Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Label-Free Diagnosis of Colorectal Cancer Using a SERS Platform Based on a Modified Glass Fiber Filter Membrane Combined with Machine Learning

Mengting Lin
Fujian Normal University
Qianmin Tu
tuqianmin985@outlook.com

Fujian Normal University
Wenlong Xiang
Fujian Normal University
Yichen He
Fujian Normal University
Sinan Feng
Fujian Normal University
Dunyan Chen
Fuzhou University Affiliated Provincial Hospital
Jianwei Wang
Fuzhou University Affiliated Provincial Hospital
Yudong Lu
Fujian Normal University
Ruiyun You
Fujian Normal University
Lizhi Li
Shengli Clinical Medical College of Fujian Medical University

---

Research Article

Keywords: Colorectal cancer, Surface-enhanced Raman scattering (SERS), 4-Mercaptophenol, Machine learning, Label-free liquid biopsy

Posted Date: July 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10018144/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Non-invasive screening is critically important for the early diagnosis of colorectal cancer (CRC). Surface-enhanced Raman scattering (SERS) has shown great potential in liquid biopsy, yet it still faces the challenge of poor signal reproducibility in complex serum matrices. In this study, a flexible three-dimensional SERS platform based on a 4-mercaptophenol (4-MP)-modified glass fiber filter membrane (GFF@4MP) was developed and combined with machine learning for label-free discrimination of CRC serum. Through sodium chloride modification, silver nanoparticles were uniformly immobilized onto the three-dimensional porous scaffold. Crucially, 4-MP, serving as a surface functionalization modifier, was able to modulate the interfacial physicochemical microenvironment and provide specific binding sites. This optimized the ordered adsorption of serum targets within SERS “hot spots”, significantly reduced non-specific interference, and improved spatial uniformity (relative standard deviation of 7.63%). In the analysis of clinical serum samples (30 CRC cases and 30 healthy controls), the substrate modified with 4-MP functional groups yielded SERS spectra with superior intergroup discrimination. Combined with principal component analysis (PCA), the support vector machine (SVM) model achieved the best classification performance on an independent test set, with a prediction accuracy of 93.33% and an area under the ROC curve (AUC) as high as 0.988. This study effectively enhanced the stability of SERS detection in complex biological samples through interfacial chemical modification, thereby providing a reliable methodological strategy for non-invasive clinical screening of colorectal cancer.

1 Introduction

Colorectal cancer (CRC) is one of the digestive tract malignancies with high incidence and mortality rates worldwide[1]. Due to the lack of specific clinical symptoms in early-stage CRC, most patients are diagnosed at intermediate or advanced stages, resulting in a poor overall prognosis. Currently, routine clinical screening methods for CRC mainly include colonoscopy, fecal occult blood test (FOBT), and serum tumor marker (e.g., carcinoembryonic antigen, CEA) detection[2]. Although colonoscopy is regarded as the "gold standard" for diagnosis, patient compliance is limited by its invasive nature and cumbersome bowel preparation [3] ; meanwhile, FOBT and serum marker tests exhibit relatively insufficient diagnostic sensitivity and specificity at early stages. Therefore, the exploration of non-invasive, highly sensitive, and easily clinically translatable liquid biopsy technologies is of significant practical importance for the early detection and clinical intervention of CRC.

Surface-Enhanced Raman Scattering (SERS) spectroscopy can provide molecular fingerprint-level vibrational information and has demonstrated application value in label-free serum analysis and biochemical detection. As a complex body fluid containing proteins, nucleic acids, lipids, and various metabolites, the overall SERS spectrum of serum can reflect subtle changes in the metabolic profile under pathological conditions[4]. However, in actual label-free detection, conventional rigid metallic substrates face certain limitations. The bare surface of noble metal nanoparticles readily induces non-specific adsorption and uncontrollable aggregation of macromolecular proteins in the serum matrix, while the distribution of electromagnetic "hot spots" within the substrate is often random. These factors

tend to increase the spatial heterogeneity of Raman signals and reduce batch-to-batch reproducibility, thereby restricting the reliability of SERS technology in the analysis of complex biological samples.

To improve the spectral stability and reproducibility of substrates, the construction of three-dimensional flexible scaffolds with uniform pores and targeted chemical modification of metal surfaces represent effective strategies in current research. Ultrafine glass fiber filter membranes (GFF) possess a high specific surface area and a porous network structure [5], providing attachment sites for the three-dimensional distribution of silver nanoparticles and reducing local nanoparticle aggregation. More importantly, surface functionalization modification can actively regulate the interfacial physicochemical environment of the substrate. In this study, 4-mercaptophenol (4-MP) was introduced as a modifying molecule. Its thiol group (-SH) forms a stable "S-Ag" covalent bond with the silver surface, thereby constructing a self-assembled monolayer on the particle surface[6]. The exposed benzene ring structure and phenolic hydroxyl group (-OH) of the 4-MP molecule not only modulate the hydrophilicity/hydrophobicity and interfacial charge of the nanoparticle surface [7], but also optimize the adsorption and localization of specific target molecules in serum (e.g., aromatic ring-containing amino acid residues or lipid molecules) within SERS "hot spots" through mechanisms such as hydrogen bonding and π-π interactions [8]. This functional group-based interfacial modification can circumvent, to some extent, disordered non-specific binding, thereby enhancing the uniformity and regularity of SERS spectral signals in complex matrix systems.

Based on the above mechanisms, a three-dimensional SERS platform based on a 4-mercaptophenol-modified glass fiber filter membrane (GFF@4MP) was constructed and combined with machine learning algorithms for rapid label-free discrimination of colorectal cancer serum samples. In this study, GFF was first modified with sodium chloride to promote the uniform immobilization of Ag NPs@4-MP, and subsequently, serum SERS spectral datasets were acquired from CRC patients and healthy controls (HC). To achieve high-dimensional data feature extraction and accurate classification, principal component analysis (PCA) was employed for spectral dimensionality reduction, and the classification performances of four machine learning algorithms—Decision Tree (DT), K-Nearest Neighbors (KNN), Random Forest (RF), and Support Vector Machine (SVM)—were systematically evaluated. This study provides a methodological reference based on interfacial chemical regulation for non-invasive clinical screening of colorectal cancer using SERS technology.

2 Materials and methods

2.1 Materials and instrumentation

Silver nitrate ($AgNO_3$) was purchased from Sinopharm Chemical Reagent Co., Ltd. Rhodamine 6G (R6G) and 4-mercaptophenol (4-MP) were obtained from Shanghai Macklin Biochemical Co., Ltd. Ultrafine glass fiber filter membranes (GFF) were supplied by Zibo Kexing Yunkai Environmental Protection Technology Co., Ltd. All experiments were carried out using deionized ultrapure water with a resistivity of 18.2 MΩ·cm.

The following instruments were used: a UV-Vis spectrophotometer (UV-5500PC, Shanghai Yuanyxi Instrument Co., Ltd.); a transmission electron microscope (TEM) (JEOL JEM-F200, Japan); an X-ray photoelectron spectrometer (XPS) (Thermo Scientific K-Alpha, USA); a scanning electron microscope (SEM) (ZEISS Sigma 360, Germany); and a portable Raman spectrometer (RMS1000, Zolix Instruments Co., Ltd., Shanghai, China).

2.2 Preparation of the modified filter membrane SERS platform

2.2.1 Preparation of Ag NPs@4-MP

In this study, the synthesis of silver nanoparticle (Ag NPs) colloidal solution followed the classical sodium citrate reduction method as described by Xu's team. The detailed procedure was as follows: 0.017 g of silver nitrate ($AgNO_3$) powder was accurately weighed and placed into a conical flask. After adding 100 mL of deionized water, the mixture was heated to vigorous boiling under continuous stirring. Subsequently, 5 mL of 0.034 M trisodium citrate dihydrate solution was rapidly added, and the heating was maintained for an additional 1 hour. The completion of Ag NP synthesis was indicated by the observation of a characteristic gray-green color in the mixture and the absence of a silver mirror byproduct on the inner walls of the reaction vessel. The resulting silver colloid was then sealed and transferred to storage at 4°C.

The construction of the Raman reporter molecule-modified composite system (Ag NPs@4-MP) was carried out as follows: 5 mL of the above Ag NP sol was accurately pipetted and purified by high-speed centrifugation at 8000 rpm for 10 minutes. After discarding the supernatant, the precipitate was redispersed in an equal volume of deionized water. Then, 100 μL of a pre-prepared ethanol solution of 4-mercaptophenol (4-MP) (at a concentration of 10^{-5} M) was slowly added to the purified silver nanoparticle colloid. The mixture was continuously stirred for 1 hour at room temperature to ensure sufficient adsorption and self-assembly of the molecules on the silver surface. The finally obtained Ag NPs@4-MP functionalized probe solution was also stored in a refrigerator at 4°C for later use.

2.2.2 Preparation of GFF@4MP

During the specific assembly process, the GFF membrane was first precisely cut into circular pieces with a diameter of 0.4 cm. Subsequently, these pieces were completely immersed in a 0.3 wt% sodium chloride (NaCl) solution for 30 minutes of incubation. This operation was aimed at achieving surface modification of the substrate, ensuring that chloride ions (Cl^-) could be effectively enriched and loaded onto the GFF network scaffold. Next, the pre-prepared Ag NPs@4-MP probe solution was uniformly drop-coated onto the surface of the Cl^--modified GFF units. Finally, the material was dehydrated using vacuum drying technology, thereby ensuring that the active Ag NPs@4-MP system could be firmly and stably anchored and immobilized within the GFF substrate.

2.3 Characterization and performance testing of GFF@4MP

2.3.1 Characterization of Ag NPs@4-MP

Regarding the physicochemical characterization of the materials, transmission electron microscopy (TEM) coupled with energy-dispersive X-ray spectroscopy (EDS) was employed to comprehensively determine and analyze the morphology, physical dimensions, and elemental composition of the Ag NPs@4-MP functionalized probes. Meanwhile, to investigate the interfacial electrical properties and dispersion state of the system, the surface charge characteristics were quantitatively evaluated via Zeta potential measurements, and the particle size distribution of this composite nanosystem was further finely probed by means of particle size analysis.

2.3.2 Characterization of GFF@4MP

For the physical characterization and performance testing of the GFF@4MP composite substrate, the specific methods were deployed as follows:

In terms of morphology and composition characterization, scanning electron microscopy (SEM) coupled with energy-dispersive X-ray spectroscopy (EDS) was employed to spatially image and determine the microscopic scaffold structure and elemental distribution of the substrate. Meanwhile, X-ray photoelectron spectroscopy (XPS) was used to perform in-depth characterization of the surface chemical composition and the valence states of the major elements within the system.

For the evaluation of the SERS performance of the substrate, the classical probe molecule R6G was selected to excite Raman signals. To rigorously assess the spatial uniformity of the signals, 50 independent detection points were randomly selected on the same GFF@4MP substrate unit for spectral scanning. Batch-to-batch reproducibility was evaluated by parallel measurements on three independently prepared batches of substrate units, with 10 consecutive Raman datasets collected from each unit. Furthermore, a series of gradient concentrations of R6G standard solutions were drop-coated onto the substrate surface for spectral collection, and a quantitative linear regression curve correlating the absolute peak intensity with the concentration was established, thereby determining the detection sensitivity of the substrate. All of the above SERS spectra were recorded under identical physical parameters: an excitation wavelength of 785 nm, a laser output power of 30 mW, an integration time of 2 s, and a spectral acquisition window covering 400–1800 cm^{-1}.

2.4 Collection of colorectal cancer serum samples

Human serum samples (30 from healthy individuals and 30 from colorectal cancer patients) were provided by the Department of Oncology, Fuzhou Provincial Hospital, Fuzhou, Fujian Province, and the study was approved by the hospital's ethics committee (K2020–09–037). The hospital notified the

volunteers and patients, and written consent was obtained. The collected serum samples were stored in Eppendorf tubes at − 80°C.

2.5 SERS data acquisition of colorectal cancer serum samples

In the specific clinical blood sample analysis procedure, 40 μL of serum sample was first transferred into a 1.5 mL microcentrifuge tube. Subsequently, one GFF@4MP substrate unit was added to each tube, and the mixture was incubated statically at room temperature for 30 minutes to allow adsorption and sampling of target biomolecules. After sampling, the GFF@4MP composite substrate was carefully removed, dried by dehydration, and then used for Raman signal acquisition. To ensure the statistical representativeness of the detection data, 10 SERS spectra from different spots were recorded in parallel for each serum sample. Ultimately, a clinical spectral database comprising a total of 600 Raman maps was constructed. In the sample matrix, the healthy control group and the colorectal cancer patient group each contributed 300 high-quality spectra. All SERS measurements were performed under identical optical conditions: an excitation wavelength of 785 nm, a laser power of 30 mW, an integration time of 5 s per signal acquisition, and an effective spectral scanning window covering 400–1800 cm^{-1}.

2.6 Machine learning algorithm analysis

In order to conduct deep mining and classification of the collected SERS spectral dataset, this research built a machine learning architecture encompassing data normalization preprocessing, Principal Component Analysis (PCA) feature dimension reduction, and multimodal algorithms including Decision Tree (DT), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Random Forest (RF). In the dataset division, to guarantee objectivity and prevent the risk of overfitting brought by the cross-exposure of spectra from the same patient, this work strictly divided the 600 internal spectra into a training set and a test set at a ratio of 8:2 according to individual patient identity. Lastly, to truly assess the clinical generalization performance of the models, this study further introduced an independent external validation set composed of an entirely new set of 42 serum samples (21 colorectal cancer patients and 21 healthy individuals). By employing the aforementioned 60 internal samples for the final training of the models and computing the classification accuracy of the four models on the external validation set that was completely un-involved in the construction, an unbiased evaluation of the clinical translational robustness of the multimodal algorithm pipeline was realized.

3. Results and discussion

3.1 Characterization of Ag NPs@4-MP

Noble metal plasmonic nanomaterials serve as the core foundation for constructing high-performance SERS sensing interfaces. Given the excellent electromagnetic field enhancement effect of Ag NPs, a classical sodium citrate reduction method was employed to prepare Ag NP sol, followed by the introduction of 4-MP molecules through surface self-assembly to construct functionalized silver probes

(Ag NPs@4-MP). UV-Vis absorption spectroscopy confirmed that both nanosystems exhibited broad and intense absorption peaks at approximately 410 nm (Fig. 1.A). This spectral feature is characteristic of a typical localized surface plasmon resonance (LSPR) response, which is in good agreement with the optical properties of silver nanomaterials reported in the literature [9, 10]. Meanwhile, dynamic light scattering (DLS) measurements further revealed the hydrodynamic size and dispersion state of the systems: the average particle sizes of Ag NPs and Ag NPs@4-MP were approximately 91.03 nm and 83.35 nm, respectively, indicating that both colloids possessed good particle size uniformity (Fig. 1.B) [11]. Regarding the interfacial electrical properties, Zeta potential measurements showed that the surface potentials of Ag NPs and Ag NPs@4-MP were − 30.33 mV and − 30.9 mV, respectively (Fig. 1.C). This significant surface negativity was mainly attributed to the high enrichment and dense adsorption of citrate anions around the silver cores during the synthesis process.

From a morphological perspective, TEM images revealed that the prepared nanoparticles generally exhibited a regular spherical structure, with the physical size of individual particles mainly concentrated around 50 nm (Fig. 1.D). During the functionalization modification stage, 4-MP molecules were firmly anchored onto the silver particle surfaces through high-affinity "S-Ag" covalent bonds. The simultaneous detection of silver (Ag) and sulfur (S) elements by EDS provided direct chemical evidence for the successful assembly of the 4-MP modification layer. It is worth emphasizing that the introduction of this surface functional group modification layer effectively regulated the physicochemical microenvironment at the particle interface, fundamentally suppressing signal fluctuations caused by complex matrices, and thus providing a crucial material foundation for the subsequent construction of a high-precision intelligent diagnostic model. In summary, the multidimensional characterization data cross-validated that the silver nanoprobes prepared in this work not only possessed excellent colloidal dispersion stability but also exhibited strong LSPR optical amplification capability.

3.2 Physicochemical characterization of the GFF@4MP substrate

To achieve efficient loading of plasmonic nanoparticles, the GFF was first subjected to surface pretreatment using a NaCl solution. This process introduced chloride ions (Cl^-) into the fiber scaffold, significantly reducing the electrostatic repulsion energy barrier between the GFF substrate and the silver nanoprobes, thereby driving the stable deposition of Ag NPs and Ag NPs@4-MP onto the filter membrane [12]. The subsequently introduced vacuum drying process not only effectively prevented the oxidative deactivation of the silver nanoparticles but also further strengthened their anchoring stability onto the solid substrate. SEM morphological observation (Fig. 2.A) intuitively confirmed the effectiveness of this strategy, showing a large number of nanoparticles densely attached to the fiber network. Concurrent EDS measurements were in good agreement with the morphological images. The detection of characteristic elements such as Cl, Ag, and S in the GFF@4MP composite system conclusively verified, from the perspective of elemental composition, the pre-modification effect of the substrate and the successful immobilization of the 4-MP functionalized silver probes.

To explore the interfacial assembly mechanism, the intrinsic material of GFF is borosilicate glass microfibers, the surface of which is naturally enriched with negative charges. Referring to the Zeta potential results measured above, the negatively charged Ag NPs or Ag NPs@4-MP also exhibited moderate electrostatic repulsion with the fiber scaffold. This like-charge repulsion effect cleverly inhibited the disordered aggregation of nanoparticles at the microscopic scale, promoting a highly uniform dispersed state within the three-dimensional porous filter membrane, thus laying the foundation for the macroscopic signal uniformity of the substrate. When silver particles with an average size of approximately 50 nm were uniformly interspersed and dispersed within the interlaced fiber network, dense nanoscale gaps were constructed between adjacent particles. These microscopic gaps served as intense electromagnetic "hot spots", greatly inducing the coupling and superposition of localized surface plasmon resonance (LSPR) field enhancement, thereby endowing the composite platform with excellent optical signal amplification efficiency [13].

To further investigate the chemical composition and elemental valence states at the material interface, in-depth X-ray photoelectron spectroscopy (XPS) characterization was performed on the samples (Fig. 2.B-C). The survey scan spectra revealed that the composite substrate system mainly contained six elements: O, C, Si, Ag, Na, and Cl. In the corresponding high-resolution fine spectra, the characteristic binding energies of the O 1s, C 1s, Na, Cl 2p, Si 2s, and Si 2p orbitals were precisely located at 531.97 eV, 284.89 eV, 497.03 eV, 198.96 eV, 153.20 eV, and 102.22 eV, respectively. It is worth emphasizing that the high-resolution spectrum of Ag 3d exhibited typical spin-orbit splitting doublets at 373.68 eV and 367.59 eV, which were assigned to the Ag $3d_{3}/_{2}$ and Ag $3d_{5}/_{2}$ orbitals, respectively [14]. The above chemical state information at the atomic level constituted strong evidence that the functionalized silver nanoprobes had been stably anchored onto the modified filter membrane scaffold.

3.3 SERS performance characterization of the GFF@4MP substrate

Detection sensitivity is a key parameter for evaluating the practical application potential and clinical translation value of SERS substrates. In this section, rhodamine 6G (R6G) was used as a standard Raman probe to comprehensively investigate the Raman enhancement activity of the modified composite membrane. Experimental data showed that both the GFF@Ag NPs and GFF@4MP substrates achieved a limit of detection (LOD) for R6G molecules as low as 5×10^{-6} M (Fig. 3.B and Fig. S2.B). Analysis of the obtained Raman spectra revealed clear fingerprint peaks of the probe molecule: the band at 616 cm^{-1}mainly originated from the in-plane bending deformation of the xanthene ring (e.g., symmetric C–C–C deformation); the peaks at 775 $cm^{-1}$and 1189 cm^{-1}were attributed to the out-of-plane and in-plane bending modes of C–H bonds, respectively; the band at 1315 cm^{-1}corresponded to C = C double bond stretching; while the prominent strong peaks at 1362 $cm^{-1}$and 1510 cm^{-1}were contributed by the stretching vibrations of C–C bonds on the aromatic ring[15]. To achieve accurate quantification of the target molecule, the characteristic peak intensity at 1510 cm^{-1} was selected to investigate its linear dependence on the logarithmic value of the R6G concentration. As shown in Fig. 3.A and Fig. S2.A, the

GFF@4MP substrate exhibited excellent linear fitting (coefficient of determination R^2 = 0.9993), while the fitting coefficient for GFF@Ag NPs (R^2 = 0.9945) also remained at a high level, fully demonstrating the highly reliable quantitative analysis capability of the functionalized substrate.

In practical detection, overcoming the random signal fluctuations caused by the uneven distribution of local electromagnetic "hot spots" is a prerequisite for eliminating matrix interference and improving data reliability. To this end, 50 random detection points were selected on the surfaces of both substrates loaded with the R6G probe for continuous spectral acquisition (Fig. 3.C, Fig. S2.C). Statistical analysis of the peak intensity at 1510 cm^{-1}showed that the relative standard deviation (RSD) of the signals on the GFF@4MP substrate was as low as 7.63% (Fig. 3.D), indicating extremely high spatial uniformity. In contrast, the unmodified GFF@Ag NPs substrate exhibited a relatively higher RSD of 10.19% (Fig. S2.D). This significant improvement in signal uniformity was essentially attributed to the enhanced electrostatic repulsion among the Ag NPs@4-MP system, which promoted a denser and more highly uniform anchoring and dispersion of the plasmonic nanoparticles on the glass fiber network scaffold.

Furthermore, evaluating the reliability of the SERS sensing platform also requires focused consideration of its batch-to-batch reproducibility and stability. In the reproducibility experiments, R6G was again used as the target analyte, and parallel comparisons of Raman response consistency were conducted on three independently prepared batches of GFF@Ag NPs and GFF@4MP substrates (Fig. 3.E, Fig. S2.E). By extracting and comparing the absolute signal intensities of three characteristic peaks (Fig. 3.F, Fig. S2.F), the results confirmed that although there were very small and reasonable signal drifts between different batches, no statistically significant differences were observed at the overall level. This phenomenon strongly supports the high controllability of the current preparation process, ensuring the excellent batch stability and high reproducibility of the developed SERS substrate in long-term detection applications.

3.4 Raman spectroscopy of the GFF@4MP substrate and statistical analysis of colorectal cancer serum

The Raman fingerprint region (400–1800 cm^{-1}) is rich in high-resolution molecular vibrational information and can comprehensively map the characteristics of the biomacromolecular network in serum samples. As shown in Fig. 4.A, the SERS spectra of healthy control (HC) and colorectal cancer (CRC) serum collected on the GFF@4MP substrate exhibited significant molecular spectral differences, with the core features mainly dominated by the vibrational modes of proteins, lipids, and nucleic acids [16, 17]. Specifically, a series of characteristic peaks closely related to amino acids and protein secondary structures were observed in the spectra: the band at 485 $cm^{-1}$was assigned to the ring deformation vibration of L-arginine; the peak at 645 cm^{-1}corresponded to the C–S single bond stretching of tyrosine; and the strong peak at 1086 cm^{-1}originated from the aromatic ring stretching of phenylalanine. In addition, the doublet peak at 1217 cm^{-1}reflected the aromatic ring vibrations of L-tryptophan and phenylalanine, while the band at 1585 cm^{-1}suggested the C = C stretching and CH_2 deformation of hydroxyproline. The characteristic signal of nucleic acid components appeared at 728

cm^{-1}, originating from the ring vibrations of adenine and hypoxanthine. In the mid-to-high frequency region, features of lipid and carbohydrate metabolites gradually emerged, including bands at 1132 cm^{-1}(C-C backbone vibration), 1341 cm^{-1}(CH_3/CH_2 twisting in collagen or lipids), and 1438 cm^{-1}(CH_2 deformation of N-acetylglucosamine)[18–20]. The intensity and shift fluctuations of these specific bands systematically revealed the metabolic reprogramming phenomenon in the tumor microenvironment, the underlying mechanisms of which may involve cancer-induced amino acid metabolic disorders, nucleic acid conformational variations, and lipid membrane remodeling accompanying malignant transformation. These differential features strongly confirmed the application potential of this interface-modified SERS platform in non-invasive tumor diagnosis.

To extract core features from the high-dimensional spectral data and achieve dimensionality reduction for visualization, principal component analysis (PCA) was introduced in this study. Figure 4 shows the three-dimensional sample distribution space constructed based on the first three principal components (PC1, PC2, and PC3). On the unmodified bare silver substrate (GFF@Ag NPs), the variance contribution rates of PC1, PC2, and PC3 were 53.79%, 15.46%, and 10.18%, respectively, with a cumulative explained variance of 79.43% (Fig. 4.B-E). However, within this principal component space, the sample points of CRC patients and HC volunteers exhibited substantial overlap. This indicated that the original substrate, due to its high susceptibility to non-specific adsorption, led to relatively large random signal fluctuations and thus had limited intrinsic discriminative ability for the two types of samples. In contrast, the GFF@4MP substrate modified with 4-MP functional groups showed significantly optimized clustering effects (Fig. 4.F-I). The cumulative explained variance of its first three principal components increased to 81.70% (with PC1 accounting for 53.02%, PC2 for 20.89%, and PC3 for 7.79%). More importantly, in this projection space, the spatial separation between the CRC and HC groups was significantly enhanced, and the interclass overlap region was greatly reduced. This notable improvement was mainly attributed to the microenvironmental regulation of the complex serum matrix by the 4-MP interfacial modification layer; it effectively suppressed disordered background noise, amplified disease-specific metabolite differences, and thus greatly enhanced the class discriminability of the spectral data, laying a solid foundation for the subsequent construction of a high-precision intelligent classification model.

Based on four machine learning algorithms, namely Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbors (KNN), and Decision Tree (DT), serum classification models were constructed for the GFF@Ag NPs and GFF@4MP substrates, respectively. The predictive performances on healthy control (HC) and colorectal cancer (CRC) samples were intuitively displayed through confusion matrices (Fig. 5 and Fig. S3). The results showed that when the unmodified bare silver substrate (GFF@Ag NPs) was used, the classification accuracies of the above four models on the independent test set were only 57.50% (DT), 66.67% (KNN), 80.83% (RF), and 88.33% (SVM), respectively. In contrast, owing to the microenvironmental regulation of the substrate by 4-MP interfacial chemical modification and the normalization of spectral signals, the diagnostic accuracy of the GFF@4MP platform was comprehensively improved. The corresponding test set accuracies were significantly increased to 82.50% (DT), 78.33% (KNN), 90.00% (RF), and 93.33% (SVM), respectively. This stark contrast

conclusively demonstrated that the functional group modification strategy could universally and effectively enhance the intrinsic separability of the spectral data, thereby improving the recognition accuracy of various algorithms. Furthermore, it indicated that this interfacial engineering played a crucial role in mitigating random signal fluctuations and optimizing model convergence performance. In particular, the SVM and RF models exhibited excellent classification robustness with the modified substrate, greatly reducing the risk of misclassification and highlighting the effectiveness of this diagnostic paradigm.

To further quantitatively evaluate the comprehensive diagnostic efficacy of each classification model in practical disease screening, the receiver operating characteristic (ROC) curve was introduced as an evaluation tool. The area under the ROC curve (AUC), as a core statistical metric, objectively and comprehensively reflects the balance between sensitivity and specificity of a classifier in distinguishing cancer patients from healthy individuals. For the high-quality spectral dataset acquired using the GFF@4MP substrate, systematic ROC analysis (see Fig. 5 and Table 1) revealed the excellent predictive boundaries of the four algorithms: the AUC values of the DT and KNN models reached 0.841 and 0.836, respectively, indicating good discriminative potential; while the RF and SVM models achieved very high AUC scores of 0.958 and 0.988, respectively. This series of evaluation metrics not only confirmed the optimal matching of SVM to the current dataset but also strongly reaffirmed the great clinical value of GFF@4MP as an intelligent SERS liquid biopsy platform for non-invasive precision screening of colorectal cancer.

To further evaluate the true clinical generalization performance and robustness of the established models, an independent external validation set comprising an entirely new set of 42 serum samples (21 CRC patients and 21 healthy controls, yielding a total of 420 SERS spectra) was tested. In this phase, the models were trained on the complete internal dataset of 600 spectra before predicting the unseen external samples. As depicted in Fig. S4, the models based on the GFF@4MP substrate maintained reliable predictive capabilities on the external cohort. The classification accuracies on the independent test set were 70.24% for DT, 75.95% for KNN, 80.00% for RF, and 85.48% for SVM. Correspondingly, the AUC values were 0.746, 0.825, 0.929, and 0.924, respectively. Consistent with the internal testing outcomes, the SVM and RF algorithms exhibited superior translational robustness in processing complex clinical spectral features, significantly outperforming DT and KNN. The strong external validation metrics, particularly the high accuracy of SVM (85.48%) and the excellent AUC of RF (0.929), firmly substantiate the reliability and clinical translatability of the GFF@4MP-based SERS platform for non-invasive colorectal cancer screening.

In summary, the deep integration of surface-enhanced Raman scattering (SERS) with machine learning algorithms demonstrates great application prospects in clinical assisted diagnosis. As shown by the complete performance metrics and statistical comparison results of the four types of models (DT, KNN, RF, SVM) detailed in Table 1, the SVM and RF algorithms exhibited superior generalization ability and classification accuracy when processing high-dimensional clinical spectral features, whereas the discriminative efficacy of DT and KNN was relatively limited. More importantly, compared with the

unmodified bare substrate, the introduction of 4-MP for interfacial functional group modification led to a systematic and comprehensive improvement in the classification efficacy and prediction confidence of all algorithms. This experimental finding indirectly confirms that the 4-MP modification layer, by actively regulating the physicochemical microenvironment of the substrate, fundamentally overcomes the problems of spectral mutation and signal distortion caused by the heterogeneous distribution of electromagnetic "hot spots" in traditional SERS substrates. This methodological innovation based on interfacial chemical engineering not only greatly enhances the reliability of data mining and the ultimate diagnostic confidence but also further highlights the practical clinical translation potential of this "spectral-algorithm" integrated detection platform for non-invasive early screening of colorectal cancer.

Table 1. Summary of results of machine learning algorithms

SERS substrate	Type		DT	KNN	RF	SVM
GFF@Ag NPs	Train	Accuracy（%）	92.50	97.29	100	95.63
	Test	Accuracy（%）	57.50	66.67	80.83	88.33
		AUC	0.504	0.670	0.822	0.859
GFF@4MP	Train	Accuracy（%）	95.83	92.29	99.79	96.46
	Test	Accuracy（%）	82.50	78.33	90.00	93.33
		AUC	0.841	0.836	0.958	0.988
	External Validation	Accuracy (%)	70.24	75.95	80.00	85.48
		AUC	0.746	0.825	0.929	0.924

4 Conclusion

In this study, a flexible three-dimensional SERS platform (GFF@4MP) based on functional modification with 4-mercaptophenol (4-MP) groups was successfully developed, and its clinical diagnostic potential for rapid label-free discrimination of colorectal cancer (CRC) serum was validated. The study confirmed that the introduction of 4-MP as a surface functionalization molecule could effectively reconstruct the interfacial physicochemical microenvironment of the silver nanoparticles. This modification layer not only optimized the charge distribution on the substrate surface but also guided the ordered enrichment

of serum target molecules within SERS "hot spots" through specific interactions, fundamentally mitigating the signal fluctuations and non-specific interference easily induced by complex biological fluid matrices. Based on this interfacial chemical regulation mechanism, the GFF@4MP substrate exhibited high spectral spatial uniformity and excellent batch-to-batch reproducibility. The results of clinical sample analysis demonstrated that the functional group-modified SERS platform, combined with principal component analysis (PCA) and machine learning algorithms, could accurately capture and discriminate the metabolic fingerprint differences in serum between healthy individuals and colorectal cancer patients. Among them, algorithm models such as the support vector machine (SVM) showed outstanding classification efficacy and generalization ability. In summary, this work not only confirmed the critical role of surface chemical engineering in improving the stability of SERS detection in complex samples but also provided a highly reliable methodological paradigm and intelligent diagnostic tool for non-invasive early liquid biopsy of colorectal cancer.

Declarations

Author information

Author notes

Mengting Lin and Qianmin Tu contributed equally to this work.

Authors and Affiliations

College of Chemistry and Materials Science, Fujian Provincial Key Laboratory of advanced Oriented Chemical Engineer, Fujian Key Laboratory of Polymer Materials, Engineering Research Center of Industrial Biocatalysis, Fujian Province Higher Education Institutes, Fujian Normal University, Fuzhou, 350007, Fujian, China

Mengting Lin, Qianmin Tu, Wenlong Xiang, Yichen He, Sinan Feng, Yudong Lu and Ruiyun You.

Department of General Surgery, Fuzhou University Affiliated Provincial Hospital, Fuzhou 350001, Fujian, China; Department of Emergency Surgery Shengli Clinical Medical College of Fujian Medical University, Fuzhou 350001, Fujian, China

Lizhi Li

Department of Clinical Laboratory, Fuzhou University Affiliated Provincial Hospital, Fuzhou, Fujian, China 350001

Dunyan Chen and Jianwei Wang

CRediT authorship contribution statement

Mengting Lin: Conceptualization, Methodology, Visualization, Validation, Data Curation and Writing–original draft. Qianmin Tu: is co-first author and contributes equally to this work. Wenlong Xiang: Visualization and Data curation. Yichen He: Validation and Investigation. Sinan Feng: Methodology and Visualization. Dunyan Chen: Visualization. Jianwei Wang: Methodology. Yudong Lu: Project administration and Validation. Ruiyun You: Supervision, Funding acquisition, Writing–review & editing and Conceptualization. Lizhi Li：Supervision, Resources, Funding acquisition.

Corresponding authors

Correspondence to Yudong Lu and Lizhi Li.

Ethics declarations

Ethics approval

In this study, the collection and use of human blood samples were approved by the Ethics Committee of Fujian Provincial Hospital. The study was conducted in accordance with the ethical principles of the World Medical Association's Declaration of Helsinki (1964) and the Council for International Organizations of Medical Sciences' International Ethical Guidelines for Biomedical Research Involving Human Subjects (2002).

Consent to Participate

Written informed consent was obtained from all participants prior to sample collection.

Clinical trial number

Not applicable.

Declaration of Competing Interest

The authors declare that there is no conflict of interest.

Funding

This work was supported by Fuzhou-Xiamen-Qian'an National Independent Innovation Demonstration Zone Collaborative Innovation Platform Project（2025E3002）2025-P-004）, Natural Science Foundation of Fujian Province (2025J01626), Fujian provincial education and scientific research special project (Min CAI No. [2024] 900, Min CAI No. [2025] 790 K20260002), Provincial Science and Technology Innovation Joint Fund Climbing Program（2024Y9020）and Fujian Normal University's "Innovation and Entrepreneurship Training Program for College Students” (cxxl-2026130).

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process.

During the preparation of this work, the author(s) used Gemini in order to improve the readability, grammar, and overall language of the manuscript. After using this tool, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

Rights and permissions

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

References

1. Wu Y, Xu P, Lin J, Li W, Lin Q, Lu H, Chen J, Liu Z, Lu Y (2026) pH-responsive sodium alginate/CMCS/CMCNa composite hydrogel beads for sustained delivery of 5-fluorouracil-β-cyclodextrin inclusion complexes. RSC Adv 16(27):25131–25143. https://doi.org/10.1039/D6RA00194G
2. Heyman H, Saraste D, Jonsson H, Blom J (2026) Colorectal Cancer Risk With Negative Colonoscopy or Nonadherence After Positive FOBT Screening. JAMA Netw Open 9(3):e262404–e262404. https://doi.org/10.1001/jamanetworkopen.2026.2404
3. Chuang MP-C, Hsu W-F, Su C-W, Ming-Fang A, Yen T, Hsiu-Hsi Chen C-Y, Hsu H-M, Chiu (2026) Fecal Hemoglobin-Guided Precision Postpolypectomy Surveillance Within the Current Framework: A Four-Million-Participant Fecal Immunochemical Test–Based Screening Study, Gastroenterology https://doi.org/https://doi.org/10.1053/j.gastro.2026.02.045
4. Wang H, Lin L, Lu S, Zhang L, Wu Y, You R, Wu Y, Lu Y (2026) Fabricating a SMVF@GO@Ag Composite Aerogel SERS Sensor for the Enrichment and Label-Free Detection of Urinary VOCs. Anal Chem 98(5):3792–3802. https://doi.org/10.1021/acs.analchem.5c05672
5. Mao J, Sun J, Wang L, Liu X, Bi J (2024) Flexible and high-strength bioactive glass fiber membrane for bone regeneration with the aid of alkoxysilane sol spinnability. Mater Today Bio 28:101224. https://doi.org/https://doi.org/10.1016/j.mtbio.2024.101224
6. You R, Huang Q, Lin Z, Wang W, Lie J, Chen J, Zhang G, Lu Y (2023) Preparation of SERS base membrane with cellulose compound dopamine and determination of hypochlorite. Microchim Acta 190(11):447. https://doi.org/10.1007/s00604-023-06006-4
7. Milani SG, Soares de Aguiar F, Júnior M, de Lemos S, Timsina R, de Sousa AG, Brolo (2025) Resonance Effects in Anomalous Anti-Stokes to Stokes Ratios in Surface-Enhanced Raman Scattering (SERS). J Phys Chem C 129(32):14542–14550. https://doi.org/10.1021/acs.jpcc.5c01830
8. Wang W, Ruan S, Su Z, Xu P, Chen Y, Lin Z, Chen J, Lu Y (2023) A novel on–off SERS nanoprobe based on sulfonated cellulose nanofiber-Ag composite for selective determination of NADH in human serum. Microchim Acta 190(7):254. https://doi.org/10.1007/s00604-023-05809-9

9. Lu D, Huang Z, Chen J, Lu Y, pH-Adjusted Liquid SERS (2025) Approach: Toward a Reliable Plasma-Based Early Stage Lung Cancer Detection. Anal Chem 97(1):508–515. https://doi.org/10.1021/acs.analchem.4c04671
10. Chen S, Wang H, Ren P, You R, Lin Q, Lu Y (2025) Green synthesis of ultra-stable silver nanoparticles modified with oxidized sodium alginate and high sensitivity detection of urine biomarkers. Int J Biol Macromol 320:145764. https://doi.org/https://doi.org/10.1016/j.ijbiomac.2025.145764
11. Yang J, Zhang B, Nie Q, Zheng R, Lin P, Wang C, Lu Y, Xu Y, You R (2024) Preparation of carboxymethyl cellulose membrane flexible SERS substrate and its application in the detection of metabolites and pH in urine. Int J Biol Macromol 283:137821. https://doi.org/https://doi.org/10.1016/j.ijbiomac.2024.137821
12. Cai L, Li M, Zan J, Wang H, Sun T (2025) Detection of thiram residues on apple surface by surface enhanced Raman spectroscopy with modified paper substrate. Food Chem 491:145302. https://doi.org/https://doi.org/10.1016/j.foodchem.2025.145302
13. Zhang B, Yan X, Lin P, Zheng R, You R, Wu Y, Chen S, Lu Y (2026) Cellulose-based hydrogel SERS sensor for label-free detection and identification of antithyroid drugs. Int J Biol Macromol 355:151515. https://doi.org/https://doi.org/10.1016/j.ijbiomac.2026.151515
14. Wang H, Nie Q, Lin L, You R, Ye Z, Zhang J, Lu Y, Lin Z, Yan Z (2026) Flexible wearable hydrogel patch integrated with a sponge-like cellulose SERS chip for sweat analysis of urea and lactate. Talanta 298:128977. https://doi.org/https://doi.org/10.1016/j.talanta.2025.128977
15. Feng P, Seo J, Li J, Jung HY, Jung YJ (2025) Advanced SERS sensor with horizontally aligned sub-5 nm silicon nanowires and high-density silver nanoparticles for ultra-sensitive molecular analysis. Biosens Bioelectron 286:117633. https://doi.org/https://doi.org/10.1016/j.bios.2025.117633
16. Zhang C, Tong Y, Wang H, Zhang S, Li Y, Niu D, Liu S, Luan D, Li P, Li J, Du L (2026) Deep learning-assisted metabolic fingerprint profiling based on V-groove and wrinkle-shaped 3D surface-enhanced Raman scattering substrate for early colorectal cancer diagnosis. Biosens Bioelectron 294:118183. https://doi.org/https://doi.org/10.1016/j.bios.2025.118183
17. Zhang C, Zhao W-H, Zuo D, Zhou T, Hou W, Wang L, Shi S, Yang Y, Liu Y, Sun S-K, Ren L, Ye Z, Liu D, Li D, Chen X, Hao J (2026) AI-Guided SERS Defines a Pan-Cancer Diagnostic Biomarker. Adv Sci 13(7):e16268. https://doi.org/https://doi.org/10.1002/advs.202516268
18. Chen J, Xu P, Ma W, Lu H, You R, Lu Y, Liu Z (2026) Integration of machine learning and SERS technology for precise classification and diagnosis of colon cancer from plasma samples. Spectrochim Acta Part A Mol Biomol Spectrosc 346:126930. https://doi.org/https://doi.org/10.1016/j.saa.2025.126930
19. Nie Q, Xu P, Liu J, Wang Y, Huang Q, You R, Liu Y, Li L, Lu Y (2025) Enhanced SERS detection of the colorectal cancer biomarker utilizing a two-dimensional silver substrate. Colloids Surf B 245:114317. https://doi.org/https://doi.org/10.1016/j.colsurfb.2024.114317
20. Zhu Q, Chen G, Fu L, Cao D, Wang Z, Yang Y, Wei W (2025) Machine learning-based SERS serum detection platform for high-sensitive and high-throughput diagnosis of colorectal precancerous

lesions. Bioeng Translational Med 10(5):e70019.
https://doi.org/https://doi.org/10.1002/btm2.70019

Figures

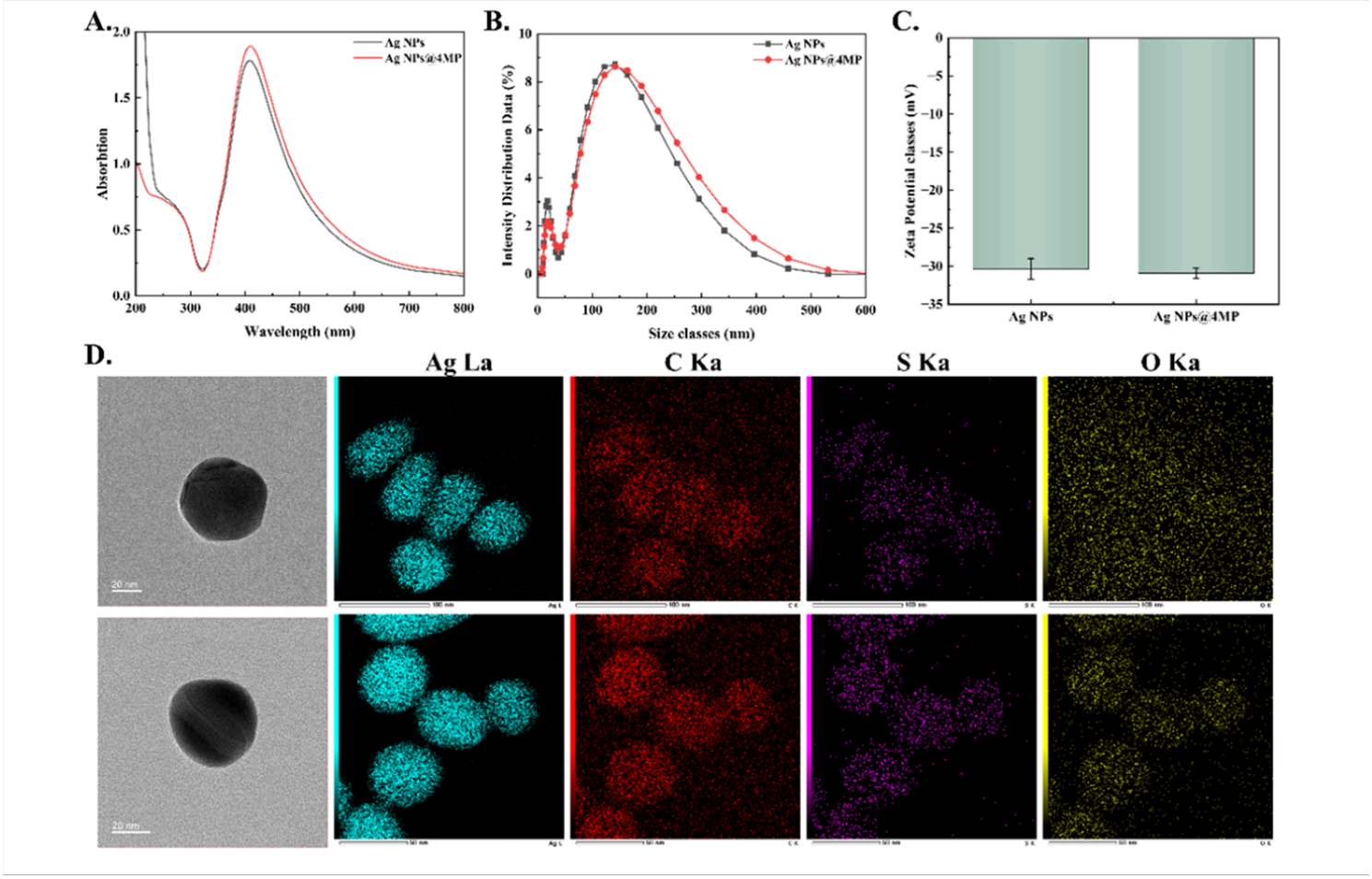


Figure 1

A. UV-Vis spectra of Ag NPs and Ag NPs@4-MP; B. Particle size distribution profiles of Ag NPs and Ag NPs@4-MP; C. Zeta potential plots of Ag NPs and Ag NPs@4-MP; D. TEM-EDS mapping of Ag NPs and Ag NPs@4-MP.

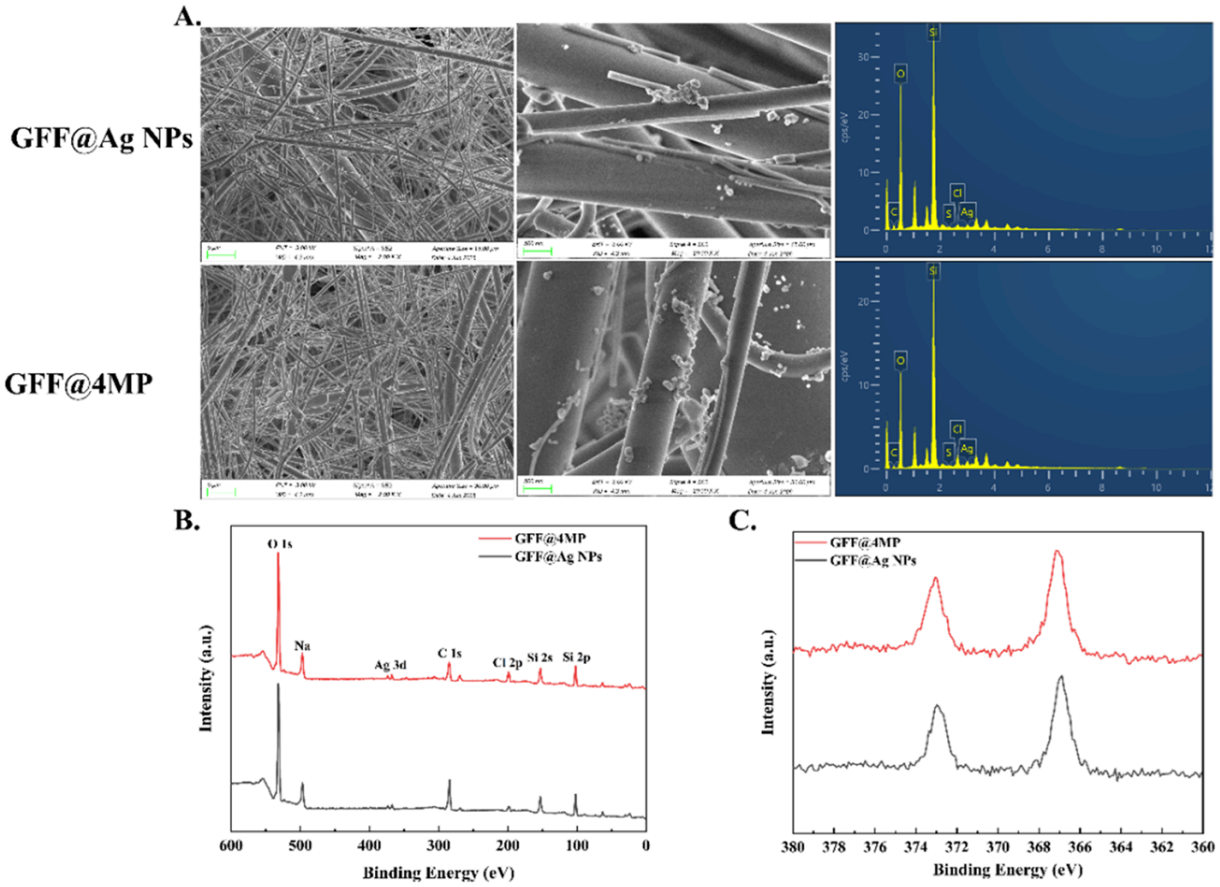


Figure 2

A. SEM images and EDS elemental spectra of GFF@Ag NPs and GFF@4MP substrates; B. XPS survey spectra of GFF@Ag NPs and GFF@4MP substrates; C. XPS high-resolution Ag 3d spectra of GFF@Ag NPs and GFF@4MP substrates.

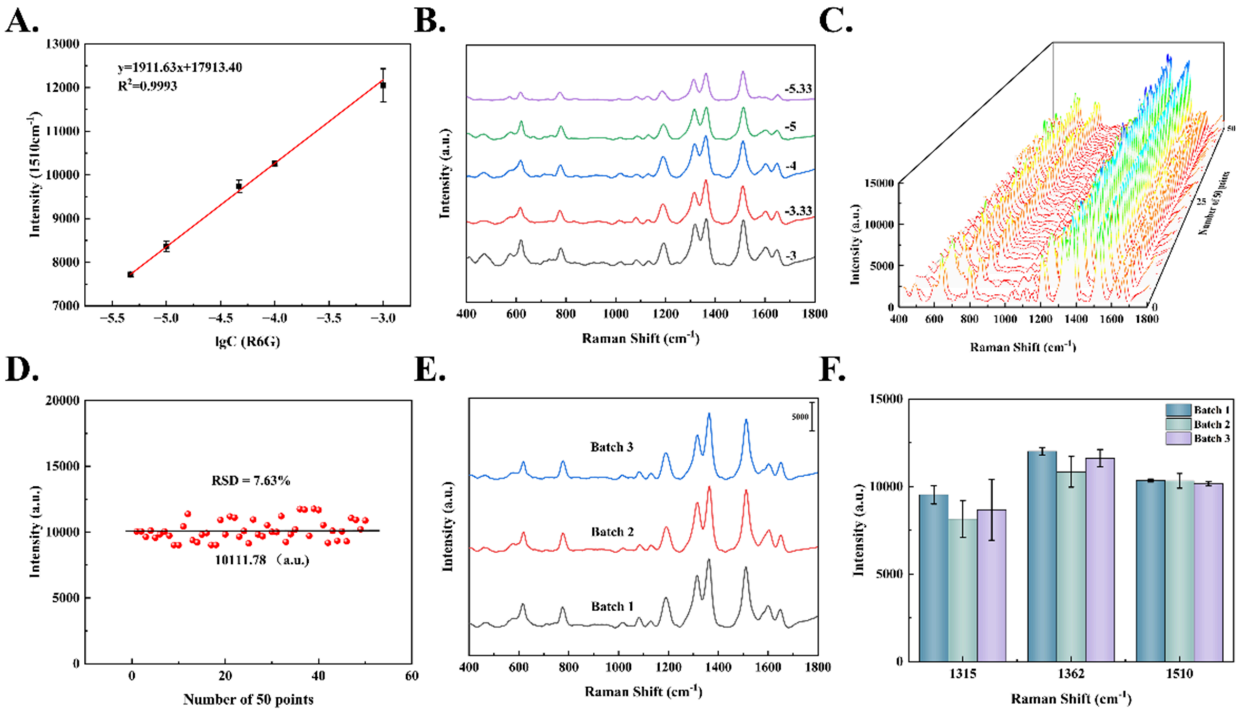


Figure 3

A-B. Sensitivity linear fitting plot and SERS spectra of the GFF@4MP substrate; C-D. SERS spectra and scatter plot of 50 random spots on the GFF@4MP substrate after adsorption of 10^{-4} M R6G; E-F. SERS spectra and SERS intensity bar chart of three random batches of GFF@4MP substrates after adsorption of 10^{-4} M R6G.

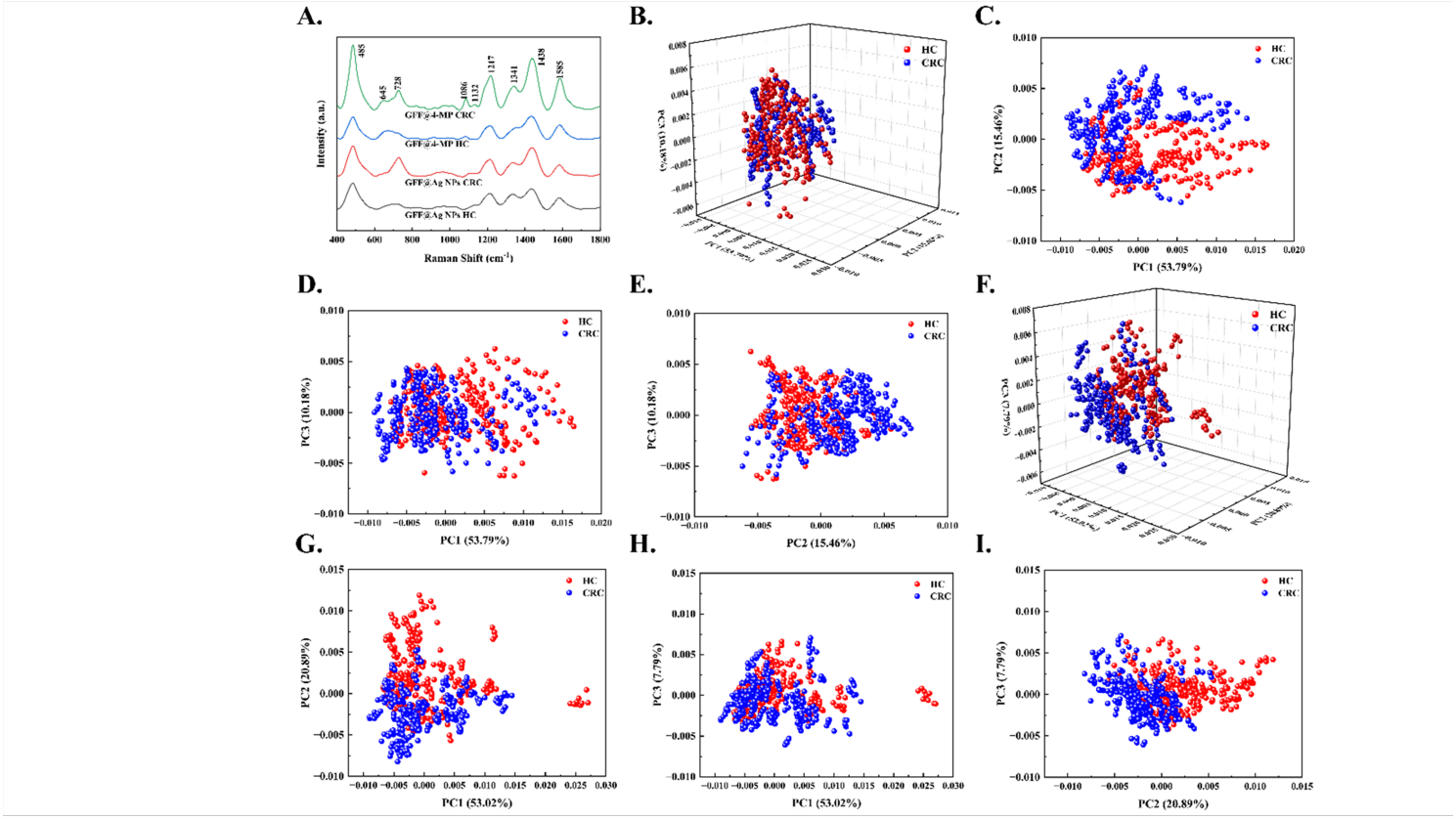


Figure 4

A. SERS spectra of HC and CRC serum detected on GFF@Ag NPs and GFF@4MP substrates; B. PCA scatter plot of the serum SERS dataset on the GFF@Ag NPs substrate, with PC1 vs. PC2 mapped in C, PC1 vs. PC3 mapped in D, and PC2 vs. PC3 mapped in E; F. PCA scatter plot of the serum SERS dataset on the GFF@4MP substrate, with PC1 vs. PC2 mapped in G, PC1 vs. PC3 mapped in H, and PC2 vs. PC3 mapped in I.

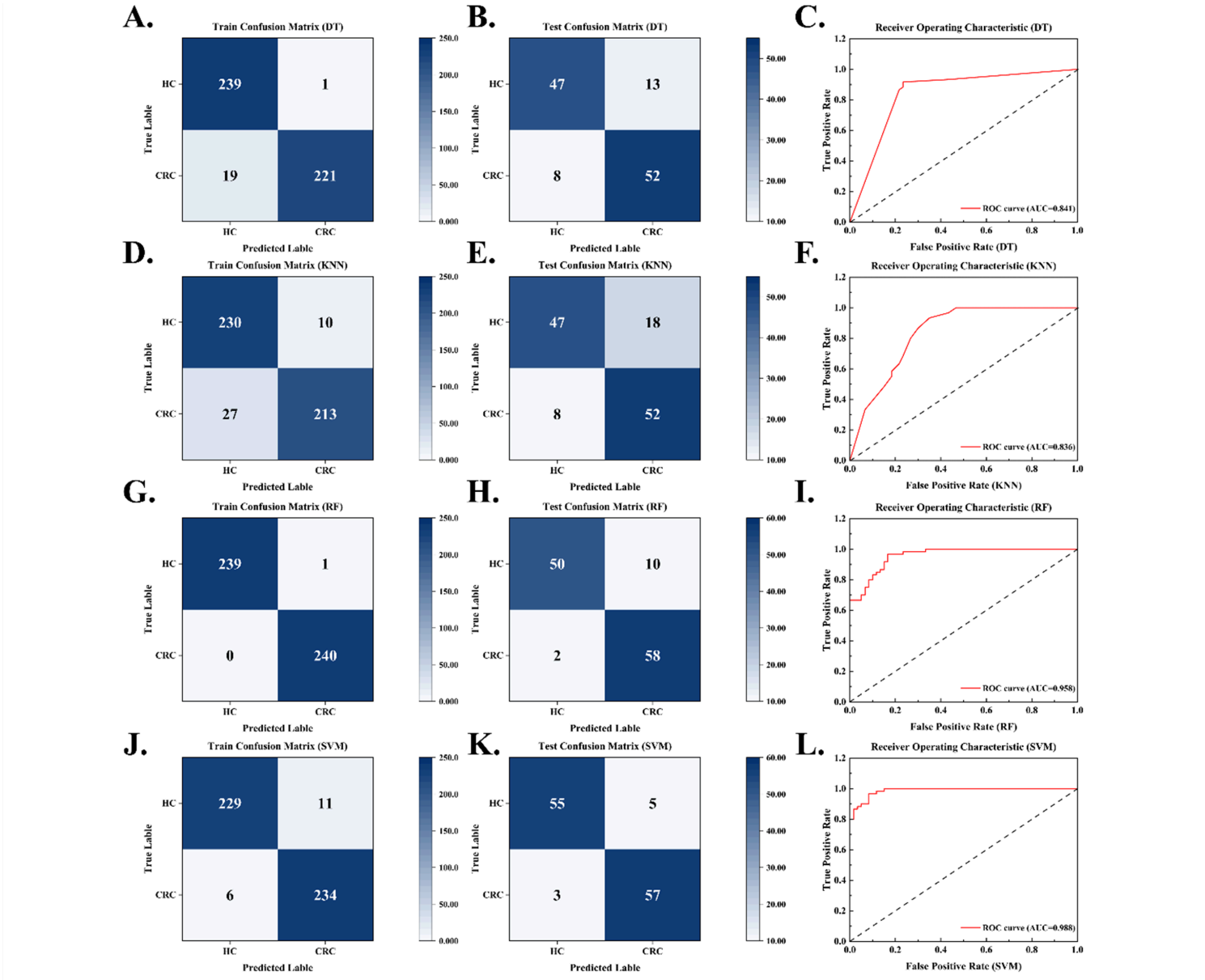


Figure 5

Confusion matrices and ROC curves of the training and test sets for the GFF@4MP substrate obtained using DT (A-C), KNN (D-F), RF (G-I), and SVM (J-L) algorithms, respectively.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterial.docx
- GA.png