

NON-INVASIVE DIAGNOSTICS OF LUNG CANCER BASED ON WHOLE BLOOD SURFACE-ENHANCED RAMAN SPECTROSCOPY AND DEEP MACHINE LEARNING

Cairou Chen ¹, Qibin Zhang ², Dechan Lu ¹, Jiatong Liu ², Yudong Lu ¹, Kunhong Liu ^{2*}

¹ College of Chemistry and Materials Science, Fujian Provincial Key Laboratory of Advanced Oriented Chemical Engineer, Fujian Key Laboratory of Polymer Materials, Fujian Normal University, Fuzhou, Fujian, China; e-mail: m17875511657@163.com

² School of Informatics, Xiamen University, Xiamen, Fujian Province, China

Combining deep machine learning with silver nanoparticle (Ag NP)-based surface-enhanced Raman spectroscopy (SERS), we have developed a novel method for whole blood analysis for cancer detection applications. The whole blood was collected from two groups: one group of patients ($n = 26$) with lung cancer and another group of healthy volunteers ($n = 45$). The logistic regression (LR), k -nearest neighbor (KNN), decision tree (DT), and random forest (RF) algorithms were employed to develop a diagnostic model using the same spectral data. The results show that the diagnostic accuracy of LR, KNN, DT, and RF models was 87, 66, 77, and 83%, respectively. LR is superior to other algorithms in the SERS spectra classification of whole blood. We therefore believe that this proposed strategy will have great clinical potential for SERS technology combined with LR and act as a complementary method for the detection of lung cancer.

Keywords: surface-enhanced Raman scattering, whole blood, lung cancer, diagnostic model.

НЕИНВАЗИВНАЯ ДИАГНОСТИКА РАКА ЛЕГКОГО НА ОСНОВЕ СПЕКТРОСКОПИИ ПОВЕРХНОСТНО-УСИЛЕННОГО КОМБИНАЦИОННОГО РАССЕЯНИЯ СВЕТА ЦЕЛЬНОЙ КРОВИ И ГЛУБОКОГО МАШИННОГО ОБУЧЕНИЯ

C. Chen ¹, Q. Zhang ², D. Lu ¹, J. Liu ², Y. Lu ¹, K. Liu ^{2*}

УДК 535.375.5:616-006

¹ Колледж химии и материаловедения Фуцзяньского педагогического университета, Фуцжоу, Фуцзянь, Китай; e-mail: m17875511657@163.com

² Школа информатики Сямэньского университета, Сямэнь, Фуцзянь, Китай

(Поступила 22 июня 2021)

На основе объединения глубокого машинного обучения с поверхностно-усиленной КР-спектроскопией (SERS) наночастиц серебра (NP Ag) разработан новый метод анализа цельной крови для обнаружения рака. Цельная кровь собрана у двух групп — больных ($n = 26$) раком легкого и здоровых добровольцев ($n = 45$). Для разработки диагностической модели с использованием одних и тех же спектральных данных применены алгоритмы логистической регрессии (LR), k -ближайшего соседа (KNN), дерева решений (DT) и случайного леса (RF). Диагностическая точность моделей LR, KNN, DT и RF составила 87, 66, 77 и 83 % соответственно. Логистическая регрессия превосходит другие алгоритмы в классификации спектров SERS цельной крови. Предложенная стратегия имеет высокий клинический потенциал для технологии SERS в сочетании с логистической регрессией и действует как дополнительный метод для обнаружения рака легких.

Ключевые слова: поверхностно-усиленное комбинационное рассеяние, цельная кровь, рак легкого, диагностическая модель.

Introduction. Lung cancer poses a serious threat to human health with high morbidity and mortality [1]. Although the incidence of lung cancer has gradually decreased, the 5-year survival rate of patients is approximately 13–15% [2]. Besides, imaging and histopathologic characteristics are considered to be the gold standard for diagnosing lung cancer, but these methods are likely to lead to the risk of false-positive and unnecessary treatments or diagnostic interventions for patients [3]. Therefore, there is an urgent need to seek more convenient diagnostic methods to carry out early diagnosis and treatment of lung cancer. Effective non-invasive early screening is a promising strategy for reducing lung cancer mortality.

Liquid biopsy has a great potential as a viable alternative diagnostic approach for cancer because of its characteristics of easily accessible samples [4], non-invasive, high reproducibility, and being unaffected by tumor heterogeneity [5]. At present, liquid biopsy is traditionally focused on various body fluids [6], such as whole blood, plasma, urine, saliva, and so on, among which plasma and serum are the most widely used in applications [7]. Yet, clinically, the acquisition of plasma or serum samples entails the addition of anticoagulants, centrifugation, and other complicated operations, destabilizing the structure and conformation of the measured biomolecules, which would severely affect the accuracy of the test, and even cause false-positive results in the end. As one of the body fluids, whole blood carries genetic and molecular information [8], including DNA, RNA, proteins, and various enzymes and amino acids from its cells of origin [9]. It is estimated that whole blood consists of a few hundred components [10], but the precise composition and content are still unknown, especially some lipids, polypeptide growth factor, hormones, etc., which poses many difficulties for current research efforts [11].

Raman scattering technology, a nondestructive and non-invasive molecular spectroscopy technology, has fingerprint characteristics, thus providing a large amount of information for molecular architectures of probe substance [12]. In effect, given this limitation of weak scattering of single and high background fluorescence [13], Raman spectroscopy is disadvantageous to practical applications. However, SERS can compensate for the deficiencies of traditional Raman; therefore, SERS has become as a great potential technique for biological detection. The technique has the unique advantages of non-invasiveness, high sensitivity, good selectivity, fast analysis speed, and little water interference, making it highly promising for clinical examination [14]. Recently, developing cancer diagnostics based on SERS has been a topic of intensive research in the field of biomedical research. SERS has been applied successfully in the diagnostics and identification of various diseases such as breast cancer [15], cervical cancer [16], colon cancer [17], and stomach cancer [18] via detecting serum, saliva, urine, and other bodily fluids [19]. Lin et al. [20] presented PCA-LDA, a machine learning algorithm, combining it with SERS technology for the diagnostics of nasopharyngeal cancer at different tumor stages. E. Vargas-Obieta et al. [21] used SERS technology to analyze the serum of breast cancer patients and healthy people and PCA-LDA to develop a diagnostic model.

At present, domestic and foreign scholars have already reached a mature stage in the application of the SERS technology in cancer, but further research is still required. On the one hand, given that the analysis of the types of cancer is not comprehensive enough, the SERS technology in whole blood has not yet been applied to some cancers with high morbidity and mortality, such as lung cancer. On the other hand, the application of SERS technology in whole blood has rarely been combined with a variety of machine learning methods to make differential diagnostics and comparison of its spectrum, so as to obtain a more suitable model for whole blood SERS spectrum diagnostics.

Not only is the whole blood SERS spectrum very sophisticated but the difference spectrum is often very small and difficult to identify too [22]. For this reason, machine learning methods, such as deep learning, can be employed to overcome the complexity and heterogeneity [23]. In particular, deep learning algorithms can classify large and complicated data based on patterns submerged inside. These algorithms have been used for a wide range of applications, such as medical images [24], audio data [25], tissue images, biological data and signals [23–26]. They have shown promising performance for analyzing spectroscopic signals, including Raman spectroscopy in complex samples, for example, biological samples [27]. SERS datasets are high-dimensional with small samples and multiple indices. Therefore, logistic regression (LR), k -nearest neighbor (KNN), decision tree (DT), and random forest (RF) are suitable for such analysis.

This article is intended to explore efficient approaches to the early diagnostics of lung cancer using the combination of SERS with machine learning approaches. SERS technology was employed to detect 26 lung cancer patients and 45 healthy volunteers. LR, KNN, DT, and RF were used to establish the diagnostic model. The performance of the four models was compared to get a more suitable model for whole-blood SERS spectrum diagnostics.

Experimental. Silver nanoparticles (Ag NPs) were fabricated by Nicolae Leopold's deoxidation method [28]. In brief, 4.5 mL of 9.99×10^{-2} M sodium hydroxide solution was mixed with 5 mL of 6.04×10^{-2} hydroxylamine hydrochloride solution. Subsequently, the mixture was poured into 90 mL of 1.11×10^{-3} M silver nitrate solution. The resulting mixture was stirred rapidly until the solution became homogeneous and turned milky gray (about 20 min).

Whole blood was collected from 45 healthy volunteers and 26 patients with lung cancer who had a confirmed clinical and histopathological diagnosis, according to the research approved by the Fujian Provincial Hospital ethics board. All participants signed a written informed consent permitting the use of whole-blood samples in this study. Prior to the whole-blood collection, all participants had fasted overnight to avoid interference caused by food consumption. Whole-blood samples of 3 mL were obtained from the subjects, stored at room temperature, and tested within 3 h.

The SERS spectra were recorded with a confocal Raman spectrometer (Renishaw in Via, UK) with a 785-nm laser excitation operating at 10% power, within the range 400–1800 cm^{-1} with a resolution of 2 cm^{-1} . Each Raman spectrum was captured by a $\times 20$ lens with an integration time of 10 s. The above-prepared Ag NPs were centrifuged for 10 min at 10,000 rpm at room temperature (RT), and the supernatant was discarded, leaving concentrated Ag NPs. Immediately after, the concentrated Ag NPs were transferred to new EP tubes to be mixed thoroughly with an equal volume of whole blood at RT for 5 min. Last, SERS spectra were recorded by applying 2 μL of the solution mixture onto an aluminum sheet. Each sample was measured three times.

Data analysis. LR [29] is a classic algorithm, which can treat one-dimensional classification and multivariate classification to depict the relationship between the independent variable X and the dependent variable Y . The dependent variable is the result that we hope to obtain, and the independent variable is the potential factor that affects the result.

KNN [30] is one of the simplest machine learning algorithms. The classification of new sample points is predicted by identifying data points that are divided into several classes. DT [31] is an effective decision-making method. A set of hierarchical rules is inferred from a group of cases without order and rules, and the probability distribution of all possible outcomes is expressed in a tree diagram to generate a DT, so as to achieve the purpose of accurate prediction or correct classification of the research object.

RF [32] is a classifier containing multiple decision trees. The RF algorithm builds a DT on a random subset. The classification results of all trees are collected for classification, which can effectively prevent the deviation of a single tree and over-fitting caused by a tree depth that is too high.

The diagnostic model of lung cancer was established based on LR, KNN, DT, and RF machine learning utilizing whole-blood spectral data. Briefly, by randomly splitting the data into a training set and a test set, we train the KNN, LR, DT, and RF classifier using the training set. After completing the training, the trained classifiers were tested using the test set. Then, the confusion matrix, containing the information of the actual category of the sample and the predicted category of the sample, was created by comparing the predicted results of the classifier with the real labels of the test set. Finally, we further calculated the sensitivity and specificity of the model through the confusion matrix.

Results and discussion. Figure 1 shows the schematic workflow of using the SERS technology combined with machine learning algorithms to analyze whole blood. The Ag NPs were successfully fabricated and concentrated as a SERS enhanced substrate. Then, after mixing the concentrated Ag NPs and whole blood, the resultant mixture was dropped on the Al sheet for drying. Then, the SERS technology was employed to analyze it. To finish off, LR, KNN, DT, and RF were used to establish the diagnostics model. As shown in Fig. 2, Ag NPs exhibited an absorption maximum at 424 nm. The inset picture shows a transmission electron microscopy (TEM) photograph of the prepared Ag NPs.

To assess the enhancement effect of Ag NPs on the whole-blood Raman scattering, we measured the regular Raman spectra and surface-enhanced Raman spectra of whole-blood samples from the lung cancer patient. Figure 3a displays the SERS spectrum of the whole blood sample with Ag NPs at a ratio of 1:1 and the regular Raman spectrum of the blood sample from the same patient without Ag NPs. By comparing curves 1 and 2, it can be clearly seen that no characteristic Raman peaks were visible for the regular Raman spectra of the whole blood sample whereas Raman bands were enhanced significantly by adding Ag NPs to the surface of the sample. The reason why there was a dramatic increase in many dominant vibration bands is that the biochemical substances of the whole-blood sample were absorbed closely on the surface of silver particles.

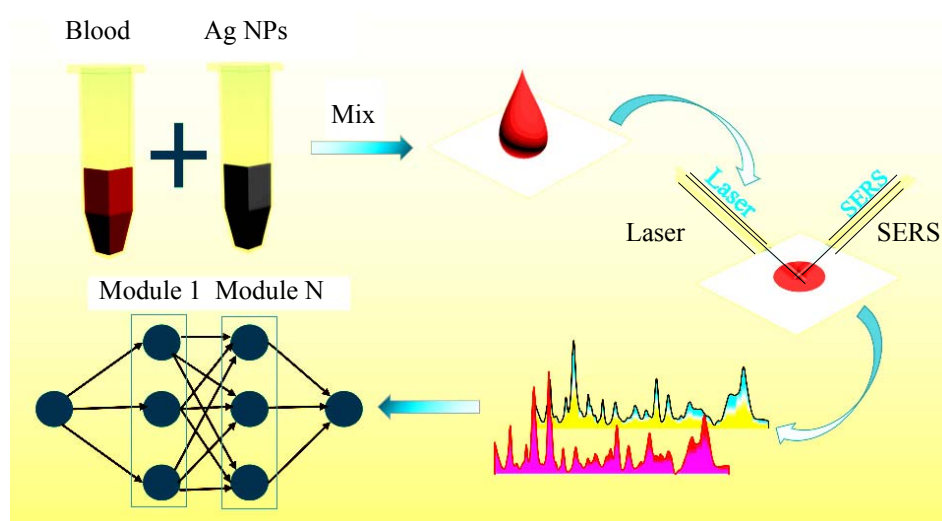


Fig. 1. Schematic diagrams of the detection procedure.

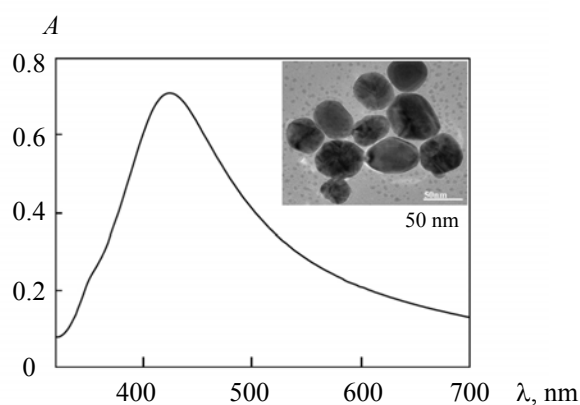


Fig. 2. UV/Visible absorption spectrum of the Ag NPs. Inset: TEM micrograph of Ag NPs.

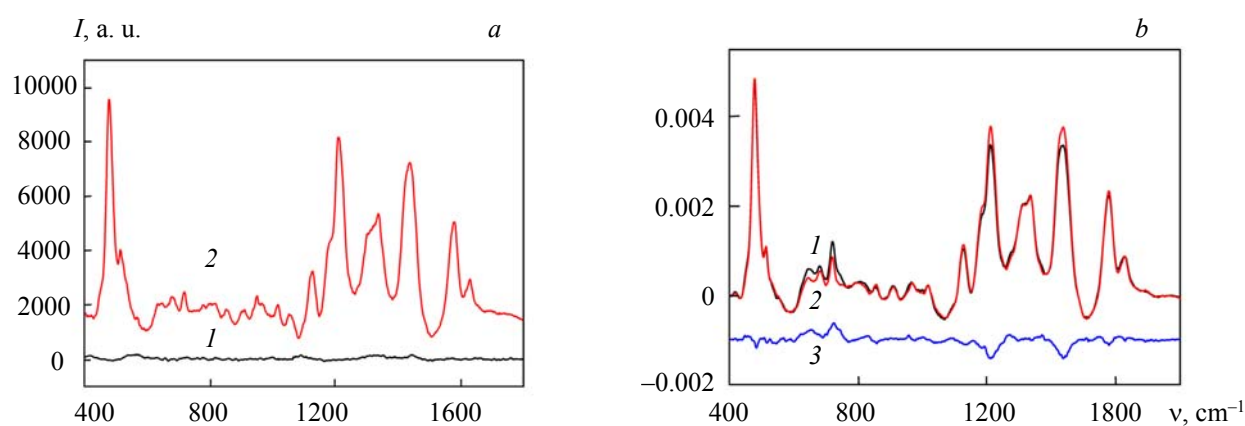


Fig. 3. (a) SERS spectrum of the whole-blood sample from a lung cancer patient by mixing the whole blood with Ag NPs at a ratio of 1:1 (1), the regular Raman spectrum of serum sample from the same patient without Ag NPs (2); (b) normalized mean SERS spectrum for the lung cancer whole-blood samples: $n = 26$ (1) and the normal whole blood samples $n = 45$ (2), and their difference spectrum (3).

For a better comparison of the spectral shape differences between the two groups, the 5th order polynomial fitting with Vancouver software was performed to remove the fluorescence background of the original SERS data yielding the whole-blood SERS spectrum alone. Next, the spectral range 400–1800 cm^{-1} was normalized for batch processing by LabVIEW software to avoid the interference caused by minor differences in experimental conditions such as laser power fluctuation or aggregation difference, etc. Figure 3b displays the mean spectrum and difference spectrum of the whole blood of 45 healthy people and 26 lung cancer patients. As can be seen from the normalized spectrum in Fig. 3b, the patients and healthy people share most of the same Raman vibration peaks, such as 477, 645, 681, 720, 906, 968, 1016, 1126, 1210, 1335, 1437, 1579, and 1628 cm^{-1} . Among them, the three relatively strong characteristic peaks are 477, 1210, and 1437 cm^{-1} . Although the peak locations do not differ greatly, the differences in peak intensity exist, which is also a rationale for subsequent classification. As a result, we assigned the characteristic peak in the whole-blood spectrum as reported in Table 1 [33]. By comparing the average SERS spectra of the two kinds of whole-blood samples, it can be found that the corresponding Raman signals at 647, 681, and 720 cm^{-1} of lung cancer patients were significantly stronger than those of healthy people. Of these, the Raman characteristic peaks located at 647 and 720 cm^{-1} in the whole-blood SERS spectra of lung cancer can be assigned to the C-C twisting mode of phenylalanine (proteins) and DNA respectively. Compared with the two characters Aistic peaks in healthy controls, the relative intensities of the peaks are dramatically stronger in the whole-blood spectrum of lung cancer patients. It is illustrated that the content of phenylalanine and free DNA

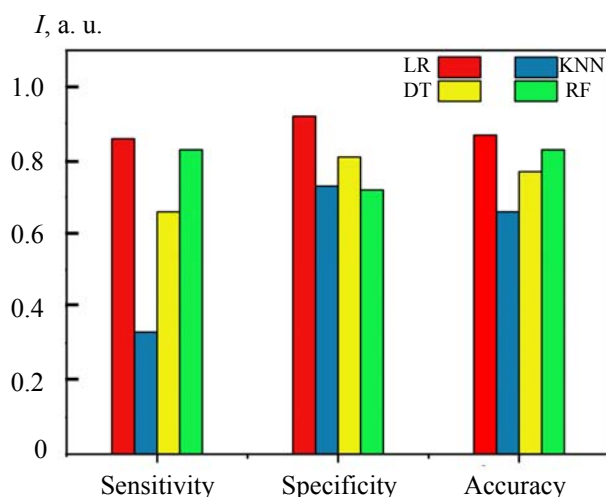


Fig. 4. Comparison of diagnostic performance of the four models based on healthy volunteers and lung cancer patients.

TABLE 1. SERS Peak Assignment of the Whole Blood

Peak position, cm^{-1}	Major assignment
477	Polysaccharides (amylase, amylopectin)
645	C-C twisting mode of phenylalanine (proteins)
681	C-C twist
720	DNA
906	Tyrosine
968	Lipids
1016	Carbohydrates peak for solids
1126	$\nu(\text{C-C})$ skeletal of acyl backbone in lipid
1210	C-C ₆ H ₅ stretching mode in tyrosine
1335	CH ₃ CH ₂ wagging, collagen (protein assignment)
1437	CH ₂ (lipids in tissue)
1579	Pyrimidine ring (nucleic acids) and heme protein
1628	C α =C α stretch

in the whole blood of patients is higher. However, the intensities of the peak at 1210 and 1434 cm^{-1} of the whole-blood SERS spectra of healthy people were stronger than those of the lung cancer whole blood. The Raman characteristic peaks located at 1210 and 1434 cm^{-1} in the whole-blood SERS spectrum of lung cancer can be ascribed to tyrosine and lipid. Knowing that tyrosine plays an important role in human metabolism, we presumed that the tyrosine levels of patients may be associated with the high consumption in patients' vigorous metabolism. Also, the lipid content in the whole blood of lung cancer patients was significantly reduced compared with that of healthy people, indicating that lipid metabolism is enhanced in response to lung injury when lung cancer occurs. The difference in the relative intensity of the spectrum peak means that SERS technology on whole blood has great potential in the diagnostics of lung cancer.

Four algorithms were used to build a diagnostic model of lung cancer, and their diagnostic performance was compared. We computed that the sensitivity and specificity of the LR model were 86 and 92% respectively, and for the KNN model they were 33 and 73%. The DT model exhibited 66% sensitivity and 81% specificity, whereas the RF model showed 83 and 72%. In addition to sensitivity and specificity as performance metrics, the ROC curves that can accurately reflect the relationship between the sensitivity and specificity were plotted and the area under the receiver-operator characteristic curve (AUC) values that can measure the performance of machine learning algorithms for binary classification problems were calculated. The AUC of LR, KNN, DT, and RF models was 0.93, 0.79, 0.63, and 0.81, respectively.

However, the vast majority of machine learning algorithms are random algorithms, which can potentially lead to different results. Therefore, 10-fold cross-validation was used to calculate the objective and stable accuracy of each classifier. It was employed to access the performance, in order to minimize the impact of abnormal circumstances, for example, the imbalance between the proportion of the two types of samples in the training set and the test set and a slight decline in the classifier performance index for many outliers in the test set. Here, we randomly divided data sets into ten equal parts, each of which was used as the test set in turn, and the remaining nine data sets as the training set, so that ten classifiers can be obtained. The accuracy of the model can be obtained by averaging the accuracy of the ten classifiers, which can greatly reduce the abnormal influence brought about by the random algorithm. In this way, the accuracy rate of the LR, KNN, DT, and RF models was 87, 66, 77 and 83%, respectively.

To sum up, the diagnostic performance of each diagnostic model for lung cancer was compared through sensitivity, specificity, accuracy, and area under the ROC curve (AUC). As can be seen from the bar chart in Fig. 4, the sensitivity, specificity, and accuracy of the LR model are much higher than those of the KNN, DT, and RF models. Meanwhile, the AUC of the LR model has the highest value (Fig. 5). Our results indicate that the LR model has good performance.

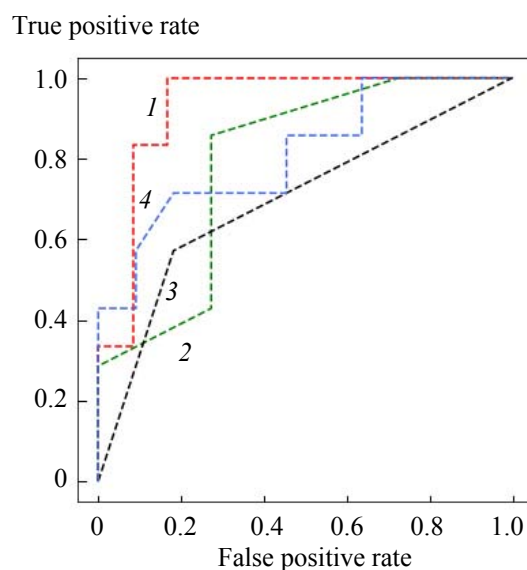


Fig. 5. ROC curves for four models of healthy volunteers and lung cancer patients: LR ROC (area = 0.93) (1), KNN ROC (area = 0.79) (2), DTC ROC (area = 0.69) (3), and RFC ROC (area = 0.81) (4).

Conclusions. For the application requirements of rapid and accurate diagnostics of lung cancer, this article carried out research into differential diagnostics of lung cancer based on whole-blood samples combining SERS technology and machine learning algorithms. The spectral data of a band measuring 400–1800 cm^{-1} wide were extracted to obtain the characteristic peaks of vibration and rotation of macromolecules in whole blood, preliminarily constructing the SERS spectrum database of serum samples. One can analyze the characteristics of the SERS spectra of normal individuals and lung cancer patients to obtain the specific changes in nucleic acids, phenylalanine, and lipids in whole blood. We need an advanced algorithm to develop and build a robust diagnostic system based on the SERS spectrum of whole-blood samples. In this work, LR, KNN, DT, and RF are used to build a diagnostic model. After comparing the performance of the four models, we conclude that the LR model has the best performance of sensitivity and specificity. Developing a model by the LR algorithm based on whole-blood SERS may have great application potential in the diagnostics and screening of lung cancer.

Acknowledgments. We are grateful for the support by the Fujian Provincial Department of Science and Technology guiding (Project No. 2020Y0019); Industry-university Cooperation Project of the Fujian Provincial Department of Science and Technology (No. 2020N5006); the National Natural Science Foundation of China (No. 61772023), the National Key Research and Development Program of China (No. 2019QY1803), and the Fujian Science and Technology Plan Industry-University-Research Cooperation Project (No. 2021H6015).

REFERENCES

1. K. A. Cronin, A. J. Lake, S. Scott, R. L. Sherman, A. M. Noone, N. Howlader, S. J. Henley, R. N. Anderson, A. U. Firth, J. Ma, B. A. Kohler, A. Jemal, *Cancer*, **124**, 2785–2800 (2018).
2. J. Bossert, M. Ludwig, P. Wronski, J. Koetsenruijter, K. Krug, M. Villalobos, J. Jacob, J. Walker, M. Thomas, M. Wensing, *NP J. Primary Care Respiratory Med.*, **31**, 2 (2021).
3. C. Fang, D. Xu, J. Su, J. R. Dry, B. Linghu, *NP J. Digital Med.*, **4**, 14 (2021).
4. N. Guibert, A. Pradines, G. Favre, J. Mazieres, *Eur. Respir. Rev.*, **29**, 190052 (2020).
5. E. Ococks, A. M. Frankell, N. Masque Soler, N. Grehan, A. Northrop, H. Coles, A. M. Redmond, G. Devonshire, J. M. J. Weaver, C. Hughes, K. Lehovskiy, A. Blasko, B. Nutzinger, O. Consortium, R. C. Fitzgerald, E. Smyth, *Ann. Oncol.*, **32**, 522–532 (2021).
6. W. Yu, J. Hurley, D. Roberts, S. K. Chakraborty, D. Enderle, M. Noerholm, X. O. Breakefield, J. K. Skog, *Ann. Oncol.*, **32**, 466–477 (2021).
7. S. Jia, L. Xie, L. Li, Y. Qian, J. Wang, S. Wang, W. Zhang, B. Qian, *Expert Rev. Mol. Diagn.*, **21**, 417–427 (2021).
8. R. Nooreldeen, H. Bach, *Int. J. Mol. Sci.*, **22** (2021).
9. Y. Moro Awelisah, G. Li, M. Ijaz, L. Lin, *Opt. Laser Technol.*, **133** (2021).
10. F. Tustum, F. R. Takeda, M. S. Viyuela, J. B. da Cruz Junior, A. Brandao, R. A. A. Sallum, U. Ribeiro, Jr., I. Ceconello, *J. Surg. Oncol.*, **121**, 784–794 (2020).
11. M. S. Wrobel, J. H. Kim, P. Raj, I. Barman, J. Smulko, *Biosens. Bioelectron.*, **180**, 113115 (2021).
12. A. Germond, T. Ichimura, T. Horinouchi, H. Fujita, C. Furusawa, T. M. Watanabe, *Commun. Biol.*, **1**, 85 (2018).
13. S. Pahlow, S. Meisel, D. Cialla-May, K. Weber, P. Rosch, J. Popp, *Adv. Drug Deliv. Rev.*, **89**, 105–120 (2015).
14. A. Kozik, M. Pavlova, I. Petrov, V. Bychkov, L. Kim, E. Dorozhko, C. Cheng, R. D. Rodriguez, E. Shemet, *Anal. Chim. Acta*, **1187**, 338978 (2021).
15. L. Shen, Y. Du, N. Wei, Q. Li, S. Li, T. Sun, S. Xu, H. Wang, X. Man, B. Han, *Spectrochim. Acta A: Mol. Biomol. Spectrosc.*, **237**, 118364 (2020).
16. V. Karunakaran, V. N. Saritha, M. M. Joseph, J. B. Nair, G. Saranya, K. G. Raghu, K. Sujathan, K. S. Kumar, K. K. Maiti, *Nanomedicine*, **29**, 102276 (2020).
17. X. Li, T. Yang, S. Li, J. Yao, Y. Song, D. Wang, J. Ding, *Laser Phys. Lett.*, **12**, 115603 (2015).
18. L. Guo, Y. Li, F. Huang, J. Dong, F. Li, X. Yang, S. Zhu, M. Yang, *J. Innov. Opt. Health Sci.*, **12**, 1950003 (2019).
19. S. Li, Z. Guo, Z. Liu, *Expert Rev. Anticancer Ther.*, **15**, 5–7 (2015).
20. D. Lin, J. Pan, H. Huang, G. Chen, S. Qiu, H. Shi, W. Chen, Y. Yu, S. Feng, R. Chen, *Sci. Rep.*, **4**, 4751 (2014).

-
21. E. Vargas-Obieta, J. C. Martinez-Espinosa, B. E. Martinez-Zerega, L. F. Jave-Suarez, A. Aguilar-Lemmarroy, J. L. Gonzalez-Solis, *Lasers Med. Sci.*, **31**, 1317–1324 (2016).
 22. X. Li, T. Yang, S. Li, D. Wang, Y. Song, K. Yu, *J. Raman Spectr.*, **47**, 917–925 (2016).
 23. J. Liu, M. Osadchy, L. Ashton, M. Foster, C. J. Solomon, S. J. Gibson, *Analyst*, **142**, 4067–4074 (2017).
 24. H. Sadeghi, H. S. Braun, B. Panti, G. Opsomer, O. Bogado Pascottini, *PloS One*, **17**, e0263409 (2022).
 25. G. Pinkas, Y. Karny, A. Malachi, G. Barkai, G. Bachar, V. Aharonson, *IEEE Open J. Eng. Med. Biol.*, **1**, 268–274 (2020).
 26. M. A. Haile, E. Zhu, C. Hsu, N. Bradley, *Struct. Health Monitor.*, **19**, 1340–1350 (2019).
 27. B. Yan, B. Li, Z. Wen, X. Luo, L. Xue, L. Li, *BMC Cancer*, **15**, 650 (2015).
 28. S. Feng, D. Lin, J. Lin, B. Li, Z. Huang, G. Chen, W. Zhang, L. Wang, J. Pan, R. Chen, H. Zeng, *Analyst*, **138**, 3967–3974 (2013).
 29. R. Zhang, Y. Chen, G. Zhou, B. Sun, Y. Li, Z. Chen, *Evid. Based Complement. Alternat. Med.*, **2020**, 8859558 (2020).
 30. M. A. Mahfouz, A. Shoukry, M. A. Ismail, *Artific. Intell. Med.*, **111**, 101985 (2021).
 31. C. Kingsford, S. L. Salzberg, *Nat. Biotech.*, **26**, 1011–1013 (2008).
 32. T. Shi, S. Horvath, *J. Comput. Graph. Stat.*, **15**, 118–138 (2012).
 33. A. C. S. Talari, Z. Movasaghi, S. Rehman, I. U. Rehman, *Appl. Spectrosc. Rev.*, **50**, 46–111 (2014).