

RESEARCH

Open Access



Rapid detection of lung cancer based on serum Raman spectroscopy and a support vector machine: a case-control study

Linfang Yan¹, Huiting Su^{1*}, Jiafei Liu¹, Xiaozheng Wen¹, Huaichao Luo², Yu Yin³ and Xiaoqiang Guo¹

Abstract

Background Early screening and detection of lung cancer is essential for the diagnosis and prognosis of the disease. In this paper, we investigated the feasibility of serum Raman spectroscopy for rapid lung cancer screening.

Methods Raman spectra were collected from 45 patients with lung cancer, 45 with benign lung lesions, and 45 healthy volunteers. And then the support vector machine (SVM) algorithm was applied to build a diagnostic model for lung cancer. Furthermore, 15 independent individuals were sampled for external validation, including 5 lung cancer patients, 5 benign lung lesion patients, and 5 healthy controls.

Results The diagnostic sensitivity, specificity, and accuracy were 91.67%, 92.22%, 90.56% (lung cancer vs. healthy control), 92.22%, 95.56%, 93.33% (benign lung lesion vs. healthy) and 80.00%, 83.33%, 80.83% (lung cancer vs. benign lung lesion), respectively. In the independent validation cohort, our model showed that all the samples were classified correctly.

Conclusion Therefore, this study demonstrates that the serum Raman spectroscopy analysis technique combined with the SVM algorithm has great potential for the noninvasive detection of lung cancer.

Keywords Raman spectroscopy, Lung cancer, Serum, SVM

Background

Lung cancer has the highest incidence rate and mortality among all malignant tumors [1]. As the stage of lung cancer progresses, the five-year survival rate gradually decreases [2]. In stage IA groups, the rates exceed 90%, while in stage IV groups, the rates are less than 10% [2].

Therefore, the early diagnosis of lung cancer is of great significance and can considerably improve lung cancer patients' therapeutic effects and prognosis. In the lung cancer screening guidelines, low-dose spiral computed tomography (LDCT) is recommended for people with high-risk factors worldwide [3]. However, in addition to radiation injury, the high false positive rate of LDCT will lead to unnecessary invasive examination and over-diagnosis [4, 5]. Pathological examination is the gold standard for lung cancer diagnosis. This technique commonly requires the help of surgical approaches such as fiber bronchoscopy, image-guided trans-thoracic needle aspiration, and thoracoscopy [6]. Additionally, these procedures are costly, prone to complications, and there is a possible need for enough samples [6]. Still, it is unsuitable

*Correspondence:

Huiting Su
huitingsu@yeah.net

¹Guang'an People's Hospital, Guang'an, Sichuan Province, China

²Sichuan Cancer Hospital & Institute, Sichuan Cancer Center, Chengdu, China

³Sichuan Institute for Brain Science and Brain-Inspired Intelligence, MOE Key Lab for Neuroinformation, University of Electronic Science and Technology of China, Chengdu, China



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

for early lung cancer diagnosis due to the inescapable invasiveness and harsh application conditions. Routine serum biomarkers are unsatisfactory due to their lower sensitivity or specificity [7–9]. Liquid biopsy is not widely used in clinical practice due to the uncertainty threshold and high testing costs [10]. Thus, a more convenient and noninvasive diagnosis tool with high sensitivity and specificity is needed to reduce mortality rates and burdens on the medical system.

Raman spectroscopy is a well-established analytical technique based on the inelastic scattering generated by rotational and vibrational modes of molecular bonds [11]. Compared with routine diagnostic methods, Raman spectroscopic techniques have the advantages of being fast, accurate, and non-destructive [12]. Serum is mainly composed of water, carbohydrates, proteins, phospholipids and polysaccharides, showing a unique Raman fingerprint profile. Metabolism of malignant cells affect the composition and content change of serum [13, 14]. Raman-based methods can effectively detect minor changes that occur during cancer development [15]. Meanwhile, serum samples are easier to obtain and the Raman detection system has the advantages of no sample preparation and non-contact measurements. Effectiveness and safety of this detection method in the previous study about the COVID-19 has been confirmed [16]. Furthermore, a review about the potential of Raman spectroscopy to analyze liquid plasma/serum shows that liquid form has potential advantages over the infrared absorption analysis of dry droplet form and will prove to be highly beneficial to clinicians for rapid screening in the future [17]. More studies have found that Raman spectroscopy has diagnostic potential in a variety of tumors [18], while few original clinical trials utilized serum Raman spectroscopy to diagnose lung cancer. Therefore, investigations on the performance of serum Raman spectroscopy in diagnosing lung cancer are of great significance.

Due to the complexity and heterogeneity of Raman spectrum data, machine learning methods are necessary for deep data mining. SVM is a machine learning algorithm that classifies data based on supervised learning, particularly suitable for small sample problems and high latitude pattern recognition [19, 20]. SVM is an effective classifier because it can be used for both linearly separable and linearly inseparable data sets [21]. Additionally, the SVM algorithm is applied most frequently in classification and prediction methods with high accuracy for disease risk prediction [22]. Notably, the combination of SVM and Raman spectroscopy has previously been used to distinguish patients with hystero myoma and cervical cancer from healthy controls and the results were satisfactory [23].

In this study, we detected the serum from lung cancer patients, benign lung lesion patients, and healthy controls using Raman spectroscopy to explore the screening value of Raman spectroscopy. Furthermore, a support vector machine (SVM) was used for model building and training.

Methods

Patients

A total of 90 patients with lung-occupying lesions were recruited in this project. Patients were selected and enrolled upon confirmation of clinical or pathological diagnosis. All enrolled participants at the time of pathological or clinical diagnosis of benign lung lesion or lung cancer reported no history of malignancy or prior treatment, such as chemotherapy or radiotherapy. The lung cancer group included patients with a pathological diagnosis of lung cancer. The stages were determined in accordance with the 8th edition of tumor node metastasis (TNM) classification [24] for lung cancer, with each lesion being individually staged and the final stage being based on the highest stage. In contrast, the benign lung lesion group included patients diagnosed with inflammatory or granulomatous changes in pathology. Thus, the lung cancer and benign lung lesion groups obtained 45 patients, respectively. The healthy control group also consisted of 45 healthy individuals recruited from the medical examination center of the participating institutes.

This study was approved by the Medical Ethics Committee of Guang'an People's Hospital (approval number: 2,022,007), Sichuan Province, China, and conducted following the principles of the Declaration of Helsinki. All the enrolled individuals signed the written informed consent.

Sample preparation

Fasting venous blood samples were collected from all participants before treatment. The serum was isolated from blood samples by centrifuging for 10 min at 3000 rpm. All the serum samples were strictly sealed in cryopreservation tubes and stored at -80 °C until being scanned. For the measurement, approximately 0.5 ml of the serum sample was prepared in cryopreservation tubes made of polypropylene with a specification of 2 ml. All blood samples were collected from March to December 2022 in this study.

Raman measurements

The Raman system is designed by the Sichuan Institute for Brain Science and Brain-Inspired Intelligence, which consists of a volume-phase holographic spectrograph (F/2ctroEMvision), deep-cooled CCD camera (at -60 °C, Andor iVac DR-316B-LDC-DD), Raman probe, and laser. The sample end uses a microscope objective (50X, NA

0.5, WD 8.0, Sunnyoptical) as a focusing lens. An internal laser line filter (Semrock, LL01-785-12.5) was applied to obtain a clean laser profile. A single-mode diode laser with wavelength 785 nm and 100 mW power was used for Raman excitation. The laser power on the sample was detected to be around 70mW. Furthermore, the spectra were recorded in the 400–1800 cm⁻¹ range. The detection process was repeated 3 times and acquired 5 spectra each time, and 15 Raman spectra were collected in total from each serum sample.

First, the ethanol spectrum was measured using an exposure time of 3s for the wavenumber calibration. Second, the cryopreservation tubes with a 5% normal saline spectrum were acquired using an exposure time of 3 s with five successive scans for every beginning and completion of the experiment. The average spectrum of the cryopreservation tube spectra was used for background subtraction. Next, the Raman spectrum of the serum samples sealed within the cryopreservation tube was measured using the same integration parameters as the cryopreservation tube measurements. Three experimenters took the Raman scan for each sample tube and repeated it five times. Following cosmic ray removal from the spectral data, we had 15 scans per serum sample. Furthermore, the cryopreservation tube was placed in the specific card slot of the Raman spectrometer, ensuring that the laser passed through the tube wall at a certain angle.

Date processing and SVM classification

Raman spectral data preprocessing steps include smoothing by automatic-weighted least squares, baseline correction based on polynomial fitting, and normalization by total area. A total of 1800 spectra from 120 individuals were preprocessed for model building. The ANOVA statistical test method was used to select relevant features. Additionally, the points that passed the ANOVA statistical test between the two groups were used as input features of the SVM. Our SVM algorithm used a non-linear radial basis function (RBF).

In this study, we used a two-level cross-validation approach. The model building data set were divided into two groups, containing 70% and 30% of the data for training and testing groups, respectively. The training and cross-validation data sets were separated by randomly selecting 70% of the total data. The remaining 30% of the data was used as unseen data to assess the predictive power of the classification models. The process mentioned above was repeated 50 times.

We externally validated the predictive model after it was built. In total, 15 serum samples were selected for verification, including 5 lung cancer patients, 5 patients with benign lung lesions, and 5 healthy controls. These samples are used as independent external datasets, and their spectra were preprocessed the same way as those used to build the model. Using an SVM model, each unlabeled spectrum was assigned to the class with the highest probability. The true classification of the samples was not revealed until after the model had made its predictions, allowing for an unbiased evaluation of the model's performance. Finally, the receiver operating curve (ROC) was used to check the diagnostic performance of the model. MATLAB was used for the preprocessing of Raman spectrum data and the calculation of SVM and ROC.

Results

Clinical characteristics

Data regarding the age, sex, pathological results, and clinical stages of the participants are shown in Table 1. There is no statistically significant difference between these three groups regarding gender and age. The lung cancer group consisted of 20 cases of adenocarcinoma, 14 cases of squamous cell carcinoma, 8 cases of small cell carcinoma, 1 case of neuroendocrine tumor, 1 case of adenosquamous cell carcinoma, and 1 case of undifferentiated carcinoma. Moreso, the benign lung lesion group included 16 cases of pneumonia, 14 cases of chronic obstructive pulmonary diseases, 11 cases of tuberculosis, 2 cases of benign lung tumor, 1 case of interstitial lung disease, and 1 case of fungal pneumonia.

Table 1 Clinical characteristics

	Lung cancer <i>n</i> = 45	Benign lung lesion <i>n</i> = 45	Healthy control <i>n</i> = 45
Age, y			
Min-max(median)	51–85(67)	55–84(67)	56–76(65)
Sex			
Male	35	29	27
Female	10	16	18
Pathological findings			
Adenocarcinoma	20		
Squamous cell carcinoma	14		
Small cell carcinoma	8		
Neuroendocrine tumor	1		
Adenosquamous carcinoma	1		
Undifferentiated carcinoma	1		
TNM Stage			
I	12		
II	9		
III	8		
IV	16		

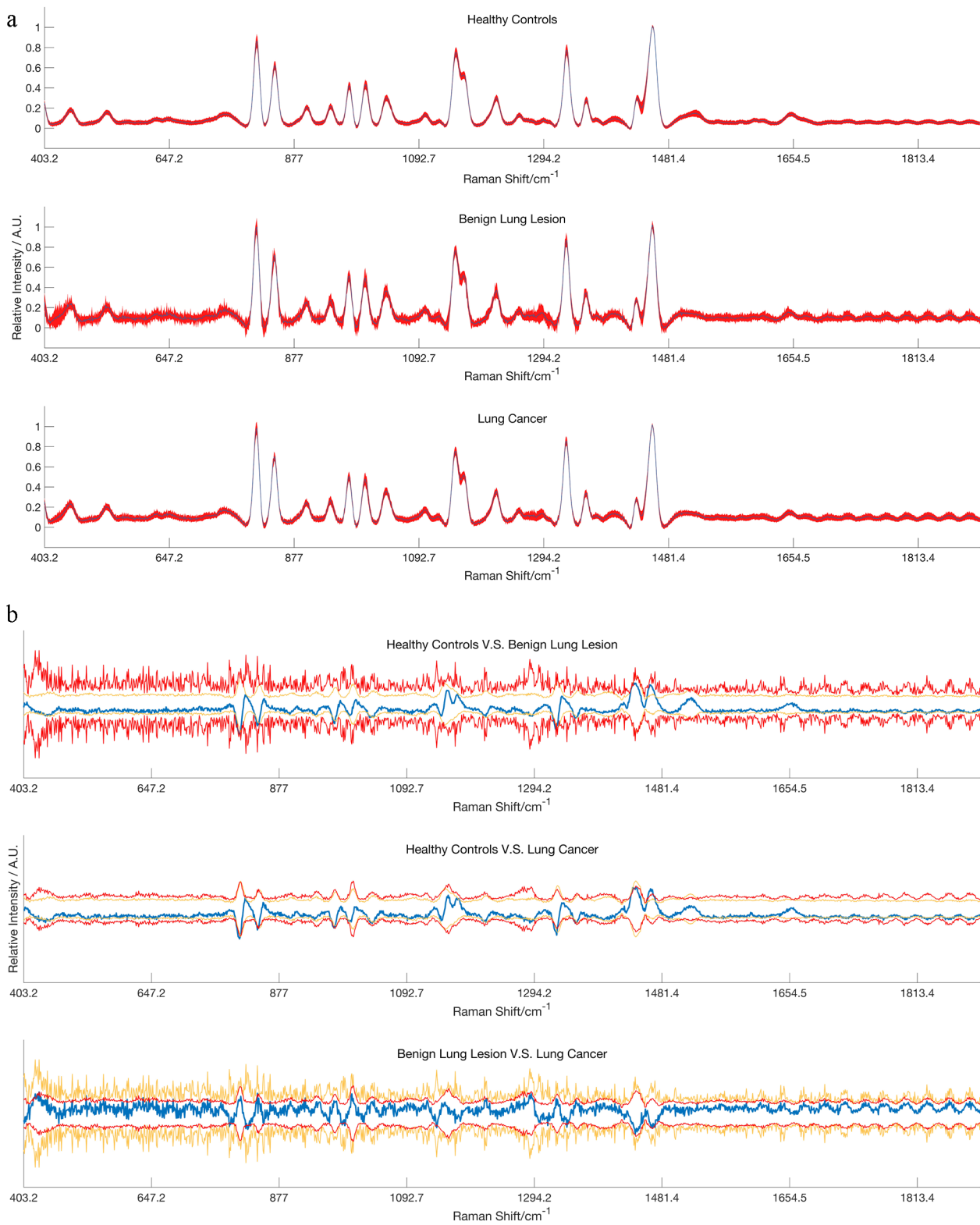


Fig. 1 The total average serum Raman of the three groups and the difference between the groups. **(a)** The total average of the three types of Raman, the color band represents the standard deviation. **(b)** The Raman difference signal between the groups (blue) and the Raman signal of the groups between ± 2 standard deviations (red and yellow)

Raman Spectra and statistical analysis

The mean preprocessed spectra of the three groups are shown in Fig. 1a. The differences between healthy controls vs. benign lung lesions, healthy controls vs. lung cancer, and benign lung lesions vs. lung cancer are depicted in Fig. 1b. The difference in the mean spectrum is shown within ± 2 standard deviations, suggesting that the mean difference between the groups is statistically insignificant. It is necessary to exploit the difference that may exist through deep learning algorithms.

The ANOVA test selected relevant features between the three experimental groups. The sample selection process

was random and needed to be repeated 100 times. Only the statistical significance of the ANOVA test over 70 times out of 100 points was selected as the feature. High inter-group consistency was shown after the ANOVA analysis, while the differences in the intra-group were random. The difference between the three compared groups was in the spectra range of 400–1800 cm^{-1} and is observed in Fig. 2a. The differences between lung cancer vs. benign lung lesions shown are significantly less than those two compared groups. However, the ANOVA test result showed no apparent consistency for the inter-group (Fig. 2b).

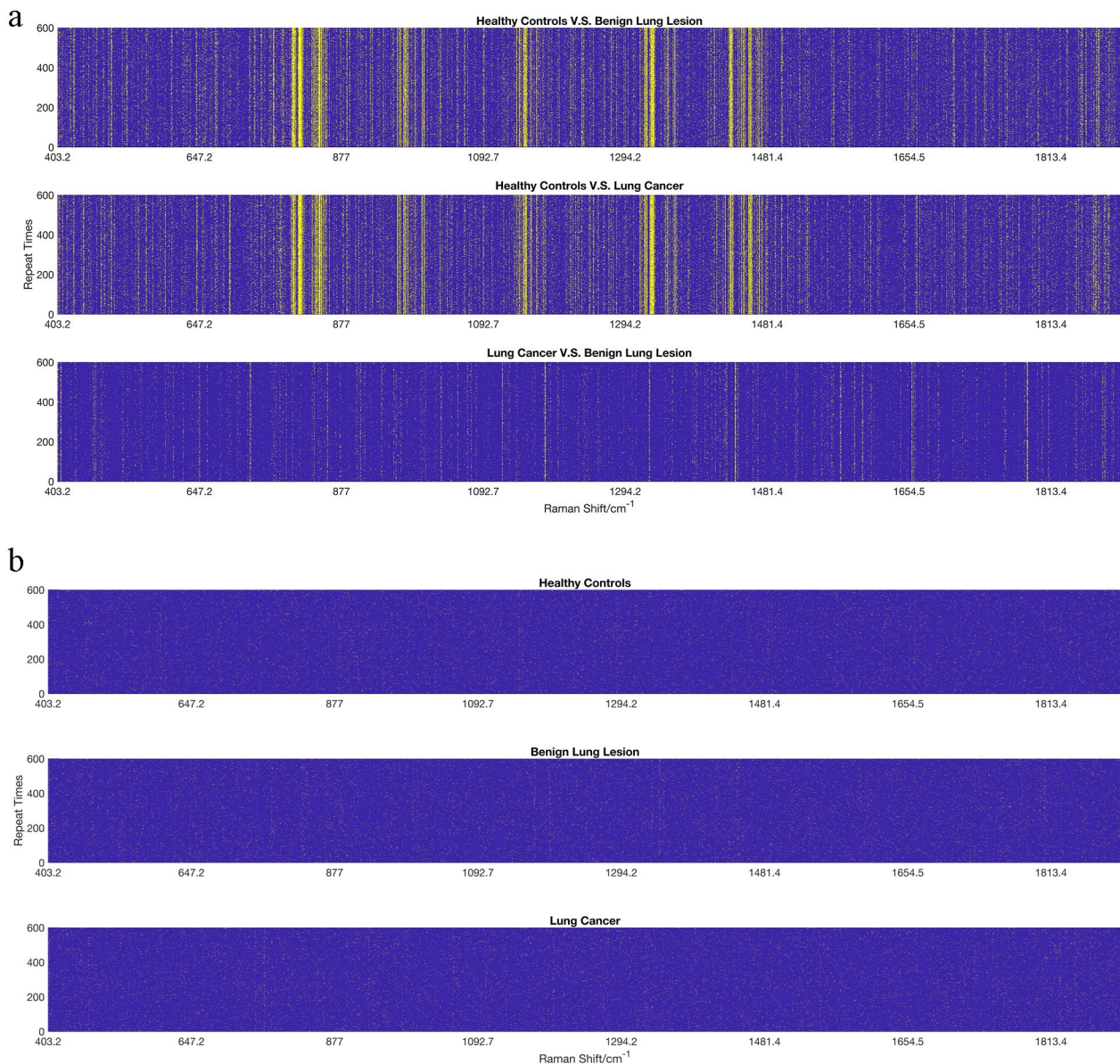


Fig. 2 The result of the ANOVA test. The spectra range without a significant difference in the ANOVA test ($p < 0.05$) was indicated in blue, while others were indicated in yellow. **(a)** The Raman shift spectrum of the difference in the p -value for the inter-group. **(b)** The Raman shift spectrum of the difference in the p -value for the intra-group

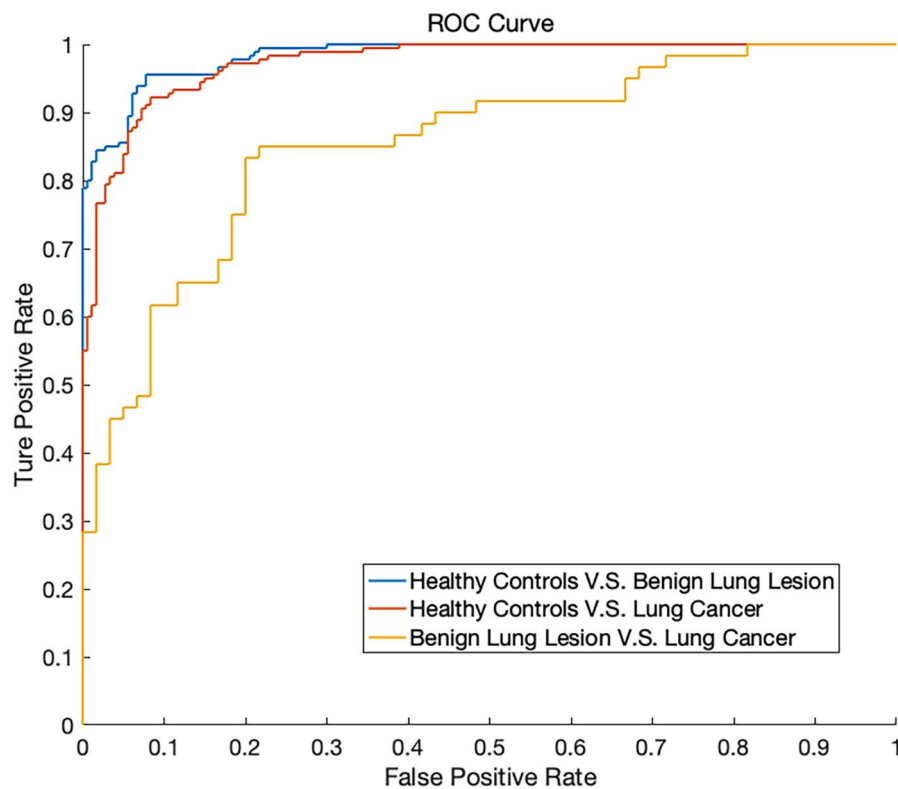


Fig. 3 The ROC curve of the SVM diagnostic algorithm for the healthy controls group vs. benign lung lesion group, healthy controls group vs. lung cancer group, and benign lung lesion group vs. lung cancer group

Table 2 Performance parameters of the SVM

Class	Value \pm SD			
	AUC	Sensitivity	Specificity	Accuracy
healthy controls vs. benign lung lesions	0.984 \pm 0.003	0.922 \pm 0.004	0.956 \pm 0.002	0.933 \pm 0.006
healthy controls vs. lung cancer	0.974 \pm 0.003	0.917 \pm 0.007	0.922 \pm 0.006	0.906 \pm 0.005
lung cancer vs. benign lung lesions	0.853 \pm 0.006	0.800 \pm 0.007	0.833 \pm 0.005	0.808 \pm 0.011

Serum Raman spectroscopy combined with support vector machine algorithm shows great diagnostic ability in lung cancer screening. The performance of the classifiers are evaluated by the ROC curve and shown in Fig. 3. All ROC analyses are based on nonparametric techniques and are conducted for the SVM analyses. For each of the three classification tasks, the area under the curve (AUC) value and the results of AUC, specificity, accuracy, and sensitivity of the SVM classification are calculated and shown in Table 2.

The overall serum-level classification of each of the 15 serum samples is shown in Table 3. A serum sample was assigned to the class that received the majority of spectra assigned to it. For the independent test, our model showed that all the samples were classified correctly.

Discussion

Raman spectroscopy measurement is an increasingly popular method of diagnosing cancer [18]. Recently, many studies have shown that Raman spectroscopy is a high-accuracy method for diagnosing lung cancer [25–30]. However, most studies mainly carried out Raman detection on tissues to screen for lung cancer [27–30]. Raman detection using tissue is not as convenient as blood detection in general physical examination. Notably, serum detection could be a more favorable and noninvasive method than tissue. Once lung cancer screening can be carried out through blood testing, early lung cancer screening can be realized in a general physical examination which is incomparable with tissue testing [17].

Our study observed significant differences between the average Raman spectrum of lung cancer patients and healthy controls. Meanwhile, the classification model of lung cancer patients and healthy controls show excellent discrimination ability with AUC values of 0.973, and the

Table 3 Results of 15 samples for the independent verification

Individual spectra predictions				External validation results		
Sample #	Number of prediction			Sample #	Predicted class	True class
	Healthy control	Benign lung lesion	Lung Cancer			
1	990	307	203	1	Healthy Control	Healthy Control
2	985	375	140	2	Healthy Control	Healthy Control
3	989	313	198	3	Healthy Control	Healthy Control
4	994	200	306	4	Healthy Control	Healthy Control
5	997	300	203	5	Healthy Control	Healthy Control
6	237	911	352	6	Benign lung lesion	Benign lung lesion
7	319	949	232	7	Benign lung lesion	Benign lung lesion
8	212	982	306	8	Benign lung lesion	Benign lung lesion
9	345	986	169	9	Benign lung lesion	Benign lung lesion
10	299	982	219	10	Benign lung lesion	Benign lung lesion
11	98	695	707	11	Lung Cancer	Lung Cancer
12	235	294	971	12	Lung Cancer	Lung Cancer
13	208	316	976	13	Lung Cancer	Lung Cancer
14	276	266	958	14	Lung Cancer	Lung Cancer
15	203	488	809	15	Lung Cancer	Lung Cancer

sensitivity and specificity were 0.917 and 0.922, respectively. Similar conclusions of serum samples detected by Raman spectroscopy were also produced in Shin et al. [31]. and Moisiou et al. [32]. studies in which diagnostic sensitivity and specificity in lung cancer were 0.84 (95% CI 0.69–0.93), 0.85 (95% CI 0.62–0.97) and 0.85 (95% CI 0.68–0.95), 0.87 (95% CI 0.73–0.96), respectively. Moreso, Lei et al. [33]. used surface enhanced Raman spectroscopy (SERS) combined with principal component analysis (PCA) and partial least-squares discriminant analysis (PLS-DA) to diagnose and distinguish lung cancer and normal serum. Importantly, the model's sensitivity improved to 100%, while the specificity decreased to 83.33% [29]. Compared with these similar studies, our classification model seems more excellent. From the studies of Ke et al. [34]. and Chen et al. [35]., the results of tissue samples detected by Raman spectroscopy were more reliable than serum samples. Tissue samples are certainly better, but they cannot be used for early screening of lung cancer because they are difficult to obtain. Therefore, if the pathology sample was unavailable, serum detection could be more favorable and noninvasive.

Different from previous studies, our study included a benign lung lesion group, which is also one of our innovative points. The majority of patients in the benign lung lesion group were diagnosed with infectious inflammation. There is a certain similarity between serum metabolites in cancers and inflammatory diseases [13, 14, 36]. Furthermore, most patients in our lung cancer groups are usually not challenged with a single disease, and they are often concurrent with chronic lung inflammatory disease. Therefore, the dual factors increase the classification difficulty of our model of lung cancer patients and benign lung lesion individuals. This result may be why

our model's diagnosis accuracy, sensitivity, and specificity were only 0.808, 0.800, and 0.833, respectively. From the lung cancer and benign lung lesion classification model, the area under the curve (AUC) value was also obviously lower than that of the other two classification models, only 0.853. Nevertheless, our results are meaningful and provide a reference for differentiating benign and malignant lung diseases. Takamori et al. analyzed salivary metabolites and built a multiple logistic regression (MLR) models for discriminating patients with lung cancer from benign lung lesions (AUC=0.729, 95%CI=0.598–0.861, $p=0.003$) [37]. Compared with this consequence, our research shows a more robust diagnostic ability. Besides the use of cancer diagnosis, Raman spectroscopy has also been applied in many studies regarding inflammatory diseases such as dengue fever [38], malaria infection [39], virus infection [40], cryptococcal infection [41], ulcerative colitis [42], and cervicitis [43]. Our results are consistent with these studies and shows more excellent diagnostic ability. Therefore, our model has a good clinical practicability.

Conclusions

This study is based on a label-free serum detection of the Raman spectrum and combined with machine learning methods to realize the rapid diagnosis of lung cancer. We used the SVM to establish the two-way (lung cancer vs. healthy control, benign lung lesion vs. healthy, and lung cancer vs. benign lung lesion) classification model. Notably, all three models demonstrated an outstanding differentiation ability. Therefore, these findings indicate that the serum Raman spectroscopy combined with a support vector machine model can be used as a standard pre-screening tool for lung cancer.

This study also has limitation. The sample size was small, and well-powered large-scale multicenter studies are needed to verify this conclusion in the future. Besides, prospective early screening clinical trial design is also needed.

Abbreviations

SVM	Support vector machine
LDCT	Low-dose spiral computed tomography
TNM	Tumor node metastasis
ROC	Receiver operating curve
AUC	area under the curve
RBF	Non-linear radial basis function
SERS	Surface-enhanced Raman spectroscopy
PCA	Principal component analysis
PLS-DA	Partial least-squares discriminant analysis

Acknowledgments

The authors thank AiMi Academic Services (www.aimieditor.com) for English language editing and review services.

Author contributions

Huiting Su conceived and designed the paper. Xiaozheng Wen and Xiaoqiang Guo recruited participants and collected the serum samples. Jiafei Liu collected and analyzed the clinical data. Huaichao Luo detected the samples. Yu Yin analyzed the data and provided technical support. Linfang Yan wrote the initial draft of the manuscript. All authors contributed to the editing, review of the manuscript and approved the final manuscript.

Funding

This study was funded by Guang'an Science and Technology Bureau (NO.2021SYF01).

Data availability

The datasets used and analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

This study was approved by the Medical Ethics Committee of Guang'an People's Hospital (approval number: 2022007), Sichuan Province, China. All participants provided written informed consent.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 25 May 2023 / Accepted: 28 June 2024

Published online: 02 July 2024

References

- Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin.* 2022;72(1):7–33.
- Goldstraw P, Chansky K, Crowley J, Rami-Porta R, Asamura H, Eberhardt WE, et al. The IASLC Lung Cancer Staging Project: proposals for revision of the TNM Stage groupings in the Forthcoming (Eighth) Edition of the TNM classification for Lung Cancer. *J Thorac Oncol.* 2016;11(1):39–51.
- Mazzone PJ, Silvestri GA, Souter LH, Caverly TJ, Kanne JP, Katki HA, et al. Screening for Lung Cancer: CHEST Guideline and Expert Panel Report. *Chest.* 2021;160(5):e427–94.
- Shen H. Low-dose CT for lung cancer screening: opportunities and challenges. *Front Med.* 2018;12(1):116–21.
- Patz EF Jr, Pinsky P, Gatsonis C, Sicks JD, Kramer BS, Tammemägi MC, et al. Overdiagnosis in low-dose computed tomography screening for lung cancer. *JAMA Intern Med.* 2014;174(2):269–74.
- Hirsch FR, Prindiville SA, Miller YE, Franklin WA, Dempsey EC, Murphy JR, et al. Fluorescence versus white-light bronchoscopy for detection of preneoplastic lesions: a randomized study. *J Natl Cancer Inst.* 2001;93(18):1385–91.
- Grenier J, Pujol JL, Guilleux F, Daures JP, Pujol H, Michel FB. Cyfra 21–1, a new marker of lung cancer. *Nucl Med Biol.* 1994;21(3):471–6.
- Okamura K, Takayama K, Izumi M, Harada T, Furuyama K, Nakanishi Y. Diagnostic value of CEA and CYFRA 21–1 tumor markers in primary lung cancer. *Lung Cancer.* 2013;80(1):45–9.
- Muraki M, Tohda Y, Iwanaga T, Uejima H, Nagasaka Y, Nakajima S. Assessment of serum CYFRA 21–1 in lung cancer. *Cancer.* 1996;77(7):1274–7.
- Li W, Liu JB, Hou LK, Yu F, Zhang J, Wu W, et al. Liquid biopsy in lung cancer: significance in diagnostics, prediction, and treatment monitoring. *Mol Cancer.* 2022;21(1):25.
- Ember KJI, Hoeve MA, McAughtrie SL, Bergholt MS, Dwyer BJ, Stevens MM, et al. Raman spectroscopy and regenerative medicine: a review. *NPJ Regen Med.* 2017;2:12.
- Wang N, Cao H, Wang L, Ren F, Zeng Q, Xu X, et al. Recent advances in spontaneous Raman Spectroscopic Imaging: Instrumentation and Applications. *Curr Med Chem.* 2020;27(36):6188–207.
- Ahmed N, Kidane B, Wang L, Nugent Z, Moldovan N, McElrea A, et al. Metabolic changes in early-stage Non-small Cell Lung Cancer patients after Surgical Resection. *Cancers (Basel).* 2021;13(12):3012.
- You L, Fan Y, Liu X, Shao S, Guo L, Noreldeen HAA, et al. Liquid Chromatography-Mass Spectrometry-based tissue metabolic profiling reveals major metabolic pathway alterations and potential biomarkers of Lung Cancer. *J Proteome Res.* 2020;19(9):3750–60.
- Zheng Q, Li J, Yang L, Zheng B, Wang J, Lv N, Luo J, Martin FL, Liu D, He J. Raman spectroscopy as a potential diagnostic tool to analyse biochemical alterations in lung cancer. *Analyst.* 2020;145(2):385–92.
- Yin G, Li L, Lu S, Yin Y, Su Y, Zeng Y, Luo M, Ma M, Zhou H, Orlandini L, Yao D, Liu G, Lang J. An efficient primary screening of COVID-19 by serum Raman spectroscopy. *J Raman Spectrosc.* 2021;52(5):949–58.
- Parachalil DR, McIntyre J, Byrne HJ. Potential of Raman spectroscopy for the analysis of plasma/serum in the liquid state: recent advances. *Anal Bioanal Chem.* 2020;412(9):1993–2007.
- Auner GW, Koya SK, Huang C, Broadbent B, Trexler M, Auner Z, et al. Applications of Raman spectroscopy in cancer diagnosis. *Cancer Metastasis Rev.* 2018;37(4):691–717.
- Gao L, Ye M, Lu X, Huang D. Hybrid method based on Information Gain and Support Vector Machine for Gene Selection in Cancer classification. *Genomics Proteom Bioinf.* 2017;15(6):389–95.
- Ozer ME, Sarica PO, Arga KY. New machine learning applications to accelerate personalized medicine in breast Cancer: rise of the support Vector machines. *OMICS.* 2020;24(5):241–6.
- Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. *BMC Med Res Methodol.* 2019;19(1):64.
- Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. *BMC Med Inf Decis Mak.* 2019;19(1):281.
- Zheng X, Wu G, Wang J, Yin L, Lv X. Rapid detection of hystero myoma and cervical cancer based on serum surface-enhanced Raman spectroscopy and a support vector machine. *Biomed Opt Express.* 2022;13(4):1912–23.
- Goldstraw P, Chansky K, Crowley J, et al. The IASLC Lung Cancer Staging Project: proposals for revision of the TNM stage groupings in the forthcoming (eighth) edition of the TNM classification for lung cancer. *J Thorac Oncol.* 2016;11(1):39–51.
- Wang H, Zhang S, Wan L, Sun H, Tan J, Su Q. Screening and staging for non-small cell lung cancer by serum laser Raman spectroscopy. *Spectrochim Acta Mol Biomol Spectrosc.* 2018;201:34–8.
- Qian K, Wang Y, Hua L, Chen A, Zhang Y. New method of lung cancer detection by saliva test using surface-enhanced Raman spectroscopy. *Thorac Cancer.* 2018;9(11):1556–61.
- McGregor HC, Short MA, McWilliams A, Shaipanich T, Ionescu DN, Zhao J, et al. Real-time endoscopic raman spectroscopy for in vivo early lung cancer detection. *J Biophotonics.* 2017;10(1):98–110.
- Sinica A, Brožáková K, Brůha T, Votruba J. Raman spectroscopic discrimination of normal and cancerous lung tissues. *Spectrochim Acta Mol Biomol Spectrosc.* 2019;219:257–66.

29. Zhang K, Hao C, Huo Y, Man B, Zhang C, Yang C, et al. Label-free diagnosis of lung cancer with tissue-slice surface-enhanced Raman spectroscopy and statistical analysis. *Lasers Med Sci.* 2019;34(9):1849–55.
30. Song D, Yu F, Chen S, Chen Y, He Q, Zhang Z, et al. Raman spectroscopy combined with multivariate analysis to study the biochemical mechanism of lung cancer microwave ablation. *Biomed Opt Express.* 2020;11(2):1061–72.
31. Shin H, Oh S, Hong S, Kang M, Kang D, Ji YG, et al. Early-Stage Lung Cancer diagnosis by Deep Learning-based spectroscopic analysis of circulating exosomes. *ACS Nano.* 2020;14(5):5435–44.
32. Moisiu V, Stefanu 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 Nanomed.* 2019;14:6165–78.
33. 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 Mol Biomol Spectrosc.* 2021;261:120021.
34. Ke ZY, Ning YJ, Jiang ZF, Zhu YY, Guo J, Fan XY, et al. The efficacy of Raman spectroscopy in lung cancer diagnosis: the first diagnostic meta-analysis. *Lasers Med Sci.* 2022;37(1):425–34.
35. Chen C, Hao J, Hao X, Xu W, Xiao C, Zhang J, et al. The accuracy of Raman spectroscopy in the diagnosis of lung cancer: a systematic review and meta-analysis. *Transl Cancer Res.* 2021;10(8):3680–93.
36. Greten FR, Grivennikov SI. Inflammation and Cancer: triggers, mechanisms, and consequences. *Immunity.* 2019;51(1):27–41.
37. Takamori S, Ishikawa S, Suzuki J, Oizumi H, Uchida T, Ueda S, et al. Differential diagnosis of lung cancer and benign lung lesion using salivary metabolites: a preliminary study. *Thorac Cancer.* 2022;13(3):460–5.
38. Mahmood T, Nawaz H, Ditta A, Majeed MI, Hanif MA, Rashid N, et al. Raman spectral analysis for rapid screening of dengue infection. *Spectrochim Acta Mol Biomol Spectrosc.* 2018;200:136–42.
39. Bilal M, Saleem M, Amanat ST, Shakoor HA, Rashid R, Mahmood A, et al. Optical diagnosis of malaria infection in human plasma using Raman spectroscopy. *J Biomed Opt.* 2015;20(1):017002.
40. Moor K, Terada Y, Taketani A, Matsuyoshi H, Ohtani K, Sato H. Early detection of virus infection in live human cells using Raman spectroscopy. *J Biomed Opt.* 2018;23(9):1–7.
41. Zhu S, Li Y, Gao H, Hou G, Cui X, Chen S, et al. Identification and assessment of pulmonary *Cryptococcus neoformans* infection by blood serum surface-enhanced Raman spectroscopy. *Spectrochim Acta Mol Biomol Spectrosc.* 2021;260:119978.
42. Veenstra MA, Palyvoda O, Alahwal H, Jovanovski M, Reisner LA, King B, et al. Raman spectroscopy in the diagnosis of ulcerative colitis. *Eur J Pediatr Surg.* 2015;25(1):56–9.
43. Barik AK, M SP, Pai NM, Upadhy MV, Pai R. A micro-raman spectroscopy study of inflammatory condition of human cervix: probing of tissues and blood plasma samples. *Photodiagnosis Photodyn Ther.* 2022;39:102948.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.