International Journal of Nanomedicine

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

A Machine Learning-Driven Surface-Enhanced Raman Scattering Analysis Platform for the Label-Free Detection and Identification of Gastric Lesions

Fengsong Chen[1], Yanhua Huang[1], Yayun Qian[2], Ya Zhao[2], Chiwen Bu[3], Dong Zhang[3]

[1]Department of Gastroenterology, Haimen People's Hospital, Nantong, 226000, People's Republic of China; [2]Institute of Translational Medicine, Medical College, Yangzhou University, Yangzhou, 225001, People's Republic of China; [3]Institute of Surgery, Guanyun People's Hospital, Guanyun, 222200, People's Republic of China

Correspondence: Dong Zhang; Chiwen Bu, Email gyxrmyykjk@163.com; gyxrmyybcw@163.com

Background: Gastric lesions pose significant clinical challenges due to their varying degrees of malignancy and difficulty in early diagnosis. Early and accurate detection of these lesions is crucial for effective treatment and improved patient outcomes.

Methods: This paper proposed a label-free and highly sensitive classification method for serum of patients with different degrees of gastric lesions by combining surface-enhanced Raman scattering (SERS) and machine learning analysis. Specifically, we prepared Au lotus-shaped (AuLS) nanoarrays substrates using seed-mediated and liquid–liquid interface self-assembly method for measuring SERS spectra of serum, and then the collected spectra were processed by principal component analysis (PCA) - multi-local means based nearest neighbor (MLMNN) model to achieve differentiation.

Results: By employing this pattern analysis, AuLS nanoarray substrates can achieve fast, sensitive, and label-free serum spectral detection. The classification accuracy can reach 97.5%, the sensitivity is higher than 96.7%, and the specificity is higher than 95.0%. Moreover, by analyzing the PCs loading plots, the most critical spectral features distinguishing different degrees of gastric lesions were successfully captured.

Conclusion: This discovery lays the foundation for combining SERS with machine learning for real-time diagnosis and recognition of gastric lesions.

Keywords: surface-enhanced Raman scattering, gastric lesions, Au lotus-shaped nanoarrays, principal component analysis, multi-local means based nearest neighbor

Introduction

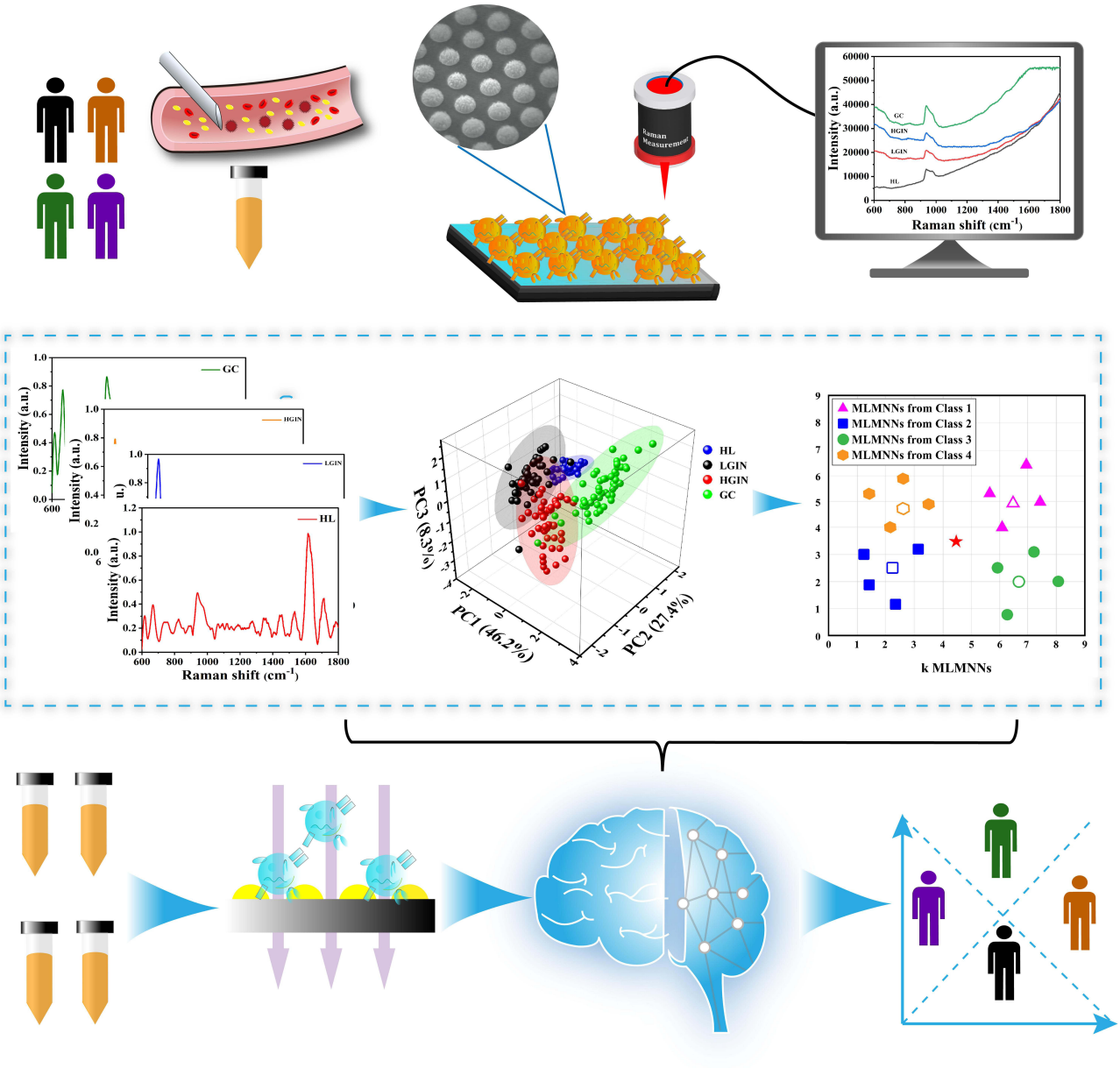
Gastric cancer (GC) is one of the three cancers with the highest incidence rate in the world.[1] According to statistics from the World Health Organization, over one million people worldwide are diagnosed with stomach cancer each year, with over 700,000 deaths from the disease.[2] However, recent reports indicate that Precancerous lesions of gastric cancer (PLGC) can detect potential gastric lesions, such as ulcers or tumors, which can help with early diagnosis and treatment and significantly improve the overall survival rate of gastric cancer patients.[3,4] Precancerous lesions of gastric cancer (PLGC) refers to abnormal changes in the structure, function, or tissue of the gastric mucosa during the pathological process before the development of gastric cancer. PLGC can include low-grade intraepithelial neoplasia (LGIN) and high-grade intraepithelial neoplasia (HGIN) based on organic and functional abnormalities.[5,6] The most common diagnostic techniques experienced by clinicals are positron emission tomography (PET), X-ray imaging, endoscopy, computed tomography (CT) and pathological biopsies.[7–9] However, their high cost and damage limit their practical application.[10–12] The content of certain biochemical components in serum may change at different stages.[13] Analyzing the changes in serum components can provide prevention and diagnosis for gastric lesions.[14,15] The current methods cannot accurately and quickly analyze the results. Therefore, it is necessary to adopt fast, sensitive, and effective methods to accurately detection and identification of gastric lesions.

Received: 29 March 2024
Accepted: 1 September 2024
Published: 10 September 2024

Raman spectroscopy (RS) is an optical scattering phenomenon used to analyze the structure and composition of substances by measuring the frequency shift of scattered light.[16–18] However, its practical application is hindered by the low signal intensity during the Raman scattering process, ie, the small Raman cross-section. This means that overcoming signal attenuation is a challenge when collecting and analyzing Raman spectra. Compared to traditional RS, surface enhanced Raman spectroscopy (SERS) introduces various shapes of noble metal nanostructures on the surface, significantly enhancing sensitivity and signal-to-noise ratio.[19] The detected signals can be amplified by factors ranging from 10^3 to 10^8. SERS promotes this enhancement and can even increase detection sensitivity to the level of individual molecules.[20,21] This is mainly due to the local surface plasmon resonance (LSPR) generated by laser irradiation on the surface of nanomaterials (such as Ag or Au nanoparticles), which leads to electric field enhancement (EM) and polarization of surrounding molecules, enhancing Raman signals.[22] The spectrum of SERS has a narrow linewidth, is not prone to sudden spectral attenuation, and is not affected by water interference. Meanwhile, due to the uniqueness of molecular characteristic vibrational spectra, SERS is highly suitable for molecular recognition, which involves obtaining "fingerprint" information of substances.[23,24] By analyzing 'fingerprint' information, it is possible to accurately determine the biochemical composition and content of substances.[25] In cases where gastric lesions lack obvious symptoms and hinder accurate diagnosis, the unique "fingerprint" information in serum SERS spectra can be used to identify cellular abnormalities.[26] This provides a basis detection for gastric lesions, especially in the absence of clear symptoms. Thus, SERS has vast application prospects in serum detection and analysis, and can as a diagnostic tool with enormous potential. AuLS nanoarrays substrate is a type of nanostructure composed of gold atoms arranged in a lotus shaped. AuLS nanoarrays substrate exhibit an excellent conductivity, thermal conductivity, and optical properties due to their well-defined geometry and high surface-to-volume ratio. The lotus shaped structure of AuLS nanoarrays substrate endows them with numerous "hotspots", making them nanomaterials with plasmonic excitation characteristics. The unique optical properties of AuLS nanoarrays substrate make it an ideal material for bioimaging, bringing new possibilities to the field of medicine and laying the foundation for the development of personalized medical and treatment approaches.

Label-free detection is a method of analyzing samples without the need for added markers or fluorescent labels.[27,28] Compared to label detection, label-free detection has the advantages of low cost and avoiding the complexity and interference associated with introducing markers. At present, some progress has been made in using SERS technology for label-free detection of serum. Simultaneously, substances exhibit slight differences in SERS spectra at different stages, leading to frequent identification errors between various spectra. Additionally, the complex biochemical composition of serum samples hinders the rapid and accurate detection, as well as precise identification of crucial substance changes in SERS spectra. In order to enhance the accuracy of disease diagnosis, multiple machine learning methods have been employed for the analysis and identification of SERS spectra.[29–31] The k-Nearest Neighbors (KNN) classification has the advantages of simplicity, efficiency, insensitivity to outliers, and suitability for small-scale datasets.[32] It has been widely studied and applied in the fields of chemical material detection, food safety detection, and disease detection.[33,34] The traditional KNN has problems such as simple majority votes and simple measurement of Euclidean distance. When making classification decisions, it is assumed that each neighbor has the same weight, but in reality, different neighbors have different contributions to the classification, and reliable neighbors have a greater contribution. This has resulted in the accuracy of KNN for complex biological samples still not being enough.

In this work, a machine learning-driven SERS analysis platform was proposed for the label-free detection and identification of gastric lesions (Scheme 1). AuLS nanoarrays substrate was fabricated by seed-mediated method and liquid–liquid interface self-assembly. AuLS nanoarrays was selected as nanomaterials due to its excellent uniformity, stability, repeatability, and enhancement effect. In order to improve diagnostic accuracy, PCA-MLMNN model was structured to analyze SERS spectra of serum of different degrees of gastric lesions. PCA simplified multiple complex variables in spectral datasets, captured the most significant features distinguishing different degrees of gastric lesions. MLMNN has the characteristics of KNN, multiple local mean vectors, and representation-based distance. The model can achieve satisfactory accuracy, sensitivity, and specificity. To our knowledge, this is the first report on the detection of gastric lesions using a machine learning-driven SERS analysis platform.



Scheme 1 Schematic representation of label-free detection and identification of gastric lesions by combining SERS and PCA-MLMNN model.

Results and Discussion

AuLS Nanoarrays Substrate Characterization

AuLS nanoarrays substrate was fabricated according to the PS single-layer sphere template etching strategy. Figure 1A–C showed that uniformly shaped AuLS nanoarrays tightly arranged on the surface of PDMS, which can significantly generate significant local field enhancement, serving as the "hot spots" for signal amplification. Uniformity of the AuLS nanoarrays substrate was quite important for the SERS signal measurement. In order to investigate the uniformity of the AuLS nanoarrays substrate, SERS signals of 10 randomly selected spots on the AuLS nanoarrays substrate was measured (Figure 1D). The basically consistent intensity and RSD value (5.87%) of peak intensity at 592 cm^{-1} (Figure 1E) indicated that the AuLS nanoarray has signified excellent signal uniformity. The electromagnetic field distribution of AuLS nanoarrays was exhibited using the finite-difference time-domain, as shown in Figure 1F. A large number of "hots pots" can be observed distributed at the edge of the tip, which can significantly enhance the SERS spectrum. In the

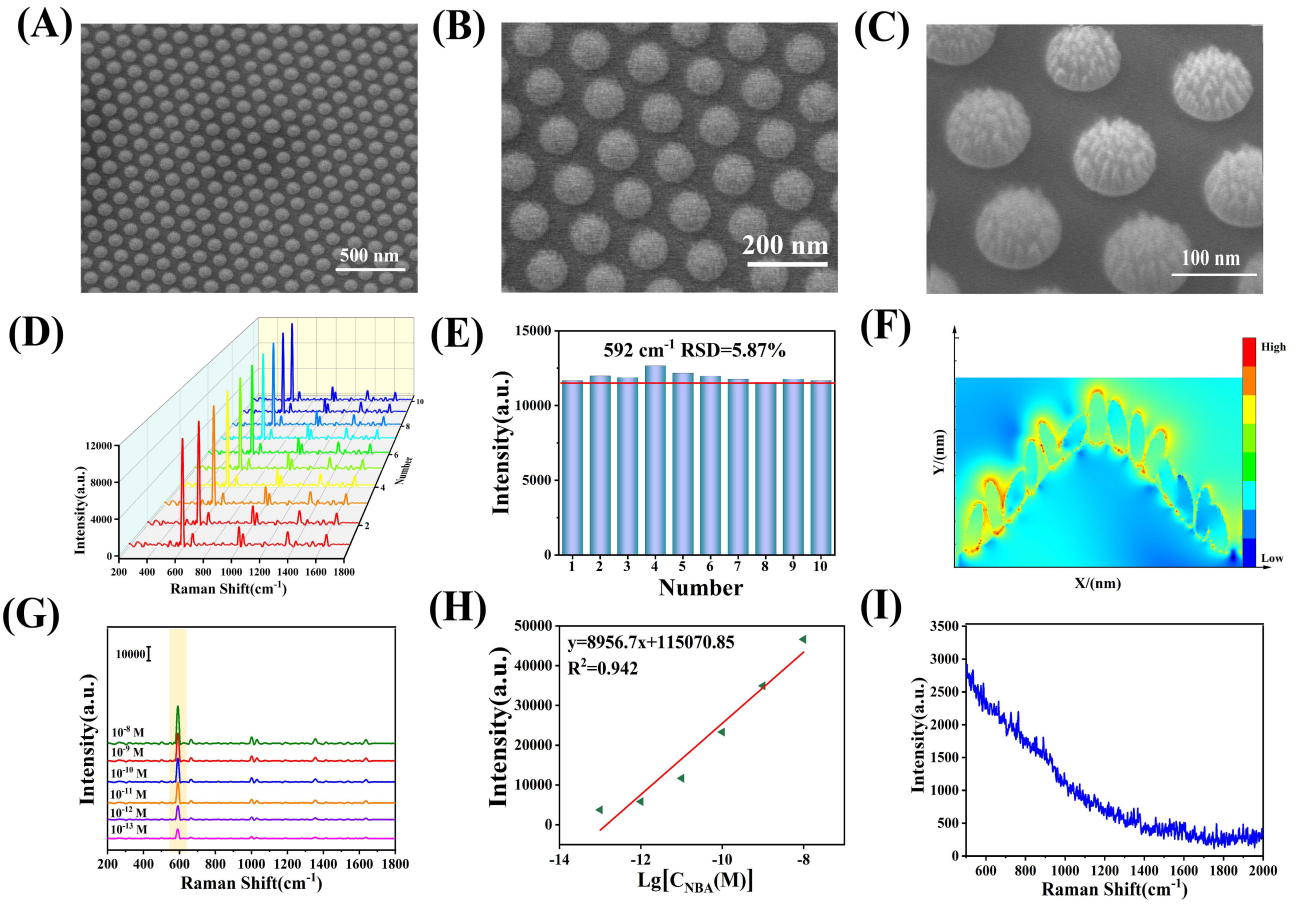


Figure 1 (**A**) SEM images of the PS colloidal monolayer template. (**B**) Low magnification and (**C**) high magnification SEM images of AuLS nanoarrays. (**D**) SERS spectral images from 10 random points on the array surface. (**E**) Histogram of the Raman peak intensity at 592 cm^{-1}. (**F**) Image depicting the electromagnetic field intensity distribution in AuLS nanoarrays. (**G**) SERS spectral measurements of AuLS nanoarrays with varying concentrations of NBA. (**H**) The corresponding linear regression curve. (**I**) SERS spectra of AuLS nanoarrays post-treatment with 0.1 M KI.

presence of NBA, SERS spectrum was recorded on AuLS nanoarrays substrate to explore the sensitivity from 10^{-8} M to 10^{-13} M (Figure 1G). NBA modification was performed on AuLS nanoarrays to detect the SERS enhancement ability of AuLS nanoarrays. The formula was as follows: $EF = (I_{SERS}/C_{SERS})/(I_{Raman}/C_{Raman})$. There was a strong linear correlation between the logarithm of NBA concentration and the SERS intensity at 592 cm^{-1} (Figure 1H). The calibration curve was y=8956.7x+115070.85, with correlation coefficient R^2 value of 0.942. For label-free SERS measurement, interference of the background signal from the substrate is unacceptable; thus, we applied 0.1 M KI to remove the residual chemicals and the SERS spectra recorded on the substrate indicated the cleanliness of the AuLS nanoarrays (Figure 1I).

Collection of Serum Samples

The experiment was performed in accordance with the guidelines "declaration of helsinki" and was approved by the Ethics Committee of Guanyun people's hospital (NO: 2023-YKL-G013). The blood samples were collected from January 2023 to December 2023. According to the ethical guidelines outlined by the Council for International Organizations of Medical Sciences, all samples were collected from blood donors who provided informed consent. Random sampling was used in the experiment to ensure the representativeness and generalizability of the results. The inclusion criteria for gastric lesions were patients aged 18 to 75, diagnosed with gastric lesions through gastroscopy, CT, and histopathological examination (HE staining). The exclusion criteria were: (1) other types of cancer, Helicobacter pylori (HP) infection, or recent surgery; (2) severe diseases such as cardiovascular diseases or liver and kidney dysfunction. The healthy subjects consisted of individuals aged 18 to 75 who underwent routine health examinations

Table 1 Demographics of Study Population

Groups	Healthy Subjects (n=50)	Low-Grade Intraepithelial Neoplasia (n=50)	High-Grade Intraepithelial Neoplasia (n=50)	Gastric Cancer (n=50)
Age (years) Mean	47	53	58	57
Gender				
Male	28	23	26	24
Female	22	27	24	26

and showed no signs of gastric lesions based on gastroscopy, CT, and histopathological examination. To ensure accuracy, we excluded individuals with Hp infection from both the healthy subjects and the patients, as Hp infection could potentially influence the SERS spectra results. The detailed information of all donors are summarized in Table 1. The gastroscopy, CT, and HE staining of all donors are shown in Figure 2.

As depicted in Figure 2, gastroscopy of healthy individuals revealed a smooth mucosa of the antrum and Angle of the stomach, presenting a light red color with no signs of bleeding, clear mucus, and absence of abnormalities. CT scan and pathological examination confirmed the absence of any anomalies. Gastroscopic findings in LGIN patients indicated permeable and reddened submucosal vessels in the pylorus, along with various-sized superficial ulcers around the pylorus exhibiting a grayish-white surface. CT imaging demonstrated thickening and reinforcement of the gastric antrum wall. Pathological sections exhibited mild-to-moderate epithelial atypia characterized by elongated, slender nuclei with dark

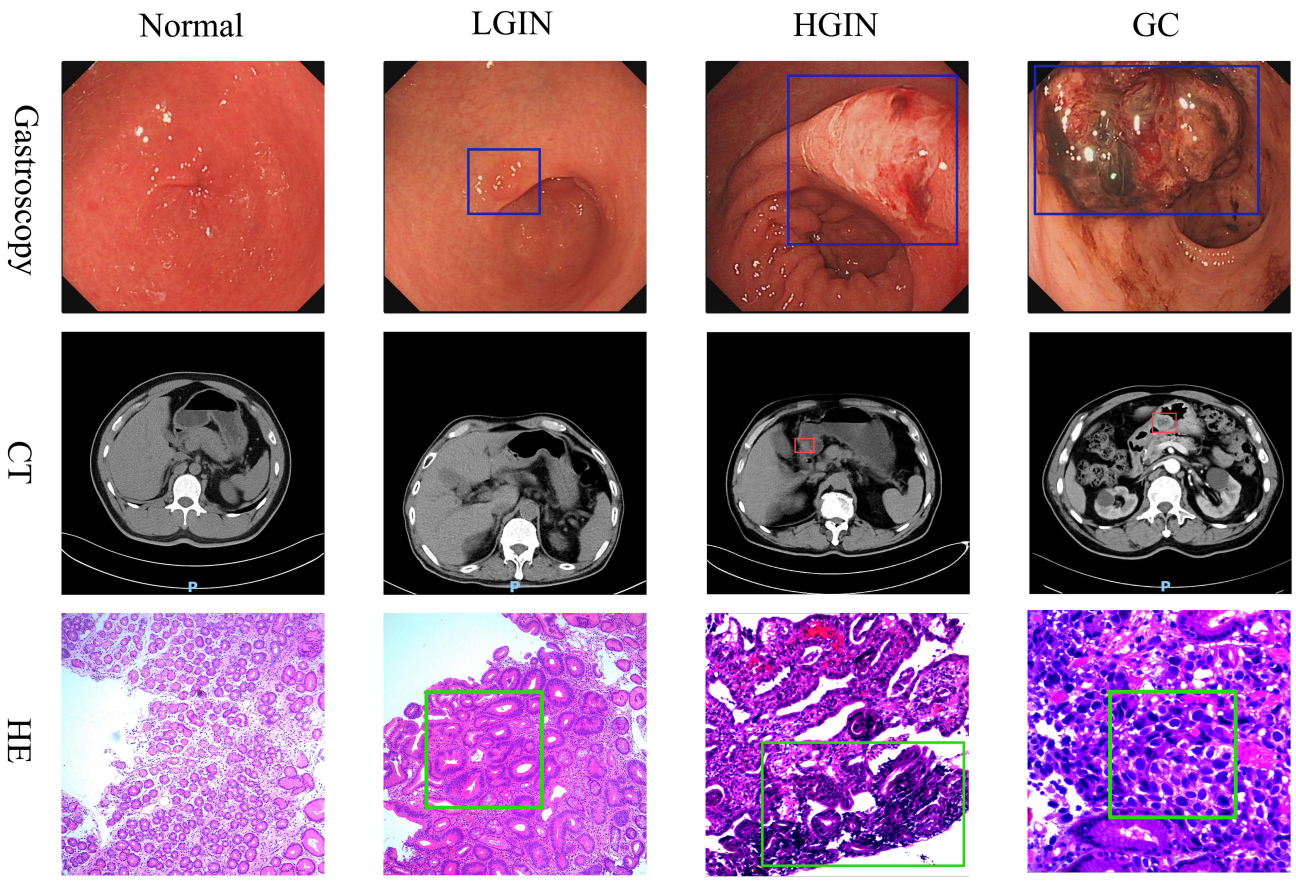


Figure 2 Gastroscopy, CT, and pathological sections (HE staining) for HL, LGIN, HGIN, and GC.

purple staining. Additionally, hypertrophic dilation was observed within structurally intact glands. Conversely, gastroscopy in HGIN patients revealed ulcers in the pyloric area accompanied by reddish mucosal staining. CT scans showed thickening of the gastric antrum. Pathological sections displayed disordered dense glands with densely stained rod-shaped or oval nuclei coexisting with endocytosis. Gastroscopic examination for GC patients' unveiled irregularly raised filth-covered pylorus. CT scan results indicated significant narrowing of the pylorus. Pathological analysis revealed highly heterogeneous hyperplasia of glandular epithelial cells, abnormal concomitant glandular ducts, as well as infiltration growth of cancer cells. We used colored boxes in the figure to indicate the locations of lesions in the stomach tissue for better observation. Specifically, blue, red, and green boxes corresponded to the lesion sites in the gastroscopy, CT, and HE staining of LGIN, HGIN, and GC patients, respectively.

Comparison of Average SERS Spectra of Collected Serum of Different Degrees

Figure 3 shows the average SERS spectra of collected serum of HL, LGIN, HGIN, and GC on AuLS nanoarrays substrate. Each spectra, so presented, have the average values of fifty independent measurements (n=50). The overall SERS spectral similarities and differences between different degrees of gastric lesions were depicted in the average SERS spectra. To better observe the similarities and differences between various stages of gastric alterations, we compared the average SERS spectra of HGIN and Stage I GC. There were 12 stage I patients among the gastric cancer patients (n=50), and we obtained the average SERS spectra from these stage I patients. Figure S1 showed that although the waveforms of HGIN and Stage I GC were very similar, there were still significant differences at certain positions. The Raman shifts reflected the various biochemical components in the serum, while the peak intensities indicated the concentration differences of these components at various stages. By determining the characteristic peaks with significant changes through Raman shifts, it was crucial to understand the chemical composition of different stages of gastric lesions. Analyzing these characteristic peaks helped reveal the dynamic changes in the serum from precancerous lesions to gastric cancer. These SERS spectra were somehow similar in peak shape and position, but the peak intensities of SERS spectra were quite different, among which 616, 659, 747, 825, 899, 941, 1032, 1099, 1124, 1172, 1203, 1280, 1342, 1404, 1450, 1536, 1605, 1621, 1691 and 1713 cm^{-1} were significantly different. The black arrows indicated the different characteristic peaks of the average SERS spectra. These detailed assignments of characteristic peaks were clearly observable in Table S1.

As mentioned above, the characteristic peaks at 747, 825, 899, 941, 1032, 1124, 1172, 1203, 1280, 1342, 1450, 1536 and 1691 cm^{-1} were found that the relative intensity gradually increased as gastric lesions worsen, which resulted from polysaccharides, protein, DNA/RNA, phospholipids and other substances. Conversely, the intensity of characteristic peaks at 1099, 1404, 1605 and 1621 cm^{-1} showed a decrease trend during gastric lesions worsen, which resulted from

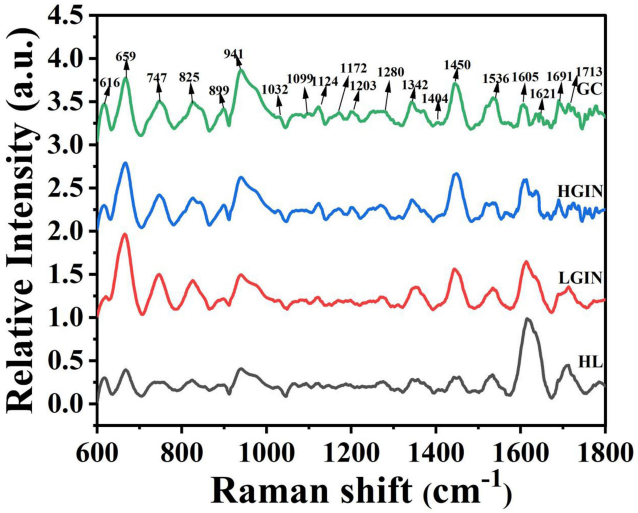


Figure 3 The average SERS spectra of serum of HL, LGIN, HGIN, and GC.

some amino acids such as phenylalanine, tyrosine, and tryptophan. Furthermore, irregular intensities of characteristic peaks were also observed at 616, 659 and 1713 cm^{-1} as gastric lesions worsen. The changes in characteristic peaks reflected changes in the content of certain biochemical components in the serum, which may be related to metabolic abnormalities during the development of gastric lesions. Overall, the development of gastric lesions was mainly caused by molecular vibrations of carbohydrates, proteins, lipids, and some amino acids. The fingerprint features of SERS spectrum were very important for the detection of different degrees of gastric lesions. However, distinguishing different degrees of gastric lesions by intensity of characteristic peaks was influenced by subjective factors of the operators and certain contingency factors.

Multivariate Analysis

In order to distinguish the SERS spectra of serum of different degrees of gastric lesions in a more effective method, the PCA-MLMNN model was constructed. PCA was an intuitive and fast dimensionality reduction tool used to reduce dimension of the original data set to a few principal components (PCs) with maximized variance within a group of data.[35] The first few PCs covered most of the important information in spectral data, while the contributions of other PCs can be ignored.[36,37] Figure 4A shows the eigenvalues and explained variance for each PCs the scree plot of. It can be observed that the first few PCs account for the vast majority of the total information in all SERS spectra, and the information decreases rapidly with the increase of the number of PCs. The first principal component (PC1) explained the most variance within the data, which was very significant. It should be noted that compared to the related variance with PC1,

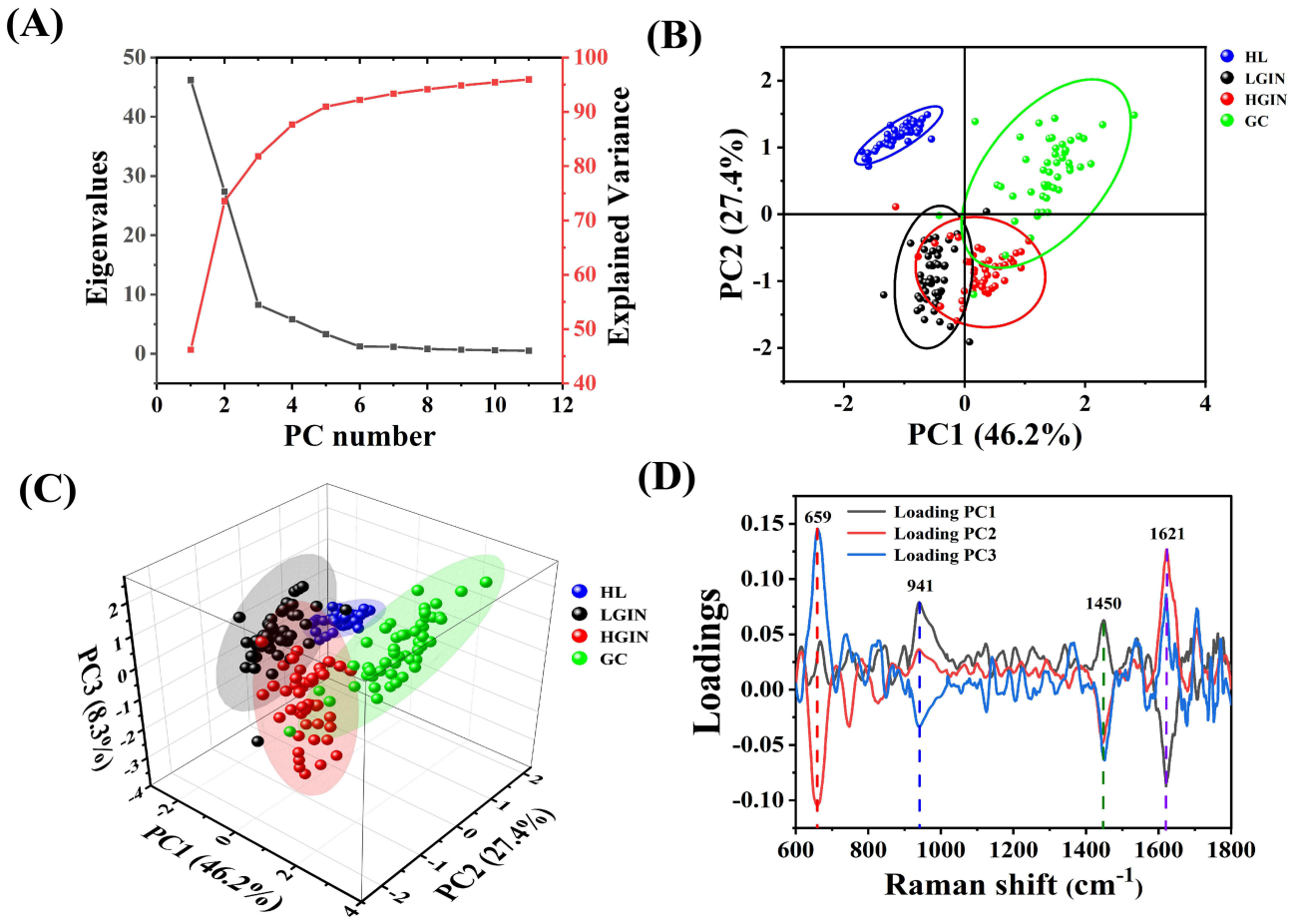


Figure 4 (**A**) Scree plot of the variation of component number with the calculated eigenvalues for SERS spectra of the serum. (**B**) 2D PCA score plot for the four groups of samples, HL, LGIN, HGIN, and GC. The ellipses correspond to 95% confidence intervals. (**C**) 3D PCA score plot for the four groups of samples, HL, LGIN, HGIN, and GC. (**D**) PC1, PC2, and PC3 loading plots.

PC2 was relatively lower. Each PCs was orthogonal to each other. The first eleven PCs explained over 95% of the changes between SERS spectra.

Clearly separated SERS spectral clusters were observed in the score plots (Figure 4B and C). The three principal components explained 81.9% of variation in spectral feature with the first PC accounting for 46.2% of variance, the second and the third PC for 27.4% and 8.3%, respectively, reflecting that the first three PCs had contained most information of the original data. Each dot in the score plot represented one Raman spectrum. The blue, black, red, and green spheres referred to HL LGIN, HGIN and GC groups, respectively. Most projection points for each group were surrounded by a clear 95% confidence ellipsoid. As shown in the graph, HL group lied in the second quadrant and was well separated from other groups. The LGIN group distributed in the third quadrant and partially overlapped with the HGIN group in the fourth quadrant. The GC group in the first quadrant was far from the HL group and partially overlaps with the LGIN and HGIN group. In this case, it was difficult to completely distinguish between different degrees of gastric disease, as the confidence ellipse partially overlaps.

By analyzing the characteristic peaks of the waveforms in the first three PCs loading plots, the most significant features distinguishing different degrees of gastric lesions were determined. The PC1, PC2, and PC3 loading plots are shown in Figure 4D. The greater the absolute value of the characteristic peak of loading plot, the greater the contribution rate to the principal components and the greater the influence on discrimination between different degrees of gastric lesions. According to the characteristic peak of the first three PCs loading plots with a contribution rate of 81.9%, the characteristic peaks at 659, 941, 1450, and 1621 cm^{-1} explained the most significant spectral difference between different degrees of gastric lesions. The characteristic peaks at 659 and 1621 cm^{-1} have been associated with cystine C-S stretching mode and tryptophan, respectively. Changes in the two peaks may be due to inflammation affecting the metabolism of amino acids in the stomach during this period. At the same time, the rapid proliferation of cancer cells required large amino acids as raw materials for protein synthesis, resulting in a decrease in the concentration of cystine and tryptophan. The significant intensity increases of peak at 941 cm^{-1} belonged to polysaccharides and amylose may be associated with the increased release of digestive enzymes, which accelerated the breakdown and absorption of carbohydrates, leading to an increase in serum sugar concentration. In addition, the increased peak at 1450 cm^{-1} in the SERS spectrum correspond to lipids C-H deformation bands and CH_2 bending. This may be due to inflammation and cancer cell activity affecting lipid peroxidation and the activity of lipid metabolism enzymes, leading to an increase in serum lipid concentration.

Overall, in different degrees of gastric lesions, the changes in serum components such as sugars, lipids, and some amino acids exhibited complex patterns. These changes reflected the physiological and metabolic responses of the body to inflammation, tissue damage, and cancer. Moreover, the box plots illustrating the most significant feature peaks in the loading plots of PC1, PC2, and PC3 demonstrated a clear classification (Figure S2, Supplementary Materials and Methods). This further validated the effectiveness of PCA assisted SERS in detection and identification of PLGC.

After PCA analysis and processing, we selected the first eleven PCs, which account for 95% of the total variance, as input variables for the model. MLMNN was a method that used a classification local mean vector to linearly represent query samples, and used the distance between the query sample and each local average vector based on representation coefficients as a similarity measure for final classification. Figure S3 illustrated the classification principle of MLMNN, wherein the relationship between the query sample and the local average vector was thoroughly considered by representation coefficients. To assess the accuracy of the model, a comparative analysis was conducted against the models PCA-KNN and KNN. The accuracy of PCA-MLMNN was determined through 5-fold cross validation, varying the number of PCs and K values, as depicted in Figure 5A. When the parameter k was 3 and the number of PCs was 4, the model has the highest accuracy. In the application of PCA-KNN, the highest accuracy was achieved when the parameter k was set to 3 and the number of PCs was 2, as illustrated in Figure 5B. The discriminative performance of the KNN model under various K values is depicted in Figure 5C. The highest accuracy was observed when K is set to 5. It can clearly observe that the accuracy of the PCA-MLMNN model was higher than the other model.

Subsequently, these optimal parameters were used to calculate the confusion matrix of the model. The rows in the confusion matrix represented the actual labels, while the columns represented the predicted labels. If the actual label was equal to the predicted label, it indicated that the SERS spectrum has been correctly classified. The diagonal in the

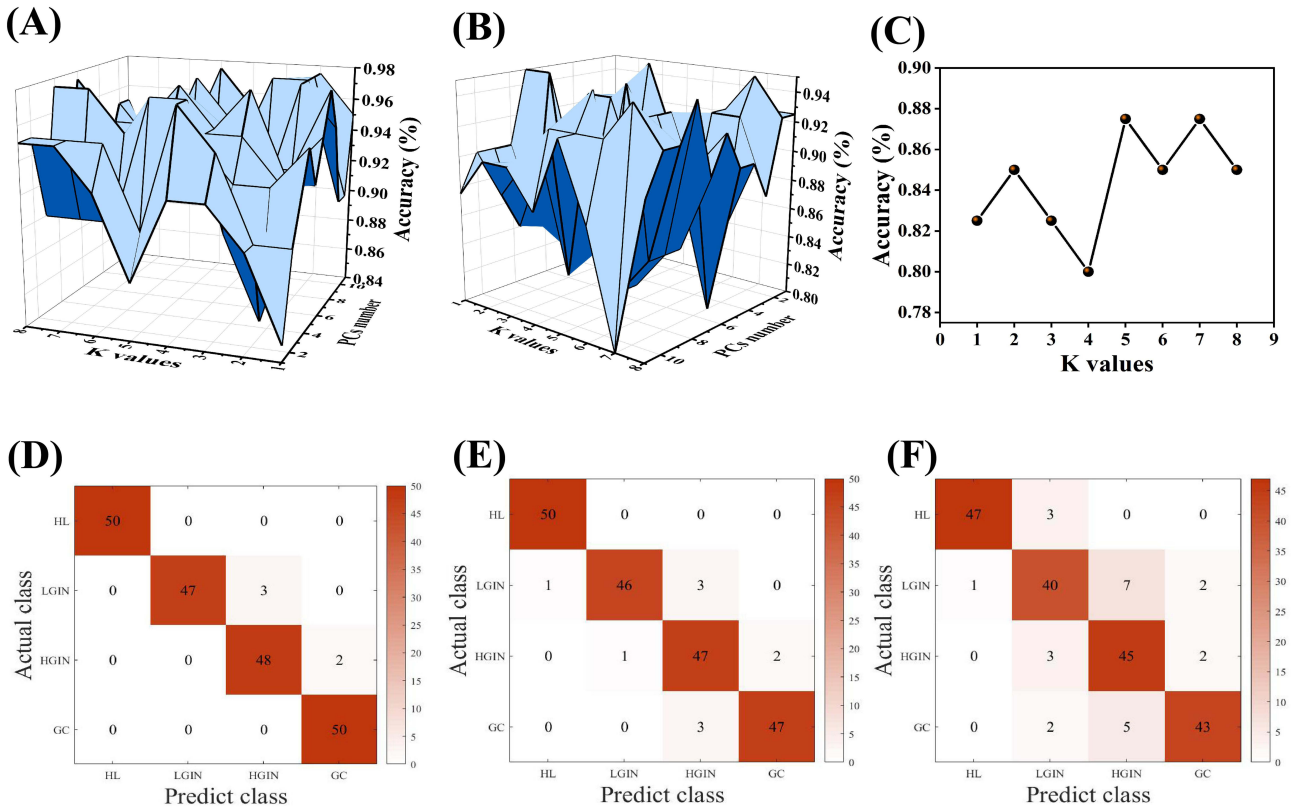


Figure 5 The under different parameters accuracy of (**A**) PCA-MLMNN, (**B**) PCA-KNN, (**C**) KNN. The confusion matrix of (**D**) PCA-MLMNN, (**E**) PCA-KNN, (**F**) KNN.

confusion matrix represented the number of correctly classified samples, while the non-diagonal represented the number of misclassified samples. The accuracy of 97.5% was achieved in the PCA-KNN confusion matrix (Figure 5D). Three (LGIN) and two (HGIN) SERS spectra were incorrectly classified as HGIN and GC, respectively. The accuracy of 95.0% was achieved in the PCA-KNN confusion matrix (Figure 5E). Ten SERS spectra were misclassified. The accuracy of 87.5% was achieved in the KNN confusion matrix (Figure 5F).

Furthermore, the study also analyzed the sensitivity and specificity for HL&LGIN,HGIN,GC, LGIN&HL,HGIN,GC, HGIN&HL,LGIN,GC, and GC& HL,LGIN,HGIN (Table S2, Supplementary Materials and Methods). In contrast to PCA-KNN and KNN, the model displayed obvious superiority in the analysis of SERS spectral data. Hence, PCA-MLMNN was utilized as the multivariate analysis method to promptly and effectively detect different degrees of gastric lesions by leveraging the unique spectral features present in the serum.

Conclusion

An integrated PCA-MLMNN model driven SERS analysis platform was established that was capable of timely, accurately detection and identification serum of patients with different degrees of gastric lesions. The prepared AuLS nanoarrays substrate exhibited good uniformity, stability, and repeatability, and can significantly amplify the extremely weak signal intensity of biological components. The PCA-MLMNN model accurately identified SERS spectra of different degrees of gastric lesions with good accuracy, sensitivity, and specificity. By performing principal component analysis (PCA) on the SERS spectrum, we identified four Raman peaks at 659 cm^{-1}, 941 cm^{-1}, 1450 cm^{-1}, and 1621 cm^{-1}, revealing the significant differences between different degrees of gastric lesions. The detailed clinical outcomes and characteristics of the patients showed complex patterns in serum components across different degrees of gastric lesions. As the lesions progressed, the concentrations of sugars, lipids, and amino acids in the serum changed significantly, reflecting the physiological and metabolic responses to lesion progression. These changes helped us better

understand and detect different degrees of gastric lesions. With the appropriate design of the machine learning-driven SERS analysis platform, we believe this method can probe the biochemical processes of gastric lesions, providing a new perspective for gastric lesion differentiation.

Data Sharing Statement

The data will be available from the corresponding author.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Project code: 81403232), the Major Programs of Natural Science Foundation of higher education in Jiangsu Province (19KJA480003), the Natural Science Foundation of Jiangsu Province (No. BK20171290), the Administration of Traditional Chinese Medicine Project of Jiangsu Province (MS2021081), Lianyungang Health Technology Project (202341), Jiangsu Key Laboratory of Experimental & Translational Non-coding RNA Research Project (202274), the "Scientific Research and Innovation Team" project of Kangda College, and Nanjing Medical University (KD2022KYCXTD013).

We confirm that the article published on the SSRN website (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4723927) is indeed a research outcome of our team. However, the study involved initial animal experiments on precancerous lesions of gastric cancer, not human studies, and the methods used are entirely different from those described in this manuscript. The relevant information has now been removed from the SSRN website.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Bang CS, Yang YJ, Lee JJ. et al. Endoscopic submucosal dissection of early gastric cancer with mixed-type histology: a systematic review. *Dig Dis Sci*. 2020;65(1):276–291. doi:10.1007/s10620-019-05761-w
2. Cao W, Chen HD, Yu YW, et al. Changing profiles of cancer burden worldwide and in China: a secondary analysis of the global cancer statistics 2020. *C M J*. 2021;134(7):783–791.
3. Etemadi A, Safiri S, Sepanlou SG. GBD 2017 Stomach Cancer Collaborators. The global, regional, and national burden of stomach cancer in 195 countries, 1990-2017: a systematic analysis for the Global Burden of Disease study 2017. *Lancet Gastroenterol Hepatol*. 2020;5(1):42–54. doi:10.1016/S2468-1253(19)30328-0
4. Fan JH, Wang JB, Wang SM, et al. Body Mass Index and Risk of Gastric Cancer: a 30-year Follow-up Study in the Linxian General Population Trial Cohort. *Cancer Sci*. 2017;108(8):1667–1672.
5. Tan HY, Wang C, Liu G, Zhou X. Long noncoding RNA NEAT1-modulated miR6 regulates gastric cancer development through targeting STAT3. *J Cell Biochem*. 2018;120(4):4827–4836. doi:10.1002/jcb.26691
6. Milne A, Offerhaus G. Early-onset gastric cancer: learning lessons from the young. *World J. Gastro Oncol*. 2010;2(2):59–64. doi:10.4251/wjgo.v2.i2.59
7. Wang W, Cao K, Han Y, et al. Computed tomographic characteristics of gastric schwannoma. *J Int Med Ras*. 2019;47(5):1975–1986. doi:10.1177/0300060519833539
8. Park H, Kim YJ, Ko SY, et al. Benign regional lymph nodes in gastric cancer on multidetector row CT. *Acta Radiol*. 2012;53(5):501–507. doi:10.1258/ar.2012.120054
9. Engin G, Korman U. Gastrointestinal lymphoma: a spectrum of fluoroscopic and CT findings. *Diagn Interv Radiol*. 2011;17(3):255–265. doi:10.4261/1305-3825.DIR.3332-10.3
10. Zhang Z, Smith R, Wang X, et al. Surface-enhanced Raman spectroscopy combined with machine learning for non-invasive detection of gastric cancer. *Anal Chem*. 2020;92(14):9988–9994.
11. Liu Y, Zhao J, Li H, et al. Machine learning assisted SERS for the accurate diagnosis of early-stage lung cancer. *Biosens Bioelectron*. 2019;145:111699. doi:10.1016/j.bios.2019.111699
12. Chen H, Liu W, Zhang H, et al. Combining SERS and deep learning for prostate cancer detection: a comparison with traditional diagnostic methods. *J Biophotonics*. 2021;14(5):e202000344.

13. Zagari R, Rabitti S, Greenwood D, et al. Systematic review with meta-analysis: diagnostic performance of the combination of pepsinogen, gastrin-17 and anti- Helicobacter pylori antibodies serum assays for the diagnosis of atrophic gastritis. *Aliment Pharmacol Ther*. 2017;46 (7):657–667. doi:10.1111/apt.14248
14. Begum A, Baten M, Begum Z, et al. Role of serum pepsinogen I and II ratio in screening of gastric carcinoma. *Mymensingh Med J*. 2017;26 (4):628–634.
15. Gormally E, Caboux E, Vineis P, et al. Circulating free DNA in plasma or serum as biomarker of carcinogenesis: practical aspects and biological significance. *Mutat Res Rev Mutat Res*. 2007;635(2):105–117. doi:10.1016/j.mrrev.2006.11.002
16. Das RS, Agrawal YK. Raman spectroscopy: recent advancements, techniques and applications. *Vib Spectrosc*. 2011;57(2):163–176. doi:10.1016/j.vibspec.2011.08.003
17. Zeng Q, Liu R, Wang N, et al. Research Progress of Raman Spectroscopy in Medical Laboratory Science (Invited). *Acta Optica Sinica*. 2021;50 (10):1017002.
18. Huang Z, Mcwilliams A, Lui H, et al. Near-infrared Raman spectroscopy for optical diagnosis of lung cancer. *Int J Cancer*. 2003;107 (6):1047–1052. doi:10.1002/ijc.11500
19. Cialla D, Marz A, Bohme R, et al. Surface-enhanced Raman spectroscopy (SERS): progress and trends. *Anal Bioanal Chem*. 2012;403(1):27–54. doi:10.1007/s00216-011-5631-x
20. Kneipp K, Kneipp H, Itzkan I. Surface-enhanced Raman scattering: a new tool for biochemistry spectroscopy. *Curr Sci*. 1999;77:915.
21. Chen Y, Zheng S, Tang X, et al. Recent research progress of surface-enhanced Raman scattering dominated analysis strategies in early diagnosis of diseases. *Chem Asian J*. 2013;18(12):e202300264. doi:10.1002/asia.202300264
22. Nie SM, Emerys R. Probing single molecules and single nanoparticles by surface-enhanced Raman scattering. *Science*. 1997;275(5303):1102–1106. doi:10.1126/science.275.5303.1102
23. Jiang H, Zhang X, Li LS. Preparation and SERS properties of Au nano-particles with different sizes. *Micronanoelect Technol*. 2019;56:107–110.
24. Yang F, Wen P, Zhang ZQ, et al. Fabrication of flexible surface-enhanced Raman spectroscopy chip. *Chin J Lasers*. 2021;48(1):0113001. doi:10.3788/CJL202148.0113001
25. Hu W, Xia H, Hu Y, et al. Recent progress on three-dimensional substrates for surface-enhanced Raman spectroscopic analysis. *Microchem*. 2022;172:106908. doi:10.1016/j.microc.2021.106908
26. Liu SP, Zhu HF, Chen N. Surface enhanced Raman scattering spectrum analysis of nude mouse serum with Au Nanoparticles active substrate. *Chin J Lasers*. 2012;39:137–140.
27. Gong LG, Ma J, Zhou HJ. Surface-enhanced Raman spectroscopy for gastric cancer and normal tissue. *Period Ocean Univ China*. 2013;43:099–103.
28. Teh SK, Zheng W, HO KY, et al. Diagnostic potential of near-infrared Raman spectroscopy in the stomach: differentiating dysplasia from normal tissue. *Br J Cancer*. 2008;98(2):457–465. doi:10.1038/sj.bjc.6604176
29. Hajjo R, Sabbah D, Bardaweel S, et al. Identification of tumor-specific MRI biomarkers using machine learning (ML). *Diagnostics*. 2021;11 (5):742. doi:10.3390/diagnostics11050742
30. Ryzhikova E, Ralbovsky NM, Sikirzhytski V, et al. Raman spectroscopy and machine learning for biomedical applications: alzheimer's disease diagnosis based on the analysis of cerebrospinal fluid. *Spectrochim Acta A*. 2020;248:1386–1425.
31. Gao NN, Wang Q, Tang J, 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):4775–4784. doi:10.1007/s00216-021-03431-3
32. Leong Y, Lee Y, Koh C, et al. Surface-Enhanced Raman Scattering (SERS) Taster: a Machine-Learning-Driven Multireceptor Platform for Multiplex Profiling of Wine Flavors. *Nano Lett*. 2021;21(6):2642–2649. doi:10.1021/acs.nanolett.1c00416
33. Ciloglu FU, Saridag AM, Kilic IH, et al. Identification of methicillin-resistant Staphylococcus aureus bacteria using surface-enhanced Raman spectroscopy and machine learning techniques. *Analyst*. 2020;145(23):7559–7570. doi:10.1039/D0AN00476F
34. Li H, Hassan M, Wang J, et al. Investigation of nonlinear relationship of surface enhanced Raman scattering signal for robust prediction of thiabendazole in apple. *Food Chem*. 2021;339:127843. doi:10.1016/j.foodchem.2020.127843
35. Hyunku S, Jeong H, Park J, Hong S, Choi Y. Correlation between cancerous exosomes and protein markers based on surface-enhanced Raman spectroscopy (SERS) and principal component analysis (PCA). *ACS Sens*. 2018;3(12):2637–2643. doi:10.1021/acssensors.8b01047
36. Guo L, Li Y, Huang F, et al. Identification and Analysis of Serum Samples by Surface-Enhanced Raman Spectroscopy Combined with Characteristic Ratio Method and PCA for Gastric Cancer Detection. *J Innov Opt Health Sci*. 2019;12(02):1950003. doi:10.1142/S1793545819500032
37. Lim J, Nam J, Shin H, et al. Identification of Newly Emerging Influenza Viruses by Detecting the Virally Infected Cells Based on Surface Enhanced Raman Spectroscopy and Principal Component Analysis. *Analy Chem*. 2019;91(9):5677–5684. doi:10.1021/acs.analchem.8b05533