SiO_2 Nanolayer Regulated Ag@Cu Core-Shell SERS Platform Integrated Machine Learning for Intelligent Identification of Jujuboside A, Saikosaponin A and Timosaponin A-III

Wenying Zhou, Xue Han, Yanjun Wu, Guochao Shi, Shiqi Xu, Mingli Wang, Wenzhi Yuan, Jiahao Cui, and Zelong Li

***Abstract*—An effective SERS-based detection method assisted by machine learning algorithm has been developed to intelligently identify the pharmacodynamic substances in traditional Chinese medicine (TCM) which aided in the quality control and identification of TCM. In this work, the ultrasensitive grating-like core-shell Ag@SiO_2@Cu@moth wing (Ag@SiO_2@Cu@MW) SERS platform was explored which can precisely control the intergap distances by adjusting the thickness of the SiO_2 layer. This SERS platform demonstrated reproducibility and reliability, with a low relative standard deviation (RSD) of 7.38% and an enhancement factor (EF) of 2.49 $\times$ 10^7. The optimized Ag_{30}@SiO_2(4)@Cu_{20}@MW substrate were then applied in the accurately analyzing pharmacodynamic substances. Specifically, the label-free SERS analysis showed the distinct spectral features for Jujuboside A, Saikosaponin A and Timosaponin A-III. Machine learning algorithms, such as principal component analysis (PCA), decision tree (DT), support vector machine (SVM), k-nearest neighbors (kNN) were employed and further in differentiating with the three pharmacodynamic substances Raman spectrum groups. These results indicate that SERS technology in combination with machine learning algorithms can not only achieve rapid and accurate detection of different types of pharmacodynamic substances, but also promote the modernization and international application of TCM.**

***Index Terms*—Grating-like core-shell structure, machine learning, pharmacodynamic substances in traditional Chinese medicine, SERS.**

Manuscript received 28 March 2024; revised 16 May 2024; accepted 3 June 2024. Date of publication 6 June 2024; date of current version 13 June 2024. This work was supported in part by Hebei Natural Science Foundation under Grant F2022406001, in part by the Graduate Student Innovation Ability Training of Education of Hebei Province Department under Grant CXZZSS2024119, in part by the Chengde Biomedicine Industry Research Institute Funding Project under Grant 202205B086, and in part by the Technology and Starting Fund for Scientific Research of High-level Talents of Chengde Medical University-Nature under Grant 202206. *(Corresponding authors: Guochao Shi; Shiqi Xu; Mingli Wang.)*

Wenying Zhou, Yanjun Wu, Guochao Shi, Shiqi Xu, Wenzhi Yuan, Jiahao Cui, and Zelong Li are with the Hebei International Research Center for Medical-Engineering, Chengde Medical University, Chengde 067000, China (e-mail: sgc@cdmc.edu.cn; xsq@cdmc.edu.cn).

Xue Han is with the Department of Neurology, Affiliated Hospital of Chengde Medical University, Chengde 067000, China.

Mingli Wang is with the 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 (e-mail: wml@ysu.edu.cn).

Digital Object Identifier 10.1109/JPHOT.2024.3410697

I. Introduction

POPULAR with the public for its mildness, long-lasting effects and fewer side effects, Chinese medicine has been used for centuries to prevent and treat disease [1]. Therefore, quality control of traditional Chinese medicine (TCM) is especially important. In the past time, most of the traditional methods, such as chromatography and mass spectrometry, have been used to detect the pharmaceutical substances in TCM [2], [3], [4]. However, these traditional methods have limitations such as complicated pre-treatment steps, the need for expensive instruments, complex operation and the lack of portability of the instruments. Therefore, there is a great need for a highly sensitive, rapid and accurate method for the detection of pharmacodynamic substances in TCM.

In recent years, the development of nanotechnology and Raman spectroscopy has made SERS technology a powerful detection technique [5], [6], [7], [8], and even developed to the stage where single molecule detection can be realized [9], [10], [11]. The SERS signal is significantly enhanced by the electromagnetic and chemical enhancement mechanism, thus overcoming the drawback of the weaks signal of Raman spectroscopy [12], [13]. The Raman signals will be significantly enhanced in the "hot spots" regions of SERS, and most of these "hot spots" appear at the nanoparticle gaps and the tips of plasma nanomaterials[14]. Therefore, it becomes crucial to prepare plasmonic nanostructures with strong electromagnetic coupling effects as SERS substrates. Noble metal nanomaterials (Au, Ag and Cu) are becoming popular for the preparation of SERS substrates due to the local surface plasmon resonance (LSPR) effect of free electrons in the visible range [15], [16], [17], [18]. Among these noble metal nanomaterials, Ag has the greatest electromagnetic enhancement effect [19], [20], [21], [22], which produces SERS signal strengths 10 to 100 times higher than those of Au-based substrates [23]. Cu, in contrast, produces a poor LSPR effect and is easy to aggregation and oxidation in air [24], and its electromagnetic enhancement and stability are weaker than Au and Ag. Therefore, in order to suppress the aggregation and oxidation of Cu nanoparticles and prepare a highly sensitive plasma nanoplatform, we used SiO_2 to modulate the Cu-Ag nanolayer spacing to prepare

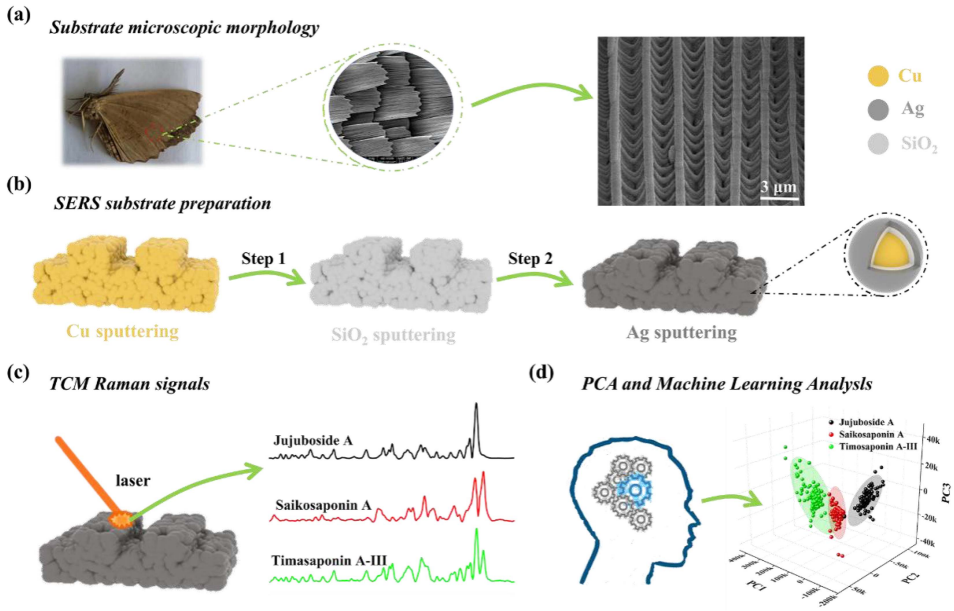

Fig. 1. Schematic illustration of the approach for preparing the grating-like core-shell structure SERS substrate. (a) Characterization of microscopic morphology of MW; (b) Preparation of the grating-like core-shell structure SERS substrate; (c) Collection of SERS signals of different pharmacodynamic substances in TCM via the SERS substrate; (d) Principal component analysis and machine learning analysis of spectral data.

this SERS substrate. SiO_2 effectively prevents the aggregation of nanoparticles and exhibits a strong plasma coupling effect between Cu and Ag nanolayers. At present, many methods have been developed to prepare highly reliable and sensitive SERS substrates. Among them, the magnetron sputtering technology does not need to introduce chemical reagents like the chemical preparation method, effectively avoiding pollution, and is an ideal technology for preparing SERS substrates [25], [26], [27].

It is well known that the choice of substrate material, the construction of morphology, and the size have an impact on the practical application of SERS substrates. In recent years, researchers have explored SERS substrates with different morphological structures using different nanotechnologies [28], [29], [30]. Among these plasmonic nanostructured structures, 3D nanostructures that can provide a large number of "hot spots" have attracted much attention [31]. Moths have evolved over time to form a series of regular grating-like structures on their wing surfaces, and the use of this bioscaffold as a substrate for template preparation greatly simplifies the preparation process and reduces costs.

Due to the complexity of Raman spectral data, traditional linear analysis methods not satisfied with the needs for intelligent data analysis and pattern recognition. In fact, machine learning, as a powerful data analysis tool, has achieved outstanding results in the field of spectral data analysis [32], [33], [34]. For example, Fang et al. have proposed principal component analysis (PCA) combined with support vector machine model to realize the trace detection of ampicillin in milk [35]. Dong et al. has combined dynamic SERS with SVM model for rapid identification of drugs in human urine [36]. Guillén et al. successfully achieved the classification of white and Iberian pork using near-infrared spectroscopy combined with neural network algorithms [37]. These examples are sufficient to show that the combination of SERS and machine learning algorithms provides an advanced method for substance identification with some potential for development.

In summary, in this paper, using a natural MW as a template, a grating-like core-shell structure SERS substrate with SiO_2 regulating the nanogaps of Cu and Ag nanolayers was prepared by magnetron sputtering technology. Combined with machine learning method, three pharmaceutical substances in TCM were identified, and the overall process was shown in Fig. 1. The results showed that the SVM model was the best predictive model for Raman spectroscopy analysis of pharmacodynamic substances with an average accuracy (ACC) of 0.924 ± 0.011. In view of the above, we conclude that machine learning methods are effective for the identification of pharmacodynamic substances in TCM, and they have great potential in the rapid and sensitive detection of pharmacodynamic substances in TCM in the near future.

II. Experimental Section

A. Chemicals and Materials

The Ag target, Cu target and SiO_2 target were purchased from Nanchang Hanchen New Material Technology Co., Ltd., Nanchang, China. Pharmacodynamic substances in TCM (Timosaponin A-III, Saikosaponin A and Jujuboside A) were obtained from Nakeli Biological Technology Co., Ltd., Chengdu, China. The MW were obtained from Beijing Jiaying Grand Life Sciences Co., Ltd. The methanol solution was purchased from Xilong Science Co., Ltd. R6G was obtained from J&K Scientific Ltd., Beijing, China. 4-aminothiophenol (4-ATP) was purchased from Aladdin Biochemical Technology Co., Ltd., Shanghai, China. The deionized water with a resistivity of 18.2 MΩ was used in all experiments.

B. Preparation of SERS Substrates

Natural moth wings were washed in ethanol solution to remove surface impurities and dried naturally. Subsequently, the substrates were prepared using a magnetron sputtering system. The magnetron sputtering process is widely regarded as an excellent substrate preparation method due to its controllability, low temperature and low damage [38]. First, the Cu@MW structure was constructed by deposition of Cu on the clean MW structure surface under experimental conditions with sputtering power of 100 W and argon flow rate set at 100 ml/m. Next, SiO_2 was modified on the Cu@MW structure to build the SiO_2@Cu@MW structure, and the sputtering power was set to 23 W at this time. Finally, the Ag nanoparticles were modified on the SiO_2@Cu@MW structure to construct the Ag@SiO_2@Cu@MW structure (the experimental parameters of magnetron sputtering were kept the same as the experiments for sputtering Cu). In summary, we have prepared a core-shell structured SERS substrate with Cu as the core, Ag as the shell and SiO_2 to regulate the nanogap on the surface of a grating-like MWs. In the experiment, the sputtering time of Cu and Ag nanoparticles was 20 min and 30 min, respectively, while the sputtering time of SiO_2 was controlled at 4 min and 8 min. Based on the above description, we finally constructed two SERS substrates: Ag_{30}@SiO_2(8)@Cu_{20}@MW and Ag_{30}@SiO_2(4)@Cu_{20}@MW.

C. SERS Measurements and Characterization

In the experiments, the field emission scanning electron microscopy (FE-SEM) (FEI-NOVA NANOSEM 230, America) and angatomic force microscopy (AFM) (Bruker Dimension ICON, America) were employed to characterize the morphology of SERS substrates. A thin gold layer was sprayed on the surface of the sample to increase its electrical conductivity before SEM measurements.

Raman measurements were performed employing a microscopic confocal Raman spectrometer (DXR2xi, Thermo Fisher Inc., USA). A 50× microscope with a wavelength of 532 nm and an effective power of 10 mW was used for measurement. The detection limit on the grating-like core-shell structure SERS substrate was determined using 10 μL droplets of 4-ATP at concentrations of 1×10^{-10} M to 1×10^{-6} M.

D. Detection of TCM Based on Machine Learning

In order to demonstrate practical applications, we evaluated the applicability of the grating-like core-shell structure SERS substrate for detecting various pharmacodynamic substances in TCM. Specifically, Jujuboside A, Saikosaponin A and Timosaponin A-III were chosen as pharmacodynamic substances representatives. Pharmacodynamic substances were dissolved in deionized water or methanol solution, respectively. 10 μL of the pharmacodynamic substances in TCM solution was dropped on the prepared SERS substrate, and the SERS spectra were collected after the solution was completely dried.

To identify the pharmacodynamic substances, we resort to the machine learning method. It is worth noting that the spectral data need to be pre-processed before the model was constructed. The specific process was shown in Fig. 2. Firstly, collect spectral data of pharmacological substances to build a mixed dataset. Secondly, preprocess the collected dataset. The Savitzky-Golay method was used for smoothing to eliminate the interference of background noise, and baseline correction was performed to eliminate the influence of baseline drift. Then, classification algorithm analysis was performed on spectral data with Raman shifts in the range of 600 cm^{-1}-1800 cm^{-1}. Finally, PCA was performed on the entire spectral dataset and the results were fed into machine learning models such as support vector machines (SVM), k-nearest neighbors (kNN), decision trees (DT) and Naïve Bayes (NB) for classification analysis. For the entire dataset, we selected 70% of the Raman spectrum data for model training and the rest for testing.

E. Performance Evaluation

For model performance evaluation, the performance of the model was assessed using the true positive/negative rate and the false positive/negative rate, which were represented in the confusion matrix. The confusion matrix visualises the computed results of the predicted and actual classes through a table. By use of these metrics, the accuracy, recall, sensitivity, specificity, precision and F_1 score of the classifier were obtained. Where accuracy represents the percentage of samples that were correctly classified, and sensitivity and specificity represent the percentage of actual positives and actual negatives that were correctly predicted, respectively. The precision, also known as the checking rate, which indicates the proportion of samples with positive predictions that are actually positive. Recall, also called the check rate, represents the number of positive cases predicted by the sample as a proportion of the number of true positive cases in the sample. Kappa coefficient is a value between 0 and 1, which is a measure of classification accuracy. The F_1-score is the reconciled mean of precision and recall, and usually, the larger the F_1-score, the better the model performance. The definition of the metrics were defined as follows [39]:

$$Sensitivity \ = \ Recall \ = \frac{true\ positive}{true\ positive \ + false\ nagative}$$

$$Specificity \ = \frac{true\ negative}{true\ negative \ + false\ negative}$$

$$Accuracy \ = \frac{true\ positive \ + true\ negative}{positive \ + negative}$$

$$Precision \ = \frac{true\ positive}{true\ positive \ + false\ positive}$$

$$F_1 - score \ = \ \frac{2 \times Precision \ \times Recall}{Precision \ + Recall}$$

In this study, we trained four machine learning models and by comparing the performance of the four models, we identified SVM as the best prediction model and calculated its confusion matrix for further analysis.

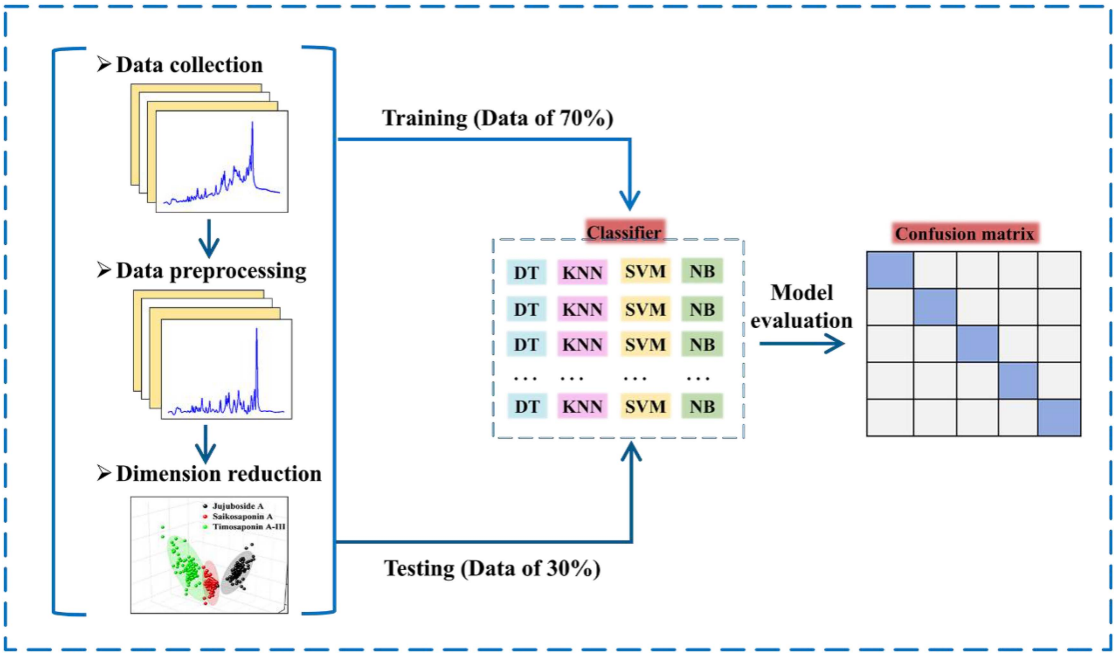

Fig. 2. Flowchart of machine learning classification algorithm.

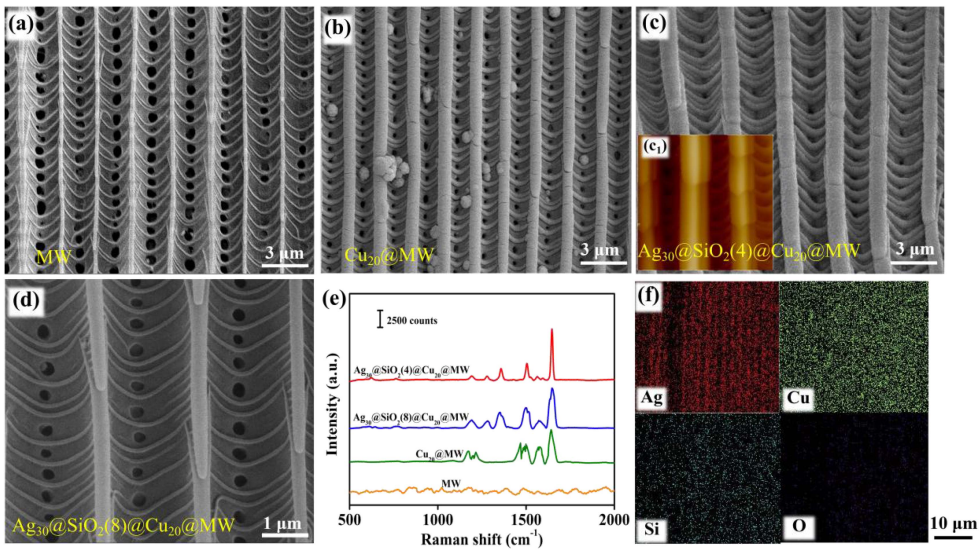

Fig. 3. (a)–(d) FE-SEM image of MW, Cu_{20}@MW, Ag_{30}@SiO_2(4)@Cu_{20}@MW (Inset (c_1) showed the AFM analysis of the Ag_{30}@SiO_2(4)@Cu_{20}@MW substrate) and Ag_{30}@SiO_2(8)@Cu_{20}@MW, respectively; (e) SERS spectra of 1×10^{-3} M R6G absorbed on various substrates; (f) Energy dispersive spectroscopy images of Ag, Cu, Si and O elements.

III. Results and Discussion

A. Morphology Characterization and Substrates Screening

Moths make up around 90% of Lepidoptera species and is the largest groups of Lepidoptera. We characterized the microstructure of MW as shown in Fig. 3(a). The surface of the MWs consists of many scales, each with regularly arranged longitudinal ribs and grooves. Adjacent longitudinal ribs are connected by nanoscale parabolic ribs. This special surface morphology greatly improved the SERS performance of the grating-like substrate. Therefore, we constructed plasmonic nanostructures based on MWs. Fig. 3(b)–(d) showed the FE-SEM images of Cu_{20}@MW, Ag_{30}@SiO_2(4)@Cu_{20}@MW and Ag_{30}@SiO_2(8)@Cu_{20}@MW, respectively. Fig. 3(e) showed the comparative Raman spectra of four substrates enriched with 1×10^{-3} M R6G molecules. It can be seen that the spectral lines on the Ag_{30}@SiO_2(4)@Cu_{20}@MW SERS substrates with cleaner and more distinct characteristic peaks, and the peaks were narrower in the positions of the characteristic peaks, which can better reflect the best SERS performance. The surface morphology of the fabricated Ag_{30}@SiO_2(4)@Cu_{20}@MW substrate was characterized using FE-SEM as shown in Fig. 3(c).

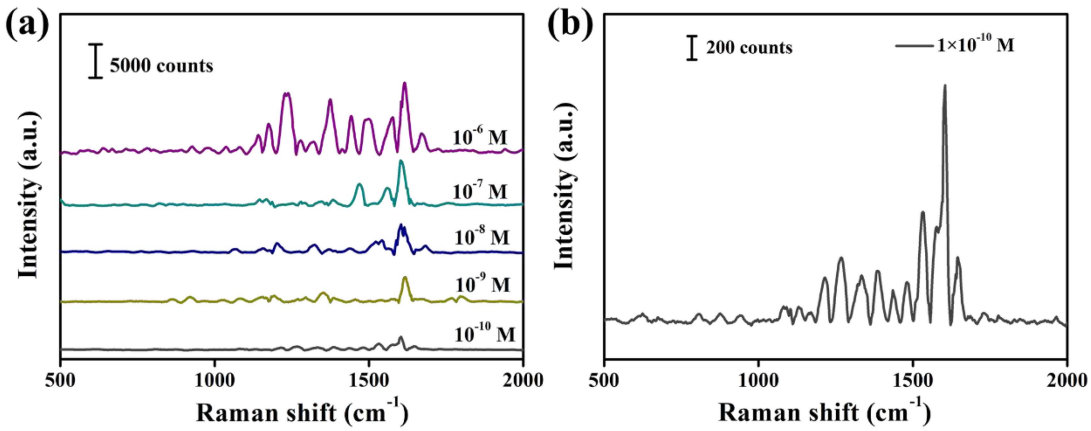

Fig. 4. (a) SERS spectra of 4-ATP solution with different concentrations on Ag_{30}@SiO_2(4)@Cu_{20}@MW SERS substrate; (b) SERS spectra of 10^{-10} M 4-ATP absorbed on Ag_{30}@SiO_2(4)@Cu_{20}@MW SERS substrate.

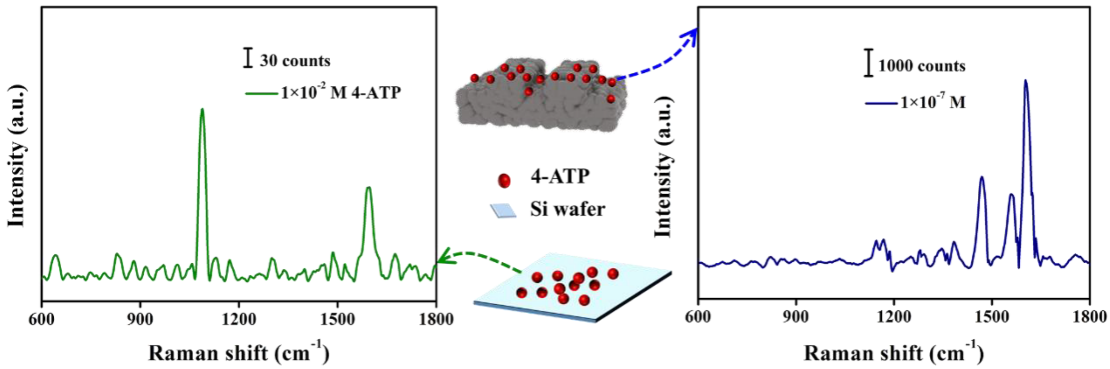

Fig. 5. (a) SERS spectrum of 10^{-2} M 4-ATP solution on Si wafer; (b) SERS spectrum of 10^{-7} M 4-ATP solution on Ag_{30}@SiO_2(4)@Cu_{20}@MW SERS substrate.

The surface of the substrate had become rough compared to the original morphology of the moth wing, the ribs and rib grooves were modified with nanoparticles, but the ribs remained regularly distribute. Furthermore, it is also clearly visible from the EDS mapping (Fig. 3(f)) that the elements were grown homogeneously on the surface. In addition, AFM analysis was also performed on the Ag_{30}@SiO_2(4)@Cu_{20}@MW substrate (Fig. 3(c_1)) for the characterization of the surface morphology of the substrate

B. SERS Performance Exploration

To verify the sensitivity and enhancement effect of the grating-like core-shell structure SERS substrate, 4-ATP was used as a probe molecule for the experiments. Fig. 4(a) exhibited the SERS spectra of the substrates adsorbed with different concentrations of 4-ATP solution, and it can be seen that the Raman signal intensity showed a decreasing trend with the decrease of 4-ATP concentration, which demonstrated that the substrate SERS signal intensity was proportional to the concentration of the probe molecule. As for the detection limit, at a concentration of 10^{-10} M, a small peak was still observed, as shown in Fig. 4(b).

In order to further evaluate the performance of the SERS substrate, the enhancement factor (EF) was calculated by (1) [40], EF is a SERS criterion for the ability to distinguish between low energy levels and SERS surface sensitivity.

$$\mathrm{EF} = \frac{I_{\mathrm{sers}}/N_{\mathrm{sers}}}{I_{\mathrm{Raman}}/N_{\mathrm{Raman}}} \tag{1}$$

where I_{sers} denotes the 1597 cm^{-1} peak intensity of 10^{-7} M 4-ATP solution on the SERS substrate and I_{Raman} denotes the intensity of 10^{-2} M 4-ATP solution at the same wavelength on Si wafer, as shown in Fig. 5(a) and (b). N_{sers} and N_{Raman} represent the number of 4-ATP molecules aggregated on the laser spot on the SERS substrate and on the Si wafer, respectively. The ratio of I_{sers}/I_{Raman} was calculated to be 29.84. The number of 4-ATP probe molecules (N: N_{Raman} and N_{sers}) was defined by (2):

$$\mathrm{N} = \frac{\mathrm{N_A}\times\mathrm{M}\times\mathrm{V_{Solution}}}{\mathrm{S_{Sub}}}\times\mathrm{S_{laser}} \tag{2}$$

where N_A is the Avogadro constant, M is the molar concentration of the 4-ATP probe molecule solution, $V_{solution}$ represents the volume of the 10 μL 4-ATP solution, S_{laser} is the laser spot area and S_{sub} is the area of the 4-ATP solution absorbed on a different substrate. In the experiment, the area of 4-ATP probe molecule solution on Si wafer was approximately 1.2 times that of 4-ATP solution on the SERS substrate. In the experiment, the laser spot diameter was about 1 μm, so, the S_{laser} was 0.785 μm^2. Finally, the calculated EF of the SERS substrate was close to 2.49×10^7.

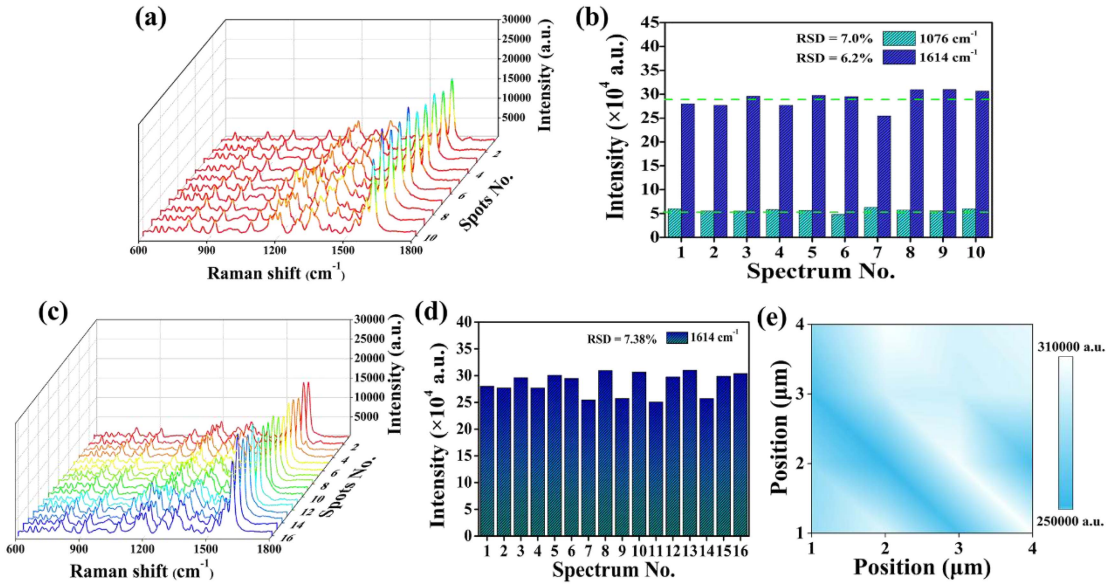

Fig. 6. (a) SERS spectra at 10 random positions on $Ag_{30}@SiO_2(4)@Cu_{20}@MW$ SERS substrate; (b) SERS peaks intensity at 1076 cm^{-1} and 1614 cm^{-1}; (c) Uniformity test of $Ag_{30}@SiO_2(4)@Cu_{20}@MW$ SERS substrates; (d) The Raman intensities of the characteristic peak at 1614 cm^{-1} on the $Ag_{30}@SiO_2(4)@Cu_{20}@MW$ SERS substrate and the corresponding RSD values; (e) The SERS mapping of the characteristic peak at 1614 cm^{-1} on $Ag_{30}@SiO_2(4)@Cu_{20}@MW$ SERS substrate.

Besides sensitivity, reproducibility is a key factors for substrate performance in SERS detection and determines the progress of SERS technology towards practical applications [41], [42]. Fig. 6(a) showed the SERS spectra of 0.01 g/l Jujuboside A at 10 different random points, which can be visualized that the SERS substrate has a high reproducibility. To further assess the reproducibility of the substrate, RSD values of 7.0% and 6.2% were calculated at the characteristic peaks at 1076 cm^{-1} and 1614 cm^{-1}, as shown in Fig. 6(b). It indicated that the substrate has good reproducibility.

In addition, to further evaluate the uniformity of the $Ag_{30}@SiO_2(4)@Cu_{20}@MW$ substrate, we randomly selected 16 locations on the SERS substrate to recorded the SERS spectra of the 0.01 g/l Jujuboside A solution, and the results were shown in Fig. 6(c). It can be seen that the peak intensities of the different spectra were almost the same for all SERS active positions. Fig. 6(d) showed the RSD of the characteristic peaks located at 1614 cm^{-1} which was based on the $Ag_{30}@SiO_2(4)@Cu_{20}@MW$ SERS substrate with a calculated value of 7.38%. The SERS mapping of the Raman characteristic peak intensity of 0.01 g/l Jujuboside A solution at 1614 cm^{-1} on a randomly selected 4×4 μm^2 area of $Ag_{30}@SiO_2(4)@Cu_{20}@MW$ SERS substrate was exhibited in Fig. 6(e), where each pixel represented its corresponding Raman signal intensity. The uniform color array indicated that the $Ag_{30}@SiO_2(4)@Cu_{20}@MW$ SERS substrate had good uniformity.

C. Identification of Pharmacodynamic Substances in TCM With Machine Learning

Timosaponin A-III is one of the main chemical components in the herb Anemarrhena asphodeloides, which has the effect of inhibiting the expression of cancer cells [43]. Jujuboside A have anti-oxidant and anti-inflammatory properties [44]. The active ingredient of Radix Bupleuri, saikosaponin A, can relieve inflammation and inhibit tumor growth [45], [46]. Anemarrhena asphophariae, Semen ziziphi spinosae and Radix Bupleuri, as the genuine medicinal materials in Hebei Province, have precious medicinal value, and the research on their pharmacodynamic substances has a profound impact on the development of TCM.

In this study, the Raman spectra of pharcodynamic substances obtained from SERS substrates of grating-like core-shell structures were evaluated using machine learning algorithms in combination with PCA. PCA was performed on the obtained spectral dataset, and PCA is commonly used as a multivariate data analysis tool for dimensionality reduction of the data. It works by extracting the most meaningful information from spectral data onto principal components (PCs), and then constructing a classification model to distinguish between different samples. Fig. 7(a) illustrated the principal component loading plot of the spectral dataset, the first three principal components account for 95% of the total Raman variation across the spectra, and thus we chose the first three principal components as the input data for the subsequent machine learning model. Based on the loadings, we plotted the scatter plots of the first three scores as shown in Fig. 7(b), and it can be seen that each potent substance was clustered around its internal centre point in three-dimensional space (here each point represents the Raman spectrum of each sample), but there were still cases of misclassification. Therefore, we further selected machine learning algorithms to analyze the spectral dataset. To be specific, the results of PCA were used as inputs to the machine learning classification models, and the spectral dataset was analyzed using four different machine learning models, including NB, DT, KNN and SVM. In the test of the machine learning model, we performed 10 data splits.

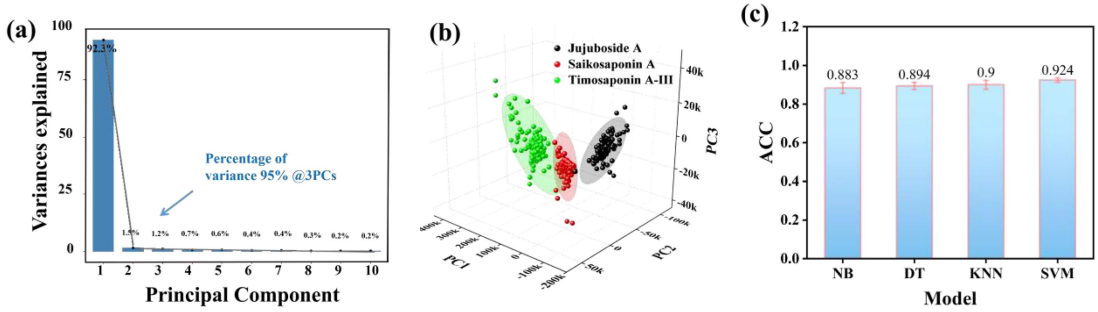

Fig. 7. (a) Principal component loading plots for spectral datasets; (b) Principal component analysis 3D scatterplot; (c) Performance of four machine learning models.

TABLE I
ACCURACY SUMMARY OF FOUR MODELS

Model	NB	DT	KNN	SVM
ACC	0.883 ± 0.028	0.894 ± 0.018	0.900 ± 0.023	0.924 ± 0.011

TABLE II
PERFORMANCE SUMMARY OF SVM MODEL

pharmacodynamic substance	Precision	Recall	F_1-score	Specificity	Sensitivity
Jujuboside A	0.955 ± 0.016	0.980 ± 0.017	0.967 ± 0.011	0.977 ± 0.009	0.980 ± 0.017
Saikosaponin A	0.916 ± 0.042	0.890 ± 0.032	0.904 ± 0.026	0.958 ± 0.024	0.890 ± 0.032
Timosaponin A-III	0.904 ± 0.019	0.903 ± 0.040	0.903 ± 0.016	0.952 ± 0.012	0.903 ± 0.040

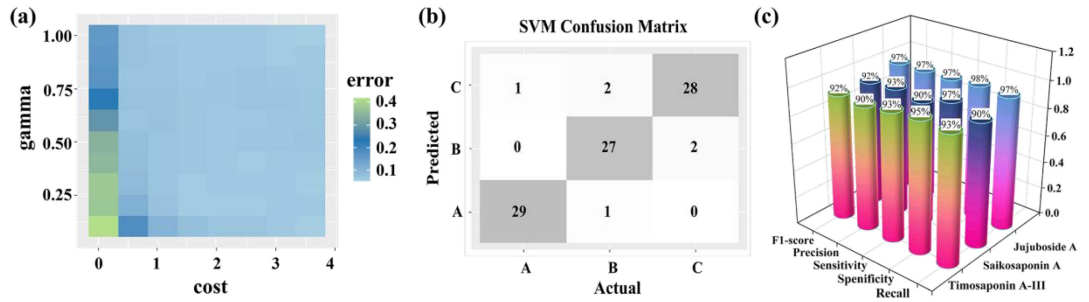

Fig. 8. (a) Penalty and gamm parameters for SVM models; (b) Confusion matrix of the pharcodynamic substances in TCM prediction using SVM (A: Jujuboside A, B: Saikosaponin A and C: Timosaponin A-III); (c) Evaluation indexes for classification of three pharmacodynamic substances based on SVM model.

The prediction performance (average accuracy and its standard deviation) of the four machine learning methods were shown in Table I. It can be seen that the SVM model achieved an accuracy of 0.924 ± 0.011 in classifying pharmacodynamic substances of TCM, as shown in Fig. 7(c). Therefore, we further analyzed and evaluated the performance of the SVM method (best performance in ten model trainings) for analyzing the spectral dataset.

SVM is a machine learning technique that relies on optimal hyperplanes or decision boundaries for classification [47], [48], where the definition of hyperplane boundaries depends on different kernel functions. The radial basis kernel function has excellent performance in model training [49], thus we adopt the radial basis kernel function as the kernel function of the SVM and the model performance is best with penalty parameter C = 3.6 and kernel function parameter gamma = 0.1 (Fig. 8(a)). Furthermore, we computed the corresponding classification confusion matrices that provide the classification details of the SVM model in terms of the prediction of pharcodynamic substance species, as shown in Fig. 8(b). The horizontal coordinate of the confusion matrix represents the actual category of the sample, while the vertical coordinate refers to the category in which the sample was predicted. According to the matrix, the SVM model showed high accuracy for all three pharcodynamic substances in TCM, and out of a total of 90 spectra, 84 spectra were correctly predicted and assigned to the corresponding pharcodynamic substances using the model, with an overall accuracy of 93.3%. The model showed excellent sensitivity, specificity, precision, recall and F_1-score for all the potent substances (Fig. 8(c)), as well as high performance in ten SVM model trainings (Table II),

which demonstrated that the SVM classifier was effective in discriminating the three pharcodynamic substances in TCM.

V. Conclusion

In this study, a grating-like core-shell structure SERS substrate with SiO_2 modulating the spacing of Cu and Ag nanolayers was proposed, which exhibited high sensitivity, low detection limit and excellent reproducibility. We combined this SERS substrate with a machine learning algorithm to interpret the SERS spectra. The SERS spectral data of three pharmacodynamic substances in TCM were downscaled using PCA. Four machine learning models, DT, KNN, NB and SVM model were constructed for classification of three pharmacodynamic substances. Among these models, the SVM model gained the best prediction performance with an average accuracy of 0.924 ± 0.011. In conclusion, our study combined machine learning algorithms with SERS spectroscopy to produce a fast, convenient, and accurate identification method, which is expected to open up a new path for the identification of pharmacodynamic substances in TCM.

Acknowledgment

The authors would like to thank the anonymous reviewers for their valuable suggestions.

References

[1] F. Yang et al., "Establishment of the thin-layer chromatography-surface-enhanced Raman spectroscopy and chemometrics method for simultaneous identification of eleven illegal drugs in anti-rheumatic health food," *Food Biosci.*, vol. 49, Oct. 2022, Art. no. 101842, doi: 10.1016/j.fbio.2022.101842.

[2] M. Yang et al., "Phytochemical analysis of traditional Chinese medicine using liquid chromatography coupled with mass spectrometry," *J. Chromatography A*, vol. 1216, no. 11, pp. 2045–2062, Mar. 2009, doi: 10.1016/j.chroma.2008.08.097.

[3] P. Xie, S. Chen, Y. Z. Liang, X. Wang, R. Tian, and R. Upton, "Chromatographic fingerprint analysis–A rational approach for quality assessment of traditional Chinese herbal medicine," *J. Chromatography A*, vol. 1112, no. 1/2, pp. 171–180, Apr. 2006, doi: 10.1016/j.chroma.2005.12.091.

[4] C. Y. Chen, Y. H. Li, Z. Li, and M. R. Lee, "Characterization of effective phytochemicals in traditional Chinese medicine by mass spectrometry," *Mass Spectrometry Rev.*, vol. 42, no. 5, pp. 1808–1827, May 2022, doi: 10.1002/mas.21782.

[5] J. F. Li, Y. J. Zhang, S. Y. Ding, R. Panneerselvam, and Z. Q. Tian, "Core-shell nanoparticle-enhanced raman spectroscopy," *Chem. Rev.*, vol. 117, no. 7, pp. 5002–5069, Apr. 2017, doi: 10.1021/acs.chemrev.6b00596.

[6] H. He, D. W. Sun, H. Pu, L. Chen, and L. Lin, "Applications of Raman spectroscopic techniques for quality and safety evaluation of milk: A review of recent developments," *Crit. Rev. Food Sci. Nutr.*, vol. 59, no. 5, pp. 770–793, Jan. 2019, doi: 10.1080/10408398.2018.1528436.

[7] X. Xie, H. Pu, and D. W. Sun, "Recent advances in nanofabrication techniques for SERS substrates and their applications in food safety analysis," *Crit. Rev. Food Sci. Nutr.*, vol. 58, no. 16, pp. 2800–2813, Aug. 2018, doi: 10.1080/10408398.2017.1341866.

[8] Y. Jiang, D. W. Sun, H. Pu, and Q. Wei, "Ultrasensitive analysis of kanamycin residue in milk by SERS-based aptasensor," *Talanta*, vol. 197, pp. 151–158, May 2019, doi: 10.1016/j.talanta.2019.01.015.

[9] L. Guerrini and D. Graham, "Molecularly-mediated assemblies of plasmonic nanoparticles for Surface-Enhanced Raman spectroscopy applications," *Chem. Soc. Rev.*, vol. 41, no. 21, pp. 7085–7107, Jul. 2012, doi: 10.1039/C2CS35118H.

[10] Z. Zhou et al., "Silver nanocubes monolayers as a SERS substrate for quantitative analysis," *Chin. Chem. Lett.*, vol. 32, no. 4, pp. 1497–1501, Apr. 2021, doi: 10.1016/j.cclet.2020.10.021.

[11] S.-H. Luo et al., "Visualization of a machine learning framework toward highly sensitive qualitative analysis by SERS," *Anal. Chem.*, vol. 94, no. 28, pp. 10151–10158, Jul. 2022, doi: 10.1021/acs.analchem.2c01450.

[12] X. Zhu et al., "A novel graphene-like titanium carbide MXene/Au–Ag nanoshuttles bifunctional nanosensor for electrochemical and SERS intelligent analysis of ultra-trace carbendazim coupled with machine learning," *Ceram. Int.*, vol. 47, no. 1, pp. 173–184, Jan. 2021, doi: 10.1016/j.ceramint.2020.08.121.

[13] W. Du, S. Wei, N. Li, Z. Hao, Y. Li, and M. Wang, "Highly sensitive fiber optic enhanced raman scattering sensor," *Opt. Laser Technol.*, vol. 168, Jan. 2024, Art. no. 109879, doi: 10.1016/j.optlastec.2023.109879.

[14] J. Zhao, L. Long, G. Weng, J. Li, J. Zhu, and J.-W. Zhao, "Multi-branch Au/Ag bimetallic core–shell–satellite nanoparticles as a versatile SERS substrate: The effect of Au branches in a mesoporous silica interlayer," *J. Mater. Chem. C*, vol. 5, no. 48, pp. 12678–12687, Oct. 2017, doi: 10.1039/C7TC03788K.

[15] C. Gong and M. S. Leite, "Noble metal alloys for plasmonics," *ACS Photon.*, vol. 3, no. 4, pp. 507–513, Feb. 2016, doi: 10.1021/acsphotonics.5b00586.

[16] T. Tan et al., "LSPR-dependent SERS performance of silver nanoplates with highly stable and broad tunable LSPRs prepared through an improved seed-mediated strategy," *Phys. Chem. Chem. Phys.*, vol. 15, no. 48, pp. 21034–21042, Oct. 2013, doi: 10.1039/C3CP52236A.

[17] Y. Li, W. Du, S. Wei, S. Guo, T. Jiao, and M. Wang, "One-step assembly of TA-FeIII supramolecular shell on AgNPs surface for regulating fluorescence intensity of quantum dots," *Appl. Phys. Exp.*, vol. 16, no. 10, Oct. 2023, Art. no. 102001, doi: 10.35848/1882-0786/acff37.

[18] Y. Li, Z. Hao, H. Cao, S. Wei, T. Jiao, and M. Wang, "Study on annealed graphene oxide nano-sheets for improving the surface enhanced fluorescence of silver nanoparticles," *Opt. Laser Technol.*, vol. 160, May 2023, Art. no. 109054, doi: 10.1016/j.optlastec.2022.109054.

[19] C. Zhang et al., "SERS detection of R6G based on a novel graphene oxide/silver nanoparticles/silicon pyramid arrays structure," *Opt. Exp.*, vol. 23, no. 19, pp. 24811–24821, Sep. 2015, doi: 10.1364/oe.23.024811.

[20] C. Zhang et al., "SERS detection of low-concentration adenosine by silver nanoparticles on silicon nanoporous pyramid arrays structure," *Appl. Surf. Sci.*, vol. 347, pp. 668–672, Aug. 2015, doi: 10.1016/j.apsusc.2015.04.170.

[21] I. Pastoriza-Santos and L. M. Liz-Marzán, "Colloidal silver nanoplates. State of the art and future challenges," *J. Mater. Chem.*, vol. 18, no. 15, pp. 1724–1737, Feb. 2008, doi: 10.1039/B716538B.

[22] L. Guo, H. Cao, L. Cao, Y. Yang, and M. Wang, "SERS study of wheat leaves substrates with two different structures," *Opt. Commun.*, vol. 510, May 2022, Art. no. 127921, doi: 10.1016/j.optcom.2022.127921.

[23] X. Zheng et al., "High performance Au/Ag core/shell bipyramids for determination of thiram based on surface-enhanced Raman scattering," *J. Raman Spectrosc.*, vol. 43, pp. 1374–1380, Oct. 2012, doi: 10.1002/JRS.4087.

[24] P. Dai, H. Li, X. Huang, N. Wang, and L. Zhu, "Highly sensitive and stable copper-based SERS chips prepared by a chemical reduction method," *Nanomaterials*, vol. 11, no. 10, Oct. 2021, Art. no. 2770, doi: 10.3390/nano11102770.

[25] W. Yuan et al., "Optimization of surface enhanced raman scattering performance based on Ag nanoparticle-modified vanadium-titanium nanorods with tunable nanogaps," *Opt. Exp.*, vol. 30, no. 21, pp. 38613–38629, Oct. 2022, doi: 10.1364/oe.474108.

[26] H. Lim et al., "A mesopore-stimulated electromagnetic near-field: Electrochemical synthesis of mesoporous copper films by micelle self-assembly," *J. Mater. Chem. A*, vol. 8, no. 40, pp. 21016–21025, Oct. 2020, doi: 10.1039/D0TA06228F.

[27] D. Chen et al., "A bio-sensing surface with high biocompatibility for enhancing Raman scattering signals as enabled by a Mo-Ag film," *Analyst*, vol. 147, no. 7, pp. 1385–1393, Mar. 2022, doi: 10.1039/d2an00008c.

[28] M. Scuderi et al., "Nanoscale study of the tarnishing process in electron beam lithography-fabricated silver nanoparticles for plasmonic applications," *J. Phys. Chem. C*, vol. 120, no. 42, pp. 24314–24323, Oct. 2016, doi: 10.1021/acs.jpcc.6b03963.

[29] M. C. Giordano, A. Foti, E. Messina, P. G. Gucciardi, D. Comoretto, and F. Buatier de Mongeot, "SERS amplification from self-organized arrays of plasmonic nanocrescents," *ACS Appl. Mater. Inter.*, vol. 8, no. 10, pp. 6629–6638, Jul. 2016, doi: 10.1021/acsami.5b11843.

[30] L. Fabris, "Bottom-up optimization of SERS hot-spots," *Chem. Commun.*, vol. 48, no. 75, pp. 9346–9348, Jul. 2012, doi: 10.1039/C2CC34068B.

[31] R. Kodiyath et al., "Silver-decorated cylindrical nanopores: Combining the third dimension with chemical enhancement for efficient trace chemical detection with SERS," *J. Phys. Chem. C*, vol. 116, no. 26, pp. 13917–13927, Jul. 2012, doi: 10.1021/jp300902u.

[32] H. Dies, J. Raveendran, C. Escobedo, and A. Docoslis, "Rapid identification and quantification of illicit drugs on nanodendritic surface-enhanced raman scattering substrates," *Sens. Actuators B Chem.*, vol. 257, pp. 382–388, Oct. 2018, doi: 10.1016/J.SNB.2017.10.181.

[33] W. Lee et al., "Label-free prostate cancer detection by characterization of extracellular vesicles using Raman spectroscopy," *Anal.Chem.*, vol. 90, no. 19, pp. 11290–11296, Oct. 2018, doi: 10.1021/acs.analchem.8b01831.

[34] Y. J. Ai et al., "Rapid qualitative and quantitative determination of food colorants by both Raman spectra and surface-enhanced Raman scattering (SERS)," *Food Chem.*, vol. 241, pp. 427–433, Feb. 2018, doi: 10.1016/j.foodchem.2017.09.019.

[35] G. Fang et al., "Surface-enhanced Raman scattering platform with triple synergistic enhancement toward label-free detection of antibiotics in milk," *Small*, vol. 18, no. 45, Nov. 2022, Art. no. e2204588, doi: 10.1002/smll.202204588.

[36] R. Dong, S. Weng, L. Yang, and J. Liu, "Detection and direct readout of drugs in human urine using dynamic surface-enhanced Raman spectroscopy and support vector machines," *Anal. Chem.*, vol. 87, no. 5, pp. 2937–2944, Mar. 2015, doi: 10.1021/acs.analchem.5b00137.

[37] A. Guillén et al., "Using near-infrared spectroscopy in the classification of white and iberian pork with neural networks," *Neural Comput. Appl.*, vol. 19, no. 3, pp. 465–470, Jan. 2010, doi: 10.1007/s00521-009-0327-2.

[38] Y. Chen and Y. Fang, "Surface enhanced Raman scattering (SERS) activity studies of Si, Fe, Ti, Al and Ag films' prepared by magnetron sputtering," *Spectrochimica Acta A Mol. Biomol. Spectrosc.*, vol. 69, no. 3, pp. 733–737, Mar. 2008, doi: 10.1016/j.saa.2007.05.030.

[39] F. Lussier, V. Thibault, B. Charron, G. Q. Wallace, and J.-F. Masson, "Deep learning and artificial intelligence methods for Raman and surface-enhanced Raman scattering," *Trends Anal. Chem.*, vol. 124, Mar. 2020, Art. no. 115796, doi: 10.1016/j.trac.2019.115796.

[40] M. Choi, T. Kang, S. H. Choi, and K. M. Byun, "Dual modal plasmonic substrates based on a convective self-assembly technique for enhancement in SERS and LSPR detection," *Opt. Exp.*, vol. 29, no. 4, pp. 6179–6187, Feb. 2021, doi: 10.1364/oe.419051.

[41] L. Zhang, C. Guan, Y. Wang, and J. Liao, "Highly effective and uniform SERS substrates fabricated by etching multi-layered gold nanoparticle arrays," *Nanoscale*, vol. 8, no. 11, pp. 5928–5937, Mar. 2016, doi: 10.1039/c6nr00502k.

[42] W. Tang, Y. An, and K. Ho Row, "Fabrication of Au nanoparticles embedded holey g-C_3N_4 as SERS substrates for sensitive and reliable detection," *Chem. Eng. J.*, vol. 402, Dec. 2020, Art. no. 126194, doi: 10.1016/j.cej.2020.126194.

[43] J. E. Gergely, A. E. Dorsey, G. P. Dimri, and M. Dimri, "Timosaponin A-III inhibits oncogenic phenotype via regulation of PcG protein BMI1 in breast cancer cells," *Mol. Carcinogenesis*, vol. 57, no. 7, pp. 831–841, Jul. 2018, doi: 10.1002/mc.22804.

[44] Y. Zhong et al., "Jujuboside A ameliorates high fat diet and streptozotocin induced diabetic nephropathy via suppressing oxidative stress, apoptosis, and enhancing autophagy," *Food Chem. Toxicol.*, vol. 159, Jan. 2022, Art. no. 112697, doi: 10.1016/j.fct.2021.112697.

[45] L. Zhao, L. Jin, and B. Yang, "Saikosaponin A alleviates Staphylococcus aureus-induced mastitis in mice by inhibiting ferroptosis via SIRT1/Nrf2 pathway," *J. Cell. Mol. Med.*, vol. 27, no. 22, pp. 3443–3450, Nov. 2023, doi: 10.1111/jcmm.17914.

[46] P. Zhang et al., "Saikosaponin A, a triterpene saponin, suppresses angiogenesis and tumor growth by blocking VEGFR2-mediated signaling pathway," *Front. Pharmacol.*, vol. 12, Oct. 2021, Art. no. 713200, doi: 10.3389/fphar.2021.713200.

[47] S. Kang, I. Kim, and P. J. Vikesland, "Discriminatory detection of ssDNA by surface-enhanced raman spectroscopy (SERS) and tree-based support vector machine (Tr-SVM)," *Anal. Chem.*, vol. 93, no. 27, pp. 9319–9328, Jul. 2021, doi: 10.1021/acs.analchem.0c04576.

[48] W. S. Noble, "What is a support vector machine?," *Nat. Biotechnol.*, vol. 24, no. 12, pp. 1565–1567, Dec. 2006, doi: 10.1038/nbt1206-1565.

[49] C. Li and P. Xu, "Application on traffic flow prediction of machine learning in intelligent transportation," *Neural. Comput. Appl.*, vol. 33, no. 2, pp. 613–624, Jan. 2021, doi: 10.1007/s00521-020-05002-6.