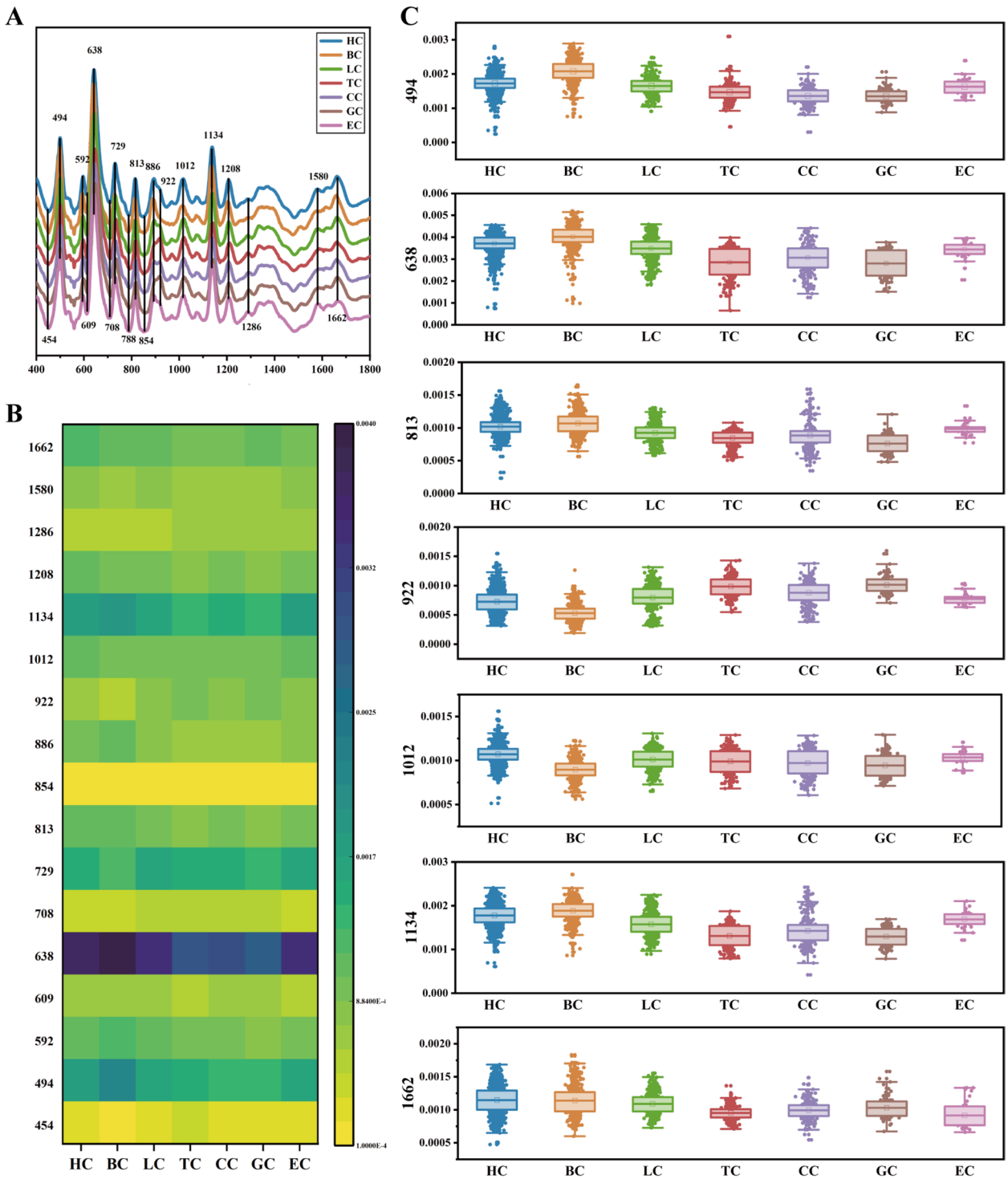

Fig. 3 The characteristic peaks of each category. **A** Spectral characteristic peak. **B** Signal intensity heatmap of characteristic peak. **C** Box and scatter diagram of specific peak position

instruct the synthesis of the minorities in an appropriate scale, without generating either insufficient number of samples that induce model-biased learning or superfluous samples that lead to overlapping classification, thus improving the generalization performance of the model. The generated dataset size after resampling with γ set to 0.6, 0.85, and 1.0 are shown in Table 2.

Table 1 Serum SERS bands positions and tentative vibrational mode assignments

Peak position (cm^{-1})	Major assignment	Peak position (cm^{-1})	Major assignment
494	Arginine	922	proline, valine, protein backbone
592	Ascorbic acid, amide-VI	1012	Phenylalanine
638	L-tyrosine, lactose	1134	D-mannos
729	Adenine, coenzyme A	1208	L-tryptophan, phenylalanine
813	L-serine, glutathione	1580	Guanine, Adenine
886	Glutathione, D-(C)-galactosamine	1662	α-helix, collagen

DNN classification based on resampling model

In order to comprehensively evaluate the influence of different resampling methods and modulation factors on classification results, we adopted deep neural network (DNN) for preliminary verification. Since the input data contained 646 wave numbers, we chose the input layer to contain 646 nodes. In the hidden layer, 1500 neurons were selected and set up four hidden layers to ensure that the model could learn and extract the complex features of the data. To train and verify the model, we conducted 1200 training epochs. In the training process, the learning rate was set to 1×10^{-5} and Adam algorithm was conducted to dynamically adjust the learning rate of each parameter to improve the convergence speed and performance of the model. Simultaneously, we employed the cross-entropy loss function to measure the closeness between the actual and expected output, and updated the gradient of the neural network by backpropagation. The probability of the output layer was then calculated by the Softmax function and the final classification label was obtained. Indicators of accuracy rate, precision rate, recall rate, and F1-score were utilized for a comprehensive assessment of the model's performance. Definition and explanation of evaluation indicators are shown in Additional file 1: S1 and Table S1. After preliminary verification, it was indicated that the BorderlineSMOTE had the best effect on oversampling the training set (Table 3). Especially when the modulation factor γ was set to 0.85, the classification results reached the most ideal state (accuracy rate: 93.15%, precision rate: 88:46%, recall rate: 85.68%, and F1-score: 86.98%), compared with the raw data set (accuracy rate: 89.12%, precision rate: 85.33%, recall rate: 85.68%, and F-1 score: 78.63%. Therefore, the BorderlineSMOTE method was selected as the final resampling strategy and the original or resampling training set with different modulation factor ($\gamma=0.6$, 0.85, and 1) was shown in Fig. 4A. For better observing the distribution of samples, Kernel principal components analysis (PCA) was applied to reduce the dimension of datasets (Fig. 4B). PC1 and PC2 principal components were acted as the horizontal and vertical coordinates, respectively. The confusion matrix of the training or test set (Fig. 4C, D) and the receiver operating the characteristic curves (ROC) (Fig. 4E) were further constructed to demonstrate the classification effect in detail. All datasets exhibited satisfying distinguishment from each type of cancer patients to healthy controls. When γ was set to 0.85, the DNN model achieved the best performance among all imbalanced classes and only a few samples were misclassified. The computed area under the curve (AUC) values were 0.991 for healthy controls, 0.995 for breast cancer patients, 0.979 for lung cancer patients, 0.996 for thyroid cancer patients, 0.994 for colorectal cancer patients, 0.982 for gastric cancer patients, and 0.941 for esophageal cancer patients, respectively.

Data dimension enhancement

Traditional Raman spectral data is one-dimensional and only some general machine learning algorithms or deep neural networks could be used, which limits the research

Table 2 The generated dataset for the resampling training set

Category	HC	BC	LC	TC	CC	GC	EC
Raw data	1896	569	513	220	215	100	38
Test set	569	171	154	66	65	30	11
Training set	1327	398	359	154	150	70	27
Over sampling $\gamma=0.6$	1327	938	765	663	593	541	501
Over sampling $\gamma=0.85$	1327	1165	1080	1023	981	948	921
Over sampling $\gamma=1$	1327	1327	1327	1327	1327	1327	1327

Table 3 DNN classification results of each resampling strategy with different modulation factor

Resampling strategy	Modulation factor γ	Accuracy	Precision	Recall	F1-score
Raw		89.12	85.33	75.34	78.63
SMOTE	0.6	91.56	85.58	85.45	85.44
	0.85	92.03	88.36	84.23	85.69
	1	92.59	88.40	85.07	86.60
ADASYN	0.6	92.03	84.64	85.23	84.88
	0.85	93.06	87.70	85.27	86.41
	1	92.31	86.18	84.85	85.50
BorderlineSMOTE	0.6	92.12	85.97	82.74	84.12
	0.85	93.15	88.46	85.68	86.98
	1	93.06	88.36	84.06	85.91
SMOTEENN	0.6	88.84	80.17	83.34	81.64
	0.85	89.96	81.93	84.17	82.96
	1	89.68	83.77	84.93	84.09
SMOTETomek	0.6	92.31	85.92	85.18	85.43
	0.85	92.87	87.80	85.93	86.80
	1	92.78	86.54	84.86	85.61

on its classification algorithms. To further improve the interpretability and classification ability of SERS spectral data in multi-cancer early detection, convolutional neural network (CNN) was utilized in current study for its favorable performance in computer vision fields. Consequently, dimensionalization was carried out to convert one-dimensional spectral data into two-dimensional images to better adapt to the characteristics of CNN. We utilized two approaches to achieve data dimension enhancement.

The first method was to obtain the corresponding two-dimensional image data by mapping the spectral data into the color space. By selecting an appropriate color mapping scheme and using a gradient color map, relative numerical values can be displayed to better highlight the features of the data. Specifically, the spectral intensity with each wave number corresponded to a color value and the horizontal coordinate was the wave number range of the spectrum (400–1800 cm^{-1}), thus generating the corresponding two-dimensional image. As shown in Additional file 1: Figure S3, the spectral signal intensity was firstly mapped to values ranging from 0 to 255, and then transformed into RGB (red, green, blue) colors using the Jet color mapping table. The transformed image will sequentially display colors from blue, cyan, green, yellow, orange to red based on the spectral signal intensity from low to high.

The second method was to process the spectral signal with a continuous wavelet transform (CWT), an algorithm derived from the Fourier transform, where the wavelet was a waveform with a finite duration and an average of zero. By decomposing signals using wavelet transforms, signals with different frequencies can be obtained, facilitating the analysis of signal time–frequency characteristics. The transformation principle is illustrated in Additional file 1: Figure S4. With different scaling coefficients, CWT could translate the spectral signal and convolve it with the signal to be measured to obtain the frequency information of the signal. This process converted the time-domain features of the original signal into frequency-domain features which effectively generated one-dimensional data into two-dimensional images. The wavelet transformation formula is defined as follows, where ψ is the wavelet signal and the third-order derivative signal of Gaussian Wavelet was selected as the wavelet signal.

$$CWT(a,b) = < f, \psi_{a,b} >= \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t) \cdot \ \psi \left(\frac{t-b}{a} \right) dt$$

$$\psi(t) = diff(e^{-\frac{t^2}{2}}, 3)$$

The two-dimensional images obtained by these two methods were transmitted to the convolutional neural network for spectral signal classification. We processed the one-dimensional spectral data of the oversampling training set and the test set respectively, resulting in 7445 training images and 1066 test images. The original spectrum is shown in Fig. 5A. By heatmap transformation, we used jet colormap from matplotlib library to transform the spectral intensity and the colors appeared blue to red from low to high (Fig. 5B). In the image generated by CWT, the horizontal coordinate was the offset time which was the same as the signal length of 400–1800

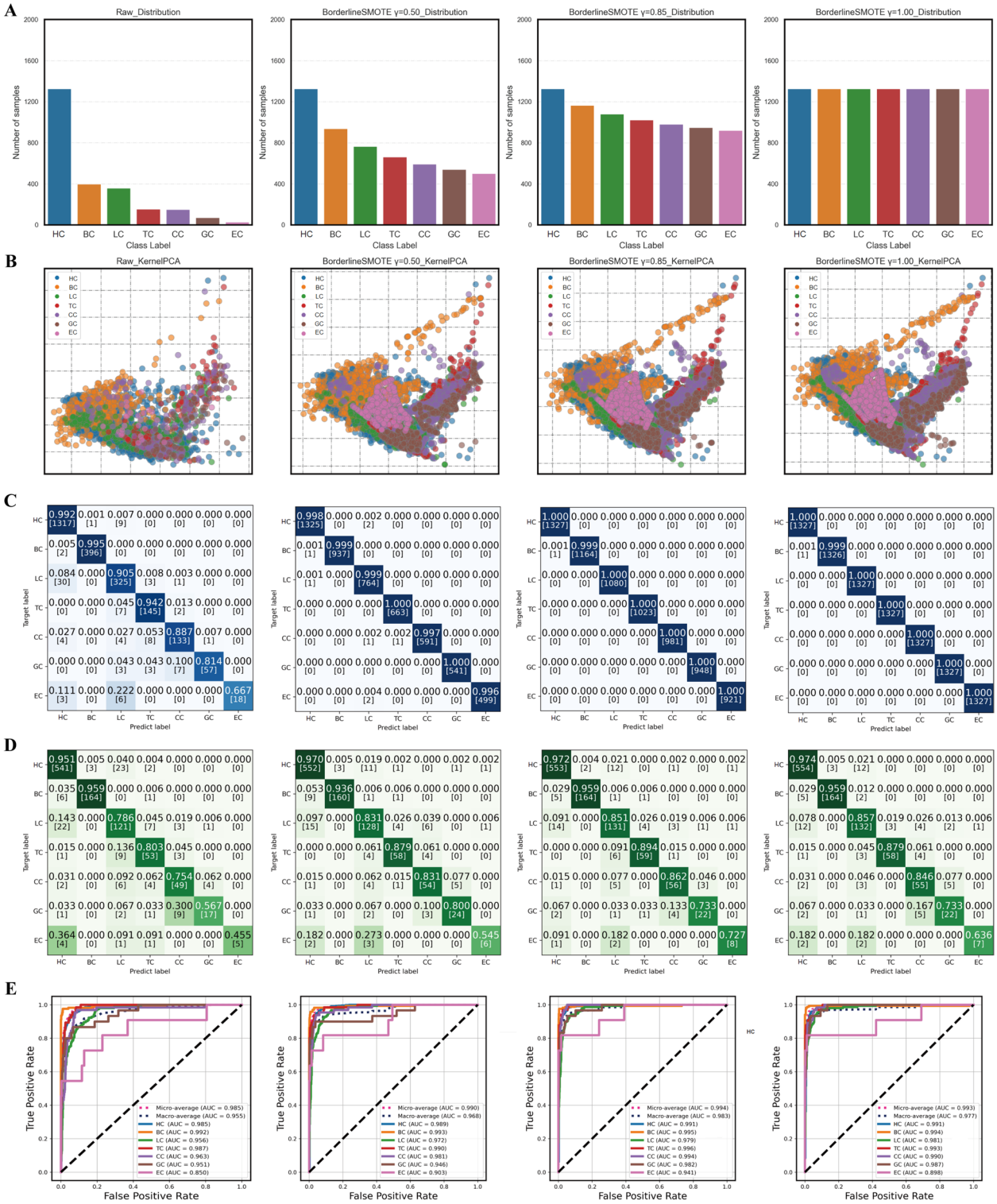

Fig. 4 Classification results by DNN with BorderlineSMOTE algorithm. **A** Bar graph shows the original or resampling training set with different modulation factor ($\gamma = 0.6$, 0.85, and 1). **B** Distribution with kernal PCA analysis. **C** The confusion matrix for training set. **D** The confusion matrix for test set. **E** ROC curves for all categories

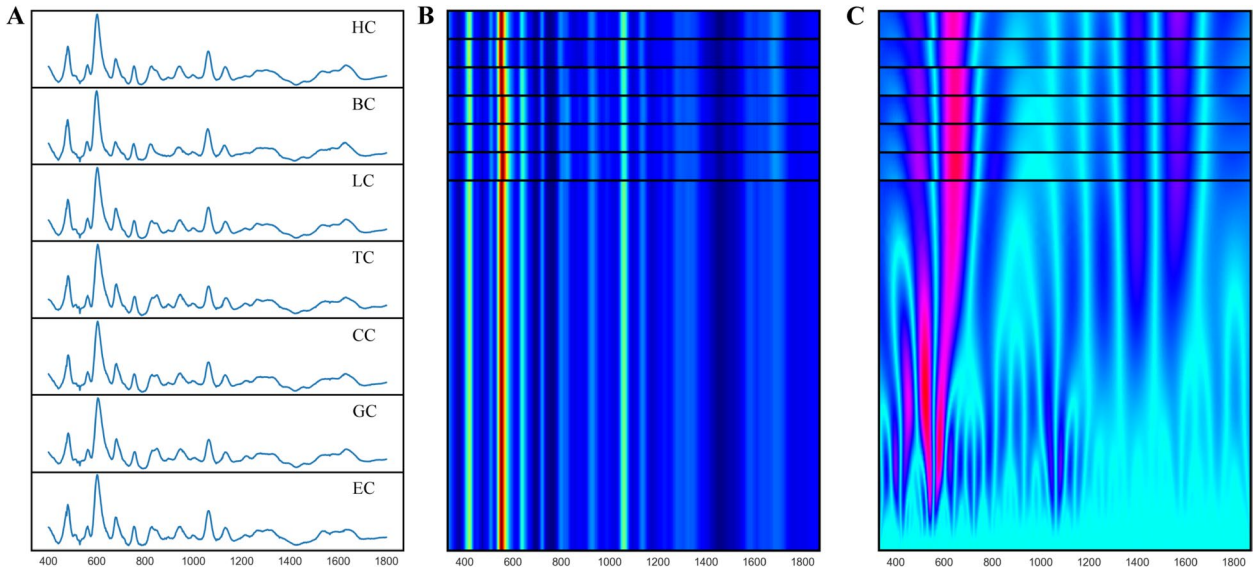

Fig. 5 Spectral data dimensionalization. **A** Raw spectral signal. **B** Heatmap after color mapping. **C** Image after continuous wavelet transform

cm^{-1}, and the vertical coordinate represented the scale (Fig. 5C). These two image processing methods provided the expression of different features of the spectral data and offered more diverse input information for the following analysis.

Residual neural network analysis

For convolutional neural network, we applied residual neural network (ResNet) to analyze the generated images. ResNet is known for introducing residual learning blocks and invented a shortcut connection for degradation, which greatly eliminated the problem of gradient vanishing and training neural networks with too much depth. The ResNet18 model was employed with a depth of 18 layers network and its network structure is shown in Additional file 1: Figure S5. We firstly set the index number of the category and build a mapping dictionary, where HC, BC, LC, TC, CC, GC, and EC were set to 0, 1, 2, 3, 4, 5, and 6 in sequence. Next, the DataLoader was defined to extract 4 images for each time. To prevent overfitting, the data order was shuffled after each epoch to ensure training diversity. Additionally, the Adam optimizer was chosen for updating model parameters. The learning rate was adjusted to half of its original value every 20 epochs, and the cross-entropy loss function was used for model optimization. A total of 160 epochs of training were conducted, and the model with the highest accuracy on the test set was ultimately selected. The classification results of the two dimension enhancing methods were verified and indicated that the use of ResNet as convolutional neural network achieved better classification performance than deep neural network in our task (Table 4). To be noticed, both methods achieve good performance, in which the image in the form of heatmap was slightly better than the image of the wavelet transform, possibly because the form of heatmap could express the characteristic information for the spectral data more intuitively. The confusion matrix for the test set of ResNet (form of heatmap) was displayed in Fig. 6A and the test process was visualized in Fig. 6B, C. The computed area under the curve (AUC) values were 0.996 for healthy controls, 0.995 for breast cancer patients, 0.988 for lung cancer patients, 0.999 for thyroid cancer patients, 0.993 for colorectal cancer patients, 0.985 for gastric cancer patients, and 0.940 for esophageal cancer patients, respectively (Fig. 6D–K). In addition, the accuracy, precision, recall, and F1-score of each category are shown in the Fig. 6L–O. The performance for the test set of ResNet (form of CWT) are shown in Additional file 1: Figure S6. In addition, we also compared ResNet18 model with five other neural network models including AlexNet, VGG16, ResNet34, DenseNet, and MobileNet for 80 epochs. It was observed that ResNet18 model demonstrated better performance when classifying with

Table 4 Performance of different spectral dimension enhancement methods with ResNet18 network

Method	Accuracy	Precision	Recall	F1-score
Heatmap	94.75	89.02	86.97	87.88
CWT	94.65	88.67	87.67	88.12

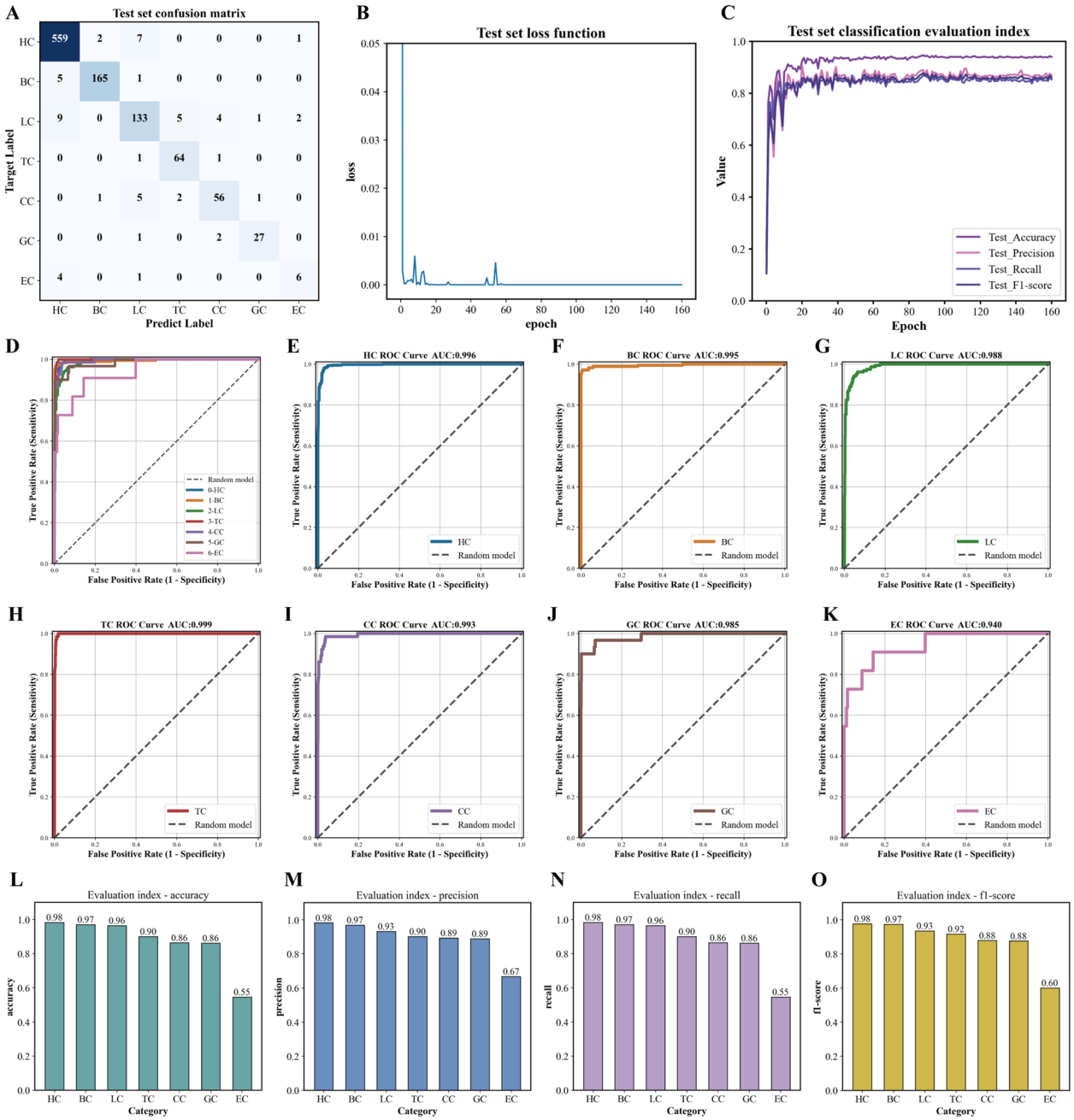

Fig. 6 Classification performance with Resnet18 network by heatmap transformation of serum SERS spectra. **A** The confusion matrix of the test set. **B** Loss function value of the test set. **C** Classification evaluation index of the test set. **D** ROC curves for all categories. **E**–**K** ROC curves and AUC values of HC, BC, LC, TC, CC, GC, and EC, respectively. **L**–**O** The classification evaluation metrics (Accuracy, Presicion, Recall, and F1-score) for each category

both form of heatmap or CWT images (Additional file 1: Table S2 and Table S3).

Class activation map for interpretability analysis

As well known, although deep neural networks or convolutional neural networks provided excellent performance in classification tasks, they still have the black-box explaining problem which was difficult to understand why they make specific predictions. Accordingly, we used class activation map (CAM) to analyze the interpretability and significance of the spectral data. This could help us understand which parts of the spectral image ResNet produced a particular prediction. The application of CAM also contributed to the in-depth understanding for

the principle of SERS serum spectroscopy in the cancer classification task. The steps for CAM acquisition were as follows: (I) feature layers which need to be visualized were extracted and the weight of each channel for the tensor was obtained; (II) the tensor was weighted and summed in the channel dimension by means of linear fusion to obtain a map of size 7*7 grid; (III) the map was normalized and resized it to the original size by interpolation. The mean values of the spectral data for HC, BC, LC, TC, CC, GC, and EC were firstly converted into heatmap (Fig. 7A) and CWT images (Fig. 7D), respectively. The weights of the output layer were then projected back into the convolutional feature map to identify the importance of the image region by CAM (Fig. 7B, E), in which the brighter areas of the 7*7 grid were more important. To intuitively understand the significance of the image area, CAM was superimposed in the input picture (Fig. 7C, F) and the red area presented the more important part,

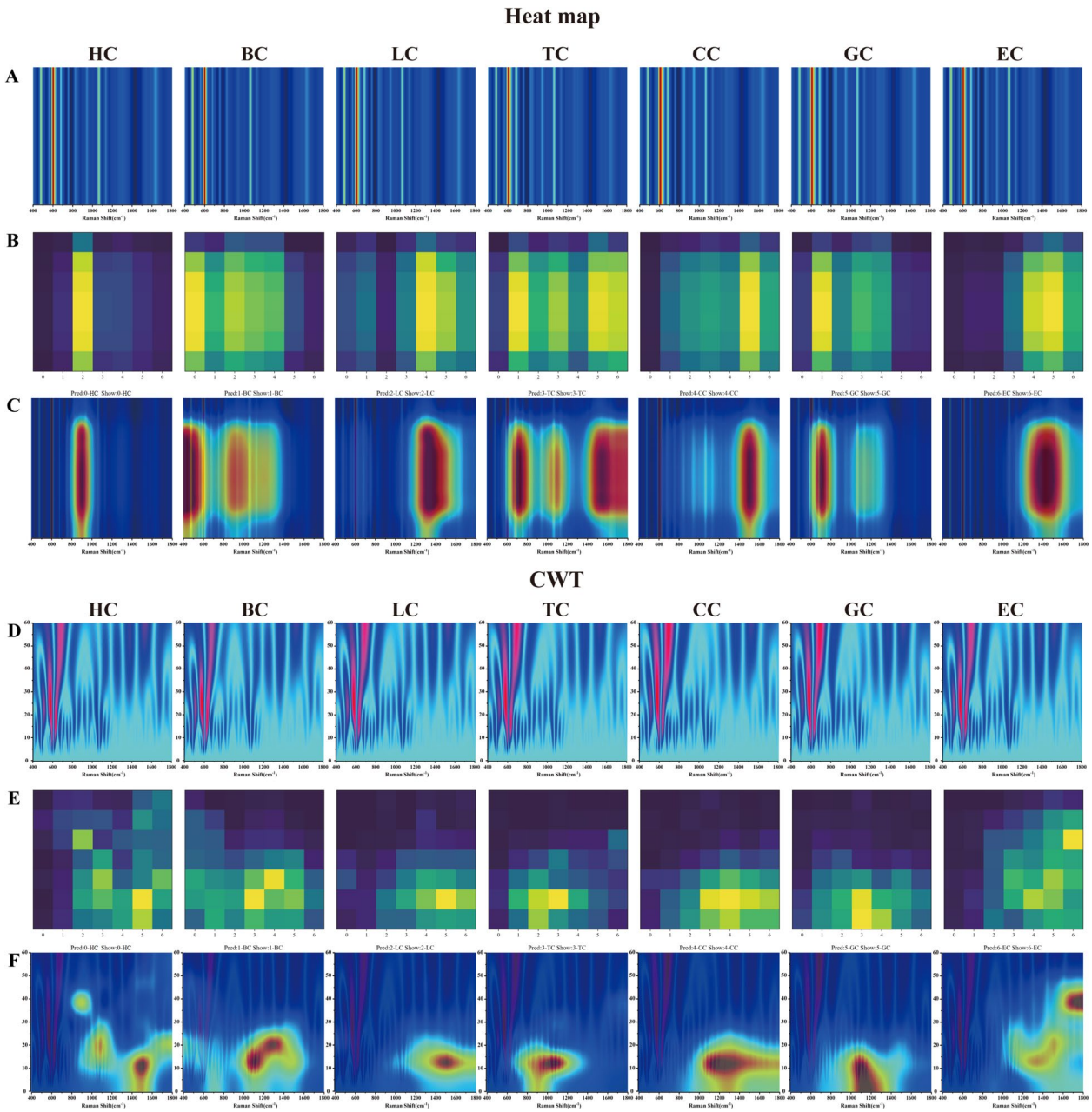

Fig. 7 CAM interpretability for serum SERS classification with the ResNet18 model. **A**, **D** Heatmap and CWT plot from transformed SERS spectra. **B**, **E** CAM feature map for Resnet18 classification. **C**, **F** CAM superimposition on the input picture

while the blue area was the less important area. Interpretative analysis showed that in the SERS spectrum of Resnet18 model based on heatmap, the signal intensity in the wave number range of 800–1000 cm^{-1}, 400–600 cm^{-1}/ 800–1000 cm^{-1}, 1200–1500 cm^{-1}, 600–800 cm^{-1}/ 1000–1200 cm^{-1}/ 1500–1800 cm^{-1}, 1400–1600 cm^{-1}, 600–800 cm^{-1}, and 1400–1600 cm^{-1} contributed significantly in HC, BC, LC, TC, CC, GC, and EC classification. For Resnet18 model based on CWT image, the signal frequency in the wave number range of 1400–1600 cm^{-1}, 1000–1400 cm^{-1}, 1400–1600 cm^{-1}, 900–1200 cm^{-1}, 1100–1500 cm^{-1}, 1000–1300 cm^{-1}, and 1600–1800 cm^{-1} played critical roles in HC, BC, LC, TC, CC, GC, and EC classification.

Discussion

Herein, we report an easier, affordable and efficient multi-cancer early detection method based on serum SERS spectra and empowered by artificial intelligence analysis. In this large-scale study containing 3551 participants and six different early-stage cancer types, serum SERS data combined with deep neural network or convolutional neural network ResNet achieved satisfying performance in the classification of healthy controls and each type of cancer patients. The number of participants, sample types, stable platform, and fitting algorithm demonstrated the robustness and generalization of our findings.

During the past decades, the non-invasive blood detection method based on immunological measurement has been effectively utilized in clinical for cancer screening. For example, carcino-embryonic antigen (CEA) for colorectal cancer [26], carbohydrate antigen125 (CA125) for ovarian cancer [27], carbohydrate antigen19-9 (CA19-9) for pancreatic cancer [28], and alpha-fetoprotein (AFP) for hepatocellular carcinoma [29, 30]. These approaches have significant advantages including the non-invasive nature, automation, and relatively low cost compared with other clinical detection methods such as imaging or endoscopy examinations. However, the low sensitivity and specificity of these methods for early cancer detection has limited their widespread use for screening purpose in a general population setting. In recent years, surface enhanced Raman spectroscopy (SERS) has attracted extensive attention in biomedical applications as a label-free and non-invasive technique [31, 32]. Numerous studies have assessed the utility of serum SERS and confirmed its important role in cancer detection and monitoring [13–22, 33, 34]. Despite high sensitivity, most of studies only recruited certain type of cancer with limited sample size. There is still a lack of large-scale dataset analysis with multiple cancer types. In our study, the 3551 participants comprised 1896 healthy controls and 1655 patients diagnosed with breast cancer (BC, $n=569$), lung cancer (LC, $n=513$), thyroid cancer (TC, $n=220$), colorectal cancer (CC, $n=215$), gastric cancer (GC, $n=100$), and esophageal cancer (EC, $n=38$). All cancer patients were stage 0, I, and II cancer according to the AJCC staging system. It is notable that we did not include a large number of cases of esophageal cancer. One of the primary reasons is that early-stage EC are often asymptomatic and a considerable proportion of EC patients are diagnosed at advanced stage, thus further illustrating the importance of SERS for early screening of esophageal cancer. Based on AgNPs incubation, we conducted Raman tests and obtained SERS average spectrum from all serum samples. By difference spectra and peak analysis, significant differences were observed in the characteristic peaks and intensity parameter from healthy controls and cancer patients, indicating cogent evidence that there were distinct signals in SERS spectrum for each type of serum sample. We then performed artificial intelligent algorithm and utilized resampling strategy since all cancer patients recruited in this study were early stage and the distribution of sample sizes for each type of cancer varied in real-world analysis. In general, most classification algorithms assume that the sample distribution is balanced. However, when the sample size is unbalanced, the classifier may be biased towards those categories with a larger sample size, making it easier to predict the category with a larger sample size, which may lead to a decrease in the accuracy of cancer screening [35, 36]. Therefore, it is necessary to employ the method of generating new data samples for small sample categories to increase the number of minority class. The power-law based distribution strategy was introduced and five resampling approaches including SMOTE, ADASYN, BorderlineSMOTE, SMOTEENN, and SMOTETomek were adopted and preliminary verified by deep neural network analysis. It was identified that the BorderlineSMOTE presented the best effect on oversampling the training set, especially when the modulation factor γ was set to 0.85 (accuracy: 93.15%, precision: 88:46%, recall: 85.68%, and F1-score: 86.98%). Therefore, the BorderlineSMOTE method was applied as the final resampling strategy and the DNN model achieved desirable performance among all imbalanced classes, with the generated AUC values of 0.991 for HC, 0.995 for BC, 0.979 for LC, 0.996 for TC, 0.994 for CC, 0.982 for GC, and 0.941 for EC, respectively. The use of power-law based BorderlineSMOTE method effectively increased the number of minority samples and the balance of data provided a positive effect on the performance of the deep neural network.

In recent years, image analysis is a hot spot in the field of artificial intelligence. Multi-layered machine learning networks including convolutional neural network (CNN),

which can extract high-level features from an input, have also been applied to analyze Raman spectra [17, 37–39]. CNN is a type of deep learning with convolutional layers to learn kernel weights and classically applied to image processing tasks. Compared to fully connected neural networks, this approach could decrease the number of filters and thus reduce its complexity. The convolutional nature of CNN also makes it insensitive to translations in the SERS spectra from instrumental and environmental differences [40, 41]. For the current research, one-dimensional SERS spectral data was processed by heatmap or continuous wavelet transform (CWT) and convert into two-dimensional images. This approach not only eliminated additional noise in the signal but also preserved both global information and local features. The residual neural network (ResNet) was then applied to analyze the generated images for the classification. By combining these two methods, we achieved an accuracy of 94.75% (precision: 89.02, recall: 86.97, F1-score: 87.88) for heatmap transformation and an accuracy of 94.65% (precision: 88.67, recall: 87.67, F1-score: 88.12) for CWT transformation. The combination use of SERS spectra data and convolutional neural network were capable of effectively distinguishing different categories and making accurate predictions. One potential problem with deep learning, including CNN, is their need for large sets of training set, which can be impractical or difficult to obtain. In this study, the large scale of dataset and power-law-based resampling strategy has adequately addressed this issue. Additionally, we conducted class activation map (CAM) to analyze the interpretability and significance of the spectral data. By identifying the strong wave number range in CAM analysis, it is possible to further understand the spectral differences among each category and provide inspiration for the biological mechanism of cancer diagnosis. One main limitation of this study was that all participants were from one single institution in China, although eligibility criteria were formulated to minimize the selective bias. Moreover, despite the broad range of cancer types captured in this study, the sample size was still small for some cancer types, precluding a full representation of heterogeneity within cancer types. Therefore, more cancer types with larger sample size should be included in future studies. In addition, there is still room for improvement in the interpretability study. If combined with relevant biological information, interpretability analysis could be further enhanced to improve the interpretation of potential information in serum SERS spectra for cancer screening.

Conclusions

In summary, the present investigation provides a sensitive platform based on SERS spectrum and reliable artificial intelligence DNN and CNN algorithms. The integrated SERS detection approach is able to distinguish major types of cancer from normal samples and construct a systematic cancer-related database for future exploration.

Abbreviations

AFP	Alpha-fetoprotein
AJCC	American Joint Committee on Cancer
AUC	Area under the curve
BC	Breast cancer
CA125	Carbohydrate antigen 125
CA19-9	Carbohydrate antigen 19–9
CAM	Class activation mapping
CC	Colorectal cancer
CCD	Charge coupled device
CEA	Carcino-embryonic antigen
cfDNA	Cell-free deoxyribonucleic acid
CNN	Convolutional neural network
CT	Computed tomography
ctDNA	Circulating tumor deoxyribonucleic acid
CWT	Continuous wavelet transform
DNN	Deep neural network
EC	Esophageal cancer
GC	Gastric cancer
HC	Healthy control
LC	Lung cancer
RGB	Red, green, blue
ROC	Receiver operating characteristic curve
SERS	Surface-enhanced Raman spectroscopy
TC	Thyroid cancer
TEM	Transmission electron microscopy

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12916-025-03887-5.

Additional file 1: S1. Definition and explanation of evaluation indicators. Table S1. Confusion matrix for binary classification problems. Table S2. Classification performance of each convolutional neural network on heatmap images for 80 epochs. Table S3. Classification performance of each convolutional neural network on CWT images for 80 epochs. Figure S1. UV-Vis absorption spectrum and transmission electron microscopy (TEM) micrograph of the AgNPs. Figure S2. The hardware components of the high-throughput detection device. Figure S3. Principle of Heatmap map transformation. Figure S4. Principle of CWT graph transformation. Figure S5. ResNet18 network structure. Figure S6. Classification performance with Resnet18 network by CWT transformation of serum SERS spectra.

Acknowledgements

Not applicable.

Authors' contributions

LY: Writing-review & editing, Writing-original draft, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. ZQ: Writing-review & editing, Formal analysis, Investigation, Conceptualization. CH: Investigation, Data curation. LS: Investigation. PK: Investigation, Data curation. WX: Investigation, Data curation. ZL: Investigation, Data curation. HJ: Investigation, Data curation. YX: Investigation, Data curation. LX: Formal analysis, Funding acquisition. HMU: Investigation. SM: Investigation. FF: Supervision, Project administration. FS: Supervision, Project administration, Funding acquisition. WC: Supervision, Project administration. All authors read and approved the final manuscript.

Funding

This work was supported by grants from National Natural Science Foundation of China (82203781, 62205179) and Fuzhou Science and Technology Major Project (2023-ZD-004).

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

This study was approved by the Research Ethics Committee of Fujian Medical University Union Hospital (No. 2021KY003, 2021KY004) and informed consent was obtained from each participant.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Breast Surgery, Fujian Medical University Union Hospital, Fuzhou, Fujian Province 350001, China. [2]Department of General Surgery, Fujian Medical University Union Hospital, Fuzhou, Fujian Province 350001, China. [3]Breast Cancer Institute, Fujian Medical University, Fuzhou, Fujian Province 350001, China. [4]Network Technology Center, Clinical Oncology School of Fujian Medical University, Fujian Cancer Hospital, Fuzhou, Fujian Province 350003, China. [5]Key Laboratory of Optoelectronic Science and Technology for Medicine of Ministry of Education, Fujian Provincial Key Laboratory of Photonics Technology, Fujian Normal University, Fuzhou, Fujian 350117, China. [6]Department of Thoracic Surgery, Fujian Medical University Union Hospital, Fuzhou, Fujian Province 350001, China. [7]Department of Colorectal Surgery, Fujian Medical University Union Hospital, Fuzhou, Fujian Province 350001, China. [8]Department of Thyroid Surgery, Fujian Medical University Union Hospital, Fuzhou, Fujian Province 350001, China. [9]Department of Laboratory Medicine, Fujian Medical University Union Hospital, Fuzhou, Fujian Province 350001, China. [10]Department of Nursing, Fujian Medical University Union Hospital, Fuzhou, Fujian Province 350001, China. [11]Department of Medical Oncology, Fujian Medical University Union Hospital, Fuzhou, Fujian Province 350001, China. [12]Fujian Provincial Key Laboratory for Advanced Micro-Nano Photonics Technology and Devices, Institute for Photonics Technology, Quanzhou Normal University, Quanzhou 362000, China. [13]Department of Clinical Medicine, Fujian Medical University, Fuzhou, Fujian 350122, China.

Received: 23 August 2024 Accepted: 19 January 2025
Published online: 21 February 2025

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71(3):209–49.
2. Crosby D, Bhatia S, Brindle KM, Coussens LM, Dive C, Emberton M, et al. Early detection of cancer. Science. 2022;375(6586):eaay9040.
3. Bleyer A, Welch HG. Effect of three decades of screening mammography on breast-cancer incidence. N Engl J Med. 2012;367(21):1998–2005.
4. Nishihara R, Wu K, Lochhead P, Morikawa T, Liao X, Qian ZR, et al. Long-term colorectal-cancer incidence and mortality after lower endoscopy. N Engl J Med. 2013;369(12):1095–105.
5. Huang KL, Wang SY, Lu WC, Chang YH, Su J, Lu YT. Effects of low-dose computed tomography on lung cancer screening: a systematic review, meta-analysis, and trial sequential analysis. BMC Pulm Med. 2019;19(1):126.
6. Cristiano S, Leal A, Phallen J, Fiksel J, Adleff V, Bruhm DC, et al. Genome-wide cell-free DNA fragmentation in patients with cancer. Nature. 2019;570(7761):385–9.
7. Heitzer E, Haque IS, Roberts CES, Speicher MR. Current and future perspectives of liquid biopsies in genomics-driven oncology. Nat Rev Genet. 2019;20(2):71–88.
8. Tan HT, Lee YH, Chung MC. Cancer proteomics. Mass Spectrom Rev. 2012;31(5):583–605.
9. Srivastava A, Creek DJ. Discovery and validation of clinical biomarkers of cancer: a review combining metabolomics and proteomics. Proteomics. 2019;19(10):e1700448.
10. Jalalian SH, Ramezani M, Jalalian SA, Abnous K, Taghdisi SM. Exosomes, new biomarkers in early cancer detection. Anal Biochem. 2019;571:1–13.
11. Kong K, Kendall C, Stone N, Notingher I. Raman spectroscopy for medical diagnostics–from in-vitro biofluid assays to in-vivo cancer detection. Adv Drug Deliv Rev. 2015;89:121–34.
12. Krafft C, Popp J. The many facets of Raman spectroscopy for biomedical analysis. Anal Bioanal Chem. 2015;407(3):699–717.
13. Gao N, Wang Q, Tang J, Yao S, Li H, Yue X, et al. Non-invasive SERS serum detection technology combined with multivariate statistical algorithm for simultaneous screening of cervical cancer and breast cancer. Anal Bioanal Chem. 2021;413(19):4775–84.
14. Nargis HF, Nawaz H, Bhatti HN, Jilani K, Saleem M. Comparison of surface enhanced Raman spectroscopy and Raman spectroscopy for the detection of breast cancer based on serum samples. Spectrochim Acta A Mol Biomol Spectrosc. 2021;246:119034.
15. Cervo S, Mansutti E, Del Mistro G, Spizzo R, Colombatti A, Steffan A, et al. SERS analysis of serum for detection of early and locally advanced breast cancer. Anal Bioanal Chem. 2015;407(24):7503–9.
16. Vargas-Obieta E, Martinez-Espinosa JC, Martinez-Zerega BE, Jave-Suarez LF, Aguilar-Lemarroy A, Gonzalez-Solis JL. Breast cancer detection based on serum sample surface enhanced Raman spectroscopy. Lasers Med Sci. 2016;31(7):1317–24.
17. Yang Y, Gao X, Zhang H, Chao F, Jiang H, Huang J, et al. Multi-scale representation of surface-enhanced Raman spectroscopy data for deep learning-based liver cancer detection. Spectrochim Acta A Mol Biomol Spectrosc. 2024;308:123764.
18. Ou Q, Jiang L, Dou Y, Yang W, Han M, Ni Q, et al. Application of surface-enhanced Raman spectroscopy to human serum for diagnosing liver cancer. Spectrochim Acta A Mol Biomol Spectrosc. 2024;308:123702.
19. Cai C, Liu Y, Li J, Wang L, Zhang K. Serum fingerprinting by slippery liquid-infused porous SERS for non-invasive lung cancer detection. Analyst. 2022;147(20):4426–32.
20. Lei J, Yang D, Li R, Dai Z, Zhang C, Yu Z, et al. Label-free surface-enhanced Raman spectroscopy for diagnosis and analysis of serum samples with different types lung cancer. Spectrochim Acta A Mol Biomol Spectrosc. 2021;261:120021.
21. Peng S, Lu D, Zhang B, You R, Chen J, Xu H, et al. Machine learning-assisted internal standard calibration label-free SERS strategy for colon cancer detection. Anal Bioanal Chem. 2023;415(9):1699–707.
22. Stefancu A, Moisoiu V, Couti R, Andras I, Rahota R, Crisan D, et al. Combining SERS analysis of serum with PSA levels for improving the detection of prostate cancer. Nanomedicine (Lond). 2018;13(19):2455–67.
23. Nicolae, Bernhar L. A new method for fast preparation of highly surface-enhanced Raman scattering (SERS) active silver colloids at room temperature by reduction of silver nitrate with hydroxylamine hydrochloride. J Phys Chem B. 2003;107(24):5723–7.
24. Zhao J, Lui H, McLean DI, Zeng H. Automated autofluorescence background subtraction algorithm for biomedical Raman spectroscopy. Appl Spectrosc. 2007;61(11):1225–32.
25. Pan C, Peng K, Chen T, Chen G, Lin Y, Zhang Q, et al. Power-law-based synthetic minority oversampling technique on imbalanced serum surface-enhanced raman spectroscopy data for cancer screening. Adv Intell Systems. 2023;5(7):2300006.
26. Novis BH, Gluck E, Thomas P, Steele GD, Zurawski VR Jr, Stewart R, et al. Serial levels of CA 19–9 and CEA in colonic cancer. J Clin Oncol. 1986;4(6):987–93.
27. Scholler N, Urban N. CA125 in ovarian cancer. Biomark Med. 2007;1(4):513–23.
28. Luo G, Jin K, Deng S, Cheng H, Fan Z, Gong Y, et al. Roles of CA19-9 in pancreatic cancer: biomarker, predictor and promoter. Biochim Biophys Acta Rev Cancer. 2021;1875(2):188409.
29. Daniele B, Bencivenga A, Megna AS, Tinessa V. Alpha-fetoprotein and ultrasonography screening for hepatocellular carcinoma. Gastroenterology. 2004;127(5 Suppl 1):S108-12.
30. Hanif H, Ali MJ, Susheela AT, Khan IW, Luna-Cuadros MA, Khan MM, et al. Update on the applications and limitations of alpha-fetoprotein for hepatocellular carcinoma. World J Gastroenterol. 2022;28(2):216–29.
31. Bonifacio A, Cervo S, Sergo V. Label-free surface-enhanced Raman spectroscopy of biofluids: fundamental aspects and diagnostic applications. Anal Bioanal Chem. 2015;407(27):8265–77.

32. Khristoforova Y, Bratchenko L, Bratchenko I. Raman-based techniques in medical applications for diagnostic tasks: a review. Int J Mol Sci. 2023;24(21):15605.
33. Rajput S, Pink D, Findlay S, Woolner E, Lewis JD, McDermott MT. Application of surface-enhanced Raman spectroscopy to guide therapy for advanced prostate cancer patients. ACS Sens. 2022;7(3):827–38.
34. Moisoiu V, Stefancu A, Gulei D, Boitor R, Magdo L, Raduly L, et al. SERS-based differential diagnosis between multiple solid malignancies: breast, colorectal, lung, ovarian and oral cancer. Int J Nanomedicine. 2019;14:6165–78.
35. Song H, Leng H, Hou Z, Gao R, Chen C, Meng C, et al. Grouped-sampling technique to deal with unbalance in Raman spectral data modeling. Photodiagnosis Photodyn Ther. 2022;40:103059.
36. Chen C, Wu X, Zuo E, Chen C, Lv X, Wu L. R-GDORUS technology: effectively solving the Raman spectral data imbalance in medical diagnosis. Chemom Intell Lab Syst. 2023;235:104762.
37. Wang Y, Qian H, Shao X, Zhang H, Liu S, Pan J, et al. Multimodal convolutional neural networks based on the Raman spectra of serum and clinical features for the early diagnosis of prostate cancer. Spectrochim Acta A Mol Biomol Spectrosc. 2023;293:122426.
38. Qian H, Shao X, Zhang H, Wang Y, Liu S, Pan J, et al. Diagnosis of urogenital cancer combining deep learning algorithms and surface-enhanced Raman spectroscopy based on small extracellular vesicles. Spectrochim Acta A Mol Biomol Spectrosc. 2022;281:121603.
39. Shao X, Zhang H, Wang Y, Qian H, Zhu Y, Dong B, et al. Deep convolutional neural networks combine Raman spectral signature of serum for prostate cancer bone metastases screening. Nanomedicine. 2020;29:102245.
40. Xiong CC, Zhu SS, Yan DH, Yao YD, Zhang Z, Zhang GJ, et al. Rapid and precise detection of cancers via label-free SERS and deep learning. Anal Bioanal Chem. 2023;415(17):3449–62.
41. Li JQ, Dukes PV, Lee W, Sarkis M, Vo-Dinh T. Machine learning using convolutional neural networks for SERS analysis of biomarkers in medical diagnostics. J Raman Spectrosc. 2022;53(12):2044–57.

Publisher's Note