Heliyon 10 (2024) e30499

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

High-performance grating-like SERS substrate based on machine learning for ultrasensitive detection of Zexie-Baizhu decoction

Wenying Zhou [a], Xue Han [b], Yanjun Wu [a], Guochao Shi [a,*], Shiqi Xu [a,**], Mingli Wang [c,***], Wenzhi Yuan [a], Jiahao Cui [a], Zelong Li [a]

[a] Hebei International Research Center for Medical-Engineering, Chengde Medical University, Chengde, 067000, Hebei, China
[b] Department of Neurology, Affiliated Hospital of Chengde Medical University, Chengde, 067000, Hebei, China
[c] State Key Laboratory of Metastable Materials Science and Technology, Key Laboratory for Microstructural Material Physics of Hebei Province, School of Science, Yanshan University, Qinhuangdao, 066004, China

ARTICLE INFO

Keywords:
SERS
Machine learning
Grating-like structure
Traditional Chinese medicine decoction
Zexie-Baizhu decoction

ABSTRACT

Rapid, universal and accurate identification of chemical composition changes in multi-component traditional Chinese medicine (TCM) decoction is a necessary condition for elucidating the effectiveness and mechanism of pharmacodynamic substances in TCM. In this paper, SERS technology, combined with grating-like SERS substrate and machine learning method, was used to establish an efficient and sensitive method for the detection of TCM decoction. Firstly, the grating-like substrate prepared by magnetron sputtering technology was served as a reliable SERS sensor for the identification of TCM decoction. The enhancement factor (EF) of 4-ATP probe molecules was as high as 1.90×10^7 and the limit of detection (LOD) was as low as 1×10^{-10} M. Then, SERS technology combined with support vector machine (SVM), decision tree (DT), Naive Bayes (NB) and other machine learning algorithms were used to classify and identify the three TCM decoctions, and the classification accuracy rate was as high as 97.78 %. In summary, it is expected that the proposed method combining SERS and machine learning method will have a high development in the practical application of multi-component analytes in TCM.

1. Introduction

The development of medication history is originated from the use of single drug, while the use of drug compatibility benefits from the continuous in-depth understanding of drugs and etiological mechanism. However, there is still a lack of knowledge about the composition of formulas and their therapeutic analyses, and the preparation of formulas of high quality and reliable therapeutic efficacy has become particularly important in order to further promote the process of modernization of TCM. Zexie-Baizhu Decoction (ZXBZD) is a Chinese typical formula that originates from the Synopsis of the Golden Chamber (Jingui Yaolue). The formula was prepared by decocting Zexie (Alisma Orientale) and Baizhu (Atractylodes macrocephala) with water according to the mass ratio of 5:2. At present, there are so many studies have shown that many monomers of Zexie and Baizhu, as well as Chinese herbal compounds containing Zexie, have therapeutic and palliative effects on lipid metabolism, insulin resistance and inflammation. Zexie has antiviral

* Corresponding author.
** Corresponding author.
*** Corresponding author.
E-mail addresses: sgc@cdmc.edu.cn (G. Shi), xsq@cdmc.edu.cn (S. Xu), wml@ysu.edu.cn (M. Wang).

https://doi.org/10.1016/j.heliyon.2024.e30499
Received 12 March 2024; Received in revised form 26 April 2024; Accepted 29 April 2024
Available online 30 April 2024

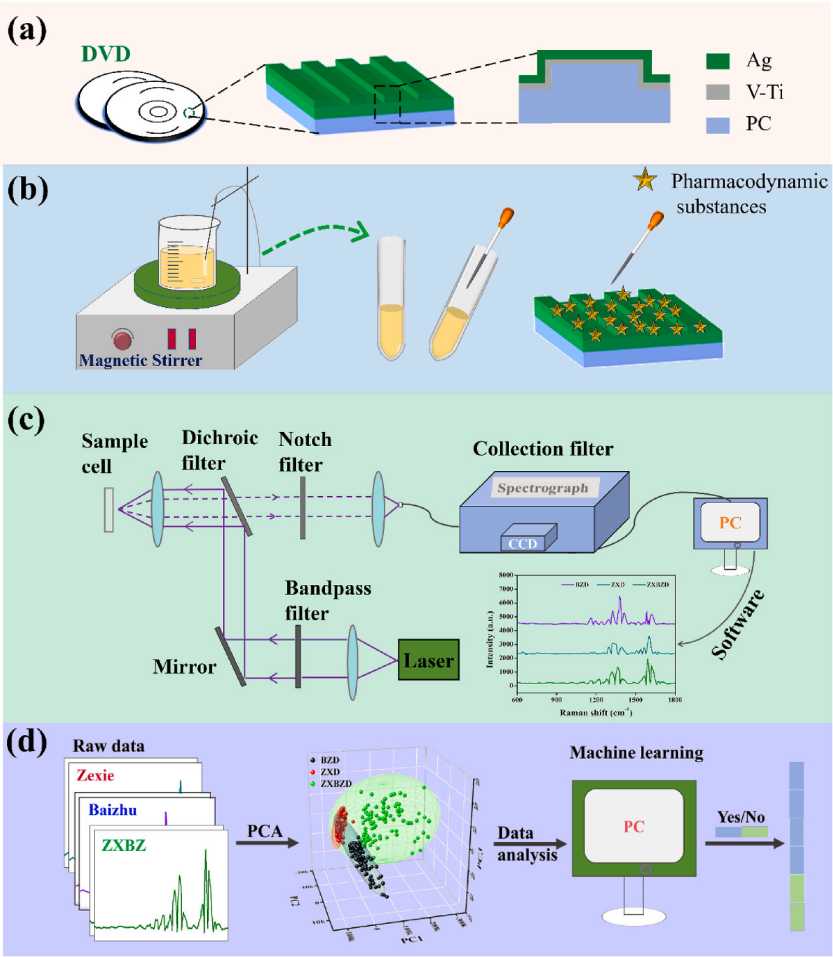

Fig. 1. (a) Preparation of the grating-like multi-layer composite structure SERS substrate; (b)Preparation of the TCM decoction; (c) Collection of SERS signals of different TCM decoction via the Raman system; (d) PCA for the spectra of TCM decoction and classification via machine learning.

[1], anti-inflammatory [2] effect and Baizhu also been shown to have antioxidant [3], inhibit cancer cells value-added role [4]. The combination of all decoctions has significant effect on the treatment of Nonalcoholic fatty liver disease (NAFLD) [5,6]. We hope that the study will advance the exploitation of ZXBZD, which will bring new enlightenment for modern development of TCM decoctions.

At present, most of the analysis of TCM prescriptions is devoted to separating and purifying compounds from the formula, and then using mass spectrometry, chromatography methods and some other coupled technologies to determine the mechanism of action between compounds [7–10]. However, the composition of TCM is complicated, and the whole process of separation and purification is complicated and laborious. Raman spectroscopy, as an advanced analytical method, which can be applied to detect vibrational modes and material composition of target analytes [11], and the mechanism of action between single and compound decoction can be explained. However, the traditional Raman scattering cross section is small, and there are some defects such as poor Raman signal response and strong fluorescence signal [12].

In recent years, the appearance of surface enhanced Raman scattering (SERS) spectrum has substantially made up for the weakness of conventional Raman spectroscopy in substance detection [13], making the Raman signal remarkably enhanced [14], which has played an important role in various domains such as biological detection, chemical analysis, food safety monitoring and environmental monitoring [15–18]. Due to the advantages of SERS, which can be used for rapid and nondestructive detection, high sensitivity, good selectivity and no interference from background matrix [19,20], SERS provides a new and powerful technical support for obtaining high-quality spectra of TCM decoctions. The amplification of SERS signal mainly depended on the electromagnetic field enhancement effect induced by the local surface plasmon resonance (LSPR) phenomenon on nobel metal nanostructure [21–23]. When laser was irradiated on the SERS substrate absorbated analyte molecules, the nobel metal nanostructure on the substrate would amplify the signal of analyte, resulting in SERS phenomenon to study the molecular vibration information, and thus obtain the molecular fingerprint information [22]. For an extended duration, researchers have dedicated efforts to preparing sensitive and reliable SERS substrates through simple and efficient preparation processes [24]. As a result, plasma nanostructures such as nanorods, nanosheet, nanocrystals and nanowires were formed [25–28], which effectively combined with the probe molecules, thus forming a strong plasma

coupling effect. Fan et al. [29] developed a functional film substrate using Si@Ag@PEI material for SO_2 identification in TCM. The LOD of the SERS sensor was 0.25 mg/kg. Chen et al. [30] designed a self-assembled MXene/Ag NPs thin film nanocarrier to study drug delivery and glutathione-triggered drug release. The SERS detection limit was 10^{-8} M. In addition, many latest sensors such as noble metal-decorated ZnO@MXene [31], Pt NPs composited TiO_2/GaN nanorods [32], noble metal mixed metal oxide [33] and graphitic carbon nitride quantum dots [34] are also expected to be applied in the field of SERS research. The SERS substrate prepared could be used for ultrasensitive recognition of the analytes. Among them, three-dimensional nanostructures have been widely favored as plasma nanostructures that can provide numerous molecular binding sites for probe molecules and provide efficient SERS sensitivity [35–38]. DVD disk, which can be found everywhere in life, have regular periodic three-dimensional nanostructures on their surfaces that greatly facilitate the adsorption of nanoparticles, and makes the coupling of surface plasmon resonance possible [39,40].

Although SERS spectrum can provide us with vibrational fingerprint information of molecules, it is often difficult to analyze SERS data due to the noise interference, experimental equipment, parameter setting and other factors affecting SERS spectrum. To obtain effective information from spectral data, we have effectively combined data analysis methods with machine learning algorithms in order to obtain objective and accurate results. As a data analysis tool, machine learning has certain practicability for processing large quantities of data [41,42]. Specifically, the feature extraction of the spectral dataset was performed by principal component analysis (PCA) method [43,44], and then selected six machine learning algorithms to classify and discriminate the single decoction of TCM and its compound decoction. Among them, the NB algorithm achieved a classification accuracy of 97.78 %. Notably, in this study, the SERS substrate based on DVD disk grating structure we prepared combined with PCA and machine learning algorithms achieved better results in the identification of TCM decoctions, and the specific flow was shown in Fig. 1. From this, we concluded that the SERS coupled with machine learning method is effective for the analysis of decoction of TCM.

2. Experimental section

2.1. Materials and chemicals

The Ag target and V–Ti target were obtained from Nanchang Hanchen New Materials Technology Co., Ltd. The DVD disk used in the experiment were sold by TPV Audio and Visual Technology (Shenzhen) Co., Ltd. Zexie and Baizhu were purchased from Jiushengtang TCM Shop. 4-aminothiophenol (4-ATP) was obtained from Aladdin Biochemical Technology Co., Ltd.,Shanghai.

2.2. SERS substrate preparation

SERS substrates were fabricated using magnetron sputtering technique, a physical vapour deposition (PVD) process that operates in a vacuum environment. It is considered to be an excellent substrate preparation technology because of its number of advantages, such as low deposition temperature, excellent coating uniformity, fast deposition rate, and the ability to prepare large-area uniform and dense films [45,46]. To construct the functionalized SERS substrate, Ag and V–Ti nanomaterials were successively deposited on the polycarbonate (PC) sheet of the DVD disk. This can be divided into two steps: (1) Acquisition of PC sheets. DVD disk was cut into uniformly sized squares after mechanical treatment (remove label layer, metal layer and top PC sheet). Next, it was sequentially rinsed in ethanol solution, nitric acid solution in turn to eliminate impurities on the PC sheets, and finally a clean PC sheet was obtained. (2) Magnetron sputtering system to prepare optimal substrates. V and Ti nanomaterials were sequentially deposited on PC sheets to obtain the V–Ti@DVD substrate (experimental parameter settings: argon flow, 120 ml/m; sputtering power: 150 W; sputtering time: 20 min). Next, deposited of Ag nanoparticles on V–Ti@DVD substrate (experimental parameter settings: argon flow, 100 ml/m; sputtering power, 100 W; sputtering time, 20 min), and finally we obtained Ag_{20}@V–Ti_{20}@DVD nanostructures. The pressure in the sputter chamber throughout the sputtering process was 3.5×10^{-3} Pa.

2.3. Preparation of TCM decoction

We prepared three kinds of traditional Chinese medicine decoction, namely Baizhu Decoction (BZD), Zexie Decoction (ZXD) and ZXBZD. The specific configurations are as follows: A certain amount of TCM (BZD contains 15 g of Baizhu, ZXD contains 15 g of Zexie. The classic ratio of Zexie and Baizhu in ZXBZD is 5:2, a total of 21 g) were weighed, mixed with 800 ml deionized water and soaked for 30min, and then maintained it boiling for 60 min, and solution 1 was obtained after filtration. The remaining dregs were combined with 600 ml of deionized water and maintained at a gentle boil for 60 min. Solution 2 was obtained after filtration, and solution 1 and 2 were mixed and placed in a test tube.

2.4. The acquisition of SERS spectra

10 μL of the traditional Chinese medicine decoction was attached to the substrates, after the samples were dried, SERS measurements were performed. For each sample, 100 positions were randomly chosen for testing. Various concentrations of 4-ATP solutions were used for determining the LOD of the substrate.

2.5. Characterizations

The topographies of the substrates was visualised by the field emission scanning electron microscopy (FE-SEM) (FEI-NOVA

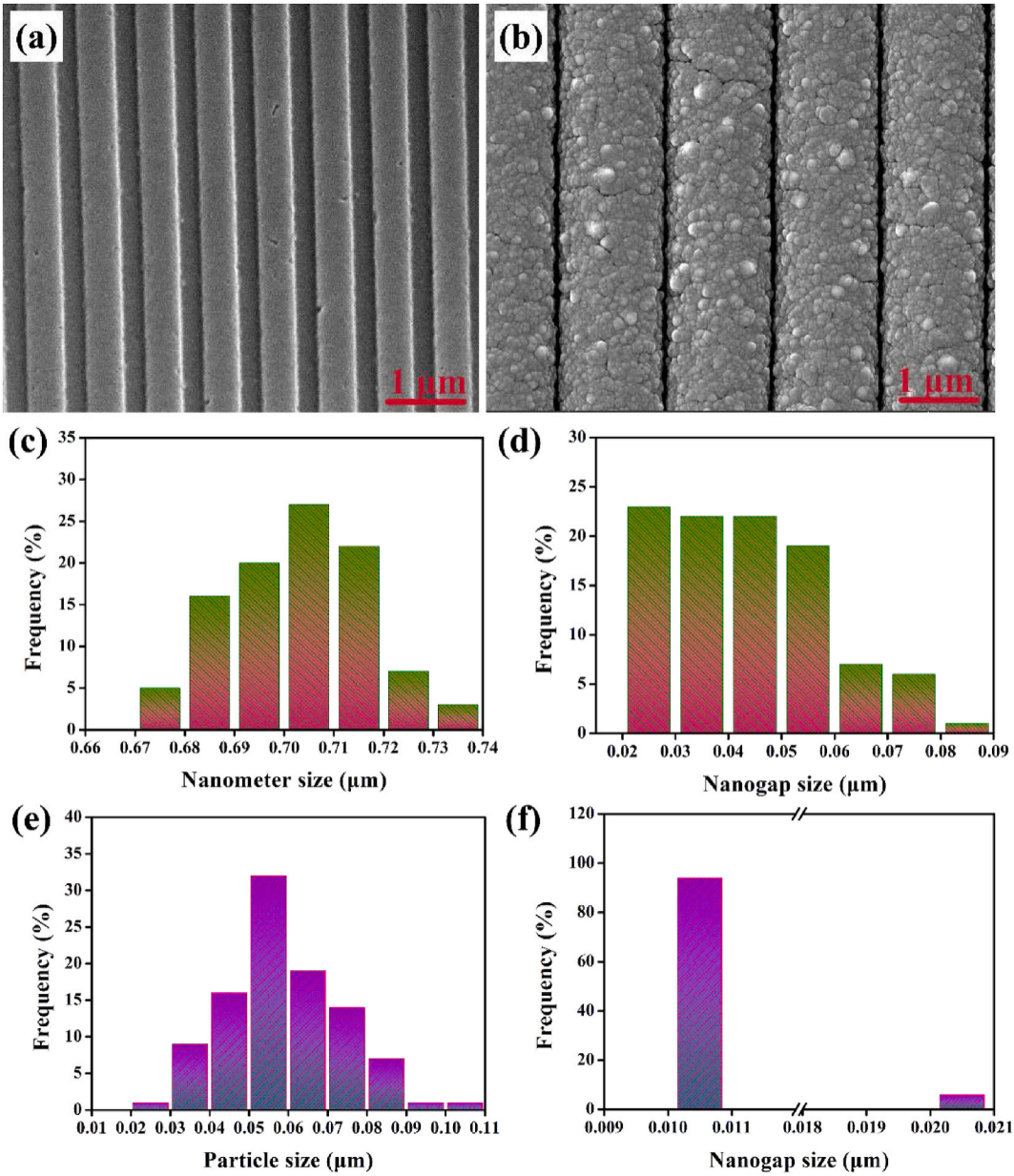

Fig. 2. (a)–(b) FE-SEM images of DVD disk and optimal substrate (Ag_{20}@V–Ti_{20}@DVD), respectively; (c)–(d) Histogram of the distribution of the nanometer size of the land and the nanogap size of the adjacent land on theAg_{20}@V–Ti_{20}@DVD SERS substrate, respectively; (e)–(f) Histogram of the distribution of particle diameter size and gap size between adjacent particles on the optimal substrate, respectively.

NANOSEM 230, America). The Raman spectra were captured using a confocal Raman spectrometer (DXR2xi, Thermo Fisher Inc., USA) with a wavelength of 532 nm, utilizing a He–Ne laser for excitation. Detection was facilitated using a 50 × objective, with an integration time of 5 s.

2.6. Machine learning models of traditional Chinese medicine decoction identification

In this study, six machine learning algorithms, DT, SVM (including Polynomial Kernel, Linear Kernel, Radial Kernel and Sigmoid Kernel) and NB were applied to all SERS spectra after PCA. When performed PCA, 1015 rows and 301 columns were used, with a total of 1015 × 301 pieces of data. DT algorithm is a kind of classification algorithm by constructing tree structure, its core idea is to recursively select the best features for data division. SVM is a kernel based modeling algorithm that constructs an optimal hyperplane by mapping data to a high latitude space so that the data is divisible in that space. Common kernel functions include polynomial kernel, linear kernel, radial kernel and sigmoid kernel. Different kernel functions correspond to different mapping methods to achieve

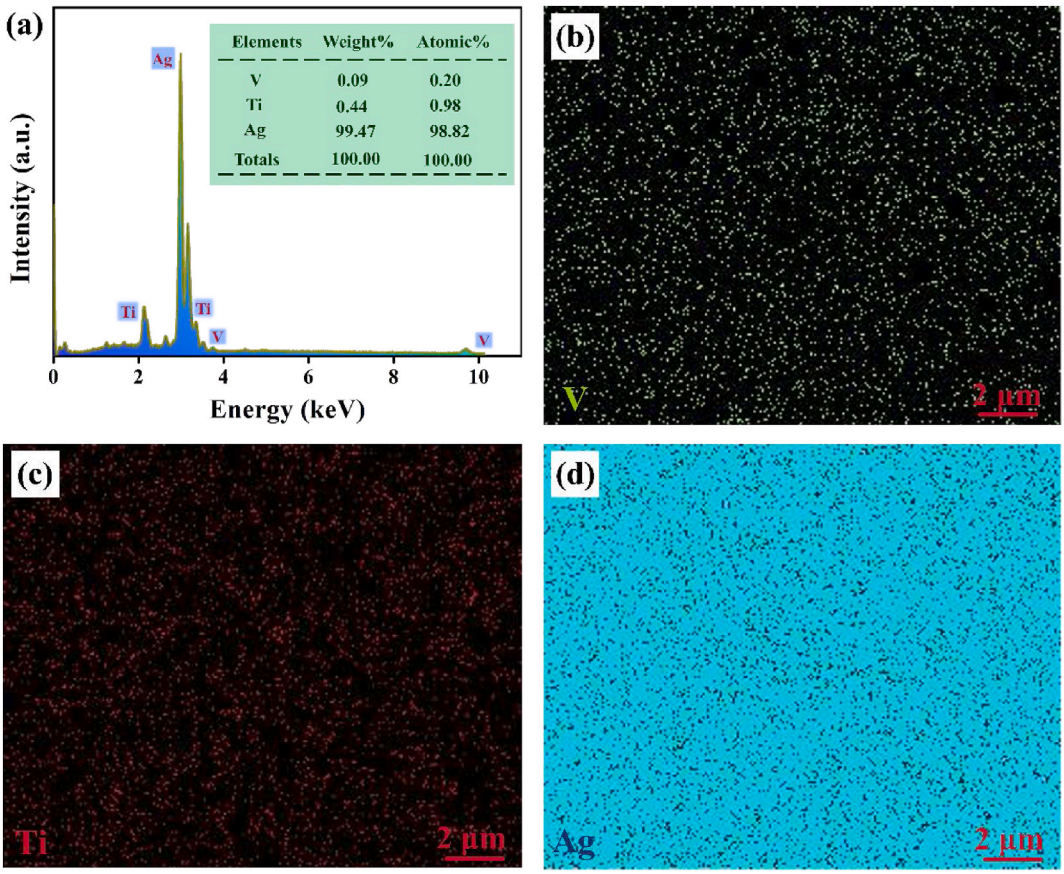

Fig. 3. (a) The EDS spectrum of the SERS substrate; (b)–(d) EDS images of three elements V, Ti and Ag on the substrate, respectively.

different classification effects. The NB algorithm is an efficient classification technique that relies on Bayes' theorem and assumes feature independence. To validate the performance of the algorithm, performance metrics were calculated using accuracy, precision, sensitivity, specificity, recall and F_1 score. Additionally, confusion matrices were generated to measure the classification performance of the machine learning algorithm. These matrices depict the accuracy of classification results and contain four fundamental indicators: true positives, false positives, true negatives, and false negatives [47].

3. Result and discussion

3.1. Morphology characterization of the DVD disk and Ag_{20}@V–Ti_{20}@DVD nanoarray

FE-SEM images were used to analyze the morphology of DVD disk structures and the prepared optimal SERS substrate. Fig. 2(a) present the microstructure of the disk. To be specific, DVD disk shows a periodic regular arrangement of grating-like structure, each of which is composed of a groove and a land. Fig. 2(b) showed the microscopic morphology of the Ag_{20}@V–Ti_{20}@DVD composite structure. It is worth notice that after sputtering the plasma nanomaterials, the surface of the disk structure becomes rough, the width of the land becomes larger and the gap between adjacent land becomes narrower. We plotted the distribution histogram for the substrate size in Fig. 2(c)–(f). Among them, the average width of the land is 0.7 μm, the average gap between adjacent land is 0.04 μm, the average diameter of the particles on the substrate is 0.02 μm, and the gap between adjacent nanoparticles is about 0.01 μm. The significance of nanogaps in influencing electromagnetic field intensity is widely acknowledged [48]. The substrate with a great deal of nanogaps around 10 nm can stimulate plenty of electromagnetic enhancement "hot spots", which greatly benefits the formation of LSPR effects.

Energy dispersive x-ray spectroscopy (EDS) can provide us with qualitative and quantitative information about the components of material. In Fig. 3(a), we found that three elements (V, Ti and Ag) were distributed on the substrate and learned the atomic and mass percentages of each element. From the obtained EDS images (Fig. 3(b)–(d)), the three elements were uniformly distributed on the substrate and represented by different colors.

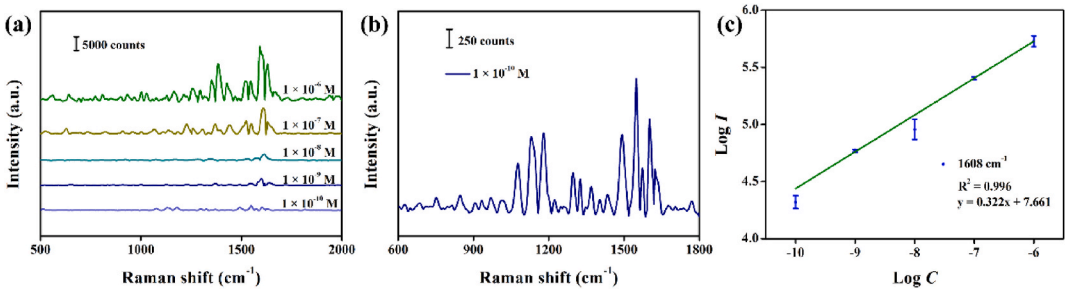

Fig. 4. (a) SERS spectra on the optimal substrate enriched with 4-ATP solution; (b) SERS spectra of 1×10^{-10} M 4-ATP absorbed on optimal SERS substrate; (c) Fitted curves for the Raman signal intensity and the 4-ATP concentration.

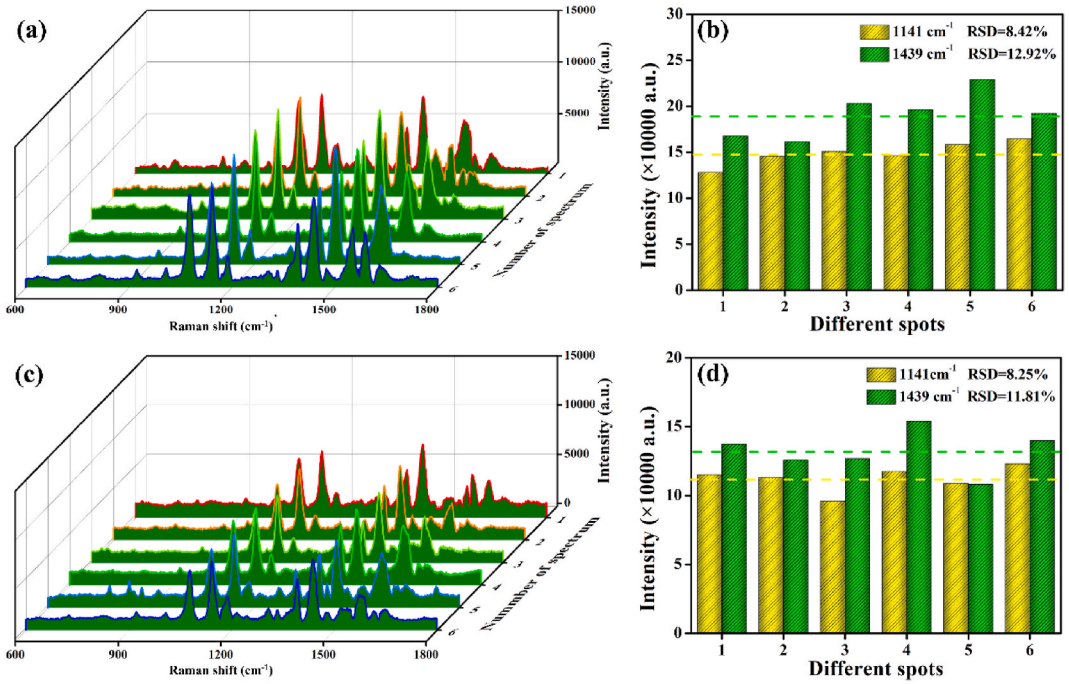

Fig. 5. (a) SERS spectra for 10^{-3} M 4-ATP collected at 6 random positions from the SERS sensor; (b) The Raman intensity of 4-ATP at 1141 cm^{-1} and 1439 cm^{-1}; (c) SERS spectra of 4-ATP solution at different spots on the SERS sensor; (d) SERS signal intensity (1141 cm^{-1} and 1439 cm^{-1}) of the 4-ATP (10^{-3} M).

3.2. Sensitivity and EF calculation of the Ag_{20}@V–Ti_{20}@DVD nanoarray

Benefiting from a great quantity of uniform nanoparticle on the substrate and the efficient LSPR response around the narrow 10 nm nanogaps, our substrate demonstrated proficient trace detection for 4-ATP in various concentrations. Fig. 4(a) displayed the SERS spectra of the substrate under various concentrations of 4-ATP. At various concentrations, the characteristic peak position of 4-ATP molecule remains stable, and which can be distinctly identified even at 1×10^{-10} M (Fig. 4(b)). Therefore, we believe that the LOD of the SERS substrate is 1×10^{-10} M. As shown in Fig. 4(c)–a curve fitting analysis was conducted following the logarithmic transformation of both the Raman signal intensity at 1608 cm^{-1} and the concentration of 4-ATP. The correlation between Log *C* and Log *I* is notably strong within the linear range. The excellent linear relationship indicated that the prepared SERS substrates have certain quantitative detection ability.

A high EF indicates a strong enhancement of the signal, and thus EF can be regarded as an important indicator for evaluating the performance of substrates. We study the EF of the substrate according to the following formula [49]:

$$EF = \frac{I_{Sers}/N_{Sers}}{I_{Raman}/N_{Raman}}$$

where I_{Sers} and I_{Raman} are the Raman signal intensities of Ag_{20}@V–Ti_{20}@DVD SERS substrate adsorbed with 1×10^{-7} M of 4-ATP

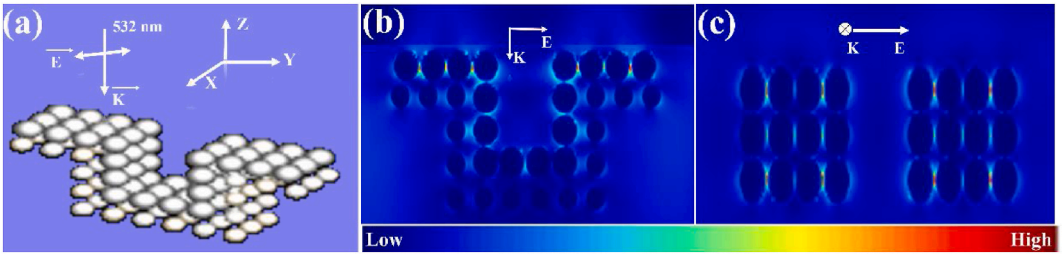

Fig. 6. (a) Model of Ag_{20}@V–Ti_{20}@DVD substrate; (b)–(c) the electromagnetic field distribution images of the substrate in the Y-Z plane and X–Y plane, respectively.

molecules and reference substrate (silica sheets enriched in 10^{-2} M 4-ATP solution) at a certain characteristic peak, respectively. N_{Sers} and N_{Raman} are the amount of 4-ATP molecules absorbed by laser spot on the substrate and the reference substrate, respectively. The N_{Raman}/N_{sers} was obtained by the following equation [50]:

$$N = \frac{N_A \times M \times V_{Soulution}}{S_{Sub}} \times S_{laser}$$

where N_A is the Avogadro constant, M is the molar concentration of the 4-ATP, $V_{solution}$ is the volume of the 4-ATP solution, S_{laser} is the laser spot area and S_{stub} is the area of the 4-ATP solution adhered on substrate. The area of 4-ATP probe molecule solution on silica sheets was approximately 1.2 times that of 4-ATP solution on the SERS substrate. In addition, the S_{laser} was 0.785 μm^2. Finally, the EF for our substrate at 1608 cm^{-1} was 1.90×10^7.

3.3. Reproducibility and uniformity of Ag_{20}@V–Ti_{20}@DVD substrate

In addition to highly reliable sensitivity, achieving a high degree of reproducibility is of vital importance for the practicality of SERS substrates [51]. Hence, to assess the repeatability of the SERS substrate, SERS spectra were acquired using a 4-ATP solution at a 10^{-3} M concentration. As shown in Fig. 5(a), SERS signals were collected from 6 randomly chosen spots on the SERS substrate. From Raman spectra, we can see that there is no obvious shift of the characteristic Raman lines, and the signal intensity does not fluctuate greatly, which indicated that the SERS substrate developed by us have a favourable reproducibility. In addition, we quantified the intensities of 1141 cm^{-1} and 1439 cm^{-1} peaks at these 6 positions to determine the relative standard deviation (RSD). The formula below was employed to compute the RSD [52]:

$$RSD = \frac{\sqrt{\sum_{i=1}^{n} (I_i - \bar{I})^2}}{\bar{I}}$$

where n denotes the quantity of Raman spectra, I_i is the Raman signal intensity, and $\bar{I}$ represents the average Raman signal intensity. In Fig. 5(b), the change in signal intensity of the 1141 cm^{-1} and 1439 cm^{-1} peak were slight, and the RSD were 8.25 % and 12.92 %, respectively. This further confirmed that the prepared substrate possesses good reproducibility.

Moreover, uniformity of the SERS signals is also evaluated by the randomly chosen 6 points on the Ag_{20}@V–Ti_{20}@DVD SERS nanoarray, in which the concentration of 4-ATP probe molecule is 10^{-3} M. As shown in Fig. 5(c)–a 3D plot of the SERS spectra showed a very uniform Raman intensity for 4-ATP. The calculated RSD of the intensity from the 4-ATP bands at 1141 cm^{-1} and 1439 cm^{-1} are 8.25 % and 11.81 %, respectively (Fig. 5(d)), further proving the remarkable uniformity of the SERS signals. These results suggested that the Ag_{20}@V–Ti_{20}@DVD SERS sensor with excellent sensitivity, good reproducibility, and uniformity is a suitable candidate for identification of TCM decoction.

3.4. 3D-FDTD

Due to the electromagnetic enhancement effect, we conducted 3D finite difference Time domain (3D-FDTD) simulations of the prepared substrates to understand the distribution of electromagnetic fields around the SERS substrate [53]. In order to simplify the model, we constructed a distribution of nanoparticles in a grating based on SEM images of Ag_{20}@V–Ti_{20}@DVD. As depicted in Fig. 6 (a), the lower stratum consists of a V–Ti nanolayer, while the uppermost layer contains an Ag nanoparticle layer. The selected source for excitation was a plane light type, operating at a laser wavelength of 532 nm. **K** was the laser irradiation direction, **E** was the laser polarization direction. Fig. 6(b)–(c) depicted how the electromagnetic field was distributed across the model in the Y-Z and X–Y planes, in that order. It can be seen that there are numerous “hot spot” regions are distributed on the silver nanolayer, forming a strong electromagnetic field coupling.

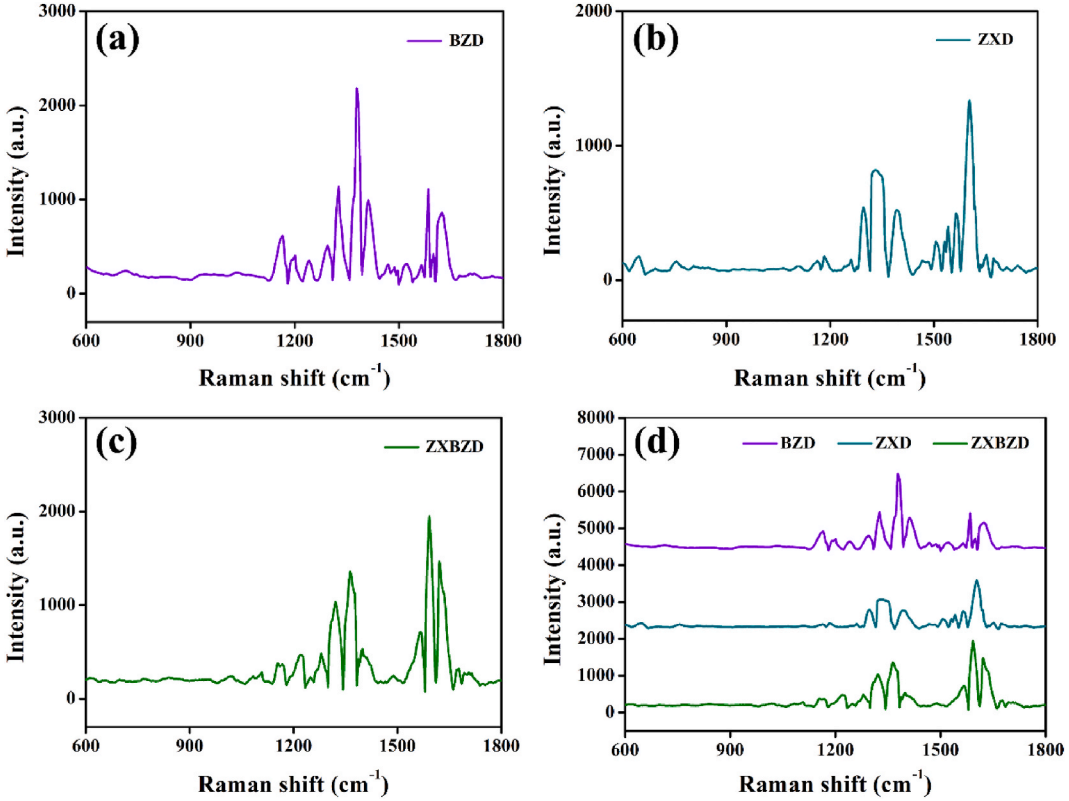

Fig. 7. (a)–(c) The SERS spectra of BZD, ZXD and ZXBZD, respectively; (d) The SERS spectrum comparison of ZXD, BZD and ZXBZD.

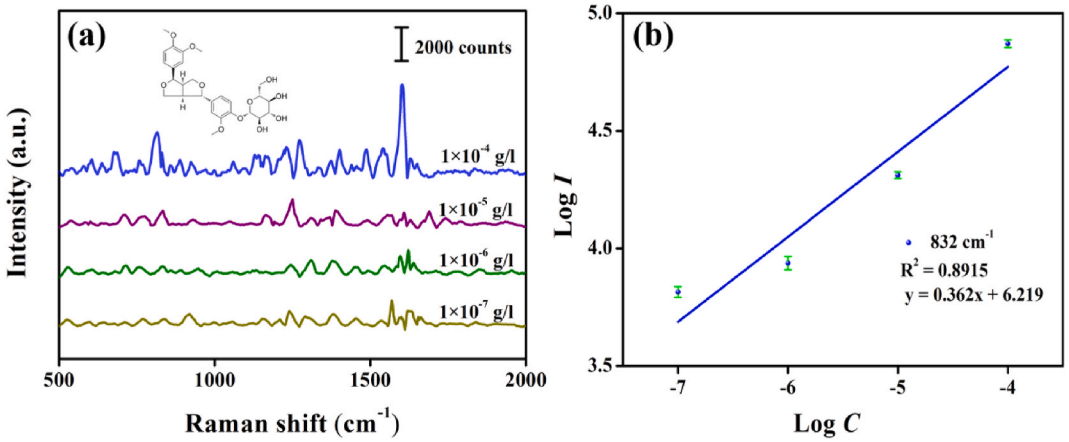

Fig. 8. (a) SERS spectra of forsythin solutions; (b) Linear calibration curve between Raman signal intensity and 4-ATP concentration under logarithmic conditions.

3.5. Detection of TCM decoction

TCM decoctions are composed of a variety of single-flavor medicines. Understanding the relationship between the biochemical components of prescriptions and single-flavor medicines is of vital meaning to the development of TCM. In this study, SERS spectrum analysis was performed on two single flavor decoction (Zexie and Baizhu) and its compound decoction (ZXBZD), and the interaction between the single flavor decoction and compound decoction was discussed through the obtained spectral data. Fig. 7(a)-(c) showed the SERS spectrum of BZD, ZXD and ZXBZD, respectively. Fig. 7(d) displayed the SERS spectrum comparison of three kinds of TCM decoction. It can be found that the spectra of these three kinds of decoctions have the same signal peak at 1161 cm^{-1} and 1598 cm^{-1}. Among these two simultaneous signal peaks, ZXBZD are not in the strongest position. This indicated that when Zexie and Baizhu were

Table 1
The developed SERS method was compared with approaches documented in existing literature.

Methods	Linear Range (mg/l)	LOD (mg/l)	Real Sample	Reference
Fingerprinting approach	3.0–94.9	0.07	Shuang-huang-lian oral liquid	[54]
CZE	3.28–131.25	0.94	Forsythina suspensa	[55]
HPLC-DAD	16.6–415.0	0.09	Forsythina suspensa	[56]
LC-ESI-MS	0.04625-18.5	0.002	Forsythia Methanol Extract	[57]
SERS	10^{-4}-0.1	10^{-4}	forsythin solution	This work

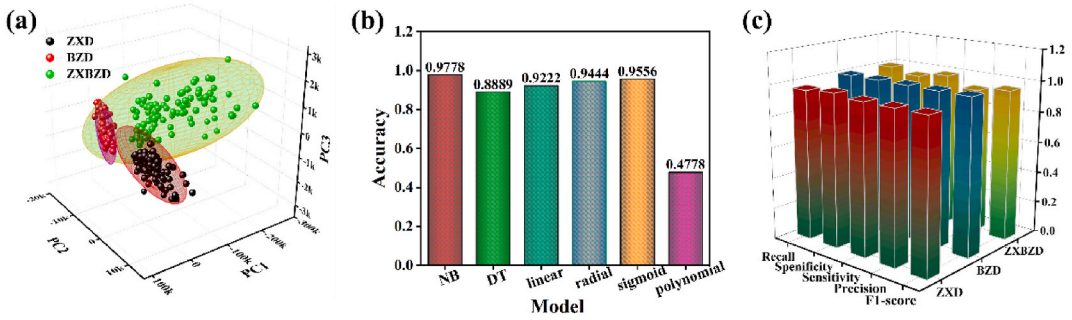

Fig. 9. Dimensionality reduction and classification of the TCM decoction Raman spectra obtained on the SERS substrate. (a) 3D plots of the principal component of TCM decoction; (b) Classification accuracies of six machine learning algorithms; (c) Evaluation indexes for classification of three TCM decoctions based on NB model.

mixed and decocted, the two drugs acted on some biochemical components represented by 1161 cm^{-1} and 1598 cm^{-1}, resulting in changes in the intensity of signal peaks, rather than the simple addition of intensity. In addition, the Raman signal peak of BZD at 1325 cm^{-1} also appeared in the spectrum of ZXBZD, and the intensity was larger than that of ZXBZD, indicating that this component acted with some components of Zexie during mixed decoction, thus increasing the intensity. The Raman signal peaks of 643 cm^{-1} vibrational peak in the decoction of Zexie alone and the peak of 711 cm^{-1} in the decoction of Baizhu alone as well as the Raman signal peaks of 1294 cm^{-1} which appeared in both the decoction of Zexie and Baizhu, which disappeared from the spectra of the ZXBZD. We suspected that it might be due to the interaction of certain components in the mixed decoction of two single-flavor medicines, resulting in the disappearance of that component. It is worth noting that due to the existence of frequency shift phenomenon, the intensity and position of Raman signal peak will change, and then the position of the same signal peak in different decoctions will be different. The study revealed the interaction between the single decoction and the prescription, which opens up a broad path for the further development of TCM.

In addition, to demonstrate the effectiveness of this approach in analyzing traditional Chinese medicine, we used the SERS substrate for ultrasensitive identification of forsythin. Different concentrations of forsythin solutions were adhered to the Ag_{20}@V–Ti_{20}@DVD substrates and dried for SERS signal acquisition, as shown in Fig. 8(a). Noticeable were the distinct peaks of forsythin at 832 cm^{-1} and 1602 cm^{-1}, and with the diminish of forsythin concentration, the corresponding Raman signal intensity also showed a downward trend. When the concentration of forsythin solution was as low as 1×10^{-7} g/l, its characteristic peaks remained noticeable. Fig. 8(b) showed the calibration curve of forsythin concentration at characteristic peak 832 cm^{-1}. Evidently, a robust linear correlation was observed between the intensity of the Raman signal and the concentration of forsythin following logarithmic analysis ($y = 0.362x + 6.219$, $R^2 = 0.8915$). This strong linear correlation demonstrated the excellent quantitative detection capability of the substrate. A comprehensive comparison was made between our method and other research methodologies, as presented in Table 1. The results of the study indicated that the developed Ag_{20}@V–Ti_{20}@DVD SERS sensor achieved or surpassed international standards in key metrics such as LOD and linear range.

3.6. Machine learning assisted TCM decoction identification

TCM decoctions have complex components, in order to achieve accurate and reliable detection of complex components, a method using SERS connected with machine learning was proposed for the identification of TCM decoctions. Based on the SERS substrate, SERS spectra of TCM decoctions were acquired and analyzed using six machine learning models, that is, DT, NB and SVM with three forms of kernel functions. PCA reduced the data dimensionality while retaining most of the spectral information. The results of PCA were shown in Fig. 9(a), in which each TCM decoctions clustered around its internal centroid in three-dimensional space, with only a few data points dispersed. Then the data after PCA were entered into six classification algorithms for automatic identification. As shown in Fig. 9(b), the NB model has the best classification accuracy of 97.78 %. In addition to the NB model, the DT, linear kernel SVM, radial kernel SVM, and sigmoid kernel SVM algorithms also achieved high prediction accuracies, with accuracies of 88.89 %, 92.22 %, 94.44 %, and 95.56 %, respectively, which are basically the same as the NB model accuracy. On the other hand, the

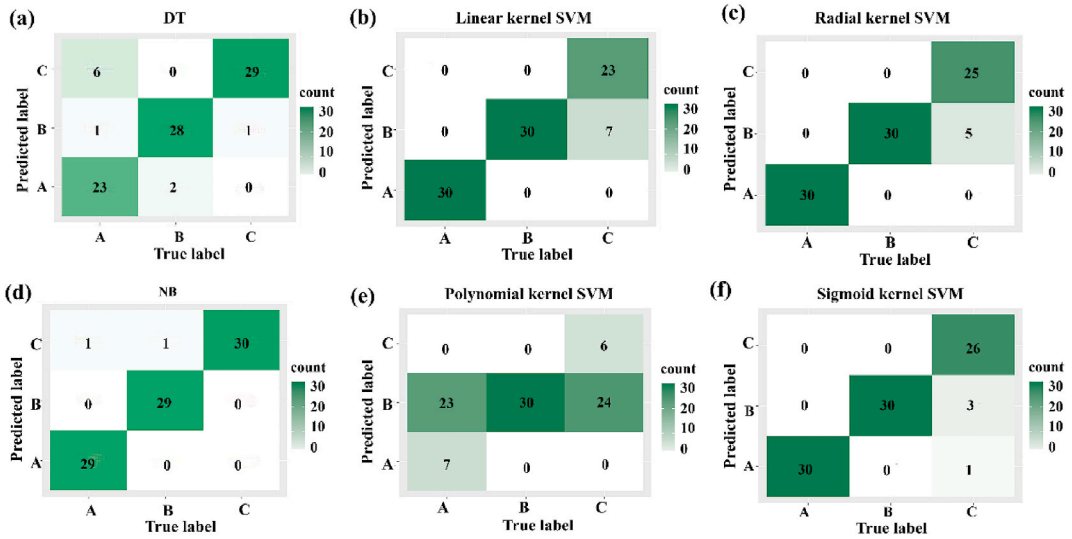

Fig. 10. Confusion matrices of the (a) DT, (b) linear kernel SVM, (c) radial kernel SVM, (d) NB, (e) polynomial kernel SVM and (f) sigmoid kernel SVM for TCM decoction (A–C: ZXD, BZD, ZXBZD).

polynomial kernel SVM algorithm had an accuracy of less than 50 %, which indicated that the algorithm performs poorly in the identification of the ZXBZD. Therefore, Fig. 9(c) further evaluated the performance of NB model. The remaining indicators of NB model were all above 90 %, showing good classification performance.

In addition, the confusion matrix serves as a valuable tool to visualize the classification results of various classifiers and intuitively grasp their performance in sample classification. We gave the confusion matrix for each machine learning algorithm, as shown in Fig. 10. Among them, Fig. 10(a)–(c) and Fig. 10(e)-(f) showed the classification results for three TCM decoctions based on the DT model, linear kernel SVM model, radial kernel SVM model, polynomial kernel SVM model and sigmoid kernel SVM model, respectively. Compared with these classification models, the NB model had the best classification results, as shown in Fig. 10(d). The prediction accuracy of ZXD was 96.67 %, and the probability of 3.33 % was predicted to be ZXBZD when the three TCM decoction were judged. For BZD, 3.33 % of the samples from BZD were incorrectly classified as ZXBZD. Furthermore, the prediction accuracy of NB algorithm for ZXBZD was 100 %. The above study demonstrates the feasibility of our proposed SERS-based machine learning method for the identification of TCM decoction.

4. Conclusion

In summary, we designed and constructed a grating-like multi-level composite structure SERS sensor for identification of TCM decoction. V–Ti and Ag nanoparticles were attached to the polycarbonate substrate of the DVD disk by magnetron sputtering technology, and the Raman signal was gradually amplified by adjusting the gap between the nanoparticles, which greatly improved the sensitivity of the prepared multi-level SERS sensor. Furthermore, the practicality of the substrate was also verified by the detection of ZXBZD. In addition, based on the spectrum data set of TCM decoction, we also introduced a machine learning algorithm to classify and identify TCM compound decoction and its single flavor decoction, and finally achieved a classification accuracy of 97.78 %. Our study provided a basis for the establishment of a multi-level SERS detection sensor, which was helpful to promote the detection and analysis of TCM decoction.

Funding

This work was supported by Hebei Natural Science Foundation(F2022406001), the Graduate Student Innovation Ability Training of Education of Hebei Province Department (CXZZSS2024119), the Chengde Biomedicine Industry Research Institute Funding project (202205B086) and the Technology and Starting Fund for Scientific Research of High-level Talents of Chengde Medical University-Nature (Grant No. 202206).

Data availability statement

Data will be made available on request.

Ethics statement

We confirm that this manuscript entitled "High-performance grating-like SERS substrate based on machine learning for ultra-sensitive detection of Zexie-Baizhu Decoction" is the authors' original work and has not been published nor has it been submitted simultaneously elsewhere.

All authors have been personally and actively involved in substantive work leading to the manuscript and we further confirm that all authors have checked the manuscript and have agreed to the submission.

CRediT authorship contribution statement

Wenying Zhou: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Xue Han:** Resources, Investigation, Data curation. **Yanjun Wu:** Visualization, Investigation, Formal analysis. **Guochao Shi:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Formal analysis. **Shiqi Xu:** Supervision, Resources, Project administration, Formal analysis. **Mingli Wang:** Supervision, Resources, Project administration, Formal analysis. **Wenzhi Yuan:** Visualization, Investigation. **Jiahao Cui:** Validation, Formal analysis. **Zelong Li:** Visualization, Software.

Declaration of competing Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

[1] Q. Zhang, Z.Y. Jiang, J. Luo, J.F. Liu, Y.B. Ma, R.H. Guo, et al., Anti-HBV agents. Part 2: synthesis and in vitro anti-hepatitis B virus activities of alisol A derivatives, Bioorganic Med. Chem. Lett. 19 (8) (2009) 2148–2153. http://10.1016/j.bmcl.2009.02.122.
[2] S.S. Liu, W.L. Sheng, Y. Li, S.S. Zhang, J.J. Zhu, H.M. Gao, et al., Chemical constituents from Alismatis Rhizoma and their anti-inflammatory activities in vitro and in vivo, Bioorg. Chem. 92 (2019) 103226. http://10.1016/j.bioorg.2019.103226.
[3] C. Xiao, C. Xu, N. He, Y. Liu, Y. Wang, M. Zhang, et al., Atractylenolide II prevents radiation damage via MAPKp38/Nrf2 signaling pathway, Biochem. Pharmacol. 177 (2020) 114007. http://10.1016/j.bcp.2020.114007.
[4] S. Chen, K. Tang, P. Hu, S. Tan, S. Yang, C. Yang, et al., Atractylenolide III alleviates the apoptosis through inhibition of autophagy by the mTOR-dependent pathway in alveolar macrophages of human silicosis, Mol. Cell. Biochem. 476 (2) (2021) 809–818. http://10.1007/s11010-020-03946-w.
[5] Y. Cao, J. Shi, L. Song, J. Xu, H. Lu, J. Sun, et al., Multi-omics integration analysis identifies lipid disorder of a non-alcoholic fatty liver disease (NAFLD) mouse model improved by Zexie-Baizhu decoction, Front. Pharmacol. 13 (2022) 858795. http://10.3389/fphar.2022.858795.
[6] S. Milić, D. Stimac, Nonalcoholic fatty liver disease/steatohepatitis: epidemiology, pathogenesis, clinical presentation and treatment, Dig. Dis. 30 (2) (2012) 158–162. http://10.1159/000336669.
[7] F. Tian, X.F. He, J. Sun, X.D. Liu, Y. Zhang, H. Cao, et al., Simultaneous quantitative analysis of nine constituents in six Chinese medicinal materials from Citrus genus by high-performance liquid chromatography and high-resolution mass spectrometry combined with chemometric methods, J Sep Sci 43 (4) (2020) 736–747. http://10.1002/jssc.201901033.
[8] H. Qiu, C. Shan, C. Fei, P. Xue, Y. Zhou, J. Yuan, et al., Geoherbalism metabolomic analysis of atractylodes lancea (thunb.) DC. By LC-triple TOF-MS/MS and GC-MS, Molecules 28 (16) (2023) 5974. http://10.3390/molecules28165974.
[9] X. Wang, X. Zhang, J. Li, J. Fu, M. Zhao, W. Zhang, et al., Network pharmacology and LC-MS approachs to explore the active compounds and mechanisms of Yuanjiang decoction for treating bradyarrhythmia, Comput. Biol. Med. 152 (2023) 106435. http://10.1016/j.compbiomed.2022.106435.
[10] Z. Wang, X. Wang, Y. Wang, C. Wu, J. Zhou, Simultaneous determination of five antiviral drug residues and stability studies in honey using a two-step fraction capture coupled to liquid chromatography tandem mass spectrometry, J. Chromatogr. A 1638 (2021) 461890, https://doi.org/10.1016/j.chroma.2021.461890.
[11] J. Zheng, L. He, Surface-enhanced Raman spectroscopy for the chemical analysis of food, Compr. Rev. Food Sci. Food Saf. 13 (3) (2014) 317–328. http://10.1111/1541-4337.12062.
[12] J. Krajczewski, E. Dumiszewska, D. Czołak, S. Turczyniak Surdacka, A. Kudelski, New, epitaxial approach to SERS platform preparation – InP nanowires coated by an Au layer as a new, highly active, and stable SERS platform, Appl. Surf. Sci. 607 (2023) 155096. http://10.1016/j.apsusc.2022.155096.
[13] M. Fan, G.F.S. Andrade, A.G. Brolo, A review on recent advances in the applications of surface-enhanced Raman scattering in analytical chemistry, Anal. Chim. Acta 1097 (2020) 1–29, https://doi.org/10.1016/j.aca.2019.11.049.
[14] J.W. Tang, Q.H. Liu, X.C. Yin, Y.C. Pan, P.B. Wen, X. Liu, et al., Comparative analysis of machine learning algorithms on surface enhanced Raman spectra of clinical Staphylococcus species, Front. Microbiol. 12 (2021) 696921. http://10.3389/fmicb.2021.696921.
[15] Y. Liu, H. Ma, X.X. Han, B. Zhao, Metal–semiconductor heterostructures for surface-enhanced Raman scattering: synergistic contribution of plasmons and charge transfer, Mater. Horiz. 8 (2) (2021) 370–382. http://10.1039/d0mh01356k.
[16] C. Zong, M. Xu, L.-J. Xu, T. Wei, X. Ma, X.-S. Zheng, et al., Surface-enhanced Raman spectroscopy for bioanalysis: reliability and challenges, Chem. Rev. 118 (10) (2018) 4946–4980. http://10.1021/acs.chemrev.7b00668.
[17] N.R. Barveen, T.-J. Wang, Y.-H. Chang, Z. Yuan-Liu, Ultrasensitive and reusable SERS probe for the detection of synthetic dyes in food industry through hybrid flower-shaped ZnO@Ag nanostructures, J. Alloys Compd. 861 (2021) 157952, https://doi.org/10.1016/j.jallcom.2020.157952.
[18] X. Zhao, C. Liu, J. Yu, Z. Li, L. Liu, C. Li, et al., Hydrophobic multiscale cavities for high-performance and self-cleaning surface-enhanced Raman spectroscopy (SERS) sensing, Nanophotonics 9 (2020) 4761–4773. http://10.1515/nanoph-2020-0454.
[19] Q. Yang, J. Wang, H. Wu, S. Qin, J. Pan, C. Li, Hierarchically rough CuO/Ag composite film with controlled morphology as recyclable SERS-active substrate, Appl. Surf. Sci. 598 (2022) 153746. http://10.1016/j.apsusc.2022.153746.
[20] X. Wang, L. Zhu, Z. Zhu, S. Chang, J. Qian, J. Jiang, et al., Simultaneously improved SERS sensitivity and thermal stability on Ag dendrites via surface protection by atomic layer deposition, Appl. Surf. Sci. 611 (2022) 155626. http://10.1016/j.apsusc.2022.155626.
[21] M. Cueto, M. Piedrahita, C. Caro, B. Martínez-Haya, M. Sanz, M. Oujja, et al., Platinum nanoparticles as photoactive substrates for mass spectrometry and spectroscopy sensors, J. Phys. Chem. C 118 (21) (2014) 11432–11439. http://10.1021/jp500190m.
[22] L.-L. Tan, M. Wei, L. Shang, Y. Yang, Cucurbiturils-mediated noble metal nanoparticles for applications in sensing, SERS, theranostics, and catalysis, Adv. Funct. Mater. 31 (2020) 2007277. http://10.1002/adfm.202007277.
[23] H. Kang, J.T. Buchman, R.S. Rodriguez, H.L. Ring, J. He, K.C. Bantz, et al., Stabilization of silver and gold nanoparticles: preservation and improvement of plasmonic functionalities, Chem. Rev. 119 (1) (2019) 664–699. http://10.1021/acs.chemrev.8b00341.
[24] B. Liu, X. Yao, S. Chen, H. Lin, Z. Yang, S. Liu, et al., Large-area hybrid plasmonic optical cavity (HPOC) substrates for surface-enhanced Raman spectroscopy, Adv. Funct. Mater. 28 (43) (2018) 1802263. http://10.1002/adfm.201802263.

[25] S.H. Park, J.G. Son, T.G. Lee, H.M. Park, J.Y. Song, One-step large-scale synthesis of micrometer-sized silver nanosheets by a template-free electrochemical method, Nanoscale Res. Lett. 8 (1) (2013) 248. http://10.1186/1556-276x-8-248.
[26] J.-J. Zhu, X.-H. Liao, X.-N. Zhao, H.-Y. Chen, Preparation of silver nanorods by electrochemical methods, Mater. Lett. 49 (2) (2001) 91–95, https://doi.org/10.1016/S0167-577X(00)00349-9.
[27] J. Choi, G. Sauer, K. Nielsch, R.B. Wehrspohn, U.P.D. Gösele, Hexagonally arranged monodisperse silver nanowires with adjustable diameter and high aspect ratio, Chem. Mater. 15 (2003) 776–779. http://10.1021/CM0208758.
[28] T.-J. Wang, H.-W. Chang, Y. Wang, H.-A. Chi, J.-S. Chen, One-Step surfactant-free photoreduction synthesis of single-crystal silver triangular nanoprisms by surface modified chemically patterned ferroelectric crystals for SERS application, Appl. Surf. Sci. 623 (2023) 157114. http://10.1016/j.apsusc.2023.157114.
[29] B. Fan, Y. Wang, Z. Li, D. Xun, J. Dong, X. Zhao, et al., Si@Ag@PEI substrate-based SERS sensor for rapid detection of illegally adulterated sulfur dioxide in traditional Chinese medicine, Talanta 238 (1) (2022) 122988. http://10.1016/j.talanta.2021.122988.
[30] Y. Chen, C. Jiang, F. Huang, Z. Yu, L. Jiang, Efficient interfacial self-assembled MXene/Ag NPs film nanocarriers for SERS-traceable drug delivery, Anal. Bioanal. Chem. 415 (22) (2023) 5379–5389. http://10.1007/s00216-023-04813-5.
[31] K.S. Pasupuleti, A.M. Thomas, D. Vidyasagar, V.N. Rao, S.-G. Yoon, Y.-H. Kim, et al., ZnO@Ti3C2Tx MXene hybrid composite-based Schottky-Barrier-Coated SAW sensor for effective detection of sub-ppb-level NH3 at room temperature under UV illumination, ACS Mater. Lett. 5 (10) (2023) 2739–2746. http://10.1021/acsmaterialslett.3c00698.
[32] K.S. Pasupuleti, S.S. Chougule, N. Jung, Y.-J. Yu, J.-E. Oh, M.-D. Kim, Plasmonic Pt nanoparticles triggered efficient charge separation in TiO2/GaN NRs hybrid heterojunction for the high performance self-powered UV photodetectors, Appl. Surf. Sci. 594 (2022) 153474, https://doi.org/10.1016/j.apsusc.2022.153474.
[33] A.B. Cueva Sola, K.S. Pasupuleti, J.H. Jeon, T. Thenepalli, N.-h. Bak, S. Sampath, et al., Sustainable solution to the recycling of spent SCR catalyst and its prospective gas sensor application, Mater. Today Sustainability 25 (2023) 100649, https://doi.org/10.1016/j.mtsust.2023.100649.
[34] K.S. Pasupuleti, S.S. Chougule, D. Vidyasagar, N.-h. Bak, N. Jung, Y.-H. Kim, et al., UV light driven high-performance room temperature surface acoustic wave NH3 gas sensor using sulfur-doped g-C3N4 quantum dots, Nano Res. 16 (5) (2023) 7682–7695. http://10.1007/s12274-023-5472-x.
[35] Y. Shao, S. Li, Y. Niu, Z. Wang, K. Zhang, L. Mei, et al., Three-dimensional dendritic Au-Ag substrate for on-site SERS detection of trace molecules in liquid phase, Nanomaterials 12 (12) (2022) 2005. http://10.3390/nano12122002.
[36] C. Zhu, D. Liu, M. Yan, G. Xu, H. Zhai, J. Luo, et al., Three-dimensional surface-enhanced Raman scattering substrates constructed by integrating template-assisted electrodeposition and post-growth of silver nanoparticles, J. Colloid Interface Sci. 608 (2) (2022) 2111–2119. http://10.1016/j.jcis.2021.10.133.
[37] C. Li, J. Yu, S. Xu, S. Jiang, X. Xiu, C. Chen, et al., Constructing 3D and flexible plasmonic structure for high-performance SERS application, Adv. Mater. Technol. 3 (11) (2018) 1800174, https://doi.org/10.1002/admt.201800174.
[38] E.Z. Tan, P.G. Yin, T.T. You, H. Wang, L. Guo, Three dimensional design of large-scale TiO(2) nanorods scaffold decorated by silver nanoparticles as SERS sensor for ultrasensitive malachite green detection, Acs. Appl. Mater.Inter. 4 (7) (2012) 3432–3437. http://10.1021/am3004126.
[39] N. Chamuah, A. Saikia, A.M. Joseph, P. Nath, Blu-ray DVD as SERS substrate for reliable detection of albumin, creatinine and urea in urine, Sens. Actuat. B Chem. 285 (2019) 108–115, https://doi.org/10.1016/j.snb.2019.01.031.
[40] R. Ahmed, M.O. Ozen, M.G. Karaaslan, C.A. Prator, C. Thanh, S. Kumar, et al., Tunable fano-resonant metasurfaces on a disposable plastic-template for multimodal and multiplex biosensing, Adv. Mater. 32 (19) (2020) 1907160. http://10.1002/adma.201907160.
[41] N. Banaei, J. Moshfegh, A. Mohseni-Kabir, J.M. Houghton, Y. Sun, B. Kim, Machine learning algorithms enhance the specificity of cancer biomarker detection using SERS-based immunoassays in microfluidic chips, RSC Adv. 9 (4) (2019) 1859–1868, https://doi.org/10.1039/C8RA08930B.
[42] A. Barucci, C. D'Andrea, E. Farnesi, M. Banchelli, C. Amicucci, M. de Angelis, et al., Label-free SERS detection of proteins based on machine learning classification of chemo-structural determinants, Analyst 146 (2) (2021) 674–682. http://10.1039/d0an02137g.
[43] A.K. Gupta, C.-H. Hsu, C.S. Lai, Enhancement of the Au/ZnO-NA plasmonic SERS signal using principal component analysis as a machine learning approach, IEEE Photonics J 12 (2020) 1–11, https://doi.org/10.1109/jphot.2020.3015740.
[44] W. Yuan, X. Han, G. Shi, M. Wang, W. Zhou, J. Cui, et al., Machine learning-driven multi-level composite SERS platform for trace detection of chlorogenic acid as pharmacodynamic substance in honeysuckle, Opt Laser Technol 169 (2024) 109911, https://doi.org/10.1016/j.optlastec.2023.109911.
[45] Y. Chen, Y. Fang, Surface enhanced Raman scattering (SERS) activity studies of Si, Fe, Ti, Al and Ag films' prepared by magnetron sputtering, Spectrochim. Acta Mol. Biomol. Spectrosc. 69 (3) (2008) 733–737. http://10.1016/j.saa.2007.05.030.
[46] H.Y. Wu, H.C. Lin, G.Y. Hung, C.S. Tu, T.Y. Liu, C.H. Hong, et al., High sensitivity SERS substrate of a few nanometers single-layer silver thickness fabricated by DC magnetron sputtering technology, Nanomaterials 12 (16) (2022) 2742. http://10.3390/nano12162742.
[47] J.H. Cabot, E.G. Ross, Evaluating prediction model performance, Surgery 174 (3) (2023) 723–726. http://10.1016/j.surg.2023.05.023.
[48] F. Shao, Z. Lu, C. Liu, H. Han, K. Chen, W. Li, et al., Hierarchical nanogaps within bioscaffold arrays as a high-performance SERS substrate for animal virus biosensing, Acs. Appl. Mater.Inter. 6 (9) (2014) 6281–6289. http://10.1021/am4045212.
[49] M.Y. Lv, H.Y. Teng, Z.Y. Chen, Y.M. Zhao, X. Zhang, L. Liu, et al., Low-cost Au nanoparticle-decorated cicada wing as sensitive and recyclable substrates for surface enhanced Raman scattering, Sens. Actuat. B Chem. 209 (2015) 820–827, https://doi.org/10.1016/j.snb.2014.12.061.
[50] M. Zhang, J. Meng, D. Wang, Q. Tang, T. Chen, S. Rong, et al., Biomimetic synthesis of hierarchical 3D Ag butterfly wing scale arrays/graphene composites as ultrasensitive SERS substrates for efficient trace chemical detection, J. Mater. Chem. C 6 (8) (2018) 1933–1943. http://10.1039/C7TC03922K.
[51] S. Lu, Z. Yin, L. Zhang, J. Du, C. Jing, On-site detection of multiple extracellular antibiotic resistance genes using SERS, Sens. Actuat. B Chem. 369 (2022) 132262, https://doi.org/10.1016/j.snb.2022.132262.
[52] W. Zhou, X. Han, G. Shi, W. Han, M. Wang, W. Yuan, et al., Machine learning-driven grating-like SERS Platform toward ultra-sensitive detection of forsythin, J. Lumin. 263 (2023) 120085. http://:10.1016/j.jlumin.2023.120085.
[53] W.-b. Li, X. Lu, R. Yang, F. Liang, W.-d. Chen, Z.-w. Xie, et al., Highly sensitive and reproducible SERS substrates with binary colloidal crystals (bCCs) based on MIM structures, Appl. Surf. Sci. 597 (2022) 153654. http://10.1016/j.apsusc.2022.153654.
[54] W. Si, Y. Qiao, Z. Liu, G. Jin, Y. Liu, X. Xue, et al., Combination of multi-model statistical analysis and quantitative fingerprinting in quality evaluation of Shuang-huang-lian oral liquid, Anal. Bioanal. Chem. 412 (26) (2020) 7073–7083. http://10.1007/s00216-020-02841-z.
[55] J. Liang, F.Q. Gong, H.M. Sun, Simultaneous separation of eight lignans in forsythia suspensa by β-cyclodextrin-modified capillary zone electrophoresis, Molecules 23 (3) (2018) 514. http://10.3390/molecules23030514.
[56] H. Guo, A.-H. Liu, L. Li, D.-A. Guo, Simultaneous determination of 12 major constituents in Forsythia suspensa by high performance liquid chromatography—DAD method, J. Pharm. Biomed. Anal. 43 (3) (2007) 1000–1006, https://doi.org/10.1016/j.jpba.2006.09.033.
[57] Y. Cui, Q. Wang, X. Shi, X. Zhang, X. Sheng, L. Zhang, Simultaneous quantification of 14 bioactive constituents in Forsythia suspensa by liquid chromatography-electrospray ionisation-mass spectrometry, Phytochem. Anal. 21 (3) (2010) 253–260. http://10.1002/pca.1194.