



Rapid and quantitative detection of respiratory viruses using surface-enhanced Raman spectroscopy and machine learning

Yanjun Yang^{a,*}, Beibei Xu^b, Jackelyn Murray^c, James Haverstick^d, Xianyan Chen^b, Ralph A. Tripp^c, Yiping Zhao^{d,**}

^a School of Electrical and Computer Engineering, College of Engineering, The University of Georgia, Athens, GA, 30602, USA

^b Department of Statistics, The University of Georgia, Athens, GA, 30602, USA

^c Department of Infectious Diseases, College of Veterinary Medicine, The University of Georgia, Athens, GA, 30602, USA

^d Department of Physics and Astronomy, The University of Georgia, Athens, GA, 30602, USA



ARTICLE INFO

Keywords:

SERS
Silver nanorods
Virus detection
SARS-CoV-2
Support vector machine
Machine learning

ABSTRACT

Rapid and sensitive pathogen detection is important for prevention and control of disease. Here, we report a label-free diagnostic platform that combines surface-enhanced Raman scattering (SERS) and machine learning for the rapid and accurate detection of thirteen respiratory virus species including SARS-CoV-2, common human coronaviruses, influenza viruses, and others. Virus detection and measurement have been performed using highly sensitive SiO₂ coated silver nanorod array substrates, allowing for detection and identification of their characteristic SERS peaks. Using appropriate spectral processing procedures and machine learning algorithms (MLAs) including support vector machine (SVM), k-nearest neighbor, and random forest, the virus species as well as strains and variants have been differentiated and classified and a differentiation accuracy of >99% has been obtained. Utilizing SVM-based regression, quantitative calibration curves have been constructed to accurately estimate the unknown virus concentrations in buffer and saliva. This study shows that using a combination of SERS, MLA, and regression, it is possible to classify and quantify the virus in saliva, which could aid medical diagnosis and therapeutic intervention.

1. Introduction

Rapid and sensitive pathogen detection is important for human health care and for the prevention and control of the disease. The current molecular diagnostic tests for viral detection assays can be classified into two categories (Afzal, 2020; Feng et al., 2020), i.e., nucleic acid tests (Bienko et al., 2013; Cherkaoui et al., 2021; Jones et al., 2021) and serological or immunological tests (Long et al., 2020). For example, the identification of SARS-CoV-2 typically involves viral RNA-based reverse transcription real-time polymerase chain reaction (RT-PCR) and nucleic acid hybridization strategies (Nolan et al., 2006; Udugama et al., 2020). RT-PCR is the “gold standard” with high selectivity and sensitivity but is laboratory-based (Alafeef et al., 2020), and requires viral RNA extraction and expertise in PCR. Similarly, immunological tests requiring serology may take days to weeks after the onset of symptoms for a patient to develop a detectable antibody level (Ravi et al., 2020). In addition, antibody-based detection lacks the sensitivity of RT-PCR for

virus detection. Thus, there is an unmet need for a rapid, reproducible, and sensitive method for direct virus detection, especially respiratory viruses in saliva.

Raman spectroscopy analyzes the molecular patterns using the scattering spectra associated with vibrational and rotational modes of chemical structures (Smith and Dent, 2019). Surface enhanced Raman spectroscopy (SERS) has the potential to achieve single-molecule detection (Kneipp et al., 1997; Nie and Emory, 1997), and is promising and attractive for multiplex virus detection (Shanmukh et al., 2006). To date, there are two different SERS measurement strategies for direct sensing and differentiation of viruses, i.e., labeled and label-free detections. For labeled detection, virus detection has been achieved using SERS tags immobilized to capture viruses, which provides good specificity and high sensitivity (Cha et al., 2022; Li et al., 2021; Sun et al., 2017), but the fabrication of SERS tag is time-consuming and expensive. Without using SERS tags, virus specimens can be detected based on spectral differences after capturing the virus using the

* Corresponding author.

** Corresponding author.

E-mail addresses: YanjunYang@uga.edu (Y. Yang), zhaoy@uga.edu (Y. Zhao).

functionalized SERS substrates with capture molecules such as aptamer (Torun et al., 2021) and proteins (Yang et al., 2021). However, these functionalized molecules are costly, and have other issues such as degradation and storage problems specific to the environment or treatment. In fact, individual analyte can also provide a unique SERS signature or spectral profile with very narrow spectral peaks, which becomes the basis for label-free virus detection. The direct SERS-based measurements have been shown to detect various biomolecules (Alvarez-Puebla and Liz-Marzán, 2010; Cao et al., 2013; Garcia-Rico et al., 2018; Kurochkin et al., 2021), including viruses and viral strains (Shanmukh et al., 2006). However, as the variations in molecule units of viruses are small, the differences in their corresponding SERS spectra could be very small, which presents a great challenge to classify or differentiate virus species or strains based on SERS spectra. Since the SERS spectra can be viewed as multi-variant data, different statistical methods can be adapted for virus classification, including supervised and unsupervised learning methods (Adir et al., 2020; He et al., 2021; Lussier et al., 2020; Ralbovsky and Lednev, 2020). Unsupervised learning methods deal with and learn patterns and structures in the data with no outputs. Two types of methods that are commonly applied in the SERS spectral analysis are dimensionality reduction, such as principal component analysis (PCA), and cluster analysis, such as hierarchical cluster analysis (HCA). PCA can help with data visualization, for example, in analyzing the intrinsic SERS spectra from respiratory syncytial virus (RSV) strains A/Long, B1, and A2, plots from PCA showed separation of the three different virus strains (Shanmukh et al., 2008). It can also help with data preparation for supervised learning. A label-free platform for virus capture and identification has been developed and utilized PCA and supervised learning to classify the viral SERS spectra of rhinovirus, influenza, and parainfluenza (Yeh et al., 2020). Cluster analysis can group similar spectra while their labels are unknown. For example, HCA was able to distinguish exquisite SERS spectral difference for RSV strains based on a single gene deletion (e.g. G gene mutant virus) (Shanmukh et al., 2008). Supervised learning methods deal with data with outputs. Depending on the two types of the output, label or continuous value, supervised learning methods can be grouped to methods for classification and regression. Supervised learning methods for classification, such as partial least square discriminant analysis (PLS-DA), partial least square regression (PLS), linear discriminant analysis (LDA), support vector machine (SVM), k-nearest neighbor (KNN), random forest (RF), and convolutional neural network (CNN), have been used extensively for SERS spectral differentiation, especially to distinguish SERS spectra of analytes in complex biological matrices (Cheng et al., 2021; Ding et al., 2021; Erzina et al., 2020). For example, spectroscopic classifications of SARS-CoV-2, influenza, Zika, and Marburg viruses based on RF have been realized on nanoimprinted SERS substrates (Paria et al., 2022). Methods for regression, such as SVM regression (SVR), have been used to predict the concentration of samples (Leong et al., 2021).

Up to date, there are only few reports on SERS and machine learning algorithm (MLA) based studies on virus detection (Paria et al., 2022; Shanmukh et al., 2008; Shao et al., 2014; Yeh et al., 2020; Zhang et al., 2022). In most of these studies, only 3–4 virus species have been investigated and analyzed. In addition, most of the SERS and MLA studies only focused on spectral differentiation. Although this is important, it is also relevant to quantitatively determine the viral concentration in biological fluids, allowing one to assess the level of infection to aid effective therapeutic treatment (Darwich et al., 2021; Mitchell et al., 2021). Despite this, no report has shown simultaneous species discrimination and quantification, particularly for respiratory virus species.

To develop robust device for achieving reliable and efficient detection of respiratory virus species for infection screening, the ideal SERS sensing platform should have the following characteristics: (1) accurate virus classification and quantification, including virus strains and variants; (2) rapid detection with the detection time significantly shorter

than the PCR assays; (3) cost-effective sensor with label-free detection; (4) point-of-care testing with good convenience and practicality. Herein, a large-area and label-free testing platform has been developed combining SERS and machine learning for the rapid and accurate detection of respiratory viruses. SERS spectra from thirteen respiratory virus species have been collected using highly reproducible and sensitive silver nanorod array SERS substrates and a portable Raman system, allowing for identification of their characteristic SERS peaks and construction of virus SERS spectra database. With an appropriate spectral pre-processing procedure, different classical MLAs, including support vector machine, k-nearest neighbor, random forest, etc., can be applied to differentiate and classify the virus species based on SERS spectra, including SARS-CoV-2 variant differentiation, and achieve a differentiation accuracy of >99%. Utilizing SVM-regression, calibration curves can be constructed to accurately determine unknown viral concentrations in buffer and in saliva. The detection and quantification of these common respiratory viruses, including the one that caused SARS-CoV-2, demonstrate the power of the SERS and MLA approach, which has the great potential for rapid virus detection and potentially point-of-care diagnostic platforms.

2. Experimental section

2.1. Materials

Silver (Kurt J. Lesker, 99.99%) and titanium pellets (Kurt J. Lesker, 99.995%) were purchased as the evaporation materials. Tetraethylorthosilicate (TEOS; Alfa Aesar, 99.9%), ammonium hydroxide (J. T. Baker, 28.0–30.0 wt%) and ethanol (EtOH; Sigma-Aldrich, 95%) were used for silica shell growth. Dulbecco's Modified Eagles Medium (DMEM; GIBCO BRL laboratories, Grand Island, NY) supplemented with 1% fetal bovine serum (FBS; Hyclone Laboratories, Salt Lake City, UT) was used as cell culture media. Pure water (Sigma-Aldrich) was used throughout all the experiments. All the reagents were used without further purification.

2.2. Silver nanorod (AgNR) array fabrication

AgNR arrays were prepared by the oblique angle deposition (OAD) as previously described (Driskell et al., 2008; Liu et al. 2009, 2010; Liu and Zhao, 2008), which is illustrated in Fig. S1. Briefly, clean glass slides (0.5 inch × 0.5 inch) were loaded into a vacuum deposition chamber with the substrate normal antiparallel to the incident vapor direction. First, two layers of 20 nm Ti and 200 nm Ag films were deposited in sequence at a rate of 0.2 nm/s and 0.3 nm/s, respectively. Then, the substrate normal was rotated to 86° relative to the incident vapor direction, and a thickness of 2000 nm Ag film was then deposited at a rate of 0.3 nm/s to fabricate the arrayed AgNRs. The entire evaporation process was conducted under a high vacuum condition (<3 × 10⁻⁶ Torr). This SERS substrate can produce very high SERS enhancement with spectra variation less than 10% (Abell et al., 2009; Singh et al., 2012).

2.3. AgNR@SiO₂ core-shell arrays fabrication

To prevent the chemical degradation of AgNR in buffer solution, a very thin layer (~2 nm) of dense silica shell was uniform coated on the AgNRs. The silica shell growth conditions with desired thickness were previously reported (Song et al., 2012). In this study, AgNR arrays were immersed into a homogeneous mixture composed of 30 mL of EtOH, 4 mL of H₂O, and 500 μL of TEOS for 20 min under stirring. Once 560 μL of ammonium hydroxide was added into the mixture, the reaction was initiated. The shell thickness was controlled by the hydrolysis time of TEOS after the accession of alkaline. The AgNR arrays were removed from the reaction solution after 5 min, followed by water rinsing and N₂ drying. Subsequently, a PDMS layer with arrayed small wells (4 wells, with a well diameter and depth of 4 mm and 1 mm, respectively) was

molded on the AgNR@SiO₂ array to restrict the effective sensing areas (Abell et al., 2009), referred to as AgNR@SiO₂ wells. According to our previous study (Song et al., 2012), the thickness of SiO₂ was 1.9 ± 0.4 nm. Scanning electron microscopy (SEM) image and energy dispersive spectrometry (EDS) analysis of AgNR@SiO₂ array were obtained by using a field emission scanning electron microscopy (S-4800, Hitachi, Japan) equipped with an Oxford INCA-350 energy dispersive X-ray microanalysis system operating at 8 kV. A typical SEM image of AgNR@SiO₂ array is shown in Fig. S2 in Supplementary Materials (SM). The surface compositions of AgNR@SiO₂ arrays were analyzed using an X-ray photoelectron spectroscopy (XPS) (K-Alpha, Thermo Scientific, USA). Infrared spectra were obtained on a Fourier-transform infrared spectroscopy (FTIR) (NICOLET 6700, Thermo Scientific, USA) equipped with Harrick Scientific's VariGATR™ grazing angle accessory.

2.4. Virus incubation

The following viruses were used in the analyses: SARS-CoV-2 (WA1/2020), SARS-CoV-2 B1.1.7 variant (SARS-CoV-2 B1), human coronavirus NL63 (CoV NL63), human coronavirus 229E (CoV 229E), human coronavirus OC43 (CoV OC43); influenza A H1N1 Brisbane (H1N1, IAV), influenza A H3N2 Hong Kong (H3N2, IAV), and influenza B (IBV); respiratory syncytial virus (RSV) from strain A2 (RSV-A2) and B1 (RSV-B1); human metapneumovirus (HMPV) from strain A (HMPV-A) and B (HMPV-B), as well as adenovirus type 5 (Ad5). These viruses are common respiratory viruses except for SARS-CoV-2 that is responsible for the COVID-19 pandemic. IAV, IBV, RSV-A2, RSV-B1, HMPV-A, and HMPV-B co-circulate in humans and may cause serious respiratory diseases (Hendaus and Jomha, 2021; Mahalingam et al., 2006; Piret and Boivin, 2022).

All viruses were propagated in Vero E6 cells which were maintained in DMEM supplemented with 10% heat-inactivated (56 °C) FBS. Briefly, cells were infected using a multiplicity of infection (MOI) = 0.1. After 48 h, the viruses were harvested in serum-free DMEM followed by freeze-thaw after which the contents were collected and centrifuged at 4000 g for 15 min at 4 °C. The virus titers were similar, i.e., 10⁵ PFU/mL, as determined by plaque assay as previously described (Boyoglu-Barnum et al., 2017; Murray et al., 2021; Tripp Ralph et al., 1999). The reference specimen for these studies was diluted in DMEM supplemented with 1% FBS. Influenza strains, H1N1 and H3N2, were propagated in embryonated chicken eggs and virus titers determined by hemagglutination assay using chicken red blood cells. The influenza virus titers ranged between 10⁷-10⁸ 50% egg infectious dose (EID₅₀). The reference specimen for these studies was naive allantoic fluid. All the experiments on SARS-CoV-2 and SARS-CoV-2 variant were conducted in a biosafety level 3 (BSL-3) lab, while others were performed in a BSL-2 lab. All the experimental operations should follow the biosafety guidelines: <https://www.cdc.gov/coronavirus/2019-nCoV/lab/lab-biosafety-guidelines.html>.

2.5. Specimen preparation and SERS characterization

To obtain the SERS signal of the virus specimens at different concentrations, the virus was diluted with pure water. 5 μL of the diluted virus specimen was dispensed onto the AgNR@SiO₂ wells and air-dried at 20 °C. Virus-inoculated saliva specimens were prepared by adding known (predetermined) concentrations (PFU) of virus specimens to achieve final concentrations ranging from 195 to 10⁵ PFU/mL for SERS measurement. Similarly, 5 μL of the virus saliva specimens were dispensed onto the AgNR@SiO₂ wells and air-dried at 20 °C. The SERS spectra were collected from multiple randomly selected locations by using a Tec5USA Raman spectrometer (Tec5USA Inc.), with a 785 nm excitation wavelength and a ~100 μm (diameter) laser spot. The laser power was 32 mW and the acquisition time was 2 s. The numbers of SERS spectra collected from viruses of different concentrations were summarized in Tables S1 and S2. The total collected SERS spectra were

16,214 from water diluted specimens and 15,600 from saliva specimens.

2.6. Data pre-processing

Based on the overall spectral features of SERS spectra obtained, a simple baseline correction method called “Gaussian-Lorentzian function fitting (GLFF)” was developed to obtain more uniform SERS spectra for MLAs (Yang et al., 2022). A typical raw SERS spectrum shown in Fig. S7 has featureless SERS spectral response in the wavenumber range of 320–400 cm⁻¹ and 1800–2500 cm⁻¹. Therefore, the baseline for each spectrum was fitted by the following function in these two wavenumber regions, i.e., 320–400 cm⁻¹ and 1800–2500 cm⁻¹,

$$I_{SERS}(\Delta\nu) = Ae^{-\frac{(\Delta\nu-\nu_g)^2}{2\sigma_g^2}} + \frac{2L\sigma_l}{4\pi(\Delta\nu-\nu_l)^2 + \sigma_l^2} + I_0, \quad (1)$$

where A is the amplitude of the Gaussian function, ν_g is the center of the Gaussian peak, σ_g is the standard deviation of the Gaussian function, L is the area of the Lorentzian function, ν_l is the center of the Lorentzian peak, σ_l relates to the width of the Lorentzian peak, I_0 is the “ground” level of the SERS spectrum. The red curve in Fig. S7 shows an example of the fitted baseline for the spectrum. Such a process guarantees a minimum disturbance for the raw data and avoids non-necessary information loss due to spectra pre-processing. The whole data pre-processing is illustrated in Fig. S8 for SERS spectra of CoV 229E at 25,000 PFU/mL. The raw spectra (Fig. S8A) were first normalized by the SERS intensity at $\Delta\nu = 320$ cm⁻¹ (Fig. S8B). Such a process was used to confine the range of fitting parameters in Eq. (1). After baseline fitting by Eq. (1), the corresponding baseline corrected spectra (i.e., original spectra subtracting the baseline) are shown in Fig. S8C. Then the mean value of each baseline corrected spectrum was calculated, and the final spectrum was normalized by the mean value of each spectrum (Fig. S8D).

2.7. Machine learning model

In this study, we have addressed the following questions: (1) Could different virus species be accurately classified regardless of virus concentrations? (2) Could the virus concentration be accurately determined based on SERS spectra in addition to classification? To answer the first question, four best performed traditional MLAs, SVM, RF, KNN and LDA (James et al., 2013), are employed to train models to predict virus species based on SERS spectra. SVM is an algorithm that separates the data points with p dimension by finding a (p -1)-dimensional hyperplane which maximizes the distance from the hyperplane to the nearest data point on each side (biggest margin). Suppose the training data consists of n pairs of (x_i, y_i) , $i = 1, \dots, n$ with $x_i \in \mathbb{R}^p$ and $y_i \in \{-1, 1\}$. Let $h(x)$ be a transformed feature vector and define a hyperplane by $\{x : \omega^T h(x) + b = 0\}$, the optimization problem of finding the hyperplane can be written as (Hastie et al., 2009):

$$\min_{\omega, b} \frac{1}{2} \|\omega\|^2 + E \sum_{i=1}^n \xi_i \quad (2)$$

$$s.t. \xi_i \geq 0, y_i [\omega^T h(x_i) + b] \geq 1 - \xi_i, \forall i$$

where ξ_i are slack variables indicating the distances of points that are on the wrong side of their margin to their correct margin boundary and E is the penalty that controls the slack term. The optimization problem with the Lagrangian dual objective can be further obtained as:

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \langle h(x_i), h(x_j) \rangle - \sum_{i=1}^n \alpha_i \\ s.t. \quad & 0 \leq \alpha_i \leq E, \sum_{i=1}^n \alpha_i y_i = 0, \forall i \end{aligned} \quad (3)$$

where $\langle h(x_i), h(x_j) \rangle$ is the inner product of $h(x_i)$ and $h(x_j)$. To simplify

the calculation of the inner product of transformation function $h(x)$, kernel function is introduced as $K(x_i, x_j) = \langle h(x_i), h(x_j) \rangle$. There are three popular choices of kernel functions in SVM: linear, polynomial, and radial basis. Once the optimization is solved, the function for the hyperplane can be determined as:

$$f(x) = \omega^T h(x) + b = \sum_{i=1}^n \alpha_i y_i K(x_i, x) + b \quad (4)$$

RF method is based on decision tree algorithm (Ali et al., 2012). It draws multiple bootstrap samples from the training set, trains tree classifier for each bootstrap sample, and makes a prediction for the new data with majority vote from all trained classifiers. KNN is a nonparametric algorithm, which assigns to a new data with the majority labels of its nearest neighbors. In LDA, the density of the data x given every class $G = k$ is assumed to follow a Gaussian distribution. LDA finds class k that maximize the posterior probability of class giving x . In particular,

$$\max_k Pr(G = k | X = x) \quad (5)$$

By applying Bayes rule, plugging in the density of Gaussian distribution, and simplifying the formula, the problem becomes

$$\max_k [x^T \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma^{-1} \mu_k + \log(\pi_k)] \quad (6)$$

where Σ is the covariance matrix of all classes, μ_k is the mean vector for class k , and π_k is the prior probability of class k .

When training the MLA models, a 10-fold cross validation was employed to tune the hyperparameters whose values control the learning process of the model, such as regularization parameter E and kernel coefficient γ in SVM. At the end of cross-validation, the best hyperparameters were chosen and the unbiased model performance was obtained using the following measures: accuracy, micro and macro precision, micro and macro recall, and micro and macro F1-score on validation sets (see Fig. S18 and Table S13 for definitions). Feature importance was obtained by applying the models with best hyperparameters on the whole training spectral set. The confusion matrix and receiver operating characteristic (ROC) curve were also obtained by implementing the models on the testing spectra set. The ROC curve illustrates the diagnostic ability of a binary classifier system, which was created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings.

To quantify the virus concentration, a two-step sequential approach is used: 1) to classify the species of the virus by a MLA, and 2) to conduct regression within the species identified in step 1 to estimate the concentration. Support vector regression (SVR) is used to predict the virus concentration. In SVR, a linear function $f(x) = \omega^T h(x) + b$ is to be determined that minimizes $\|\omega\|^2$ subject to all residuals less than a tolerance ϵ : $|y_i - [\omega^T h(x_i) + b]| \leq \epsilon$, where y_i is the real virus concentration, and $f(x)$ is the predicted virus concentration. Their difference are residuals in the constraints in the optimization problem. Like in SVM, by adding slack variables in the objective function, the optimization problem becomes,

$$\begin{aligned} & \min_{\omega, b} \frac{1}{2} \|\omega\|^2 + E \sum_{i=1}^n (\xi_i + \xi_i^*) \\ & \text{s.t. } y_i - [\omega^T h(x_i) + b] \leq \epsilon + \xi_i, \\ & [\omega^T h(x_i) + b] - y_i \leq \epsilon + \xi_i^*, \\ & \xi_i, \xi_i^* \geq 0, \forall i \end{aligned} \quad (7)$$

where ξ_i, ξ_i^* are slack variables depending on whether the prediction is above or below the ϵ tube. Following similar process in SVM, the solution for SVR can be obtained. SERS spectra were randomly chosen as the training spectral set and testing spectral set with a ratio of 7:3 in a stratified way. The hyperparameters for SVR, such as kernel,

regularization parameter E , and tolerance parameter ϵ , were tuned by minimizing the mean absolute errors (MAE). All MLAs are trained with library *scikit-learn* in Python 3.8.3.

3. Results and discussion

3.1. Detection and classification strategy

The method for using SERS and MLA to accurately differentiate and classify the different virus specimens is illustrated in Fig. 1. Initially, a SERS spectral database of virus specimens is constructed by collecting spectra from AgNR@SiO₂ SERS substrates. Then, based on the spectral feature, a simple baseline correction method is used to obtain highly reproducible SERS spectra for MLA. Finally, by applying a MLA, in particular the support vector machine (SVM) method, the virus species associated with corresponding SERS spectra are discriminated with a high accuracy, and a regression model (SVR) is used to quantify the viral concentration.

3.2. SERS substrate description

Using a highly sensitive, reproducible, and uniform SERS substrate for SERS measurement, it is essential to obtain high signal-to-noise ratios (SNRs) spectra with good reproducibility, which is required to achieve high identification accuracies (Ho et al., 2019). The AgNR array has been shown to possess good SERS reproducibility (~10% relative variation), high SERS enhancement factor (up to 10⁹), and large uniformity (Driskell et al., 2008; Liu et al. 2009, 2010; Liu and Zhao, 2008). To overcome issues of surface contamination, stability, and biocompatibility, the AgNR arrays were coated with a uniform and thin silica layer by the hydrolysis of tetraethylorthosilicate to form AgNR@SiO₂ core-shell arrays which can serve as an ideal SERS substrate for direct virus detection. Fig. S2 shows a representative SEM image of AgNR@SiO₂ array. The surface is composed of tilted nanorods with a wide range of morphologies, such as corrugations, bifurcations, and protrusions. The nanorod density is 12 ± 1 rods/μm². The average diameter of the AgNR@SiO₂ is 100 nm and the length is approximated as ~1000 nm with a tilted angle 77 ± 1°. To analyze the components in the surface coating of SiO₂ on the AgNR substrate, XPS, EDS and FTIR were employed. The atomic percentages of Si and O increase with coating time while the atomic percentage of Ag decreases, as detailed in Section S1 in SM, which confirms the success of SiO₂ coating.

3.3. Peak assignments for virus SERS spectra

All viruses except for Ad5, are enveloped RNA viruses. Ad5 is a nonenveloped, double-stranded DNA virus. Fig. 2 shows the representative average SERS spectra of 13 virus specimens at 50,000 PFU/mL and 3 reference specimens. The SERS peaks of all viral spectra have been carefully identified and assigned, and are shown and discussed in detail in Sections S4 - S10 in SM. The SERS spectra of different respiratory viruses share some common spectral features but also are distinguishable based on some characteristic peaks. For example, as shown in Section S4, the most prominent spectral peaks of SARS-CoV-2 which has led to COVID-19 pandemic are at Δv = 656 cm⁻¹, 726 cm⁻¹, and 1330 cm⁻¹, respectively, which are corresponding to nucleic acid bases vibrational modes of guanine, the adenine (ring), and adenine. The SERS peaks at Δv = 1003 cm⁻¹ and 1031 cm⁻¹ arise from phenylalanine. The SERS peaks between Δv = 1580 cm⁻¹ and 1700 cm⁻¹ can be attributed to vibrations of carbonyl groups on the amino acid side chains and the Amide I vibration. Interestingly notice that the SERS spectra of SARS-CoV-2 and B1 variant show a slight difference at Δv = 1660 cm⁻¹ corresponding to the Amide I vibration. Such a difference provides a potential of variant differentiation capability. However, as shown in Section S5, common coronaviruses exhibit distinctive peaks at Δv = 658 cm⁻¹, 725 cm⁻¹, 1004 cm⁻¹, 1034 cm⁻¹, 1329 cm⁻¹, 1450 cm⁻¹,

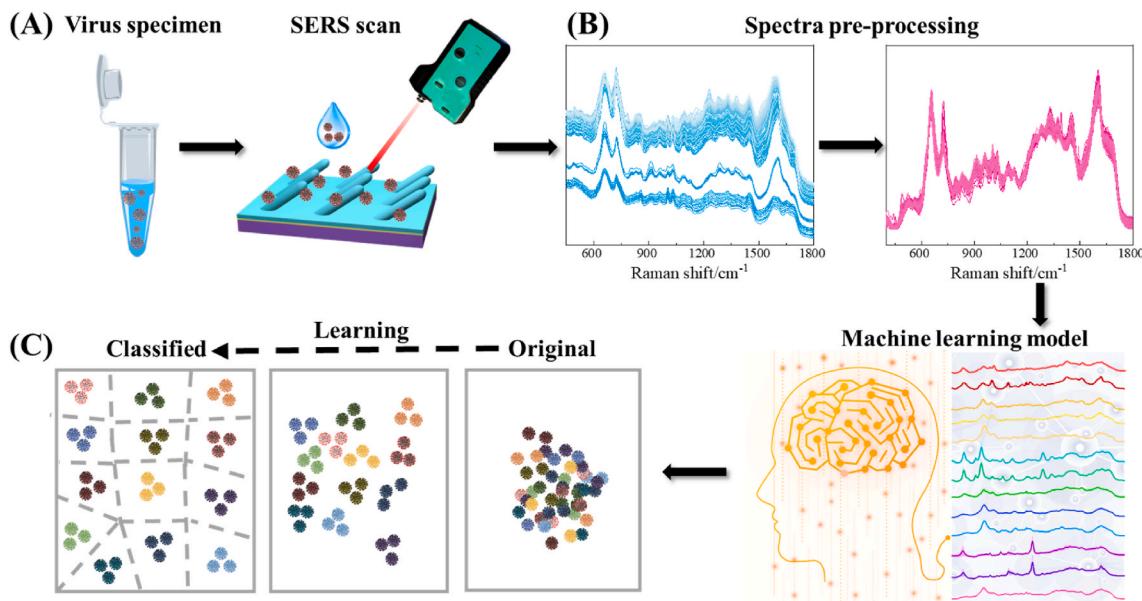


Fig. 1. Schematic illustration of machine learning-based virus differentiation: (A) specimen preparation and SERS measurements to obtain SERS spectra; (B) spectra pre-processing, and (C) classification using machine learning models.

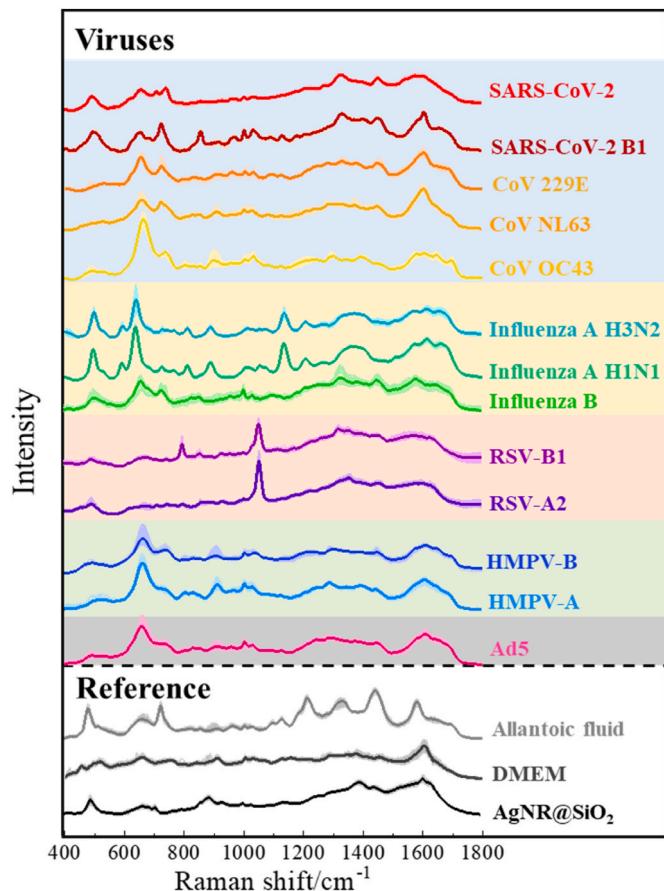


Fig. 2. SERS spectra of specimens from 13 viruses with 50,000 PFU/mL and three reference specimens. The average SERS spectra are shown by the solid curves, and all the individual SERS spectra are shown by the shadow.

and 1603 cm⁻¹ and have different relative peak intensities. In comparison, SARS-CoV-2 and B1 variant have obvious peaks at $\Delta\nu = 858$ cm⁻¹ corresponding to Tyrosine (Tyr, β -sheet).

The overall SERS peak assignments of viruses and comparison are summarized in Table S10. The number of virus specimens with shared common peaks vary from 1 to 7. The most two common peaks are the peaks at $\Delta\nu = 1003$ cm⁻¹ and $\Delta\nu = 1576$ cm⁻¹ corresponding to the symmetric ring breathing mode of Phe and the carboxylate stretching vibration ($\nu_{\text{a}}\text{COO}^-$) of Trp. Also, there are some unique peaks only belonging to specific virus. The similarity and difference in SERS peaks are due to the intrinsic molecular structure similarity and difference among these viruses.

It is important to note that the material use for virus propagation will also contribute to the SERS spectra. For example, allantoic fluid has prominent SERS peaks at $\Delta\nu = 659, 723, 1215, 1330, 1443$, and 1582 cm⁻¹, mainly from amino acids. DMEM SERS spectra contain multiple small peaks, mainly from amino acids, vitamins, and other components. All virus spectra contain a peak at around $\Delta\nu = 492$ cm⁻¹, which may be from the SERS substrate.

3.4. Virus classification using MLAs

Statistic characterizations of pre-processed viral SERS spectra are detailed in Section S12. To determine a beneficial MLA, all the SERS spectra were distributed into two parts, i.e., 90% of the spectra of different viruses at different concentrations were used for modeling spectral set, while 10% of the spectra were used for testing (Fig. S17). To classify different viral species, all the spectra within the same species but with different concentrations were grouped into same label. In the training process, a 10-fold cross-validation was employed to determine the best hyperparameters, such as, E (regularization) and γ (kernel coefficient) for SVM. In particular, for SVM, three commonly used kernels, linear, polynomial, and radial basis function (RBF), were used, and the kernel coefficient γ is a constant for the linear kernel. The cross-validation performance for different MLAs is summarized in Table S12. It is shown that almost all the MLAs give relatively good accuracy >99% in predicting the virus species. The lowest accuracy is $99.2 \pm 0.3\%$ for LDA, and the highest is $99.91 \pm 0.07\%$ for SVM with the polynomial kernel. Such high accuracies are due to the highly reproduced spectra collected from the AgNR@SiO₂ substrates. There are still some small differences in the performance measures from the cross-validation. In particular, the SVM generally gives slightly higher accuracy than other MLAs, as it introduces kernels that could make the model

more flexible and capable of dealing with complex and high dimensional data set. In addition to the high accuracy, SVM with polynomial kernel also shows smaller variation in all measures which could be an implication for its stability. Considering the model efficiency and the risk of overfitting for SVM, linear kernel is faster due to computation simplicity, and is proved to be more flexible and stable (Ben-Hur et al., 2008). Therefore, for later spectra testing and classification, we used SVM with a linear kernel and the optimized regularization parameter $E = 1$. After the training, the remain 10% spectra were tested by the trained SVM classifier. The classification results on the testing set are summarized as a confusion matrix shown in Fig. 3. Among the 13 viruses and 3 reference specimens, all labels are classified with 100% accuracy. The corresponding ROC curve is shown in Fig. S19. For all 16 specimens (13 viruses and 3 references), the mean value of the areas under the ROC curves (AUC) is greater than 99%. The important structures/peaks on the SERS spectra that contribute most to a classifier were also investigated by applying the MLA to the training spectral set and obtaining the feature importance of Raman shift. The most prominent features are overlapped with either the unique peak(s) from certain virus or the common peaks as shown in Fig. S20 and in highlight in Table S10. These results suggest that the SVM not only classifies different viruses and reference specimens with high specificity and sensitivity based on the SERS spectra, but also separates different viral strains or variants, especially SARS-CoV-2 and SARS-CoV-2 B1.

3.5. Classification and quantification of viral infections using SVM and SVR

To quantify the virus concentrations using classification strategy, we created “virus + concentration” labels, and applied MLA to the spectral

data with new labels. As shown in Table S1, by counting both the virus species and concentration, a total of 59 labels were constructed. In each label spectral set, a ratio of 9:1 of SERS spectra were randomly chosen as the training spectral set and testing spectral set and analyzed by SVM. The confusion matrix is shown in Table S14, and the classification accuracy of the test spectra is 99.8%. Among the 59 labels, 57 labels are classified with 100% accuracy, 5.71% spectra from CoV229E-25000 label were misclassified as CoV229E-50000 label, 6.06% spectra from CoV229E-6250 label were misclassified as CoV229E-12500 label. Applying classification method for quantification with multiple labels, we do not need to build a regression model to solve the quantification problem, which is simple and convenient. However, this method may not cover all viral concentrations, and the accuracy may depend on how the concentration grid is selected, coarse or fine. If one can pre-collect training spectral set with a very fine concentration-grid (depending on the lowest and highest concentrations of the viruses found in clinic specimens) and use them as modeling set, practically this model should work well in determining both the virus species and concentration. Otherwise, such a method is impractical.

A more credible method to quantify viral concentration is to use a regression model and calibration curve. Unlike the direct classification method discussed above, the regression method is a two-step method: First, the SVM will be used to identify the viral species based on the SERS spectra. Once identified, a regression, i.e., SVR, is performed under the known training spectral set for that identified species. Here we use 6 viruses with more concentration dependent spectra as examples to demonstrate the feasibility of SVR. Concentration dependent SERS spectra of 6 viruses in buffer were shown in Fig. 4 and Fig. S21. For example, for CoV 229E from 100,000 PFU/mL to 195 PFU/mL (Fig. 4A), the obvious virus SERS peaks are shown in high concentration SERS

Actual	Predicted															
	SARS-CoV-2	SARS-CoV-2 B1	CoV 229E	CoV NL63	CoV OC43	IBV	H1N1	H3N2	HMPV-A	HMPV-B	RSV-A2	RSV-B1	Ad5	DMEM	AF	AgNR@SiO ₂
SARS-CoV-2	100															
SARS-CoV-2 B1		100														
CoV 229E			100													
CoV NL63				100												
CoV OC43					100											
IBV						100										
H1N1							100									
H3N2								100								
HMPV-A									100							
HMPV-B										100						
RSV-A2											100					
RSV-B1												100				
Ad5													100			
DMEM														100		
AF															100	
AgNR@SiO ₂																100

Fig. 3. Confusion matrix of the SVM model for 13 virus specimens and 3 reference specimens. Entries in the matrix represents the percentage of test spectra that were predicted by the SVM model as class (first row) given a ground truth of class (first column); entries along the diagonal represent the accuracies for each class.

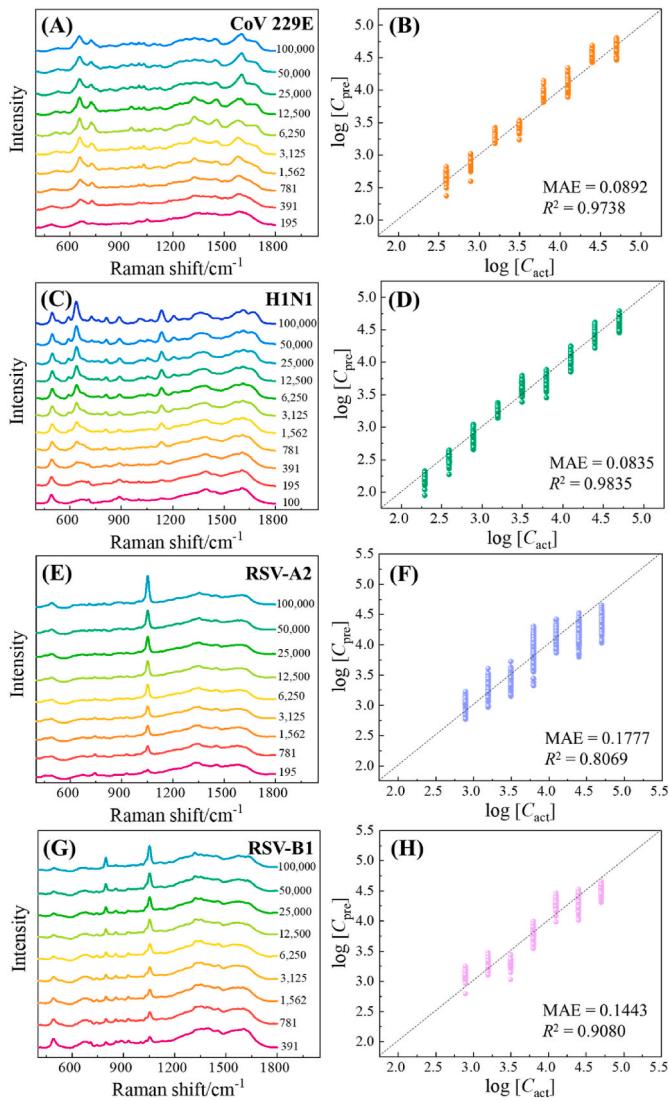


Fig. 4. Average SERS spectra of 4 viruses in buffer with different concentrations for (A) CoV 229E, (C) H1N1, (E) RSV-A2, and (G) RSV-B1. The concentrations are labeled beside the SERS spectral curves with the unit of PFU/mL. Regression results of SVR model for virus detection with unknown concentration in buffer for (B) CoV 229E, (D) H1N1, (F) RSV-A2, and (H) RSV-B1. X-axis shows $\log_{10}C_{\text{act}}$ of testing spectra, and y-axis is $\log_{10}C_{\text{pre}}$. The dash line shows $\log_{10}C_{\text{act}} = \log_{10}C_{\text{pre}}$.

spectra. When decreasing the virus concentrations, the characteristic SERS peaks of CoV 229E gradually disappear. Similar observations are obtained for Fig. 4C, 4E, 4G, S21A, and S21B. All these spectra were used for classification as shown in Section 3.4 and achieved an accuracy of 100% of species classification. For the second step, SVR is used to obtain the concentration calibration curves based on logarithms ($\log_{10}C$) of concentrations. Regression results of the predicted concentration C_{pre} and actual concentration C_{act} from the SVR model are plotted in Fig. S22 in log-log scale, and the associated SVR parameters are summarized in Table S15. The predicted $\log_{10}C_{\text{pre}}$ and the actual $\log_{10}C_{\text{act}}$ for 6 virus species shown in Figs. S22A–S22F all follow the relationship $\log_{10}C_{\text{pre}} = \log_{10}C_{\text{act}}$, with small mean absolute errors (MAEs) and the coefficient of determination $R^2 \sim 1$, which demonstrates the accuracy of the regression. In particular, high quantification accuracies can be observed from CoV 229E, CoV NL63, H1N1 and H3N2 with R^2 of 0.9932, 0.9903, 0.9930 and 0.9823, respectively, while RSV-A2 and RSV-B1 show slightly larger prediction range with a slightly smaller R^2 of 0.9387 and 0.9753, respectively. These results show that the proposed SERS + MLA

method can obtain high quantification performance for virus specimens.

Typically, specimens with unknown viral concentrations occur frequently. Therefore, it is very important to determine whether above two-step strategy can classify and quantity virus specimens with unknown concentrations. In order to evaluate above strategy, the following procedure is carried out: from the 59 sets of SERS spectra with different virus and concentration shown in Table S1, the entire set of the SERS spectra from one virus at one concentration was taken out as the testing spectral set, and the rest 58 sets of SERS spectra were used as the training spectral sets, i.e., the training spectral sets did not contain any spectra from the testing spectral set. This procedure was repeated for all the 37 sets of SERS spectra listed in Table S16, i.e., total 37 tests were conducted. The SERS spectra sets from the highest and lowest concentration of each individual virus were removed for further evaluation since they set the decision boundaries. Thus, only 37 sets of spectra data were used for classification and quantification evaluation. The accuracies in the classification step (i.e., SVM) are listed in Table S16. Accuracies for most labels are 100%, except for CoV 229E (391 PFU/mL) with 97.33%, CoV NL63 (6250 PFU/mL) with 99.06%, and CoV NL63 (25,000 PFU/mL) with 99.34%. Regression results of 6 viruses from SVR model are shown in Fig. 4 and Fig. S23. Compared to the results with known concentration regression, i.e., results shown in Fig. S22, the corresponding MAEs for 6 viruses are around 1.5 times larger, and the corresponding R^2 is ranging from 0.97 to 0.80. This slightly worse performance compared to those in Fig. S22 is understandable since there are no SERS spectra of the corresponding concentrations in modeling spectral set. The results still have very high quantification accuracies, especially more than enough for therapeutic treatment. Thus, the final strategy for virus diagnostics using SERS is a two-step process, classification, and quantification with a regression. In practical application, an unknown SERS spectrum from diagnostic device will be tested by the SVM model to identify the infected virus, then use this SVR model to predict its concentrations.

3.6. Classification and quantification of virus in saliva

To demonstrate the practicability of the SERS + MLA method for virus detection, inoculated saliva-virus mixtures were used as a proof-of-concept. Traditionally, respiratory specimens such as nasopharyngeal swabs are used to virus diagnostics. These are invasive methods requiring operators to have close contact with the patient during specimen collection, which increases the risk of virus transmission. In contrast, saliva specimens can be self-collected by the patients to use for virus diagnostics – a feature that has attracted substantial attention (Pan et al., 2020; Wyllie et al., 2020). Further, saliva can be stored in the ambient for 48 h and has been demonstrated to be sufficiently sensitive for diagnosis of mild and asymptomatic COVID-19 cases (Teo et al., 2021). Therefore, we decide to test the virus diagnostics performance using saliva specimens. Virus-inoculated saliva specimens at different concentrations were prepared, and their SERS spectra were obtained as shown in Fig. 5A, 5C, 5E, 5G, and Fig. S24. The corresponding number of SERS spectra are listed in Table S2. The characteristic virus SERS peaks are shown in high concentration SERS spectra. For example, SERS peaks at $\Delta\nu = 725, 1331$, and 1607 cm^{-1} can be observed in the spectra of CoV 229E at 100,000 PFU/mL as shown in Fig. 5A. These peaks are consistent with peaks assigned in Table S4. Similar observations are obtained for other ten viruses in inoculated saliva specimens as shown in Fig. 5C, 5E, 5G, and Fig. S24. When decreasing the virus concentrations, the SERS spectra gradually evolve to the spectra similar to the saliva spectra with the distinctive peak at 1004 cm^{-1} . These SERS spectra were used to classify the inoculated saliva specimens. According to Section 3.5, SERS spectra were randomly chosen from Table S2 as the training spectral set and testing spectral set with a ratio of 7:3 in a stratified way. The SVM with a linear kernel (with $E = 1$) was also used and the resulting confusion matrix is shown in Fig. S25. The overall classification accuracy is 99.95%. Except for CoV OC43 and H3N2, the classification accuracies for other nine viruses and reference specimens are 100%, while

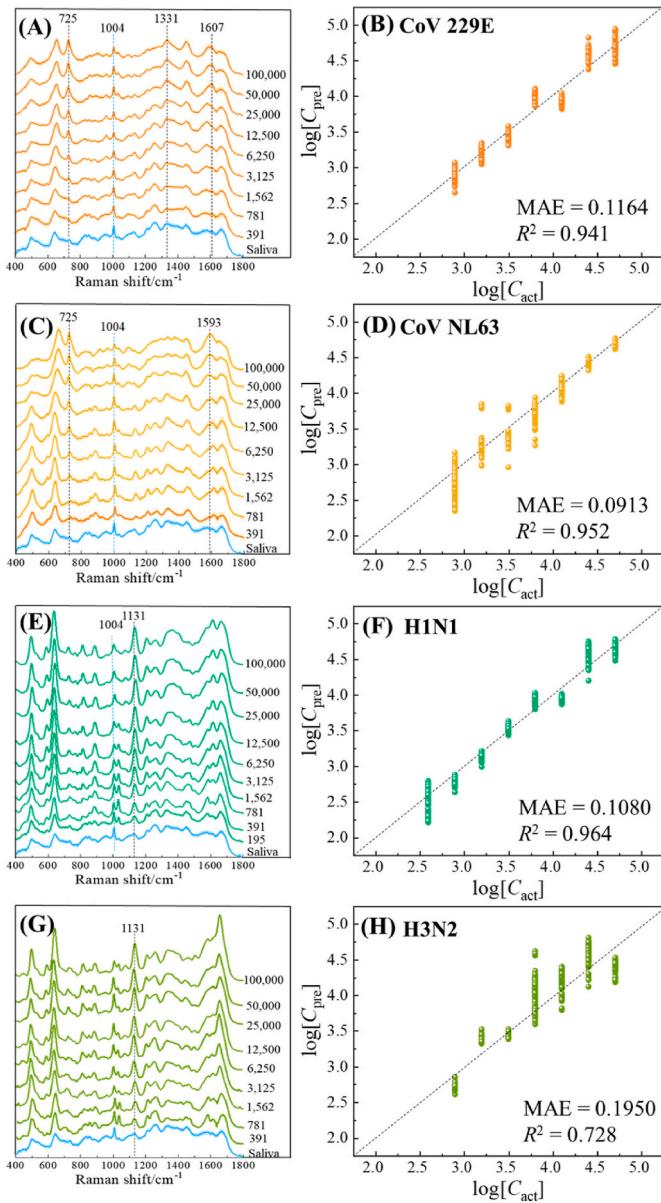


Fig. 5. Average SERS spectra and the standard deviations of 4 viruses in saliva with different concentrations for (A) CoV 229E, (C) CoV NL63, (E) H1N1, and (G) H3N2. The average SERS spectra are shown by the solid curves, and the standard deviations are marked by the shadow. Regression results of SVR model for unknown concentrations of (B) CoV 229E, (D) CoV NL63, (F) H1N1, and (H) H3N2. X-axis shows log₁₀C_{act} of testing spectra, and y-axis is log₁₀C_{pre}. The dash line shows log₁₀C_{act} = log₁₀C_{pre}.

0.3% of CoV OC43 SERS spectra were misclassified as Ad5, and 0.4% of H3N2 SERS spectra were misclassified as H1N1. Subsequently, the SVR quantification was also performed and the regression results of 11 virus-inoculated saliva specimens are shown in Fig. S26 (Due to the limited access to BLS-3 lab, we did not perform concentration dependent measurements on SARS-CoV-2 and SARS-CoV-2 B1 variant). The overall predicated concentration C_{pre} matches well with the actual concentration C_{act}, with the MAEs between 0.0332 and 0.1414, and R² to be 0.996 to 0.933. The high quantification performance clearly demonstrates that the proposed SERS + MLA method is effective for inoculated saliva specimens.

The SVR tests for unknown concentrations of virus in saliva were conducted in a fashion similar to Section 3.5. Classification and quantification performance was evaluated for the 79 sets of SERS spectra

listed in Table S17. The first-step classification gives 100% accuracy for most specimens with different viral concentration as shown in Table S17, except for Ad5 at 1562 PFU/mL, with an accuracy of 99.43%. And the second-step regression results are shown in Fig. 5B, 5D, 5F, 5H, and Fig. S27. All the predicted concentrations C_{pre} align very well with the actual concentration C_{act} with MAEs between 0.0332 and 0.1950, and R² to be 0.994 to 0.728. The MAEs in unknown concentration predictions are generally larger than those MAEs of the known concentration situation, compared to Fig. S26. However, the differences are very small, which indicates that the two-step method to classify and quantify the unknown saliva specimen is very effective. According to the confusion matrix of the SVM model for virus specimens with low concentrations as shown in Section S15 in SM, the limit of detection (LOD) is determined to be 195 PFU/mL for H1N1 and Ad5, and 391 PFU/mL for other nine viruses in saliva, and the detection ranges are listed in Table S19.

4. Conclusions

In summary, we developed a label-free diagnostic platform that combines SERS and machine learning for the rapid and accurate detection of thirteen respiratory virus species including SARS-CoV-2, common human coronaviruses, influenza viruses, and others. The SERS spectra of thirteen virus species and three reference specimens were measured using SiO₂ coated AgNR array substrates, and their characteristic SERS peaks were identified. These spectra share many similar spectral features but also have distinguished peaks for different viruses. The most two common peaks are the peaks at $\Delta\nu = 1003\text{ cm}^{-1}$ and $\Delta\nu = 1576\text{ cm}^{-1}$ corresponding to the symmetric ring breathing mode of Phe and the carboxylate stretching vibration ($\nu_{\text{a}}\text{COO}^-$) of Trp. Different classic MLAs were applied to differentiate and classify the viruses based on SERS spectra and have achieved >99% accuracy with an appropriate spectral pre-processing procedure. In particular, the SVM with a linear kernel can achieve an overall classification accuracy of 99.95%. Except for CoV OC43 and H3N2, the classification accuracies for other nine viruses and reference specimens are 100%, while 0.3% of CoV OC43 SERS spectra were misclassified as Ad5, and 0.4% of H3N2 SERS spectra were misclassified as H1N1. This high classification accuracy includes the identification viral strains or variants from the same virus species, in particularly, the SARS-CoV-2. Furthermore, to quantify the virus concentration, a two-step classification & quantification strategy is proposed using SVM (classification) and SVR (regression). It has been shown that such a strategy can accurately predict the concentration of virus specimens in both buffer and saliva. For our current study, the detection range is from 400 PFU/mL to 10⁵ PFU/mL with LOD to be 195 PFU/mL for H1N1 and Ad5, and 391 PFU/mL for other nine respiratory viruses in saliva. Importantly, once the modeling spectral set is determined and the MLA is finalized, the proposed virus diagnostic strategy is rapid, accurate, and highly quantification. With more spectral data sets from different patients and human fluids, as well as more respiratory viruses, a better and comprehensive modeling spectral set with appropriate MLA optimization, a fast, reliable and point-of-care viral diagnostic platform can be achieved. The sensing platform has the following capabilities: (1) accurate virus classification and quantification, including virus strains and variants; (2) rapid detection within 20 min (Although the detection time is not discussed above, from sample preparation to diagnostic results, the whole process does not exceed 20 min); (3) cost-effective sensor with label-free detection enabled by machine learning methods; (4) point-of-care and on-site mass screening with good convenience and practicality, combining with a portable Raman spectrometer and large-area and multi-well SERS substrate, which can be benefit for medical diagnosis and therapeutic decision, which could aid in medical diagnosis and therapeutic decision.

CRediT authorship contribution statement

Yanjun Yang: Conceptualization, Methodology, Investigation, Writing – original draft. **Beibei Xu:** Methodology, Investigation. **Jackelyn Murray:** Methodology, Investigation. **James Haverstick:** Methodology, Investigation. **Xianyan Chen:** Conceptualization, Methodology, Writing – review & editing, Funding acquisition. **Ralph A. Tripp:** Conceptualization, Methodology, Writing – review & editing, Funding acquisition, Supervision. **Yiping Zhao:** Conceptualization, Methodology, Writing – review & editing, Funding acquisition, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work is supported by the Presidential Interdisciplinary Seed Grant from the University of Georgia.

Appendix. BSupplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.bios.2022.114721>.

References

- Abell, J.L., Driskell, J.D., Dluhy, R.A., Tripp, R.A., Zhao, Y.P., 2009. Fabrication and characterization of a multiwell array SERS chip with biological applications. *Biosens. Bioelectron.* 24 (12), 3663–3670.
- Adir, O., Poley, M., Chen, G., Froim, S., Krinsky, N., Shklover, J., Shainsky-Roitman, J., Lammers, T., Schroeder, A., 2020. Integrating artificial intelligence and nanotechnology for precision cancer medicine. *Adv. Mater.* 32 (13), 1901989.
- Afzal, A., 2020. Molecular diagnostic technologies for COVID-19: limitations and challenges. *J. Adv. Res.* 26, 149–159.
- Alafeef, M., Dighe, K., Moitra, P., Pan, D., 2020. Rapid, ultrasensitive, and quantitative detection of SARS-CoV-2 using antisense oligonucleotides directed electrochemical biosensor chip. *ACS Nano* 14 (12), 17028–17045.
- Ali, J., Khan, R., Ahmad, N., Maqsood, I., 2012. Random forests and decision trees. *Int. J. Comput. Sci. Issues* 9 (5), 272.
- Alvarez-Puebla, R.A., Liz-Marzán, L.M., 2010. SERS-based diagnosis and biodetection. *Small* 6 (5), 604–610.
- Ben-Hur, A., Ong, C.S., Sonnenburg, S., Schölkopf, B., Rätsch, G., 2008. Support vector machines and kernels for computational biology. *PLoS Comput. Biol.* 4 (10), e1000173.
- Bienko, M., Crosetto, N., Teytelman, L., Klemm, S., Itzkovitz, S., van Oudenaarden, A., 2013. A versatile genome-scale PCR-based pipeline for high-definition DNA FISH. *Nat. Methods* 10 (2), 122–124.
- Boygol-Barnum, S., Todd, S.O., Meng, J., Barnum, T.R., Chirkova, T., Haynes, L.M., Jadhao, S.J., Tripp, R.A., Oomens, A.G., Moore, M.L., Anderson, L.J., 2017. Mutating the CX3C motif in the G protein should make a live respiratory syncytial virus vaccine safer and more effective. *J. Virol.* 91 (10) e02059-02016.
- Cao, C., Zhang, J., Wen, X., Dodson, S.L., Dao, N.T., Wong, L.M., Wang, S., Li, S., Phan, A. T., Xiong, Q., 2013. Metamaterials-based label-free nanosensor for conformation and affinity biosensing. *ACS Nano* 7 (9), 7583–7591.
- Cha, H., Kim, H., Joung, Y., Kang, H., Moon, J., Jang, H., Park, S., Kwon, H.-J., Lee, I.-C., Kim, S., Yong, D., Yoon, S.-W., Park, S.-G., Guk, K., Lim, E.-K., Park, H.G., Choo, J., Jung, J., Kang, T., 2022. Surface-enhanced Raman scattering-based immunoassay for severe acute respiratory syndrome coronavirus 2. *Biosens. Bioelectron.* 202, 114008.
- Cheng, N., Chen, D., Lou, B., Fu, J., Wang, H., 2021. A biosensing method for the direct serological detection of liver diseases by integrating a SERS-based sensor and a CNN classifier. *Biosens. Bioelectron.* 186, 113246.
- Cherkaoui, D., Huang, D., Miller, B.S., Turbé, V., McKendry, R.A., 2021. Harnessing recombinase polymerase amplification for rapid multi-gene detection of SARS-CoV-2 in resource-limited settings. *Biosens. Bioelectron.* 189, 113328.
- Darwiche, A.S., Polasek, T.M., Aronson, J.K., Ogungbenro, K., Wright, D.F.B., Achour, B., Reny, J.-L., Daali, Y., Eiermann, B., Cook, J., Lesko, L., McLachlan, A.J., Rostami-Hodjegan, A., 2021. Model-informed precision dosing: background, requirements, validation, implementation, and forward trajectory of individualizing drug therapy. *Annu. Rev. Pharmacol. Toxicol.* 61 (1), 225–245.
- Ding, J., Lin, Q., Zhang, J., Young, G.M., Jiang, C., Zhong, Y., Zhang, J., 2021. Rapid identification of pathogens by using surface-enhanced Raman spectroscopy and multi-scale convolutional neural network. *Anal. Bioanal. Chem.* 413 (14), 3801–3811.
- Driskell, J.D., Shanmukh, S., Liu, Y., Chaney, S.B., Tang, X.J., Zhao, Y.P., Dluhy, R.A., 2008. The use of aligned silver nanorod arrays prepared by oblique angle deposition as surface enhanced Raman scattering substrates. *J. Phys. Chem. C* 112 (4), 895–901.
- Erzina, M., Trelin, A., Guselnikova, O., Dvorankova, B., Strnadova, K., Perminova, A., Ulbrich, P., Mares, D., Jerabek, V., Elashnikov, R., Svorcik, V., Lyutakov, O., 2020. Precise cancer detection via the combination of functionalized SERS surfaces and convolutional neural network with independent inputs. *Sens. Actuators, B* 308, 127660.
- Feng, W., Newbigging, A.M., Le, C., Pang, B., Peng, H., Cao, Y., Wu, J., Abbas, G., Song, J., Wang, D.-B., Cui, M., Tao, J., Tyrrell, D.L., Zhang, X.-E., Zhang, H., Le, X.C., 2020. Molecular diagnosis of COVID-19: challenges and research needs. *Anal. Chem.* 92 (15), 10196–10209.
- Garcia-Rico, E., Alvarez-Puebla, R.A., Guerrini, L., 2018. Direct surface-enhanced Raman scattering (SERS) spectroscopy of nucleic acids: from fundamental studies to real-life applications. *Chem. Soc. Rev.* 47 (13), 4909–4923.
- Hastie, T., Tibshirani, R., Friedman, J.H., Friedman, J.H., 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.
- He, H., Yan, S., Lyu, D., Xu, M., Ye, R., Zheng, P., Lu, X., Wang, L., Ren, B., 2021. Deep learning for biospectroscopy and biospectral imaging: state-of-the-art and perspectives. *Anal. Chem.* 93 (8), 3653–3665.
- Hendaus, M.A., Jomha, F.A., 2021. Can virus–virus interactions impact the dynamics of the covid-19 pandemic? *J. Biomol. Struct. Dyn.* AHEAD-OF-PRINT 1–5.
- Ho, C.S., Jean, N., Hogan, C.A., Blackmon, L., Jeffrey, S.S., Holodniy, M., Banaei, N., Saleh, A.A.E., Ermon, S., Dionne, J., 2019. Rapid identification of pathogenic bacteria using Raman spectroscopy and deep learning. *Nat. Commun.* 10 (1), 4927.
- James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. *An Introduction to Statistical Learning*. Springer.
- Jones, L., Bakre, A., Naikare, H., Kolhe, R., Sanchez, S., Mosley, Y.-Y.C., Tripp, R.A., 2021. Isothermal amplification and fluorescent detection of SARS-CoV-2 and SARS-CoV-2 variant virus in nasopharyngeal swabs. *PLoS One* 16 (9) e0257563-e0257563.
- Kneipp, K., Wang, Y., Kneipp, H., Perelman, L.T., Itzkan, I., Dasari, R.R., Feld, M.S., 1997. Single molecule detection using surface-enhanced Raman scattering (SERS). *Phys. Rev. Lett.* 78 (9), 1667–1670.
- Kurochkin, I.N., Eremenko, A.V., Evtushenko, E.G., Nechaeva, N.L., Durmanov, N.N., Guliev, R.R., Ryzhikov, I.A., Boginskaya, I.A., Sarychev, A.K., Ivanov, A.V., Lagarkov, A.N., 2021. SERS for bacteria, viruses, and protein biosensing. In: Rai, M., Reshetilov, A., Plekhanova, Y., Ingle, A.P. (Eds.), *Macro, Micro, and Nano-Biosensors: Potential Applications and Possible Limitations*. Springer International Publishing, Cham, pp. 75–94.
- Leong, Y.X., Lee, Y.H., Koh, C.S.L., Phan-Quang, G.C., Han, X., Phang, I.Y., Ling, X.Y., 2021. Surface-enhanced Raman scattering (SERS) taster: a machine-learning-driven multireceptor platform for multiplex profiling of wine flavors. *Nano Lett.* 21 (6), 2642–2649.
- Li, J., Wuethrich, A., Edwardraja, S., Lobb, R.J., Puttick, S., Rose, S., Howard, C.B., Trau, M., 2021. Amplification-free SARS-CoV-2 detection using nanoyeast-scFv and ultrasensitive plasmonic nanobox-integrated nanomixing microassay. *Anal. Chem.* 93 (29), 10251–10260.
- Liu, Y.-J., Zhang, Z.-Y., Zhao, Q., Dluhy, R., Zhao, Y.-P., 2009. Surface enhanced Raman scattering from an Ag nanorod array substrate: the site dependent enhancement and layer absorbance effect. *J. Phys. Chem. C* 113 (22), 9664–9669.
- Liu, Y.J., Chu, H.Y., Zhao, Y.P., 2010. Silver nanorod array substrates fabricated by oblique angle deposition: morphological, optical, and SERS characterizations. *J. Phys. Chem. C* 114 (18), 8176–8183.
- Liu, Y.J., Zhao, Y.P., 2008. Simple model for surface-enhanced Raman scattering from tilted silver nanorod array substrates. *Phys. Rev. B* 78 (7), 075436.
- Long, Q.-X., Tang, X.-J., Shi, Q.-L., Li, Q., Deng, H.-J., Yuan, J., Hu, J.-L., Xu, W., Zhang, Y., Lv, F.-J., Su, K., Zhang, F., Gong, J., Wu, B., Liu, X.-M., Li, J.-J., Qiu, J.-F., Chen, J., Huang, A.-L., 2020. Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nat. Med.* 26 (8), 1200–1204.
- Lussier, F., Thibault, V., Charron, B., Wallace, G.Q., Masson, J.-F., 2020. Deep learning and artificial intelligence methods for Raman and surface-enhanced Raman scattering. *TrAC Trends Anal. Chem.* (Reference Ed.) 124, 115796.
- Mahalingam, S., Schwarze, J., Zaid, A., Nissen, M., Sloots, T., Tauro, S., Storer, J., Alvarez, R., Tripp, R.A., 2006. Perspective on the host response to human metapneumovirus infection: what can we learn from respiratory syncytial virus infections? *Microb. Infect.* 8 (1), 285–293.
- Mitchell, M.J., Billingsley, M.M., Haley, R.M., Wechsler, M.E., Peppas, N.A., Langer, R., 2021. Engineering precision nanoparticles for drug delivery. *Nat. Rev. Drug Discov.* 20 (2), 101–124.
- Murray, J., Hogan, R.J., Martin, D.E., Blahunka, K., Sancilio, F.D., Balyan, R., Lovern, M., Still, R., Tripp, R.A., 2021. Probencid inhibits SARS-CoV-2 replication in vivo and in vitro. *Sci. Rep.* 11 (1), 18085.
- Nie, S., Emory, S.R., 1997. Probing single molecules and single nanoparticles by surface-enhanced Raman scattering. *Science* 275 (5303), 1102–1106.
- Nolan, T., Hands, R.E., Bustin, S.A., 2006. Quantification of mRNA using real-time RT-PCR. *Nat. Protoc.* 1 (3), 1559–1582.
- Pan, Y., Zhang, D., Yang, P., Poon, L.L.M., Wang, Q., 2020. Viral load of SARS-CoV-2 in clinical samples. *Lancet Infect. Dis.* 20 (4), 411–412.

- Paria, D., Kwok, K.S., Raj, P., Zheng, P., Gracias, D.H., Barman, I., 2022. Label-free spectroscopic SARS-CoV-2 detection on versatile nanoimprinted substrates. *Nano Lett.* 22 (9), 3620–3627.
- Piret, J., Boivin, G., 2022. Viral interference between respiratory viruses. *Emerg. Infect. Dis.* 28 (2), 273–281.
- Ralbovsky, N.M., Lednev, I.K., 2020. Towards development of a novel universal medical diagnostic method: Raman spectroscopy and machine learning. *Chem. Soc. Rev.* 49 (20), 7428–7453.
- Ravi, N., Cortade, D.L., Ng, E., Wang, S.X., 2020. Diagnostics for SARS-CoV-2 detection: a comprehensive review of the FDA-EUA COVID-19 testing landscape. *Biosens. Bioelectron.* 165, 112454.
- Shanmukh, S., Jones, L., Driskell, J., Zhao, Y., Dluhy, R., Tripp, R.A., 2006. Rapid and sensitive detection of respiratory virus molecular signatures using a silver nanorod array SERS substrate. *Nano Lett.* 6 (11), 2630–2636.
- Shanmukh, S., Jones, L., Zhao, Y.-P., Driskell, J.D., Tripp, R.A., Dluhy, R.A., 2008. Identification and classification of respiratory syncytial virus (RSV) strains by surface-enhanced Raman spectroscopy and multivariate statistical techniques. *Anal. Bioanal. Chem.* 390 (6), 1551–1555.
- Shao, F., Lu, Z., Liu, C., Han, H., Chen, K., Li, W., He, Q., Peng, H., Chen, J., 2014. Hierarchical nanogaps within bioscaffold arrays as a high-performance SERS substrate for animal virus biosensing. *ACS Appl. Mater. Interfaces* 6 (9), 6281–6289.
- Singh, J.P., Chu, H., Abell, J., Tripp, R.A., Zhao, Y., 2012. Flexible and mechanical strain resistant large area SERS active substrates. *Nanoscale* 4 (11), 3410–3414.
- Smith, E., Dent, G., 2019. Modern Raman Spectroscopy: a Practical Approach. John Wiley & Sons.
- Song, C., Chen, J., Abell, J.L., Cui, Y., Zhao, Y., 2012. Ag-SiO₂ core-shell nanorod arrays: morphological, optical, SERS, and wetting properties. *Langmuir* 28 (2), 1488–1495.
- Sun, Y., Xu, L., Zhang, F., Song, Z., Hu, Y., Ji, Y., Shen, J., Li, B., Lu, H., Yang, H., 2017. A promising magnetic SERS immunoassay for sensitive detection of avian influenza virus. *Biosens. Bioelectron.* 89, 906–912.
- Teo, A.K.J., Choudhury, Y., Tan, I.B., Cher, C.Y., Chew, S.H., Wan, Z.Y., Cheng, L.T.E., Oon, L.L.E., Tan, M.H., Chan, K.S., Hsu, L.Y., 2021. Saliva is more sensitive than nasopharyngeal or nasal swabs for diagnosis of asymptomatic and mild COVID-19 infection. *Sci. Rep.* 11 (1), 3134.
- Torun, H., Bilgin, B., Ilgu, M., Yanik, C., Batur, N., Celik, S., Ozturk, M., Dogan, O., Ergonul, O., Solaroglu, I., Can, F., Onbasli, M.C., 2021. Machine learning detects SARS-CoV-2 and variants rapidly on DNA aptamer metasurfaces. *medRxiv*, 2021, 21261749, 2008.2007.
- Tripp Ralph, A., Moore, D., Jones, L., Sullender, W., Winter, J., Anderson Larry, J., 1999. Respiratory syncytial virus G and/or SH protein alters Th1 cytokines, natural killer cells, and neutrophils responding to pulmonary infection in BALB/c mice. *J. Virol.* 73 (9), 7099–7107.
- Udugama, B., Kadriresan, P., Kozlowski, H.N., Malekjahani, A., Osborne, M., Li, V.Y.C., Chen, H., Mubareka, S., Gubbay, J.B., Chan, W.C.W., 2020. Diagnosing COVID-19: the disease and tools for detection. *ACS Nano* 14 (4), 3822–3835.
- Wyllie, A.L., Fournier, J., Casanova-Massana, A., Campbell, M., Tokuyama, M., Vijayakumar, P., Warren, J.L., Geng, B., Muenker, M.C., Moore, A.J., Vogels, C.B.F., Petrone, M.E., Ott, I.M., Lu, P., Venkataraman, A., Lu-Culligan, A., Klein, J., Earnest, R., Simonov, M., Datta, R., Handoko, R., Naushad, N., Sewanan, L.R., Valdez, J., White, E.B., Lapidus, S., Kalinich, C.C., Jiang, X., Kim, D.J., Kudo, E., Linehan, M., Mao, T., Moriyama, M., Oh, J.E., Park, A., Silva, J., Song, E., Takahashi, T., Taura, M., Weizman, O.-E., Wong, P., Yang, Y., Bermejo, S., Odio, C. D., Omer, S.B., Dela Cruz, C.S., Farhadian, S., Martinello, R.A., Iwasaki, A., Grubaugh, N.D., Ko, A.I., 2020. Saliva or nasopharyngeal swab specimens for detection of SARS-CoV-2. *N. Engl. J. Med.* 383 (13), 1283–1286.
- Yang, Y., Peng, Y., Lin, C., Long, L., Hu, J., He, J., Zeng, H., Huang, Z., Li, Z.-Y., Tanemura, M., Shi, J., Lombardi, J.R., Luo, X., 2021. Human ACE2-functionalized gold “virus-trap” nanostructures for accurate capture of SARS-CoV-2 and single-virus SERS detection. *Nano-Micro Lett.* 13 (1), 109.
- Yang, Y., Xu, B., Haverstick, J., Ibtehaz, N., Muszyński, A., Chen, X., Chowdhury, M.E.H., Zughair, S.M., Zhao, Y., 2022. Differentiation and classification of bacterial endotoxins based on surface enhanced Raman scattering and advanced machine learning. *Nanoscale* 14 (24), 8806–8817.
- Yeh, Y.-T., Gulino, K., Zhang, Y., Sabestien, A., Chou, T.-W., Zhou, B., Lin, Z., Albert, I., Lu, H., Swaminathan, V., Ghedin, E., Terrones, M., 2020. A rapid and label-free platform for virus capture and identification from clinical samples. *Proc. Natl. Acad. Sci. U.S.A.* 117 (2), 895–901.
- Zhang, Z., Li, D., Wang, X., Wang, Y., Lin, J., Jiang, S., Wu, Z., He, Y., Gao, X., Zhu, Z., Xiao, Y., Qu, Z., Li, Y., 2022. Rapid detection of viruses: based on silver nanoparticles modified with bromine ions and acetonitrile. *Chem. Eng. J.* 438, 135589.