## Title

Label-free SARS-CoV-2 Detection Platform Based on Surface-enhanced Raman Spectroscopy with Support Vector Machine Spectral Pattern Recognition



## Authors

Li, Tieyi
Srivastava, Siddharth
Liu, Jun
et al.

## Publication Date

2023

## DOI

10.30919/es8d862

Peer reviewed

# Label-free SARS-CoV-2 Detection Platform Based on Surface-enhanced Raman Spectroscopy with Support Vector Machine Spectral Pattern Recognition

Tieyi Li[1], Siddharth Srivastava[1], Jun Liu[1], Feng Li[2], Yong Kim[2], David T.W. Wong[2], Aaron Carlin[3], Ya-Hong Xie[1*]

[1]*Department of Materials Science and Engineering, University of California Los Angeles, Los Angeles, CA 90095, USA*

[2]*UCLA School of Dentistry, 10833 Le Conte Ave. Box 951668, Los Angeles, CA 90095-1668, USA*

[3]*School of Medicine, University of California San Diego, San Diego, CA 92093, USA*

*To whom correspondence should be addressed:

Ya-Hong Xie

Department of Materials Science and Engineering,

University of California Los Angeles, Los Angeles, CA 90095

E-mail: yhx@ucla.edu

Phone number: (310) 259-6946

Keyword: SARS-CoV-2 detection, SERS identification of molecule, machine learning, label-free, gold nano-pyramidal platform

## Abstract

We introduce a biosensing platform combining surface-enhanced Raman spectroscopy (SERS) and machine learning for combating COVID-19 and potentially future occurrence of similar pandemics of viral infection in nature. Compared to the RT-PCR and rapid antigen test, our platform can detect SARS-CoV-2 in human saliva with reliable accuracy and in a short time duration. Cross-validation and blind test are performed to identify SARS-CoV-2 virus against close-related particles including SARS-CoV-1 and extracellular vesicles. Simulated clinical samples with SARS-CoV-2 spiked saliva specimens are tested for building the SARS-CoV-2 identifier, 90% sensitivity and 80% specificity are achieved respectively. Clinical samples composed of 5 COVID patients and 5 healthy controls are tested blindly and render 100% sensitivity and 80% specificity based on the trained classifier. Targeting to become a better public pandemic monitoring tool, our platform simplifies the sample harvest and processing procedures and can release test results within five hours. Our study indicates the possibility of inventing a better rapid test compared with RT-PCR and more accurate test compared with antigen test with less cost and complexity.

## 1. Introduction

Since the emergence of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) in December 2019, more than 620 million cases and 6 million deaths have been reported till November 2022, as declared by World Health Organization (WHO) [1]. The typical symptoms include fever, fatigue, severe respiratory illness, pneumonia as well as dyspnea. Recently, long-term damage to brain and heart have also been reported [2]. More SARS-CoV-2 variants have been emerging globally, such as the ones in the United Kingdom (B.1.1.7), the United States (B.1.429, Washington, B.1.1.529 or Omicron and Omicron BA.2) and India (B.1.617.2 or Delta) causing more rapid and wider spread of the pandemic around the world [3]. Currently, the SARS-CoV-2 strain Omicron BA.5 makes up around 62% of the COVID cases [4]. Though the mortality of the more recent variants has been much lower than the original strains [5], the transmissibility has significantly increased [6, 7].

SARS-CoV-2 belongs to the family of coronavirus of 60-140nm in vesicle size. It is composed of single-strand RNA, lipid bilayer membrane and structural proteins (spike protein, envelop protein, membrane protein and nucleocapsid protein) [3]. Currently the prevalent diagnostic technologies are RT-PCR and antigen test, which detect the viral RNA and the protein biomarkers (e.g., spike protein) [8]. As SARS-CoV-2 belongs to the family of the single-stranded RNA viruses, RT-PCR is the most widely used detection tool due to its high accuracy, sensitivity, and Limit of Detection (LoD). The LoD of around 100 particles/mL, sensitivity above 80% and specificity above 95% have been reported [8, 9]. It is worth noting that there are drawbacks of RT-PCR preventing it from becoming the optimal diagnostic technology for targeting highly mutable and contagious viruses. For most of the nucleic acid-based tests, highly

specific primers are required in the reverse transcription step, therefore specific new primers are needed to deal with the mutated variants [10]. RT-PCR is also extremely sensitive to the viral load of the samples thus the viral concentration fluctuation of Nasopharyngeal swab specimens or salivary specimens could result in false positive/negative cases [11]. Moreover, sophisticated equipment, costly regents as well as professional operators are required for collection and analysis, which inevitably increases the time and consumption cost. In contrast, the faster test tool, antigen test, could generate results in 15-30 minutes. However, it is less reliable due to worse sensitivity and specificity (around 50% and 90%, respectively) [12]. Fast, accurate and non-invasive detection tools are still needed to monitor the pandemic and potentially identify other highly infectious viruses in the future. In this report, we present the feasibility of applying surface-enhanced Raman spectroscopy (SERS) for rapid identification of viruses. A schematic procedure is provided in figure 1.

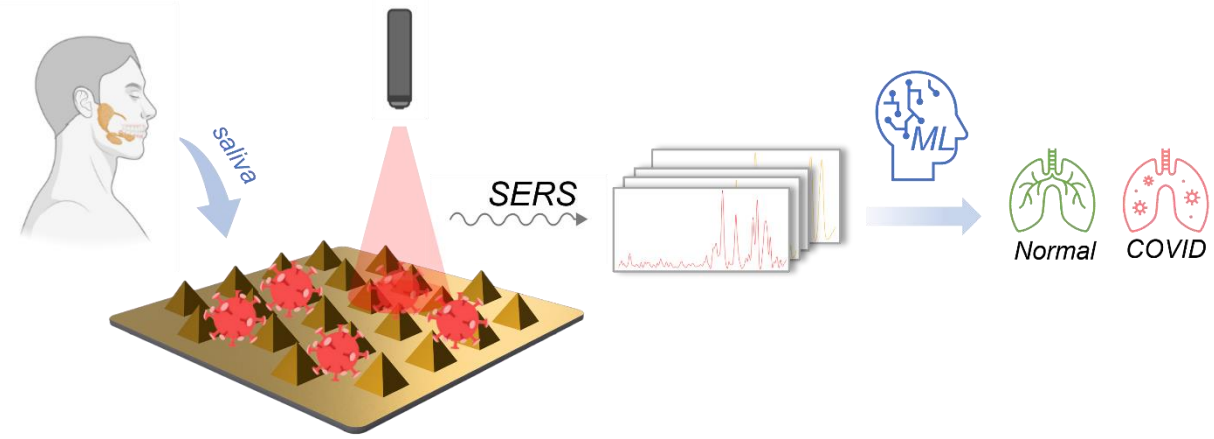



**Figure 1** Schematic of SERS-based biosensing platform for virus detection

The development of Surface-enhanced Raman spectroscopy (SERS) in biosensing has attracted a lot of attention due to its fingerprinting capability, excellent sensitivity, label-free

properties, and biocompatibility [13]. SERS has demonstrated its ability to fingerprint small molecules such as chemical dyes [14], mineral ingredients [15], larger molecules such as peptides and DNAs/RNAs [16, 17, 18]. The single molecule characterization capability with specially designed SERS substrate attracts people's attentions [19]. Moreover, in the past few decades, SERS-based profiling has been utilized for investigating biological specimens including cells [20], bacteria [21], viruses [22], and extracellular vesicles [23]. The application of SERS-based technologies for detecting multiple types of viruses, such as influenza virus [24], Hepatitis B virus [25], respiratory virus [26], have been demonstrated recently with competitive detection accuracy.

Compared to antigen test, SERS extracts SARS-CoV-2 biomarkers from multiple components, including structural protein, lipid bilayer and RNA strand [27]. Hereby, SERS has the advantages of drawing a more thorough picture over antigen test. Unlike nucleic acid based detecting technologies, SERS does not require complicated primers and regents nor special specimen treatment, therefore the estimated cost per test would be lower. Besides, SERS specimens can be isolated from different biofluids such as saliva, serum, urine and bronchoalveolar fluid, allowing for simple and non-invasive sample harvesting. Furthermore, SERS characterization for each sample requires a maximum of 1 to 6 hours, which makes it a more feasible "rapid-testing" method for SARS-CoV-2 compared to RT-PCR [9, 27]. Label-free of SERS-based test also makes it more amenable to scale up and adaption to more SARS-CoV-2 variants study.

SERS-based detection has been implemented for COVID detection. Improved detecting efficiency and limit of detection have been reported with uniquely designed biosensor setup [28]. To prepare highly concentrated virus samples for SERS characterization, Sequential

centrifugation and filtration are typically applied to isolate viruses from cell culture media [29]. It has been reported that exosomes have similar size and density as viruses (30-150nm, 1.08–1.19 g/ml) [30, 31], therefore It is inevitable to exclude exosomes during virus isolation, which could lead to confusion in fingerprinting viruses. To establish the genuine fingerprint, exosomes' signatures need to be subtracted during either sample preparation or data processing.

In this paper, we demonstrate the feasibility of our SERS and machine learning- based fingerprinting and signature identification platform as being a potentially accurate and rapid saliva-based SARS-CoV-2 detection technique that could replace the current antigen test as a pandemic monitoring tool. Figure 2 demonstrates the basic workflow. Briefly, SARS-CoV-2 virus samples were compared with SARS-CoV-1 virus and Vero-TMPRSS2 cell line- derived exosome samples and were successfully identified with 80% accuracy. We subsequently evaluated the diagnostic capabilities by comparing SARS-CoV-2 spiked human salivary samples versus healthy control. 10 SARS-CoV-2 spiked human salivary samples and 10 healthy controls salivary samples were applied to build the identifier. 90% sensitivity and 80% specificity were achieved afterward in blind test with the 20 samples. Using the above identification model, 5 COVID patients versus 5 healthy controls saliva samples were tested and 9 out of the ten individuals are identified correctly. Finally, we provide detailed estimation of the advances and theoretical analysis of the feasibility of our platform.

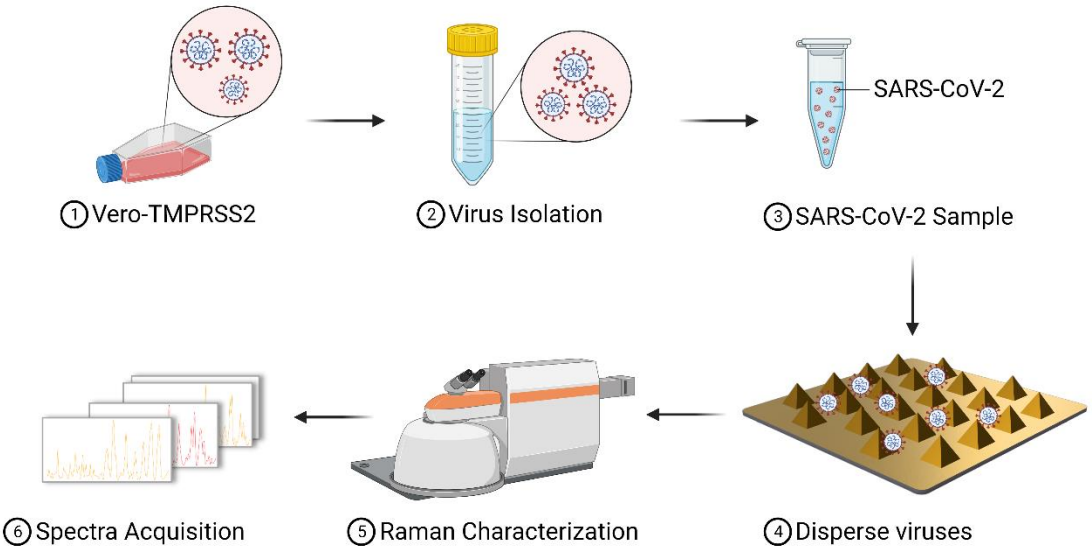



**Figure 2** Schematic working flow of SERS characterization of SARS-CoV-2 specimens

## 2. Methods and materials

### 2.1 Virus samples preparation

The virus samples were produced, inactivated, and validated by the Institutional Biosafety Committee (IBC) for the University of California, San Diego. Vero-TMPRSS2 cells are infected with viruses (either SARS-CoV-2 or SARS-CoV-2). Sequential centrifuge and filtration were used to isolate and purify the virus from cell culture media then the viruses were diluted in cell culture media (DMEM + 1% FBS + 10mM HEPES + 50 units/ml Penicillin and 50 µg/ml Streptomycin). Virus samples were then inactivated by heat (65°C for 30 minutes) [32] or UV (400 mJ/cm$^2$ delivered at UV 254nm) [33]. After inactivation, $10^8$ to $10^{10}$ viruses per ml were estimated by ddPCR (RNA). Figure 3 shows a typical Transmission Electron Microscopy (TEM, FEI TF20 High-resolution EM, USA) image of the specimen. Individual virus particles of about 50 nm diameter with the characteristic corona are clearly visible.

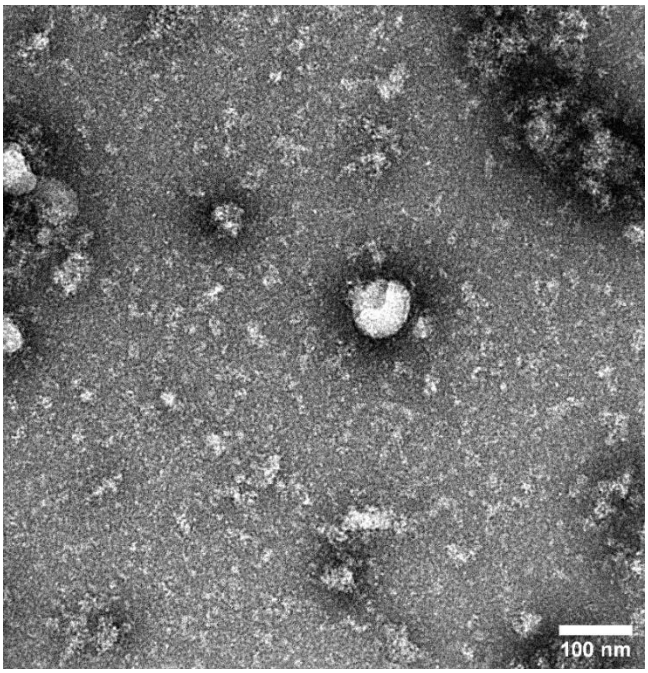



**Figure 3** TEM image of SARS-CoV-2 specimen

## 2.2 SARS-CoV-2 spiked human salivary samples preparation

The isolated and purified virus samples were used for preparing the SARS-CoV-2 spiked human salivary samples. The virus samples and salivary samples of healthy control were mixed with the volume ratio that keeps the viral concentration around $10^8$ particles/mL. Then the spiked salivary samples were aliquoted for multiple SERS testing.

## 2.3 SARS-CoV-2 clinical samples preparation

Archived saliva samples were obtained from an observational cohort study of hospitalized patients with COVID-19 from April 2020 until February 2021. The study was approved by the UCLA Institutional Review Board (#20-000473). Informed consent was obtained from all study participants. Patients with confirmed positive SARS-CoV-2 RT-PCR nasopharyngeal swabs were enrolled in an observational cohort study within 72 hours of admission. Exclusion criteria included pregnancy, hemoglobin < 8g/dL, or inability to provide informed consent. Blood specimens, nasopharyngeal swabs, and saliva were collected

throughout hospitalization for up to 6 weeks. Demographic and clinical data, including laboratory results and therapeutics, were collected from the electronic medical records. Clinical severity was scored using the NIAID 8-point ordinal scale. A total of 10 samples were included in this study. Whole saliva was collected by passive drool into a cryovial. Samples were transported to the laboratory and immediately placed in -80 °C freezer for storage.

## 2.4 Surface-enhanced Raman spectroscopy

SERS biosensing platform are based on Raman scattering, in which the incident photon undergoes inelastic scattering on interaction with the target analyte that produces unique vibrational modes from its components [34]. The localized surface plasmon resonance (LSPR) on the SERS substrate surface originates from the interaction between electromagnetic field of the incident light and electrons in metal, which significantly enhance the detectability of low concentration components in the analytes. Due to the high specificity of excitation-emission photon energy shift during Raman scattering process, the analyte of interest is able to generate a unique spectrum as the fingerprint, which can serve as a reference in the identification.

## 2.5 SERS substrate fabrication

The platform implementing surface enhancement is fabricated primarily based on polystyrene sphere lithography [35], the product possesses a 2D periodic pyramidal structure that allows for a significantly enhanced electromagnetic field to be localized at the ‘waist’ of each Au pyramid. Polystyrene spheres (Thermo Fisher Scientific, USA) were first applied to construct a monolayer on SiO2/Si wafer (MSE Supplies, USA) surface via self-assembly to create hexagonal patterns. Subsequently, the substrate was dry etched by O2 plasma under 200W for 50 s to shrink the polystyrene sphere size. The reduced polystyrene spheres act as the mask in the plasma etching process to remove the SiO2 layer under exposure. Subsequently, the substrate

was etched in 60% KOH solution (Sigma Aldrich, USA) for 2 mins to form periodic pyramidal reciprocal structures on the Si layer with patterned SiO2 as a mask. A 200nm Au film was deposited on the mode and finally, epoxy was used to peel off the Au film which was attached to a new Si wafer. On the fabricated platform, Au nano-pyramids with base length of 200 nm, height of 200 nm were obtained. These were utilized for profiling of exosomal and viral liquid biopsies. Figure 4a shows the Scanning Electron Microscopy (SEM, FEI Nova NanoSEM 230, USA) image of the SERS substrate. A periodic hexagonal pattern is formed. Considering the dimension of the nano-pyramid and the spacing between them, our platform provides appropriate room for exosomes/virus to fit in the hot-spots, which mainly lay on the lateral faces. Figure 4b indicates the landing position of analytes on the pyramidal surface, the obscure imaging is due to the presence of crystallization after the sample buffer (mostly PBS) evaporates. SEM images are taken with FEI Nova NanoSEM 230.

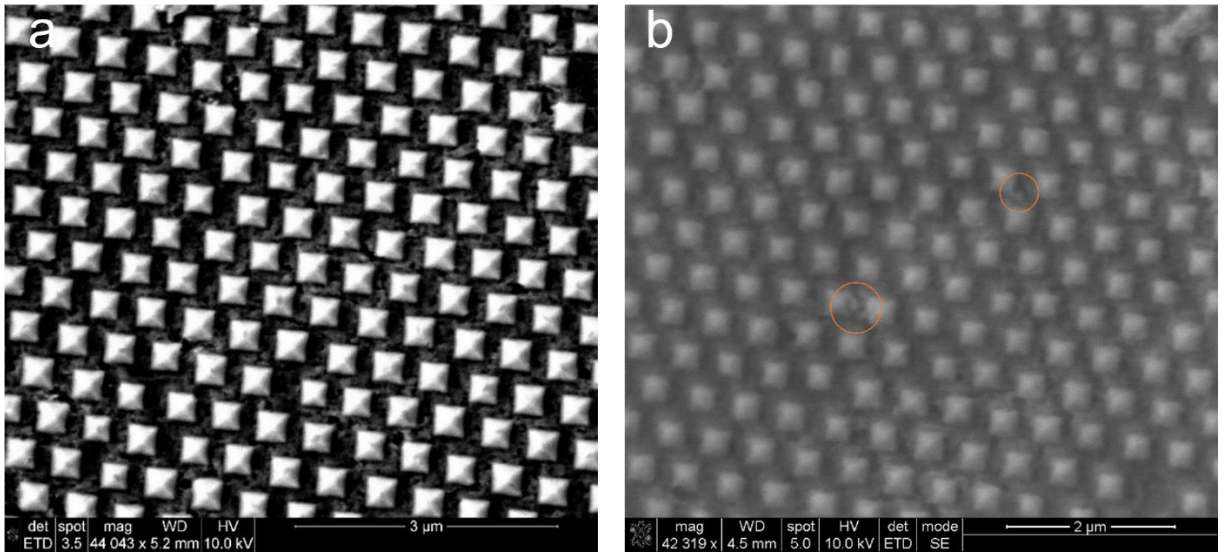



**Figure 4** SEM imaging of the platform. a) SEM image of the SERS gold nanopyramids substrate; b) SEM image showing the existence of viruses (orange circles) on the substrate after specimen solution drying

## 2.6 Method of obtaining SERS spectral signatures

In terms of the acquisition of SERS spectral signatures of the specimen, we implemented a single bioparticle scanning protocol. Specifically, a droplet of about 5 μL of the liquid sample was pipetted onto the surface of the SERS platform and dried under room ambient or in a vacuum desiccator typically within 15 minutes. Raman spectral data were immediately recorded using Raman spectrometer (Renishaw inVia Confocal Raman spectrometer, UK) under ambient conditions (20 °C, 1 atm), which is manually controlled by WiRE4.4 PC software. A laser with excitation wavelength of 785 nm was selected to suppress fluorescence background while maintaining a strong localized surface plasmon resonance. The map image acquisition function incorporated in the software was primarily used to collect numerical spectral data. A large square map (searching map) covering an area of 300 μm x 300 μm with each pixel dimension of 10 μm x 10 μm was implemented to search for the positions with micro-vesicles. Those positions were then characterized by a small square map (obtaining map) of 5 μm x 5 μm with 1 um x 1 um pixel size. Laser power of 50mW, and acquisition time of 0.1s was chosen for the searching map while 10mW, 0.5s were for the obtaining map to avoid overheating and acquire spectra with high signal-to-noise ratio. The obtaining map yielded candidate spectra through which a spectra-selecting program traverses for establishing the spectral database. The rate of characterizing analytes is around 10-40 analytes/hour. According to our current spectral dataset size, approximately 1-6 hours are needed. As demonstrated by figure 5, the spectra obtained have explicit Raman ranges with high signal-to-noise ratios. Peak assignments are given in Table S1.

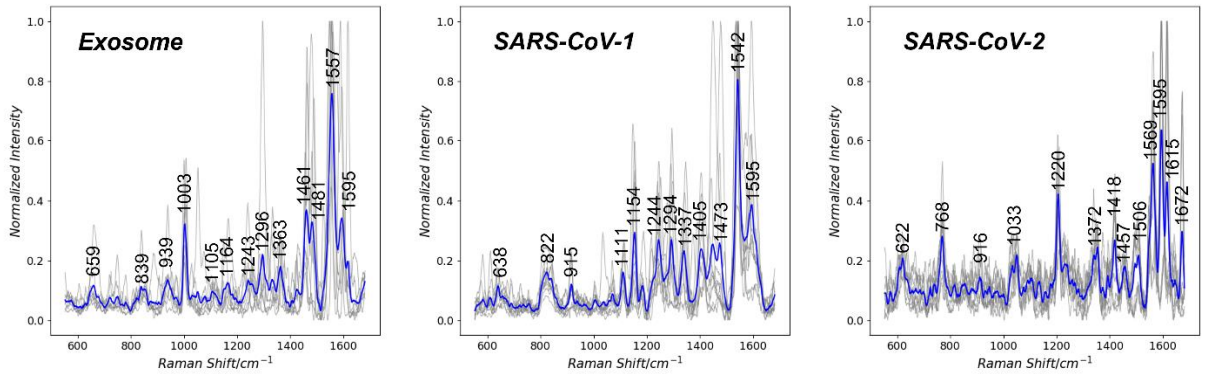



**Figure 5** Spectra of Vero-TMPRSS2 exosome, SARS-CoV-1, SARS-CoV-2. Highly uniform spectra from the particles (gray lines) and averaged spectra (blue lines) demonstrate different patterns of different particles.

## 2.7 Method of spectral processing and data analysis

Approximate 50 to 300 signal spots (depending on the particle concentration) were obtained for each sample to produce spectra that have 1023 Raman shifts in the range from 553 to 1581 $cm^{-1}$. Preprocessing steps are applied to alleviate the spectral signature fluctuations caused by sample variations, SERS platform heterogeneity, and instrument fluctuation. To elaborate, Fluorescence background subtraction and noise reduction are performed by batch processing based on asymmetric least square fitting [36] and Savitzky-Golay filtering [37], followed by min-max normalization that proportionally compresses the original intensity range to [0, 1]. A predictive model established by supervised learning or classification is the core of the proposed technology. It requires appropriate complexity of the classifier to prevent both underfitting and overfitting for the purpose of generalizing the characteristic signature effectively. We use the conventional but powerful algorithm Support Vector Machine (SVM) for the classification tasks. Unsupervised learning or clustering analysis by Hierarchical Clustering Analysis was also used as an auxiliary tool. Cross-validations are then applied to pre-evaluate

our methodology given the labels and optimize the model settings, followed by tests for evaluating diagnostic capability. All the analyses are realized with Python using NumPy, SciPy and Scikit-learn modules and take less than 20 minutes to complete.

## 3. Results and Discussion

### 3.1 Single-vesicle techniques for viral detection

The single-vesicle detectability of SIM brings advantages in COVID detection. There are also several challenges originating from the working principle of single-vesicle detection. Most importantly, the feasibility of single-vesicle detection is determined by the standard signature of the target analyte (e.g., SARS-CoV-2) that we can refer to. The presence of EVs could potentially impact the procedure of obtaining the standard SERS spectral signature of SARS-CoV-2, as shown in figure 6. The sample preparation step, the sample loading step, and the characterization step are all supposed to be conducted rigorously to prevent any possibility of contamination. The subsequent data processing step is also needed to get rid of irrelevant target analyte signatures. Secondly, though SERS dramatically increases the signal intensity of the analyte which facilitates much more sensitive detection, the inherent biological variabilities are also amplified. The signatures of SARS-CoV-2 from different SERS characterizations instances might fluctuate to some extent. Therefore, the intra-class (such as SARS-CoV-2) fluctuations versus the inter-class (such as SARS-CoV-2/EVs) differences must be validated to support the decision boundary. In addition, Single-vesicle characterization is usually performed in the manner of individual scanning, which greatly limits the data throughput. Much effort needs to be done to boost the data harvest rate and determine the characterization data size to make a sufficiently reliable diagnosis conclusion. Due to the above concerns, we have performed the following experiments to establish the capability of SIM for SARS-CoV-2 detection.

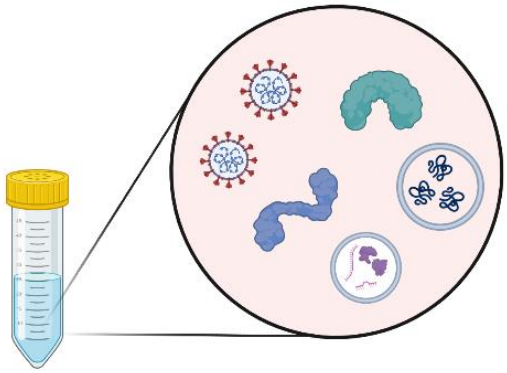


**Figure 6** Schematic plot of multiple types of contents in SARS-CoV-2 specimen. Virus (particles with spike protein), exosomes (double-layer particles without spike protein) and free-floating protein molecules are shown.

## 3.2 Differentiation of SARS-COV-2 vs SARS-COV-1 virion in mixture of cell lysate

As a prerequisite step for establishing SIM identification of SARS-CoV-2 signature, we first evaluated the proposed platform in differentiating SARS-CoV-2 from other closely related virus types, including other types of virions and extracellular vesicles, of which the dimensions are close to the SARS-CoV-2 virus. SARS-CoV-1 is reported to share more than 70% genetic similarity with SARS-CoV-2 [38], leading to highly similar structural components such as single-stranded RNA and spike protein, while the mutations make the latter less deadly but much more transmissible. With SARS-CoV-1 as a candidate, 10 SARS-CoV-1 specimens and 10 SARS-CoV-2 specimens were prepared and then characterized by SERS following our SERS map protocol. 50 to 70 spots rendering spectral signatures with high signal-to-noise ratio were collected for each sample, multiple spectra were saved per spot to account for the information of spectral intensity fluctuations, which allows for comprehensive training of the model by making it less sensitive to the slight changes.

In total, 1929 spectra from SARS-CoV-1 samples and 1559 from SARS-CoV-2 samples were recorded. Figure 5 are three examples of spectra set belonging to a single particle of Vero-TMPRSS2, SARS-CoV-1, SARS-CoV-2, respectively, in which multiple Raman 'snapshots' on different positions of a single particle and the average spectrum are presented. The peak assignment information is given in the supplementary material. Peaks in the spectra typically originate from the molecular bonds within amino acids, nucleic acid, Amide, C-C stretching or $CH_n$ deformation etc. Multiple spectral patterns were discovered within each type of specimen (e.g., SARS-CoV-1) though the spectral signatures from a single particle are uniform, therefore a standard representative signature is lacking. A possible reason is that SERS platform renders a superior sensitivity in detecting particles with extremely low concentration, the spectral signature is also prone to fluctuate due to the minor structural change of the molecule and the analyte-hotspot interaction. Hereby, we implemented the supervised and unsupervised learning model for building the viral fingerprints, which would be used as a standard for virus identification.

The virus samples were purified from Vero-TMPRSS2 cells by sequential centrifugation, other biological particles with a similar dimension as the virus might be retained, leading to the non-ideal purity which could confuse the identifying model. Therefore, we implemented a control sample of Vero-TMPRSS2 cells under the same preparation manner expecting infection. The spectral signatures from the control act as background signals of the SARS-CoV-1 and SARS-CoV-2 spectral datasets. Linear discriminant analysis (LDA) was implemented to reduce the dimension of the spectra for clearer visualization of the datapoints distribution, in which the original spectra dataset was transformed into points with two-dimensional coordinates. LDA tries to group the spectra by maximizing the distance between the centroid of each group to the global centroid meanwhile minimizing intra-group variance. The inter-group distance conceptually

represents the similarity between the corresponding spectra, as shown in figure 7. It can be concluded that SARS-CoV-1 and SARS-CoV-2 clouds overlap with the Vero-TMPRSS2 in small portions, which are believed to be the non-virus particles examined in virus samples. Subsequently, Hierarchical Clustering Analysis (HCA) was used to cluster similar particles in Vero-TMPRSS2 and virus samples. Based on the groups clustered, we label the particles originally belonging to virus samples but clustered into Vero-TMPRSS2 as negative (i.e., non-SARS-CoV-2). We call this “label-correction process”, as shown in figure 8(a), 8(b). Figure 8(c), 8(d), 8(d) present three similar SERS spectral signatures from different particles belonging to the same cluster. The spectrum in Figure 8(c) was originally mislabeled by SARS-CoV-2 which would be corrected. Peak assignments are given in Table S2.

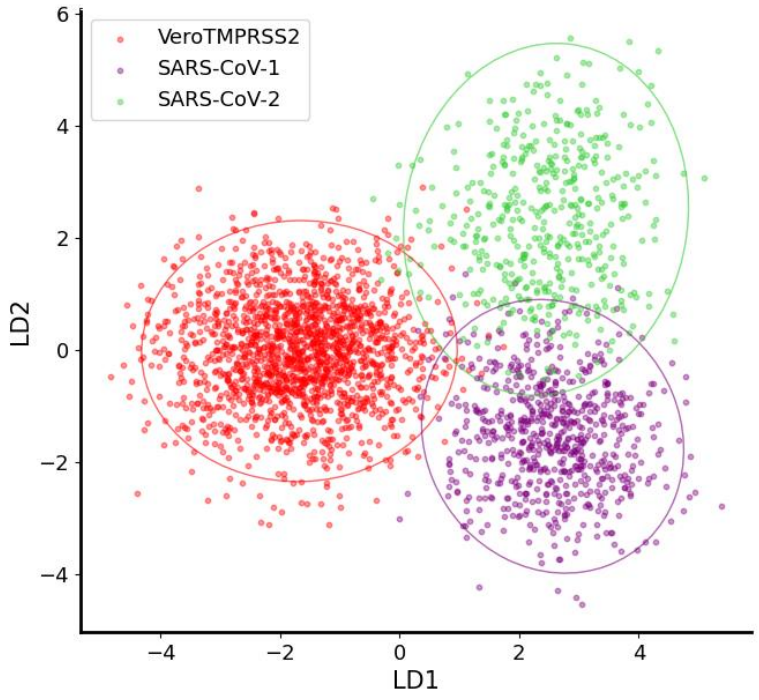


**Figure 7** Linear Discriminant Analysis for dimensionality reduction. Spectral signatures of SARS-CoV-1, SARS-CoV-2, exosomes are processed by dimensionality reduction and visualized in the 2-dimensional plot.

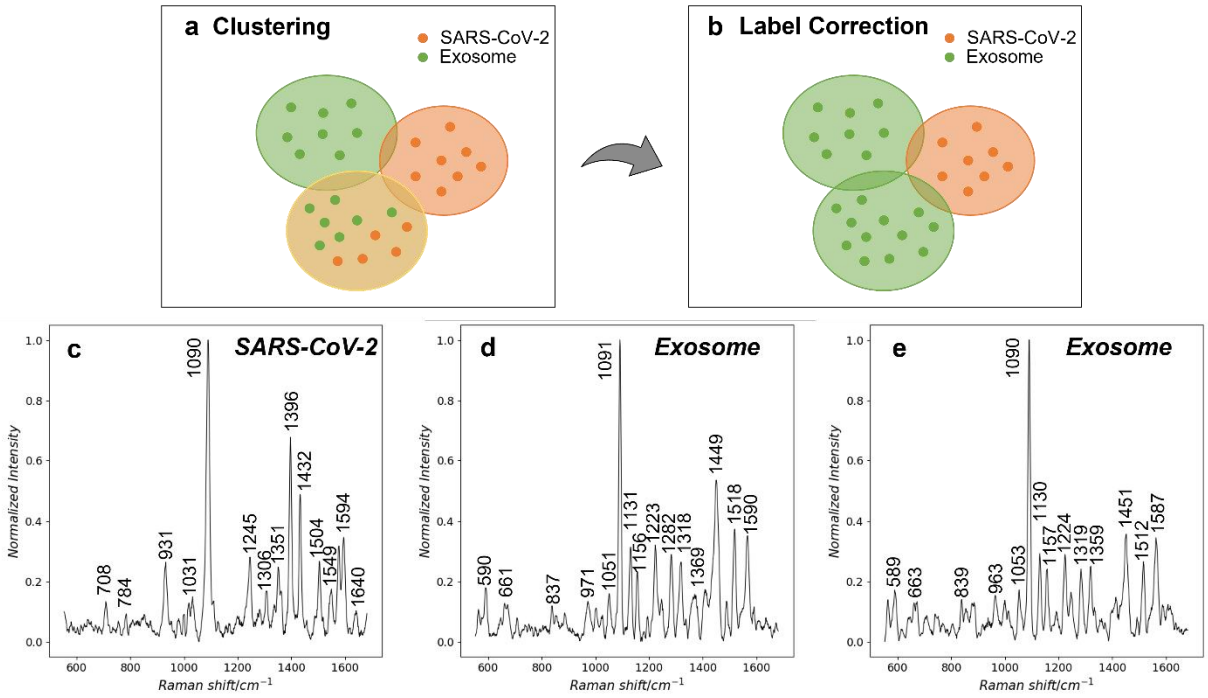



**Figure 8** HCA for correcting the mislabeled exosomes. a) Colored oval is the clusters generated by HCA. Those clusters mixed by SARS-CoV-2 and exosome denotes the existence of exosomes in SARS-CoV-2 specimen. b) Exosomes' labels in the mixed clusters are corrected. c), d), e) are three spectra attributed to different particles from the same cluster, where similar patterns are shown.

A binary classification model using support vector machines (SVM, RBF kernel, soft margin applied) was used in learning the characteristic fingerprints of SARS-CoV-1 and SARS-CoV-2. Due to the binary learning and predicting manner, the testing or validation spectra were

either recognized as SARS-CoV-1 or SARS-CoV-2, based on the relative population ratio of SARS-CoV-1 and SARS-CoV-2 for each sample. Without loss of generality, we chose SARS-CoV-2 percentages (e.g., 50 found among 200 thus, 40.0%) as the score. Considering the various viral concentrations and non-virus particles in the specimens, we assigned the binary labels to non-SARS-CoV-2 (or negative) and SARS-CoV-2 (or positive) to avoid confusion and applied a threshold to draw the boundary between the score of two types of virions. It is important to mention that the threshold was determined practically to maximize the cross-validation performance, also the sample threshold will be further applied or updated whenever more learning and predicting duties come.

During the training process, as more training instances are input, the model gradually learns the distinguishable features between the positive and the negative. Figure 9 shows the training error starts from 35% when 10% of the training process is done, and finally ends up with less than 5% after the training process is finished. Additionally, Figure 10 demonstrates a gradual separation between the scores of negative instances and positive instances.

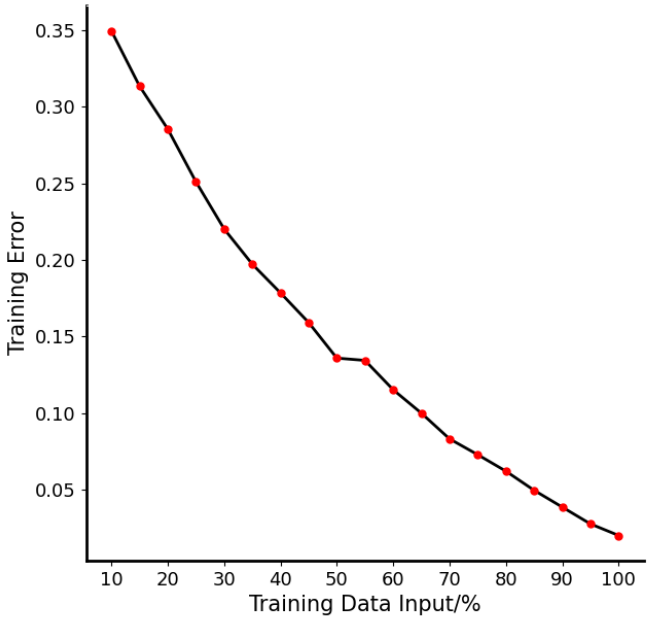



**Figure 9** Model training process; training error gradually decreases as training instances being input.

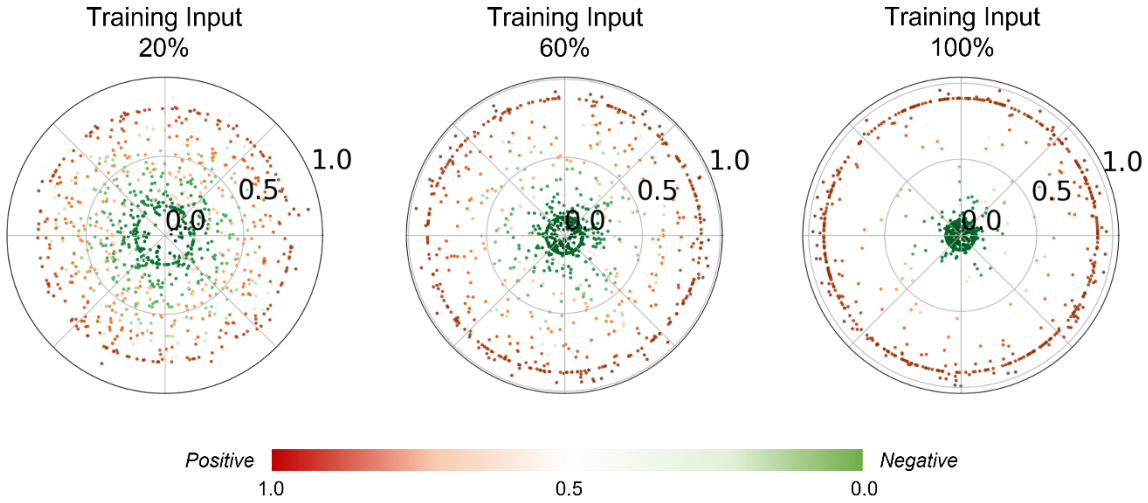



**Figure 10** Model training process; scores of negative and positive instances gradually segregate.

As stated before, we incorporated cross-validation for optimizing the classifier hyperparameters as well as choosing an appropriate threshold that generates the best predictive capability. Furthermore, to genuinely evaluate the predictive capability by alleviating the overfitting problem during validation, we applied 'leave pair of samples out' (LPSO) cross-validation. Demonstrated by figure 11, In each round of validation, a pair of samples, one each from positive and negative groups respectively, are left out as the validation set while the remaining are the training set. The 'pair' manner is to ensure the sample balance in both training and validation. This process continues until every sample is traversed once as the validation set. A score list for all the samples is built once the cross-validation is completed, then the ROC curve is plotted together with the information of the true labels by adjusting the threshold.

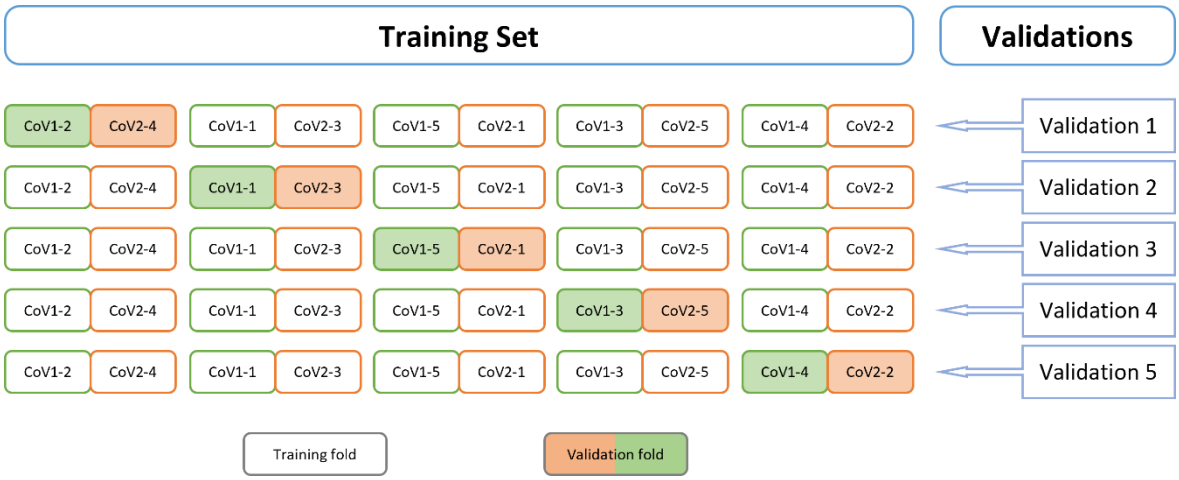



**Figure 11** cross validation; Five rounds of cross-validation are conducted; In each round, training folds (unfilled blocks) and validation folds (filled blocks) are assigned for training and validating respectively.

Following the above protocol, the ROC curve is calculated and shown in figure 12, which demonstrates an overall good pattern recognizing capability across all types of viruses.

Accordingly, the scores of the samples were shown in the box plot of chart 1, based on the statistical properties of each cross-validation round, we applied the mean of positive sample quantile Q1 and negative sample quantile Q3 as the threshold to maximize the 'margin'. Chart 3 shows the fluctuations of the threshold in cross validations. As indicated in table 1 and chart 1, a threshold of 0.300 was finalized which maximizes the average margin in cross-validations.

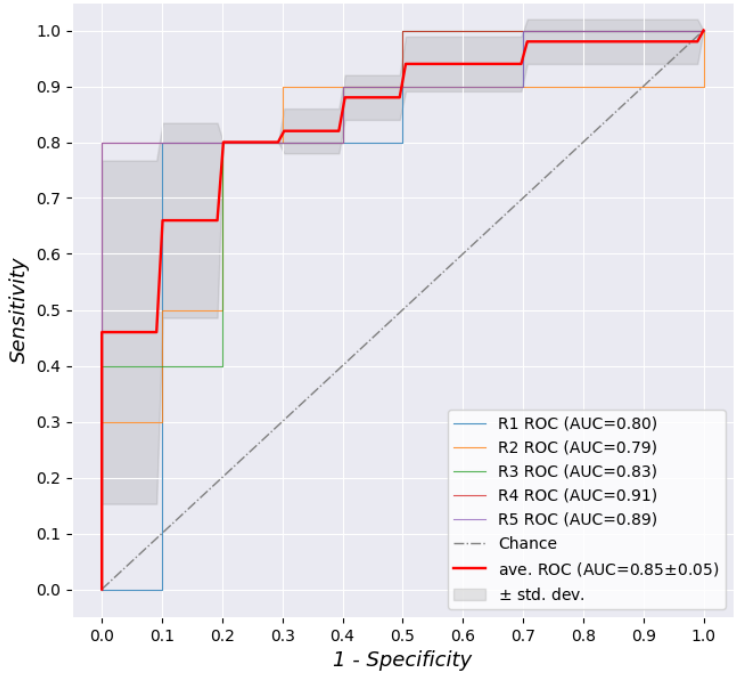



**Figure 12** Individual and mean ROC curves of cross validations

**Chart 1** sample scores (positive vesicle rate of a sample) distribution in the validation folds of cross validation rounds

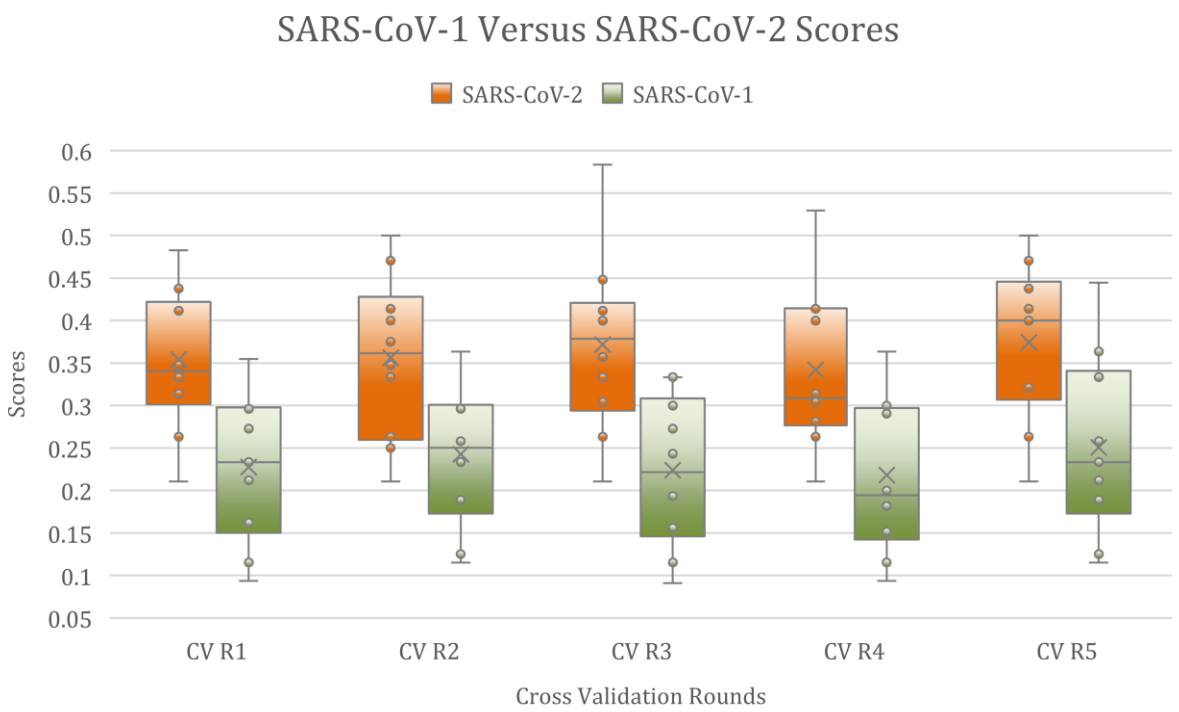
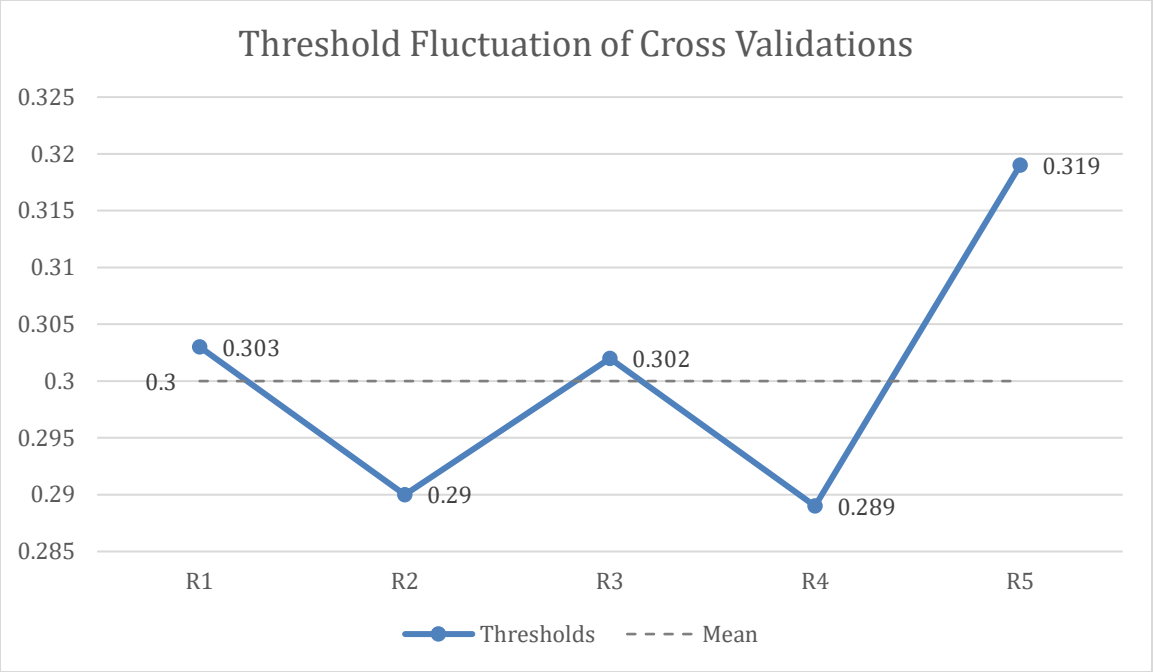



**Table 1** Q1 and Q3 values of cross validations

| ***Cross-validation*** | ***Non-SARS-CoV-2(Q3)*** | ***SARS-CoV-2(Q1)*** | ***Q1&Q3 Mean*** |
|---|---|---|---|
| R1 | 0.290 | 0.316 | 0.303 |
| R2 | 0.299 | 0.281 | 0.290 |
| R3 | 0.293 | 0.312 | 0.302 |
| R4 | 0.295 | 0.282 | 0.289 |
| R5 | 0.314 | 0.322 | 0.319 |
| AVE. | - | - | 0.300 |

**Chart 2** Fluctuations of threshold (mean of Q1 and Q3) versus cross validation rounds




A blind test is subsequently performed after the classification model is optimized. 5 SARS-CoV-2 virus specimens versus 5 SARS-CoV-1 virus specimens were blinded to be given predictions. Promising performance was given by the threshold equal to 30.0% and the sensitivity/specificity turned out to be 80%/80%. Table 2 shows the test results and chart 2 shows the positive ratio generated by the classifier.

This result combined with the LDA grouping demonstrates the feasibility of utilizing machine learning classifier and SERS to build a SARS-CoV-2 identifier, given that the specimen has a low diversity of the content (i.e., viruses and extracellular vesicles from Vero-TMPRSS2) and high viral load ($10^8$ – $10^{10}$ particles/mL).

We also evaluated the ambiguity of the classifier combined with the threshold by visualizing the sample score distribution. Except for the two incorrectly predicted samples, each correctly predicted sample has a fair distance from the decision threshold, which means our

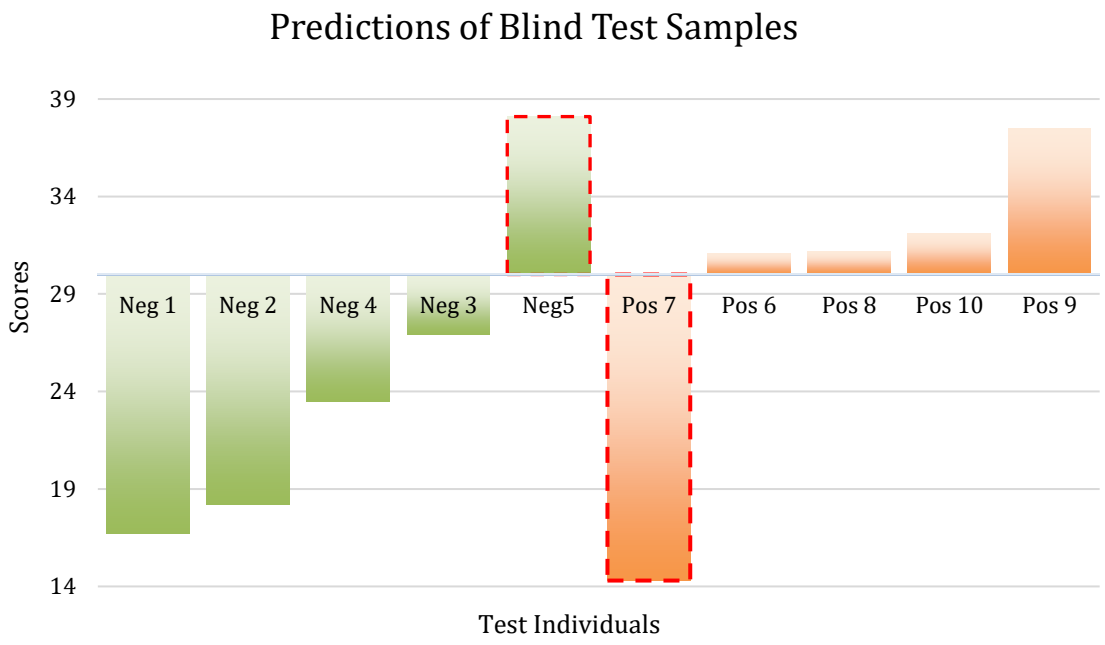
platform is able to maintain the original level of detection performance given a certain amount of fluctuation of the sample viral load as well as the model training.

**Table 2** Blind test results of SARS-CoV-1 versus SARS-CoV-2

| ***Sample ID*** | ***Negative*** | ***Positive*** | ***P.R.*** | ***Predictions*** | ***Ground truth*** |
|---|---|---|---|---|---|
| **1** | 50 | 10 | 16.7 | Non-CoV-2 | Non-CoV-2 |
| **2** | 54 | 12 | 18.2 | Non-CoV-2 | Non-CoV-2 |
| **3** | 38 | 14 | 26.9 | Non-CoV-2 | Non-CoV-2 |
| **4** | 39 | 12 | 23.5 | Non-CoV-2 | Non-CoV-2 |
| **5** | 39 | 24 | 38.1 | Cov-2 | Non-CoV-2 |
| **6** | 43 | 19 | 31.1 | Cov-2 | Cov-2 |
| **7** | 48 | 8 | 14.3 | Non-CoV-2 | Cov-2 |
| **8** | 33 | 15 | 31.2 | Cov-2 | Cov-2 |
| **9** | 40 | 24 | 37.5 | Cov-2 | Cov-2 |
| **10** | 38 | 18 | 32.1 | Cov-2 | Cov-2 |

Negative: predicted Non-SARS-CoV-2 particles; Positive: predicted SARS-CoV-2 particles; P.R.: Positive ratio (%)

**Chart 3** Sample scores of blind test in distinguishing SARS-CoV-1 versus SARS-CoV-2




### 3.3 Detection of SARS-CoV-2 in virus spiked saliva

Given the capability of identifying SARS-CoV-2, we further evaluated our SERS fingerprinting plus SVMs protocol on the specimens with higher biological content complexity and closer to the clinical specimens, i.e., virus spiked saliva samples. Specifically, we introduced SARS-CoV-2 virus spiked saliva samples and healthy controls saliva samples as negative control. The preparation protocol of virus spiked saliva samples is given in the Materials and Methods section. A new SVMs classifier was trained using 10 SARS-CoV-2 virus spiked saliva samples versus 10 healthy control saliva samples. Around 50 analytes are collected for each sample, therefore the training dataset is composed of 999 analytes with 9689 spectra.

Like the data cleaning step in SARS-CoV-1 and SARS-CoV-2 study, the non-SARS-CoV-2 particles were subtracted from the SARS-CoV-2 spiked saliva training set by finding the spectral signatures overlapping between healthy control and SARS-CoV-2 spiked saliva. HCA was again implemented in this background removal process. To ensure the objectivity of the

classification and avoid information leakage, background removal is only done to the training set, excluding both the validation set and blind test set. The training set compositions before and after background removal were compared and shown in figure 13.

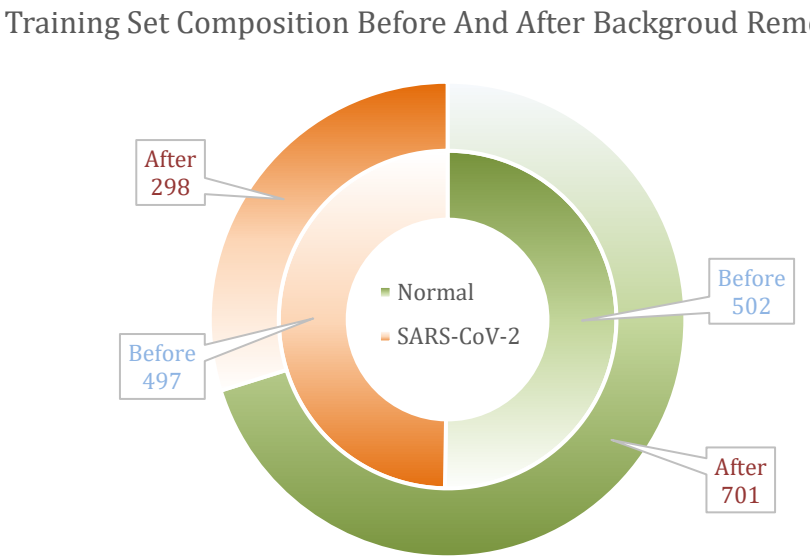



**Figure 13** Number of training instances before and after label correction by clustering analysis

Before launching into the blind test, LPSO cross-validation was done with SARS-CoV-2 spiked saliva (or positive) and healthy control (or negative) as the binary groups. As indicated by the ROC curve in figure 14, 0.83 AUC was achieved in cross-validation, which showed reasonable performance. As the previous cross validations, the statistical analyses of the sample scores of cross-validations were presented in chart 4 and table 3, and the mean of positive quantile Q1 and negative quantile Q3 was chosen as the threshold that maximizes the margin between the two types. Chart 5 shows the threshold fluctuation. The trained model by ten virus spiked saliva and ten healthy control individuals were used as classifier, together with a 0.259 as the score threshold.

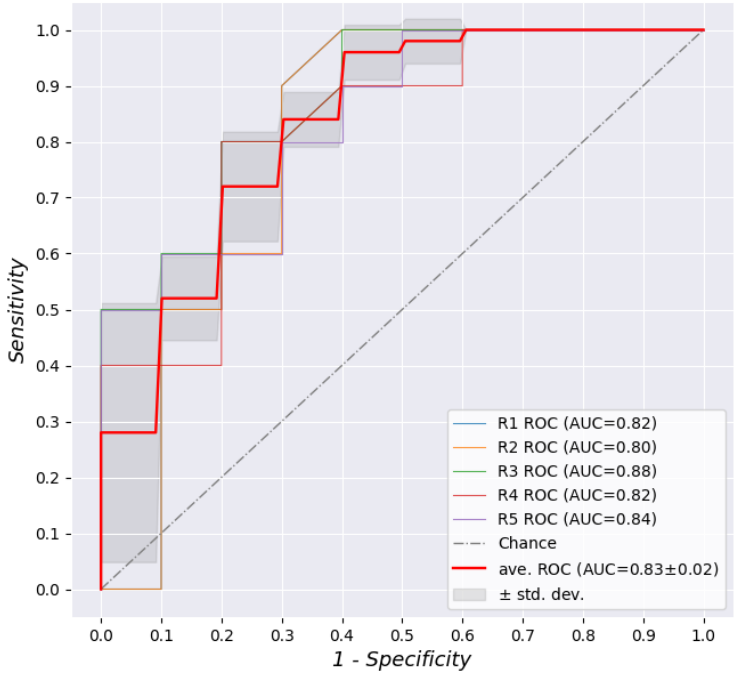



**Figure 14** Individual and mean ROC curves of cross validations

**Chart 4** sample scores (positive vesicle rate of a sample) distribution in the validation folds of cross validation rounds

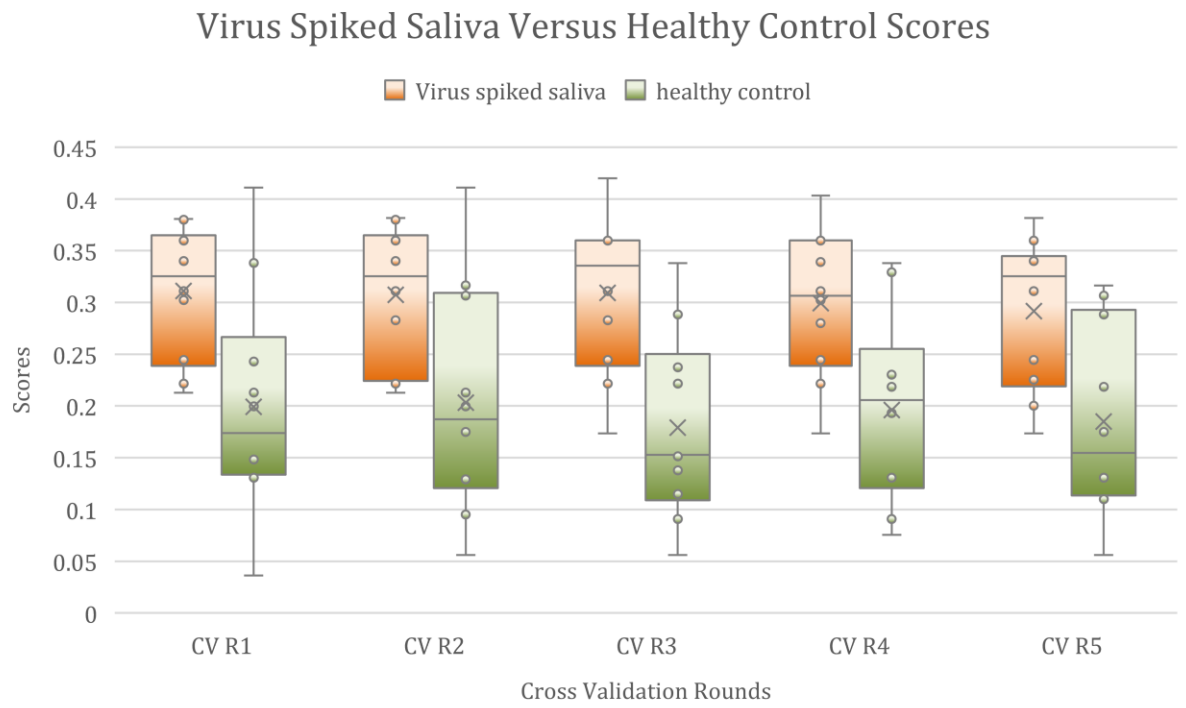
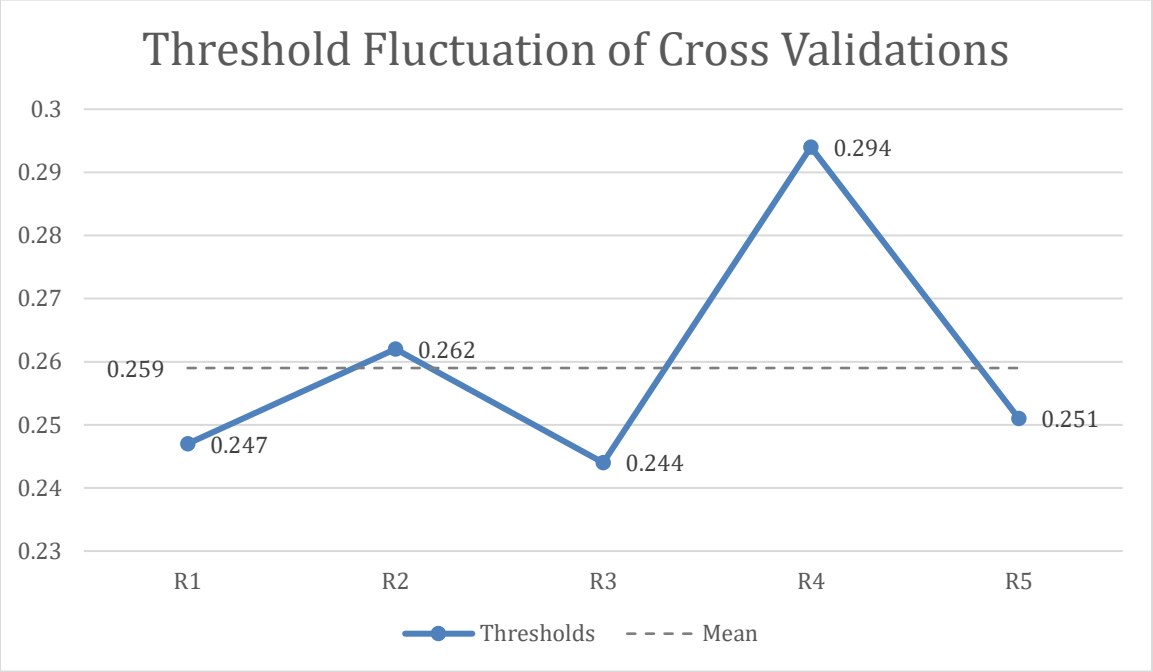



**Table 3** Q1 and Q3 values of cross validations

| ***Cross-validation*** | ***Virus Spiked Saliva*** | ***Healthy Control*** | ***Q1&Q3 Mean*** |
|---|---|---|---|
| R1 | 0.259 | 0.235 | 0.247 |
| R2 | 0.240 | 0.283 | 0.262 |
| R3 | 0.254 | 0.233 | 0.244 |
| R4 | 0.360 | 0.228 | 0.294 |
| R5 | 0.230 | 0.271 | 0.251 |
| AVE. | - | - | 0.259 |

**Chart 5** Fluctuations of threshold (mean of Q1 and Q3) versus cross validation rounds




Having trained the classifier, a blind test round with ten virus spiked saliva samples and ten healthy control saliva samples was then conducted. The virus spiked saliva samples were prepared following the same protocol as the cross-validation round, but with different healthy saliva backgrounds for mixing. This process is to simulate the various non-virus contents in human salivary specimens. The predictions and unblinding results are shown in table 4 and chart 6, and the corresponding decision matrix is presented in table 5. 90% sensitivity and 80% specificity were achieved with one virus spiked individual and two healthy control individuals predicted incorrectly. The blind test outcome indicates a reasonable performance while trying to apply our platform in diagnosis.

We do also notice some potential pitfall. First, samples 5, 16, 17, 20 are right at the threshold decision line as shown in chart 6, which decreases the robustness of the platform since the tolerance for statistical fluctuations is limited. Second, a blurrier decision boundary between

the positive/negative groups is present in the spiked saliva study compared to the virus in cell lysate study. This is demonstrated by the more positive/negative group scores overlapping, making it harder to draw an unambiguous decision boundary. The above potential pitfalls are due to the higher bioparticle complexity after spiking virus in the human salivary specimens. Therefore, decisive SARS-CoV-2 signatures are indispensable in improving the accuracy and robustness of our platform.

**Table 4** Blind test results of SARS-CoV-2 spiked saliva samples versus healthy control saliva samples

| ***Sample ID*** | ***Negative*** | ***Positive*** | ***P.R.*** | ***Predictions*** | ***Ground truth*** |
|---|---|---|---|---|---|
| **1** | 41 | 12 | 22.6 | Control | Control |
| **2** | 38 | 16 | 29.1 | Virus | Virus |
| **3** | 34 | 16 | 30.2 | Virus | Virus |
| **4** | 53 | 14 | 20.6 | Control | Control |
| **5** | 42 | 16 | 26.7 | Virus | Virus |
| **6** | 42 | 7 | 13.7 | Control | Control |
| **7** | 38 | 12 | 23.1 | Control | Control |
| **8** | 41 | 7 | 13.7 | Control | Virus |
| **9** | 25 | 11 | 29.7 | Virus | Control |
| **10** | 32 | 13 | 27.7 | Virus | Virus |
| **11** | 36 | 16 | 30.8 | Virus | Virus |
| **12** | 28 | 18 | 39.1 | Virus | Control |
| **13** | 35 | 10 | 22.2 | Control | Control |

| | | | | | |
|---|---|---|---|---|---|
| **14** | 35 | 15 | 30.0 | Virus | Virus |
| **15** | 38 | 11 | 22.4 | Control | Control |
| **16** | 37 | 13 | 26.0 | Virus | Virus |
| **17** | 37 | 13 | 26.0 | Virus | Virus |
| **18** | 36 | 13 | 26.5 | Virus | Virus |
| **19** | 40 | 11 | 21.6 | Control | Control |
| **20** | 35 | 12 | 25.5 | Control | Control |

**Chart 6** Sample scores of clinical test in distinguishing SARS-CoV-1 versus SARS-CoV-2

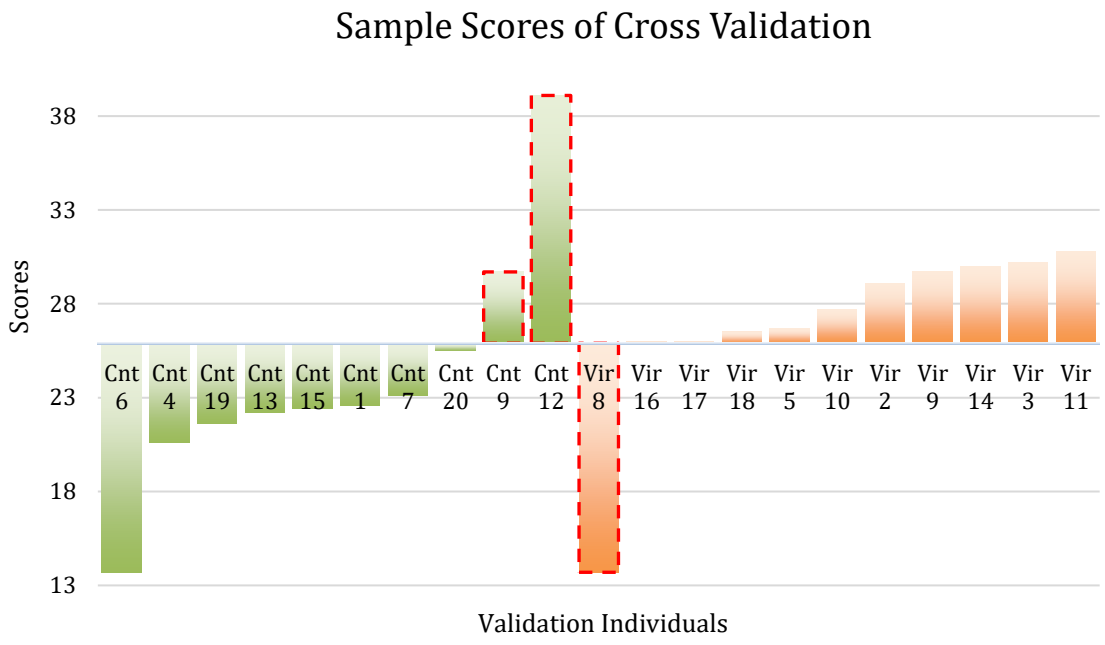



**Table 5** Confusion matrix of blind test with SARS-CoV-2 spiked saliva samples

| | ***Predicted Virus*** | ***Predicted Healthy Control*** |
|---|---|---|
| ***True Virus*** | 9 | 1 |

| | | |
|---|---|---|
| ***True Healthy Control*** | 2 | 8 |

## 3.4 Detection of SARS-CoV-2 in human saliva

All the aforementioned studies are the prerequisites for successfully utilizing our platform in clinical diagnosis. Both, the SARS-CoV-2 purified from Vero-TMPRSS2 cell media or SARS-CoV-2 spiked salivary specimens are simpler laboratory cases compared to the COVID patients' salivary specimens. Therefore, an additional test with clinical samples is necessary to evaluate the practical diagnostic capability.

Since SARS-CoV-2 spiked saliva samples can serve as a 'standard' repository for building the training set due to the presence of both SARS-CoV-2 virions and non-SARS-CoV-2 bioparticles (e.g., proteins, EVs), we applied the same trained classifier in the virus spiked saliva study based on the already proven predicting performance. The same threshold of 0.259 is used as well.

The detailed sample scores are shown in table 6 and chart 7. The final sensitivity and specificity turn out to be 100% and 80%, with only one healthy control predicted incorrectly. Among the correctly predicted samples, SN36's score is right at the decision boundary which will be sensitive to the whole training-predicting system, the remaining are clearly far from the decision boundary, as shown in chart 7. Even though the small test set might be prone to statistical fluctuations, the preliminary success presents a promising application of the SERS platform in SARS-CoV-2 diagnosis. Table 7 is the confusion matrix of the clinical test and figure 15 shows the corresponding ROC curve.

**Table 6** Results of blind test with clinical samples

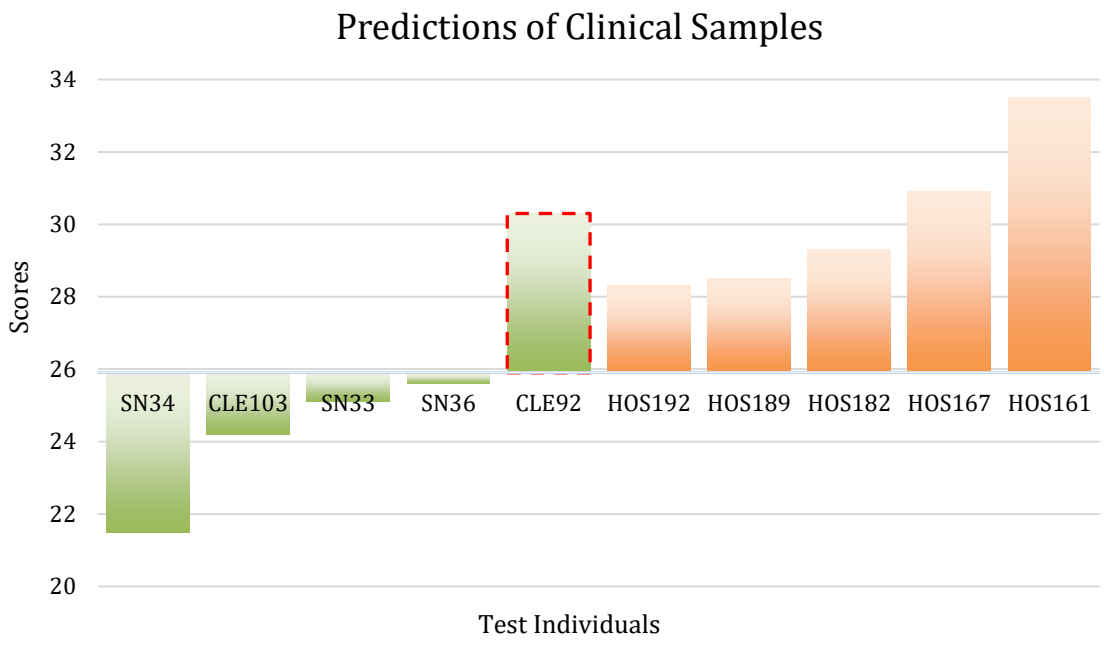
| ***Sample ID*** | ***Negative*** | ***Positive*** | ***P.R.*** | ***Predictions*** | ***Ground truth*** | ***Ct Value*** |
|---|---|---|---|---|---|---|
| CLE92 | 177 | 77 | 30.3 | Patient | Control | ND |
| CLE103 | 241 | 77 | 24.2 | Control | Control | ND |
| HOS192 | 190 | 75 | 28.3 | Patient | Patient | 33.43 |
| SN36 | 107 | 37 | 25.6 | Control | Control | ND |
| HOS167 | 306 | 137 | 30.9 | Patient | Patient | ND |
| HOS182 | 285 | 118 | 29.3 | Patient | Patient | 31.84 |
| SN33 | 137 | 46 | 25.1 | Control | Control | ND |
| HOS161 | 159 | 80 | 33.5 | Patient | Patient | 36.42 |
| HOS189 | 118 | 47 | 28.5 | Patient | Patient | 29.36 |
| SN34 | 244 | 67 | 21.5 | Control | Control | ND |

ND: Not detected

**Table 7** Confusion matrix of blind test with clinical samples

| | ***Predicted Virus*** | ***Predicted Healthy Control*** |
|---|---|---|
| ***True Virus*** | 5 | 0 |
| ***True Healthy Control*** | 1 | 4 |

**Chart 7** Sample scores of blind test in distinguishing SARS-CoV-1 versus SARS-CoV-2



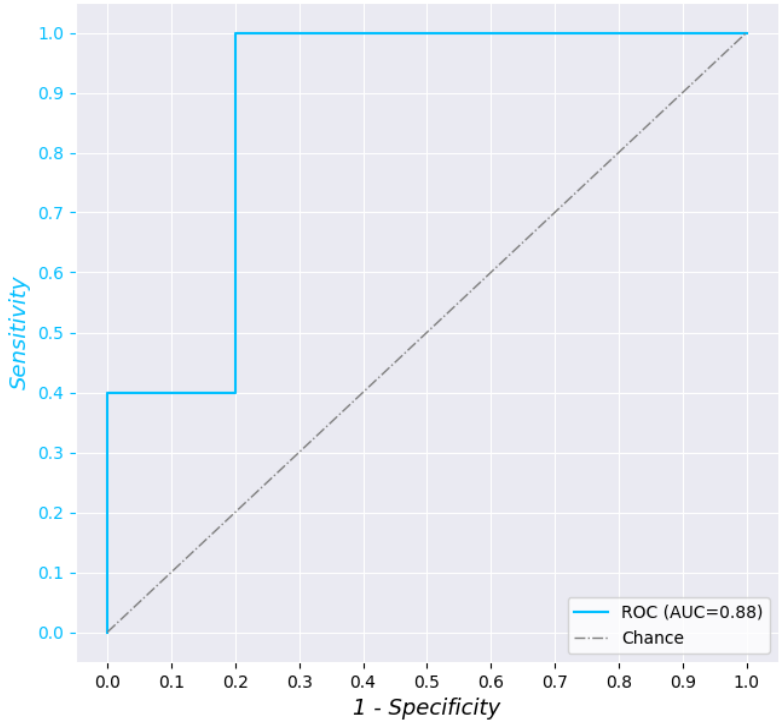




**Figure 15** ROC curve of clinical sample blind test

## 4. Discussion

In this study, we utilized support vector machine incorporated with Radial Basis Function (RBF) kernel and soft margin regularization. For the purposes of illustrating the fundamental working principle in identifying SARS-CoV-2 SERS spectral signatures, we consider the mathematical definition of the RBF and the training process under the hood. Within the RBF expression given in equation 1,

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\gamma \left\| \boldsymbol{x}_i - \boldsymbol{x}_j \right\|^2\right), \boldsymbol{x}_i, \boldsymbol{x}_j \; represent \; spectrum \tag{1}$$

Where $\gamma$ is a constant. The SERS spectrum term $\left\| \boldsymbol{x}_i - \boldsymbol{x}_j \right\|^2$ is recognized as the square of Euclidean distance. The exponential term allows for attenuation to assign a higher weight to closely separated training samples, and to normalize the original squared Euclidean distance to zero and one. Therefore, the SVM algorithm essentially searches for an optimal decision boundary that minimizes the intra-group distance score (given by the kernel function), and at the same time maximizes the inter-group distance score. Consequently, the fundamental principle is essentially to analyze the similarity represented by the spectral peak property, which is determined by the biochemical content of the analyte. The final classifier is trained to build a distinguishing criterion to identify SARS-CoV-2 presence versus other non-SARS-CoV-2 content such as SARS-CoV-1 or extracellular vesicles.

In addition to the working principle of support vector machine classifier, one more prerequisite for successful classification is the need for intra-SARS-CoV-2 group spectral differences to be less prominent than the ones between SARS-CoV-2 group and non-SARS-CoV-2 group. SARS-CoV-2 is believed to have developed many variants with slightly different components. Among our studies, Washington strain was used to prepare virus spiked saliva

samples while clinical samples were introduced without considering the mutant variant. The preliminary test performance provides indirect proof of our assumption.

Additionally, we translated the spectrum-level predictions given by the support vector machine classifier to a sample-level prediction by summarizing the instances belonging to each group. Then we chose a rather practical way to set up the decision boundary, which is based on cross-validation performance. The implicit reason is that we have quite limited knowledge about the viral load as well as the ratio of SARS-CoV-2 versus other particles. Fortunately, we could make the initial assumption that the genuine target (i.e., SARS-CoV-2) is present and only present in the virus spiked saliva specimens and patient specimens. Therefore, the positive group is bound to give higher score than the negative group as long as sufficient number of analytes are characterized, due to the presence of the extra distinct SARS-CoV-2 group compared with the control group. This initial conclusion ensures that we are able to find the approximate position of the decision threshold via 'big data strategy', which is the one that optimizes the validation performance including 20 specimens in our study. Correspondingly, the threshold contains the information on the implicit ratio of the target particles versus non-target particles. It is believed that a larger sample set is more advantageous to diagnostic accuracy.

## 5. Conclusion

In conclusion, we demonstrated the feasibility of applying SERS and machine learning pattern recognition on SARS-CoV-2 detection by harvesting and analyzing SARS-CoV-2 isolated from cell culture media and virus spiked saliva samples. Clinical testing with 5 patients versus 5 healthy controls was completed with only one false positive, rendering 100% sensitivity and 80% specificity.

In terms of the advantages of our platform, firstly the label-free manner in fingerprinting and identifying SARS-CoV-2 greatly simplifies the reagent, equipment, and specialist requirement. Our well-established SERS platform fabrication protocol and automatic Raman characterization allow for less human involvement. Therefore, a simpler COVID test procedure and lower cost test could be expected compared with RT-PCR. Additionally, like rapid antigen tests, the saliva-based specimen harvest protocol is fast and non-invasive. Virus isolation and purification are also not needed, which makes the preparation procedure for characterization simpler. The whole test duration using our platform is between 1-6 hours, mainly due to Raman scanning. Consequently, our platform offers a more accurate test performance than antigen test and a more rapid result yield than RT-PCR, those features could enable it to be a better pandemic monitoring technique.

Having demonstrated the feasibility in identifying SARS-CoV-2 Washington strain, SERS shows potential in contributing to distinguishing different variants. Multiclass classification will be conducted in place of binary classification. We have prepared multiple SARS-CoV-2 variants samples including B.1.351, B.1.1.7, BA.1, BA.5.1 etc. and are working on designing a supervised learning model appropriate to the multiclass classification task. Many algorithms have been reported to be efficient and accurate, such as Random Forest [39], K-nearest Neighbors [40], Neural Networks [41]. Foreseeing the challenges in differentiating SARS-CoV-2 variants with high similarity and the uniqueness of SERS spectrum, the collection of representative spectral data, the choice of classifier, model's parameters and even feature selections are supposed to be carefully organized.

As we mentioned, the clinical test sample size is small, which could only provide a preliminary indication of the potential of our platform's application for COVID tests. More

COVID patient samples are definitely required, and appropriate rounds of double-blind tests are needed to validate the feasibility. More importantly, due to training data consideration, the classifier is built mainly on simulated samples - SARS-CoV-2 spiked saliva samples. Model parameters might vary while we are using clinical sample data for the training. Another key metrics to evaluate a detection technology is the Limit of Detection, repetitive studies of samples with different viral loads have been planned. As a single particle characterization technique, a reliable throughput of data collection is needed to ensure the rate of capturing the target analyte. We are working on customizing the Raman spectrometer hardware and designing computer controlling software to enable automatic single particle characterization. All the above factors present challenges along the path of implementing SERS's advantages in COVID tests. Corresponding improvements and validations are being conducted.

## Author Contribution

Tieyi Li and Ya-Hong Xie designed the project. Tieyi Li carried out the experiments. Tieyi Li and Ya-Hong Xie had extensive discussions throughout the process of this work. Aaron Carlin prepared and isolated the SARS-CoV-2 samples. California NanoSystems Institute (CNSI) of UCLA took the SARS-CoV-2 image. Feng Li prepared the SARS-CoV-2 spiked saliva samples and healthy control saliva samples. Tieyi Li developed and customized the data analyses. Jun Liu helped SERS substrate fabrication. Tieyi Li and Ya-Hong Xie discussed and co-wrote the manuscript.

## Conflict of Interest

There are no conflicts to declare.

## Acknowledgements

This work was supported by the National Center for Advancing Translational Sciences at the National Institutes of Health under award numbers U18TR003778-01. We thank Laser Spectroscopy Labs at University of California, Los Angeles for granting the access to Raman spectroscopy machine and CNSI of UCLA for conducting the TEM imaging.

## References

1. “Who Coronavirus (COVID-19) Dashboard.” World Health Organization, World Health Organization, https://covid19.who.int/.
2. Lopez-Leon, Sandra, et al. "More than 50 long-term effects of COVID-19: a systematic review and meta-analysis." *Scientific reports* 11.1 (2021): 1-12.
3. Vasireddy, Deepa, et al. "Review of COVID-19 variants and COVID-19 vaccine efficacy: what the clinician should know?" *Journal of Clinical Medicine Research* 13.6 (2021): 317.
4. Grewal, Ramandip, et al. "Effectiveness of a fourth dose of covid-19 mRNA vaccine against the omicron variant among long term care residents in Ontario, Canada: test negative design study." *bmj* 378 (2022).
5. Adjei, Stacey. "Mortality Risk Among Patients Hospitalized Primarily for COVID-19 During the Omicron and Delta Variant Pandemic Periods—United States, April 2020–June 2022." *MMWR. Morbidity and Mortality Weekly Report* 71 (2022).
6. Challen, Robert, et al. "Risk of mortality in patients infected with SARS-CoV-2 variant of concern 202012/1: matched cohort study." *bmj* 372 (2021).

7. Araf, Yusha, et al. "Omicron variant of SARS-CoV-2: genomics, transmissibility, and responses to current COVID-19 vaccines." *Journal of medical virology* 94.5 (2022): 1825-1832.
8. Chau, Cindy H., Jonathan D. Strope, and William D. Figg. "COVID-19 clinical diagnostics and testing technology." *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy* 40.8 (2020): 857-868.
9. Chung, Yoon-Seok, et al. "Validation of real-time RT-PCR for detection of SARS-CoV-2 in the early stages of the COVID-19 outbreak in the Republic of Korea." *Scientific Reports* 11.1 (2021): 1-8.
10. Freeman, Willard M., Stephen J. Walker, and Kent E. Vrana. "Quantitative RT-PCR: pitfalls and potential." *Biotechniques* 26.1 (1999): 112-125.
11. Tahamtan, Alireza, and Abdollah Ardebili. "Real-time RT-PCR in COVID-19 detection: issues affecting the results." *Expert review of molecular diagnostics* 20.5 (2020): 453-454.
12. Yamayoshi, Seiya, et al. "Comparison of rapid antigen tests for COVID-19." *Viruses* 12.12 (2020): 1420.
13. Wang, Pu, et al. "Ultra-sensitive graphene-plasmonic hybrid platform for label-free detection." *Advanced Materials* 25.35 (2013): 4918-4924.
14. Kneipp, Katrin. "Chemical contribution to SERS enhancement: an experimental study on a series of polymethine dyes on silver nanoaggregates." *The Journal of Physical Chemistry C* 120.37 (2016): 21076-21081.
15. El Amri, Chahrazade, Marie-Hélène Baron, and Marie-Christine Maurel. "Adenine and RNA in mineral samples.: Surface-enhanced Raman spectroscopy (SERS) for picomole

detections." *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 59.11 (2003): 2645-2654.

16. Bruzas, Ian, et al. "Advances in surface-enhanced Raman spectroscopy (SERS) substrates for lipid and protein characterization: sensing and beyond." *Analyst* 143.17 (2018): 3990-4008.
17. Kneipp, Katrin, et al. "Detection and identification of a single DNA base molecule using surface-enhanced Raman scattering (SERS)." *Physical Review E* 57.6 (1998): R6281.
18. Bell, Steven EJ, and Narayana MS Sirimuthu. "Surface-enhanced Raman spectroscopy (SERS) for sub-micromolar detection of DNA/RNA mononucleotides." *Journal of the American Chemical Society* 128.49 (2006): 15580-15581.
19. Lin, Chenglong, et al. "Visualized SERS imaging of single molecule by Ag/black phosphorus nanosheets." Nano-Micro Letters 14.1 (2022): 75.
20. Palonpon, Almar F., et al. "Raman and SERS microscopy for molecular imaging of live cells." *Nature protocols* 8.4 (2013): 677-692.
21. Mosier-Boss, Pamela A. "Review on SERS of Bacteria." *Biosensors* 7.4 (2017): 51.
22. Luo, Shyh-Chyang, et al. "Nanofabricated SERS-active substrates for single-molecule to virus detection in vitro: A review." *Biosensors and Bioelectronics* 61 (2014): 232-240.
23. Yaraki, Mohammad Tavakkoli, Anastasiia Tukova, and Yuling Wang. "Emerging SERS biosensors for the analysis of cells and extracellular vesicles." *Nanoscale* (2022).
24. Chen, Hao, et al. "SERS imaging-based aptasensor for ultrasensitive and reproducible detection of influenza virus A." *Biosensors and Bioelectronics* 167 (2020): 112496.

25. Kamińska, Agnieszka, et al. "Detection of Hepatitis B virus antigen from human blood: SERS immunoassay in a microfluidic system." *Biosensors and bioelectronics* 66 (2015): 461-467.
26. Shanmukh, Saratchandra, et al. "Rapid and sensitive detection of respiratory virus molecular signatures using a silver nanorod array SERS substrate." *Nano letters* 6.11 (2006): 2630-2636.
27. Sharma, Bhavya, et al. "SERS: Materials, applications, and the future." *Materials today* 15.1-2 (2012): 16-25.
28. Chen, Hao, et al. "Sensitive detection of SARS-CoV-2 using a SERS-based aptasensor." *ACS sensors* 6.6 (2021): 2378-2385.
29. Stelzer-Braid, Sacha, et al. "Virus isolation of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) for diagnostic and research purposes." *Pathology* 52.7 (2020): 760-763.
30. Bar-On, Yinon M., et al. "Science Forum: SARS-CoV-2 (COVID-19) by the numbers." *elife* 9 (2020): e57309.
31. Zhang, Pan, Joo Chuan Yeo, and Chwee Teck Lim. "Advances in technologies for purification and enrichment of extracellular vesicles." *SLAS TECHNOLOGY: Translating Life Sciences Innovation* 24.5 (2019): 477-488.
32. Pastorino, Boris, et al. "Heat inactivation of different types of SARS-CoV-2 samples: what protocols for biosafety, molecular detection and serological diagnostics?." Viruses 12.7 (2020): 735.
33. Biasin, Mara, et al. "UV-C irradiation is highly effective in inactivating SARS-CoV-2 replication." Scientific Reports 11.1 (2021): 6260.

34. Stiles, Paul L., et al. "Surface-enhanced Raman spectroscopy." *Annu. Rev. Anal. Chem.* 1 (2008): 601-626.
35. Wang, Pu, et al. "Ultra-sensitive graphene-plasmonic hybrid platform for label-free detection." *Advanced Materials* 25.35 (2013): 4918-4924.
36. Peng, Jiangtao, et al. "Asymmetric least squares for multiple spectra baseline correction." *Analytica chimica acta* 683.1 (2010): 63-68.
37. John, Arlene, Jishnu Sadasivan, and Chandra Sekhar Seelamantula. "Adaptive Savitzky-Golay filtering in non-Gaussian noise." *IEEE Transactions on Signal Processing* 69 (2021): 5021-5036.
38. Cai, Zena, et al. "Identification and characterization of circRNAs encoded by MERS-CoV, SARS-CoV-1 and SARS-CoV-2." *Briefings in bioinformatics* 22.2 (2021): 1297-1308.
39. Chaudhary, Archana, Savita Kolhe, and Raj Kamal. "An improved random forest classifier for multi-class classification." Information Processing in Agriculture 3.4 (2016): 215-222.
40. Haixiang, Guo, et al. "BPSO-Adaboost-KNN ensemble learning algorithm for multi-class imbalanced data classification." Engineering Applications of Artificial Intelligence 49 (2016): 176-193.
41. Lin, Minlong, Ke Tang, and Xin Yao. "Dynamic sampling approach to training neural networks for multiclass imbalance classification." IEEE Transactions on Neural Networks and Learning Systems 24.4 (2013): 647-660.