

NEWS & VIEWS

Open Access

Early cancer detection by SERS spectroscopy and machine learning

Lingyan Shi¹✉, Yajuan Li¹ and Zhi Li¹

Abstract

A new approach for early detection of multiple cancers is presented by integrating SERS spectroscopy of serum molecular fingerprints and machine learning.

Cancer is a major public health problem and the second leading cause of death worldwide, with 10 million deaths in 2020¹. Early detection of cancer or precancerous change allows for early intervention and can improve survival rates. However, early detection of many cancers, such as esophageal and ovarian cancers, is still poor, which are often diagnosed at advanced stages². This is more likely due to multiple remaining challenges, such as finding and validating biomarkers for multicancer types, and technologies applied for cancer detection².

Current approaches for cancer screening and detection include blood tests, histopathology tests, imaging tests (for example, mammogram for breast cancer and computerized tomography for lung cancer), and genetic tests. Compared to histopathology and imaging tests, blood tests, which rely on the analysis of specific analytes and tumor biomarkers such as tumor DNA/protein and circulating tumor cells³, are noninvasive and more efficient. However, due to the challenge of validating many potential biomarkers for different cancer types, only a few biomarkers have been validated and used in clinic⁴.

Since the past decade, Raman and infrared spectroscopies have emerged for cancer diagnosis by detecting molecular vibrational fingerprints⁵. Studies utilized infrared spectroscopy and detected breast, bladder, prostate, and lung cancers via molecular fingerprints in blood plasma and serum^{6,7}. Compared with infrared spectroscopy, various types of Raman spectroscopy are more

commonly used for cancer detection. Among these, surface-enhanced Raman spectroscopy (SERS) measurement on blood samples is most frequently used in early cancer detection due to its high sensitivity and specificity^{8–12}. Nevertheless, these studies were constrained by their utilization of a very limited set of biomarkers, rendering them unsuitable for detecting multiple types of cancer. Additionally, they lacked efficient data analysis methods. In contrast, the application of machine learning techniques has proven valuable in classifying serum biomarkers to detect cancer and other diseases using omics data¹³. However, machine learning requires a large independent dataset, while in the previous study only 10 out of a thousand dimensions could be selected for the machine learning algorithm¹³.

In their article in *Light*¹⁴, Shilian Dong and colleagues presented a method for early cancer detection that uses label-free SERS spectroscopy and machine learning for cancer screening (SERS-AICS) of serum biomarkers. The workflow of SERS-AICS is shown in Fig. 1. The authors analyzed 382 healthy controls and 1582 patients across five cancer types, including lung, colorectal, hepatic, gastric, and esophageal cancers. The SERS-AICS method successfully distinguished cancers from healthy controls with a high overall accuracy (95.81%), sensitivity (95.40%), and specificity (95.87%). Moreover, the method was able to separate precancerous samples from other diseases. This method has important implications for early cancer detection.

In this study, only 15 μ l of serum sample each for control and five cancers was used for SERS pan-cancer

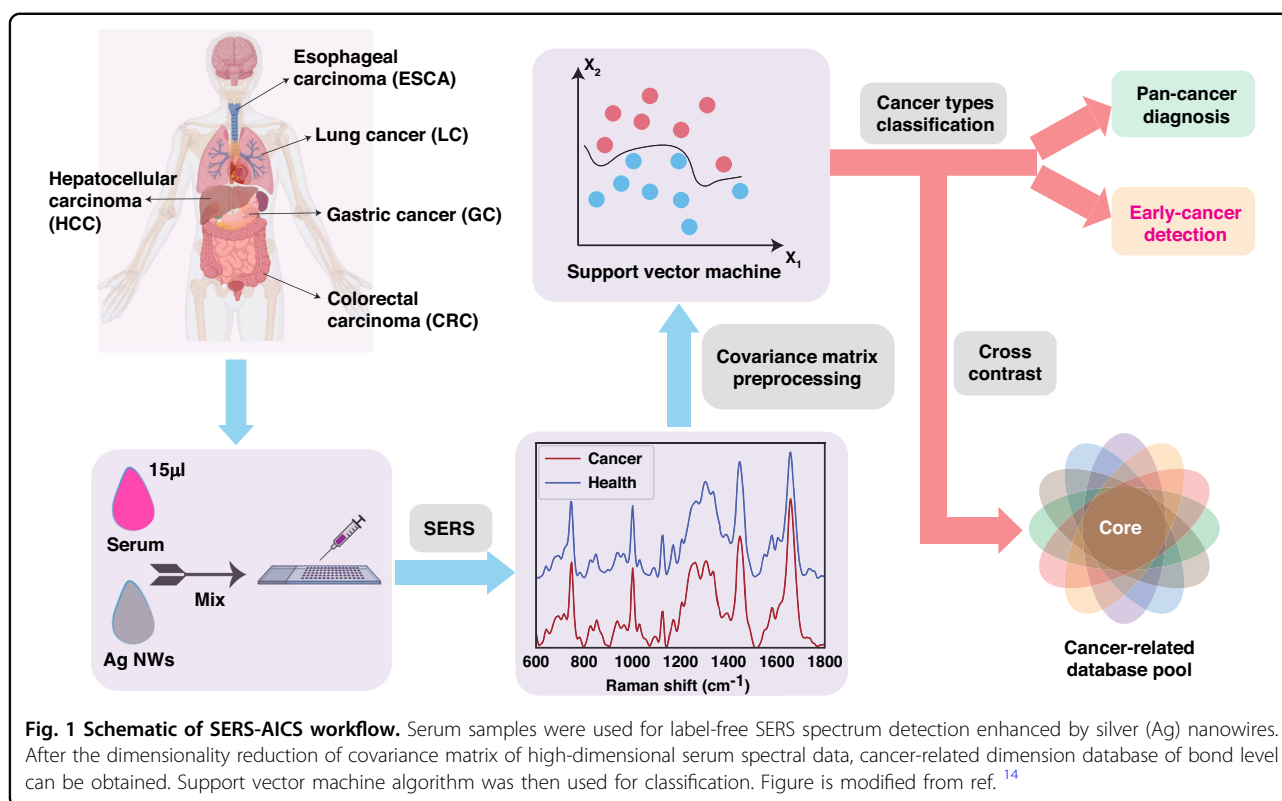
Correspondence: Lingyan Shi (l2shi@ucsd.edu)

¹Shu Chien-Gen Lay Department of Bioengineering, UC San Diego, La Jolla, CA 92093, USA

© The Author(s) 2023



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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.



screening, in which the best SERS enhancement was achieved by using 70–120 nm-diameter silver nanowires. After Raman spectra of a large number of serum samples were collected, Raman feature dimension was trimmed by using covariance matrices, and the dataset of different cancers was classified by support vertical machine (SVM). In the SVM model, 50 intervals were chosen within the Raman fingerprint range (600–1800 cm^{-1}). Thus, this method can screen up to 50 Raman feature dimensions, compared to 10 in the previous study. The SVM model could precisely distinguish subtle differences between spectral data and differentiate all five cancers from control with a high accuracy (ranging 94.10–98.25%), sensitivity (91.84–98.75%), and specificity (90.63–98.57%).

Another challenge of early cancer detection is the discrimination of cancer from other diseases of the same organs. By using the same strategy on serum from four types of diseases (lung, colorectal, gastric, hepatic) and corresponding cancers, gastric and hepatic (diseases/cancer) groups were separated with high accuracy (>92.3%), sensitivity (100% and 85.7%) and specificity (85.7 and 100%); lung and colorectal groups could also be separated except with a slightly lower specificity (77.78%) and sensitivity (66.67%), respectively. Moreover, the SVM classification is able to analyze a large amount of data

(~2000 cases). This large sample pool allows for more reliable spectral data across different cancers. Hence, shared Raman peaks by all five types of cancer, combinations of any types of cancer, or precancerous and diseased samples can be characterized and potentially used as common features for differentiating cancers.

This study combines SERS analysis of serum for cancer screening with machine learning techniques for cancer classification. This innovative approach offers a highly sensitive and selective method for distinguishing between multiple types of cancers, not only from normal samples but also from non-cancerous diseased samples. By utilizing a large database model training and validation, spectra of all molecules present in the serum can be comprehensively collected and analyzed, enabling multiple accurate discrimination of subtle differences among various cancers and controls. The newly proposed approach offers a new venue for identifying these cancers at their earliest stages. If this innovative method can replicate its outstanding performance on prospectively collected samples in a meticulously double-blinded manner, it can be implemented as a screening test in the clinics and ultimately save lives.

References

1. Ferlay, J. et al. Global cancer observatory: cancer today. Lyon: International Agency for Research on Cancer. <https://gco.iarc.fr/today> (2020).
2. Crosby, D. et al. Early detection of cancer. *Science* **375**, eaay9040, <https://doi.org/10.1126/science.aay9040> (2022).
3. Wan, J. C. M. et al. Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat. Rev. Cancer* **17**, 223–238, <https://doi.org/10.1038/nrc.2017.7> (2017).
4. Poste, G. Bring on the biomarkers. *Nature* **469**, 156–157 (2011).
5. Zhang, S. Y. et al. Molecular fingerprint detection using Raman and infrared spectroscopy technologies for cancer detection: a progress review. *Biosensors* **13**, 557, <https://doi.org/10.3390/bios13050557> (2023).
6. Huber, M. et al. Infrared molecular fingerprinting of blood-based liquid biopsies for the detection of cancer. *eLife* **10**, e68758, <https://doi.org/10.7554/eLife.68758> (2021).
7. Kepesidis, K. V. et al. Breast-cancer detection using blood-based infrared molecular fingerprints. *BMC Cancer* **21**, 1287, <https://doi.org/10.1186/s12885-021-09017-7> (2021).
8. Guerrini, L. et al. Surface-enhanced Raman scattering (SERS) spectroscopy for sensing and characterization of exosomes in cancer diagnosis. *Cancers* **13**, 2179, <https://doi.org/10.3390/cancers13092179> (2021).
9. Liu, K. Y. et al. Label-free surface-enhanced Raman spectroscopy of serum based on multivariate statistical analysis for the diagnosis and staging of lung adenocarcinoma. *Vib. Spectrosc.* **100**, 177–184, <https://doi.org/10.1016/j.vibspec.2018.12.007> (2019).
10. Wang, X. et al. Detection of circulating tumor cells in human peripheral blood using surface-enhanced Raman scattering nanoparticles. *Cancer Res.* **71**, 1526–1532, <https://doi.org/10.1158/0008-5472.CAN-10-3069> (2011).
11. Chen, S. et al. Identifying non-muscle-invasive and muscle-invasive bladder cancer based on blood serum surface-enhanced Raman spectroscopy. *Biomed. Opt. Express* **10**, 3533–3544, <https://doi.org/10.1364/BOE.10.003533> (2019).
12. Xia, L. et al. Identifying benign and malignant thyroid nodules based on blood serum surface-enhanced Raman spectroscopy. *Nanomed. Nanotechnol. Biol. Med.* **32**, 102328, <https://doi.org/10.1016/j.nano.2020.102328> (2021).
13. Zhang, L. H. et al. Raman spectroscopy and machine learning for the classification of breast cancers. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **264**, 120300, <https://doi.org/10.1016/j.saa.2021.120300> (2022).
14. Dong, S. L. et al. Early cancer detection by serum biomolecular fingerprinting spectroscopy with machine learning. *eLight* **3**, 17, <https://doi.org/10.1186/s43593-023-00051-5> (2023).