


## RESEARCH ARTICLE

# An efficient primary screening of COVID-19 by serum Raman spectroscopy

Gang Yin<sup>1,2</sup>  | Lintao Li<sup>1,2</sup> | Shun Lu<sup>1,2</sup> | Yu Yin<sup>3</sup> | Yuanzhang Su<sup>4</sup> | Yilan Zeng<sup>5</sup> | Mei Luo<sup>5</sup> | Maohua Ma<sup>5</sup> | Hongyan Zhou<sup>1,2</sup> | Lucia Orlandini<sup>1,2</sup> | Dezhong Yao<sup>3</sup> | Gang Liu<sup>6</sup> | Jinyi Lang<sup>1,2</sup>

<sup>1</sup>Department of Radiation Oncology, Sichuan Cancer Hospital & Institute, Chengdu, China

<sup>2</sup>Physical Engineering Laboratory, Radiation Oncology Key Laboratory of Sichuan Province, Chengdu, China

<sup>3</sup>Sichuan Institute for Brain Science and Brain-Inspired Intelligence, MOE Key Lab for Neuroinformation, University of Electronic Science and Technology of China, Chengdu, China

<sup>4</sup>School of Foreign Languages, University of Electronic Science and Technology of China, Chengdu, China

<sup>5</sup>Clinical Laboratory, The Public Health Clinical Center of Chengdu, Chengdu, China

<sup>6</sup>Department of Clinical Laboratory, The First Affiliated Hospital of Chengdu Medical College, Chengdu, China

## Correspondence

Jinyi Lang

Department of Radiation Oncology,  
Sichuan Cancer Hospital & Institute,  
Sichuan Cancer Center, Chengdu, China.  
Email: langjy610@163.com

Dezhong Yao

Sichuan Institute for Brain Science and  
Brain-Inspired Intelligence, MOE Key Lab  
for Neuroinformation, University of  
Electronic Science and Technology of  
China, Chengdu, China.

Email: dyao@uestc.edu.cn

Gang Liu

The First Affiliated Hospital of Chengdu  
Medical College, Chengdu, China.

Email: lee\_scch@yeah.net

## Funding information

Department of Science and Technology of  
Sichuan Province, Grant/Award Number:  
2019YFG0185; Chengdu Science and  
Technology Bureau, Grant/Award  
Number: 2020-YF05-0014-SN

## Abstract

The outbreak of COVID-19 coronavirus disease around the end of 2019 has become a pandemic. The preferred method for COVID-19 detection is the real-time polymerase chain reaction (RT-PCR)-based technique; however, it also has certain limitations, such as sample-dependent procedures with a relatively high false negative ratio. We propose a safe and efficient method for screening COVID-19 based on Raman spectroscopy. A total of 177 serum samples are collected from 63 confirmed COVID-19 patients, 59 suspected cases, and 55 healthy individuals as a control group. Raman spectroscopy is adopted to analyze these samples, and a machine learning support-vector machine (SVM) method is applied to the spectrum dataset to build a diagnostic algorithm. Furthermore, 20 independent individuals, including 5 asymptomatic COVID-19 patients and 5 symptomatic COVID-19 patients, 5 suspected patients, and 5 healthy patients, were sampled for external validation. In these three groups—confirmed COVID-19, suspected, and healthy individuals—the distribution of statistically significant points of difference showed highly consistency for intergroups after repeated sampling processes. The classification accuracy between the COVID-19 cases and the suspected cases is 0.87 (95% confidence interval [CI]: 0.85–0.88), and the accuracy between the COVID-19 and the healthy controls is 0.90 (95% CI: 0.89–0.91), while the accuracy between the suspected cases and the healthy control group is 0.68 (95% CI: 0.67–0.73). For the independent test

Gang Yin, Lintao Li, and Shun Lu contributed equally to this manuscript.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The Authors. *Journal of Raman Spectroscopy* published by John Wiley & Sons Ltd.

dataset, we apply the obtained SVM model to the classification of the independent test dataset to have all the results correctly classified. Our model showed that the serum-level classification results were all correct for independent test dataset. Our results suggest that Raman spectroscopy could be a safe and efficient technique for COVID-19 screening.

#### KEYWORDS

COVID-19, machine learning, Raman spectroscopy, screening, support vector machine

## 1 | INTRODUCTION

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), causing COVID-19 coronavirus disease in 2019, became a pandemic.<sup>[1]</sup> Clinical manifestation of COVID-19 is highly nonspecific, including fever, fatigue, cough, myalgia, dyspnea, and headache.<sup>[2]</sup> Unfortunately, no effective treatment was available, until now. It already outpaced SARS in 2003 and Middle East Respiratory Syndrome (MERS) in 2012 in incidence and mortality.<sup>[2,3]</sup> The rapid and accurate detection of viruses is of great importance in controlling SARS-CoV-2 spread. Currently, a positive for the SARS-CoV-2 viral nucleic acid by polymerase chain reaction based on the technique or deep sequencing of specimens from the respiratory tract or serum is confirmed for COVID-19.<sup>[4,5]</sup> However, false negative results for patients with insufficient amounts of the virus at the site of sample collection, causing the COVID-19 patients to spread the virus, have become a major setback in containing viral transmission.<sup>[6]</sup> Therefore, it is of practical significance to develop a safe and efficient diagnostic method.

Raman spectroscopy is a method that detects molecular vibration and molecular rotation energy levels. It detects components on the biomolecule level with a high sensitivity for distinguishing vibrations and conformations of proteins, peptides, and nucleic acid.<sup>[7]</sup> The characteristic peaks in Raman spectrum are called “fingerprints,” representing the biochemical composition of the sample, and the noncontact detection process is fast, repeatable, specific, and highly sensitive.<sup>[8]</sup> Although the spectrum of SARS-CoV-2 is not revealed, the activation of a humoral response by the virus may induce a series of immune reactions, releasing the cytokines and resulting in the change of spectrum in the serum.<sup>[9]</sup> Machine learning has proven efficient for the data analysis of the complicated change of Raman spectrum in samples. Because it is a simple process of sample preparation and it is free of sample loss, Raman spectroscopy is highly recommended for diagnosing a diverse range of diseases, including dengue fever,

hepatitis C virus, cancer, HIV, Alzheimer's disease, and endometriosis.<sup>[10–16]</sup> Raman can also be used to detect RNA viruses.<sup>[17]</sup>

In this study, we screen the serum from COVID-19 patients, suspected cases, and healthy patients (control) by Raman spectroscopy to evaluate its detection capability.

## 2 | MATERIALS AND METHOD

### 2.1 | Study objects

The study protocol was in accordance with principles of the Declaration of Helsinki and approved by the Ethics Committee of Sichuan Cancer Hospital & Institute (Chengdu, China). None of the authors had access to information that could identify individual participants during or after data collection. In this study, all blood samples were taken from February 10 to May 10, 2020.

The COVID-19 group contained 63 patients who were recruited at the Chengdu Public Health Clinical Medical Center, including 58 symptomatic patients and 5 asymptomatic COVID-19 patients. All of the COVID-19 patients were positive using viral nucleic acid by real-time polymerase chain reaction (RT-PCR) detection with specimens from the respiratory tract. For the 58 symptomatic patients, the median ( $\pm$  standard deviation [SD]) interval time between symptom onset and sampling is 3.0 ( $\pm$ 4.2) days. The suspected group contained 59 patients with flu symptoms similar to COVID-19. All of the investigated “suspected” patients were also isolated in the COVID-19 designated hospital until they were confirmed negative for COVID-19. All of them showed negative viral nucleic acid tests at least twice by RT-PCR detection, and the interval time between the two detections is less than 1 week. This group of patients was followed up until no COVID-19 patients were found in this group. The healthy control group consisted of 55 healthy individuals recruited at

**TABLE 1** Clinical characteristics of the investigated individuals

Characteristic	Characteristic	COVID-19		Suspected	Healthy control
		Symptomatic	Asymptomatic		
Total		58	5	59	55
Age (media, range)		47.6 (20–78)		45.8 (21–74)	45.5 (24–65)
Gender					
Male		26	2	36	24
Female		32	3	23	31
Distribution of temperature (blood sampling)					
<37.5°C		27	5	13	5
37.5–38.0°C		19		17	
38.1–39.0°C		7		21	
>39.0°C		5		8	
Symptoms					
Cough		36		24	
Fatigue		21		16	
Myalgia or arthralgia		9		6	
Headache		11		20	
Shortness of breath		6		3	
Disease severity					
Nonsevere		51		58	
Severe		7		1	
Abnormalities on chest CT					
Ground-glass opacity		29		19	
Local patchy shadowing		21		34	
Bilateral patchy shadowing		28		17	
Interstitial abnormalities		9		17	

Abbreviation: CT, computed tomography.

the Sichuan Cancer Hospital and Chengdu Public Health Clinical Medical Center. General information for all of the investigated individuals, such as gender, age, clinical characteristics, and detailed information, is shown in Table 1.

All of the enrolled individuals gave informed consent to the research purpose and have signed the informed consent. The project was approved by the ethics committee of the participating institutes.

## 2.2 | Sample preparation

Blood samples were taken from the COVID-19 patients and the suspected cases upon admission from February 10 to May 10, 2020. One-hour repose of blood sampling, the serum was isolated from blood samples by centrifuging at 3000 rpm for 10 min. All the serum samples were

stored at 4°C and measured within 36 h after the collection. For the measurement, approximately 0.5 ml of the serum sample was prepared in cryopreservation tubes (specification: 2 ml; material: polypropylene) and strictly sealed for the Raman scan. Additional spectra data were also collected from cryopreservation tubes with saline solution inside.

## 2.3 | Experimental setup

The system consists of a volume-phase holographic (VPH) spectrograph, deep-cooled CCD camera, Raman probe, and laser, designed by the Sichuan Institute for Brain Science and Brain-Inspired Intelligence, and specific hardware composition and parameters; see Supporting Information S1. A single-mode diode laser (real-light) with wavelength 785-nm and 100-mW power

was used for Raman excitation. The laser power on the sample was detected around 70 mW. The spectra were recorded in the range of 600–1800  $\text{cm}^{-1}$ . The Raman spectra were collected 15 times per sample, with 3-s accumulation taken at each sample. Collecting multiple spectra per sample ensures an accurate representation of the heterogeneous composition of a sample.

First, before each measurement, the ethanol spectrum was measured using an exposure time of 3 s for the wavenumber calibration. Second, the cryopreservation tubes with a 5% normal saline spectrum were acquired using an exposure time of 3 s with five successive scans for every beginning and completion of the experiment. The average spectrum of the cryopreservation tubes was used to investigate the influence of tube wall materials. Next, the Raman spectrum of the serum samples sealed within the cryopreservation tube was measured using the same integration parameters as the cryopreservation tube measurements. Three experimenters took the Raman scan for each sample tube and repeated it five times. The cryopreservation tube was placed in the special card slot of the Raman spectrometer, ensuring the laser passes through the tube wall at a certain angle. After cosmic ray removal from the spectral data, we had 15 scans conducted by each experimenter of each serum sample.

## 2.4 | Data processing steps

A total of 2355 spectra from 157 individuals were subjected to preprocessing steps, including smoothing by automatic-weighted least squares, baseline correction based on polynomial fitting, and normalization by total area.<sup>[18]</sup> These spectra data were used for feature selection and to build the classification model. Analysis of variance (ANOVA) statistical analysis was used to select relevant features included in the training of support-vector machine (SVM) models.

The model predicts unknown samples and classifies them accordingly. The performance of the proposed model has been evaluated using a cross-validation method, dividing the whole dataset into 70% for training and the remaining 30% for testing. To test the robustness of the model, we conducted a “blind” validation. Here, 70% of the random sampling of the total sample was used to establish the model, and the remaining 30% of the samples as hold-out set for SVM model test. To ensure the independence of the data, the random sampling process guaranteed that the spectra data were used to establish the model and for model test from completely different samples.<sup>[19]</sup> The aforementioned process was repeated 50 times.

Furthermore, 20 independent individuals, including 5 asymptomatic and 5 symptomatic COVID-19 patients, 5 suspected patients, and 5 healthy controls, were sampled for model test. These 20 serum samples (corresponding to 300 spectra) were set aside to create an independent external hold-out dataset. After the prediction model was built, the 20 serum samples were preprocessed in the same way and used to validate the SVM model. The detailed procedure of data analysis is provided in Supporting Information S2.

## 3 | RESULTS

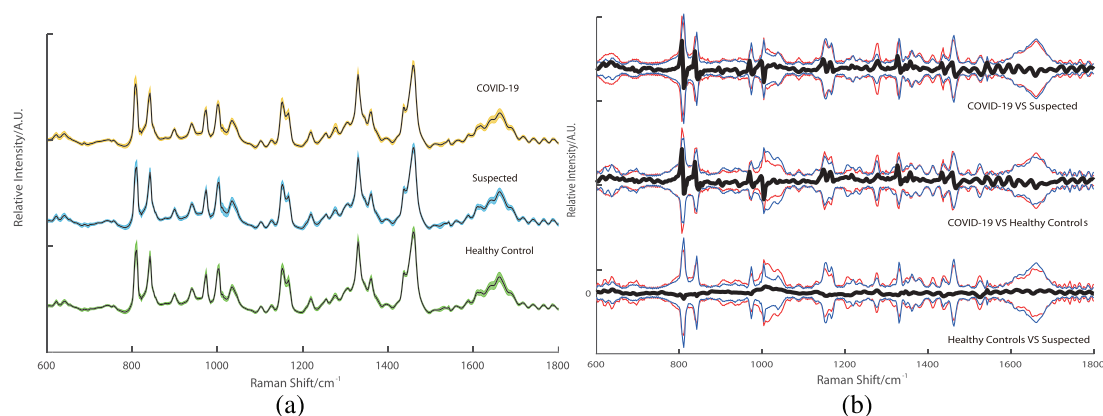
### 3.1 | Clinical characteristics

Among all of the COVID-19 patients, 55.6% are female, and the median age is 47.6 years (range: 20–78). In the suspected group, 38.9% are female, and the median age is 45.8 years (range: 21–74). Moreover, 31 (56.4%) participants in the healthy control group are female, and the median age is 45.5 years (range: 24–65). No statistically significant difference was observed between these three groups in terms of gender and age.

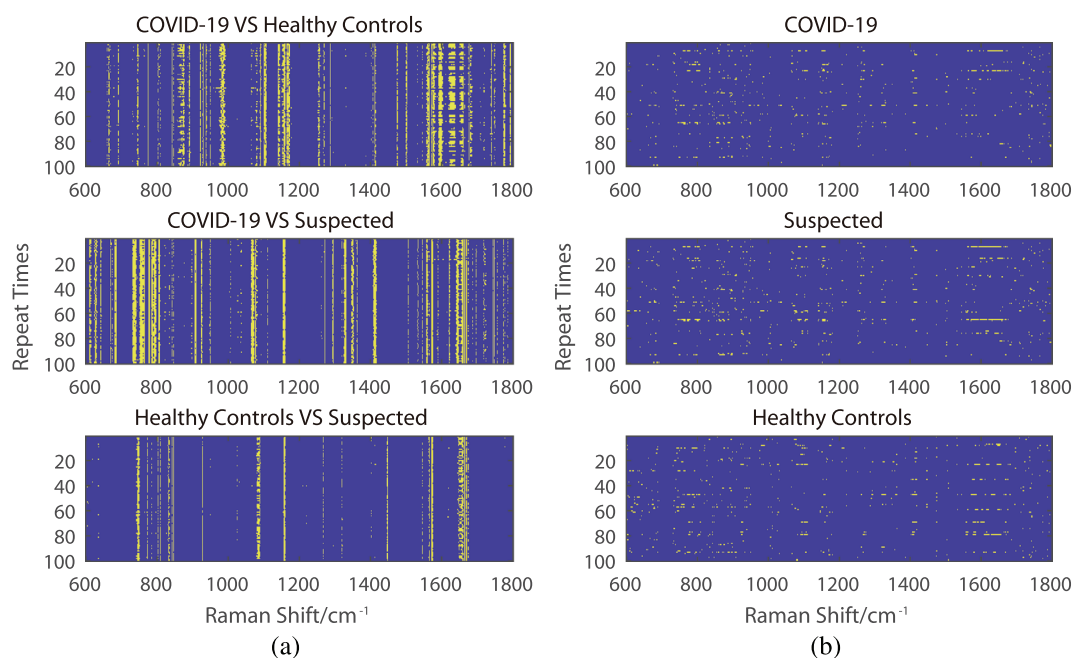
The most common symptom is a cough (62.1%) in symptomatic COVID-19 patients upon admission, followed by a fever (53.4%). A fever is present in 78.0% of all of the suspected cases upon admission, and the second most prevalent symptom is a cough (40.1%). The computed tomography (CT) scans were performed upon admission for all COVID-19 patients and the suspected cases. In the COVID-19 group, 76.5% showed abnormal results. The most common patterns on the chest CT were ground-glass opacity (50.0%) and bilateral patchy shadowing (48.3%). In the suspected group, 65.4% of the patients showed abnormal results, the most common patterns on chest CT were local patchy shadowing (57.7%) and ground-glass opacity (32.2%). According to our follow-up data, these 59 suspected cases include 13 cases of flu, 6 cases of bacterial infections, and 40 cases of unknown causes.

### 3.2 | Raman spectra and statistical analysis

Figure 1a shows the average of each group's preprocessed spectra. The differences among the average spectrum of the COVID-19 group, the healthy control group, and the suspected group are depicted in Figure 1b. The difference in the mean spectrum is shown within  $\pm 2$  SDs, suggesting that the mean difference between the groups is statistically insignificant. To build the diagnostic



**FIGURE 1** The total average serum Raman of the three groups and the difference between the groups. (a) The total average of the three types of Raman, the color band represents the standard deviation. (b) The Raman difference signal between the groups (black) and the Raman signal of the groups between  $\pm 2$  standard deviations (red and blue)



**FIGURE 2** The result of the ANOVA test. The spectra range without a significant difference in the ANOVA test ( $p < 0.05$ ) was indicated in blue, while others were indicated in yellow. (a) The Raman shift spectrum of the difference in  $p$  value after a 70% random sampling and repeated training 100 times for the intergroups. (b) The Raman shift spectrum of the difference in  $p$  value after a 70% random sampling and repeated training 100 times intragroups

algorithms, multivariate analysis is required to uncover useful variability within the individual spectrum.

An ANOVA test was applied for both the intergroup and intragroup after random sampling, and it was repeated 100 times. In these three groups, the distribution of statistically significant points of difference shows a high consistency for intergroups after repeating the sampling processes. Nevertheless, the distribution consistency of statistically significant points is not found for the intragroup.

After the ANOVA test, the difference between the COVID-19 group, the suspected group, and the healthy control group was in the spectra range of  $600\text{--}1800\text{ cm}^{-1}$  and is observed in Figure 2a. The differences between the healthy control group and the suspected group shown in the spectra range are significant less than those between the COVID-19 group and the suspected or healthy control group. However, for the intergroup, the ANOVA test result had no apparent consistency, similar to random noise (Figure 2b).



### 3.3 | Training and validation of model

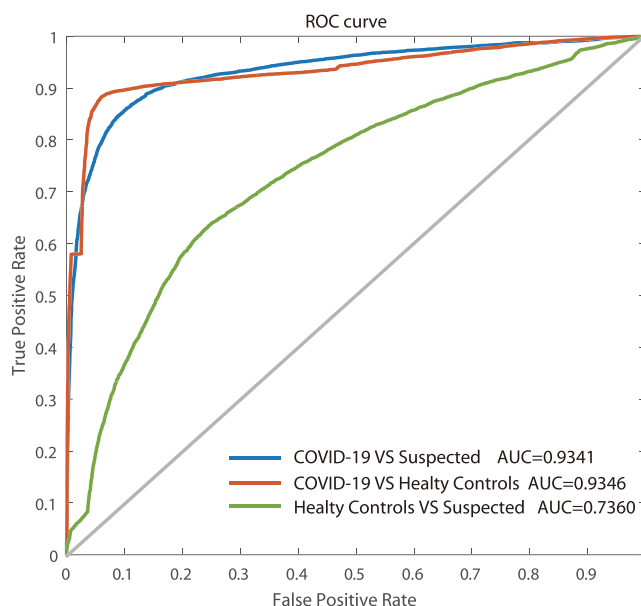
In the training phase, the SVM learns the relationship between independent and dependent variables. Specifically, it learns to make associations between intensity values at specific wavelengths of the input spectrum data with a sample class membership, which are the designated outputs of the SVM. Wave points with significant differences in ANOVA test results are used as the input of the SVM. Here, the performance parameters for each class, determined by the SVM, are shown. The sensitivity (true positive rate), specificity (true negative rate), and accuracy (percentage of correctly predicted spectra among total cases) for each group's prediction are reported together with 95% CIs. All three groups are successfully classified with average sensitivity, specificity, and accuracy (Table 2). Figure 3 shows the performance of the classifiers through a receiver operating characteristic (ROC) curve. All ROC analyses are based on nonparametric techniques and are conducted for the SVM analyses. For each of the three classification tasks, the area under curve (AUC) value is calculated and shown in Figure 3.

Table 2 shows the results of specificity, accuracy, and sensitivity of the SVM classification; the non-brackets indicate the specificity, sensitivity, and accuracy of individual spectra, and the brackets indicate the specificity, sensitivity, and accuracy of each serum sample.

For independent test datasets, all of the unlabeled spectra are assigned to the class with the highest probability by SVM model. The classification of each individual spectrum from each of the serum samples is depicted in Table 3 (left-hand side), demonstrating an overall average accuracy of 90% for all classes.

The overall serum-level classification of each of the 20 serum samples, based on spectral-level predictions, is

shown on the right-hand side of Table 3. A serum sample was assigned to the class (confirmed COVID-19, suspected COVID-19, and healthy control groups) that received the majority of spectra assigned to it. The lowest maximum percentage was observed for sample 15 who was assigned to the suspected class with 62.7% of spectra correctly assigned. The true classification of the samples was only revealed after the model had made its predictions. For the independent test dataset, our model showed that the serum-level classification results were all correct.



**FIGURE 3** The ROC curve of the SVM diagnostic algorithm for the COVID-19 group versus the suspected group, the COVID-19 group versus healthy control group, and the suspected group versus the healthy control group

**TABLE 2** Performance parameters of the SVM

Class	Performance parameter	Value $\pm$ SD	95% CI
COVID-19 versus suspected	Sensitivity	$0.89 \pm 0.08$ ( $0.90 \pm 0.08$ )	0.87–0.91 (0.87–0.92)
	Specificity	$0.86 \pm 0.09$ ( $0.88 \pm 0.09$ )	0.83–0.88 (0.85–0.90)
	Accuracy	$0.87 \pm 0.05$ ( $0.89 \pm 0.06$ )	0.86–0.89 (0.88–0.90)
COVID-19 versus healthy control	Sensitivity	$0.89 \pm 0.07$ ( $0.89 \pm 0.079$ )	0.90–0.92 (0.87–0.91)
	Specificity	$0.93 \pm 0.06$ ( $0.94 \pm 0.06$ )	0.91–0.94 (0.93–0.96)
	Accuracy	$0.91 \pm 0.04$ ( $0.91 \pm 0.04$ )	0.90–0.92 (0.90–0.93)
Suspected versus healthy control	Sensitivity	$0.70 \pm 0.09$ ( $0.72 \pm 0.11$ )	0.68–0.73 (0.69–0.75)
	Specificity	$0.66 \pm 0.09$ ( $0.71 \pm 0.11$ )	0.64–0.69 (0.68–0.74)
	Accuracy	$0.69 \pm 0.05$ ( $0.71 \pm 0.07$ )	0.68–0.70 (0.70–0.73)

Note: Brackets: serum-level classification results for each serum samples.  
Abbreviations: CI, confidence interval; SVM, support-vector machine.

**TABLE 3** Results of 20 samples for the independent verification

Individual spectra predictions						
Sample #	Predicted class			External validation results		
	COVID-19	Suspected	Healthy control	Sample #	Predicted class	True class
1	2771	/	529	1	COVID-19	COVID-19
2	2570	/	730	2	COVID-19	COVID-19
3	2642	/	658	3	COVID-19	COVID-19
4	2424	/	876	4	COVID-19	COVID-19
5	2631	/	669	5	COVID-19	COVID-19
6 <sup>a</sup>	3300	0	/	6 <sup>a</sup>	COVID-19	COVID-19
7 <sup>a</sup>	3271	29	/	7 <sup>a</sup>	COVID-19	COVID-19
8 <sup>a</sup>	3300	0	/	8 <sup>a</sup>	COVID-19	COVID-19
9 <sup>a</sup>	2811	489	/	9 <sup>a</sup>	COVID-19	COVID-19
10 <sup>a</sup>	3300	0	/	10 <sup>a</sup>	COVID-19	COVID-19
11 <sup>a</sup>	76	3224	/	11 <sup>a</sup>	Suspected	Suspected
12 <sup>a</sup>	196	3104	/	12 <sup>a</sup>	Suspected	Suspected
13 <sup>a</sup>	891	2409	/	13 <sup>a</sup>	Suspected	Suspected
14 <sup>a</sup>	93	3207	/	14 <sup>a</sup>	Suspected	Suspected
15 <sup>a</sup>	1229	2071	/	15 <sup>a</sup>	Suspected	Suspected
16	/	5	3295	16	Healthy controls	Healthy controls
17	/	115	3185	17	Healthy controls	Healthy controls
18	/	0	3300	18	Healthy controls	Healthy controls
19	/	0	3300	19	Healthy controls	Healthy controls
20	/	16	3284	20	Healthy controls	Healthy controls

<sup>a</sup>Symptomatic.

## 4 | DISCUSSION

Early detection plays a vital role in infectious diseases and controlling an outbreak. COVID-19 is difficult to diagnose from the flu through clinical symptoms, proven by clinical characteristics analysis.<sup>[2]</sup> In this study, we analyze the Raman spectrum data of serum from 177 individuals, which consisted of confirmed COVID-19 patients, suspected COVID-19 patients, and healthy controls group.

The current standard method for detecting COVID-19 is the RT-PCR.<sup>[4,20]</sup> However, the false negative rate is high related to the time between detection and disease onset, because of the difficulty of acquiring SARS-CoV-2 specimens from the respiratory tract in the early phase.<sup>[21]</sup> Currently, antibodies, including IgM and IgG, reverse transcription multiple cross displacement amplification assay, and other techniques are under development for diagnosing COVID-19.<sup>[22–26]</sup>

By using a laser to focus directly on the serum sample, previous studies have demonstrated that Raman

spectroscopy may differentiate virus infections from healthy individuals by detecting the characteristic peaks of Raman spectrum, which respond to specific carbohydrates of the viral glycoprotein.<sup>[10,12,27,28]</sup>

Because SARS-CoV-2 is a virus causing severe infectious disease, for protection, our method seals the serum sample in a polypropylene biological tube for preservation. The laser light goes through the transparent tube wall before approaching the serum sample, and it greatly reduces the signal-to-noise ratio of the Raman signal for the serum sample. Multivariate analysis is required to uncover useful variability within individual spectrum and to build diagnostic algorithms. Hence, ANOVA and SVM are used to analyze the spectrum data to obtain qualified results. In the meantime, the nonexposed samples are maintained during the entire procedure, which is a crucial step to ensure safety during COVID-19 detection. The difference in the mean spectrum is shown as within  $\pm 2$  SDs in Figure 1b, suggesting that the mean difference between the groups is statistically insignificant. High intergroup consistency is shown after the ANOVA

analysis, while the differences in the intragroups are random. Additionally, the data of the spectrum include the signal detected from the tubes and are also in accordance with previous studies.<sup>[29]</sup>

The feature extraction is one significant procedure of machine learning. State-of-art feature extraction methods, such as filter, wrapper, and embedding, train the models and determine the performance of features. In this study, we use ANOVA analysis to select the features. Furthermore, a low  $p$  value does not necessarily mean that a strong feature and conversely a high  $p$  value do not mean a weak feature. Furthermore, the feature extraction methods may increase classification accuracy. Also, we can use deep learning to avoid feature extraction problems when we accumulate more data.

To prevent the virus from spreading, it is recommended that suspected cases should be isolated and diagnosed. This group consists of patients with flu symptoms or individuals in tight contact with COVID-19 patients. Therefore, we recruit three groups, namely, the COVID-19 group, the suspected cases group, and the healthy control group. In this study, compared to the healthy control group and the suspected group, the COVID-19 group shows a significant difference in the same spectrum range of 600–1800  $\text{cm}^{-1}$ . The suspected cases were experiencing flu-like symptoms during the blood draw, a similar clinical manifestation to COVID-19. However, they showed a negative result for the SARS-CoV-2 viral nucleic acid through RT-PCR detection. The spectrum range of 600–1800  $\text{cm}^{-1}$  is previously reported to have a variety of proteins in the serum.<sup>[30,31]</sup>

Interestingly, the spectrum results are consistent with a recent study conducted to determine potential biomarker panels for diagnosing COVID-19.<sup>[32]</sup> Additionally, the difference between the healthy control individuals and the suspected cases shows a lower significance, other than the difference between the COVID-19 and the suspected or healthy control groups. The following are possible reasons for the difference detected by Raman spectroscopy in the serum of COVID-19 patients.

To begin with, SARS CoV-2 may express a specific protein that differs from the healthy samples or other diseases.<sup>[33]</sup> Meanwhile, during the pathological process of COVID-19, activation of the humoral response leads to the generation of certain antibodies,<sup>[9,32]</sup> suggesting the biochemical composition changed in the serum of COVID-19 patients and is specific and detectable by Raman spectroscopy.

Second, most of the patients in the COVID-19 group are non-severe (57 in 63) when drawing blood, indicating that Raman spectroscopy has a high specificity for detecting COVID-19. Moreover, asymptomatic infections currently account for approximately 20% of all COVID-19

cases, which is important for prevention measures.<sup>[34]</sup> In this study, the results of the independent test dataset showed that this model has a high sensitivity for asymptomatic infection detection. Recommending serum Raman spectroscopy may contribute to pandemic prevention because of its high sensitivity for both asymptomatic and symptomatic patients.

Because of the complicated biochemical components within the serum, the single spectrum collected from each sample is not identical, and it contains information for local concentrations of biomarkers and biochemical components. This increases the possibility for individual spectrum to be misclassified. Bo et al. collected the data from 28 healthy subjects, 25 patients with negative COVID-19 PCR test result but have similar clinic features with COVID-19, and 25 COVID-19 patients. They found the severity of the diseases is highly related with 204 serum metabolites from COVID-19 patients, and there are 105 protein difference expressions in serum of COVID-19 patients and subjects without COVID-19 (Shen et al.<sup>[9]</sup>). Liu et al. demonstrated that COVID-19 patients have increasing inflammatory cytokines level in their serum. The cytokine storm is known as one important reason on death of heavy and critical COVID-19 patients.<sup>[35]</sup> Hence, the difference on serum composition of COVID-19 patients comparison to no-COVID-19 is exists and can be detected. Raman spectrum has high accuracy and high sensitivity. Hence, Raman spectrum is one powerful tool for COVID-19 testing. Additionally, the relatively small sample size only represents a tiny fraction of the Chinese population. The detected data for a larger, independent population of different races would help build a better model and standard, which may profoundly improve the accuracy of Raman spectroscopy for screening.

False negative results may occur due to the influence of COVID-19 nucleic acid detection extraction method, kit sensitivity, variation in detection rate from different manufacturers, insufficient viral material in the specimen, low patient viral load, and other factors.<sup>[36,37]</sup> Researches of Lauren et al. show that on the day of symptom onset, the median false negative rate for initial RT-PCR was 38% (CI, 18% to 65%), And 3 days after symptom onset, the median false negative rate decreased to 20% (CI, 12% to 30%), and then increased again, from 21% (CI, 13% to 31%) on day 4 days after symptom onset to 66% (CI, 54% to 77%) on 16 days after symptom onset.<sup>[21]</sup> Another research find that 21.4% patients experienced a “turn positive” of nucleic acid detection by RT-PCR test for SARS-CoV-2 after two consecutive negative result. Although the false negative rate of our model is lower than RT-PCR, our model is based on the gold standard of RT-PCR and the detection performance cannot



be higher than the gold standard. This is because the confirmed cases and the suspected cases are all determined with many times RT-PCR tests in our study, while the false negative rate of RT-PCR test from literature reports is usually based on the first-time RT-PCR test.

The large-scale global epidemic of COVID-19 has brought tremendous pressure to medical institutions. As the gold standard, RT-PCR requires a lot of manpower and material resources. The serum detection based on Raman spectroscopy proposed in this paper is low-cost, fast, and low manpower requirements. Doctors are provided with more testing methods. Serum testing, as a routine testing item in medical institutions, can provide low-cost and rapid screening of patients in hospitals. In applications, medical institutions perform Raman testing in routine serum testing items. Once high-risk patients are found, they are immediately quarantined and then further tested with RT-PCR, thereby reducing the risk of infection in medical institutions.

With the spread of SARS-CoV-2, an increasing number of covert patients could seed new outbreaks.<sup>[8]</sup> To contain the virus, tests for more individuals without strict limitations may generate a massive demand for screening. Sealing the serum sample in a cryopreservation tube, portable detection devices, and a simple procedure for testing indicates that Raman spectroscopy is a safe and convenient detection method. Moreover, the accuracy will be higher with additional screening data for input into the SVM, making this a promising method for application in the future.

## 5 | CONCLUSION

In conclusion, our results suggest that Raman spectroscopy may be powerful, effective, and convenient for COVID-19 screening. Additional studies with a larger independent population are needed to verify our findings and to investigate the particular spectrum bands of the serum spectrum that corresponds to the suspected biomarkers of COVID-19.

## ACKNOWLEDGMENTS

This study was funded by the grant agency of the Chengdu Science and Technology Bureau (No. 2020-YF05-0014-SN) and Department of Science and Technology of Sichuan Province, Science and Technology Support Project (ID: 2019YFG0185).

## CONFLICT OF INTEREST

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

## AUTHOR CONTRIBUTIONS

Lintao Li, Hongyan Zhou, and Yilan Zeng carried out the collection of sample. Gang Yin and Yu Yin carried out data analysis. Yuanzhang Su offered the technique support. Bo Sun participated in data analysis. Mei Luo and Maohua Ma participated in the serum samples collection. Gang Yin, Lintao Li, Shun Lu, and Lucia Orlandini had contributions to literatures review and draft writing. Dezhong Yao, Gang Liu, and Jinyi Lang had contributions to the conception and design of the study. All authors have read and approved the final manuscript.

## ORCID

Gang Yin  <https://orcid.org/0000-0001-9995-4629>

## REFERENCES

- [1] <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>
- [2] W. J. Guan, Z. Y. Ni, Y. Hu, W. H. Liang, C. Q. Ou, J. X. He, L. Liu, H. Shan, C. L. Lei, D. S. C. Hui, B. Du, L. J. Li, G. Zeng, K. Y. Yuen, R. C. Chen, C. L. Tang, T. Wang, P. Y. Chen, J. Xiang, S. Y. Li, J. L. Wang, Z. J. Liang, Y. X. Peng, L. Wei, Y. Liu, Y. H. Hu, P. Peng, J. M. Wang, J. Y. Liu, Z. Chen, G. Li, Z. J. Zheng, S. Q. Qiu, J. Luo, C. J. Ye, S. Y. Zhu, N. S. Zhong, *N. Engl. J. Med.* **2020**, 382, 1708.
- [3] <https://www.who.int/csr/sars/postoutbreak/en/>;
- [4] CDC website
- [5] P. Zhou, X. L. Yang, X. G. Wang, B. Hu, L. Zhang, W. Zhang, H. R. Si, Y. Zhu, B. Li, C. L. Huang, H. D. Chen, J. Chen, Y. Luo, H. Guo, R. D. Jiang, M. Q. Liu, Y. Chen, X. R. Shen, X. Wang, X. S. Zheng, K. Zhao, Q. J. Chen, F. Deng, L. L. Liu, B. Yan, F. X. Zhan, Y. Y. Wang, G. F. Xiao, Z. L. Shi, *Nature* **2020**, 579, 270..
- [6] C. Rothe, M. Schunk, P. Sothmann, G. Bretzel, G. Froeschl, C. Wallrauch, T. Zimmer, V. Thiel, C. Janke, W. Guggemos, M. Seilmaier, C. Drosten, P. Vollmar, K. Zwirgmaier, S. Zange, R. Wolfel, M. Hoelscher, *N. Engl. J. Med.* **2020**, 382, 970.
- [7] J. G. Duguid, V. A. Bloomfield, J. M. Benevides, G. J. Thomas Jr., *Biophys. J.* **1995**, 69, 2623.
- [8] J. Qiu, *Nature* **2020**.
- [9] B. Shen, X. Yi, Y. Sun, X. Bi, J. Du, C. Zhang, S. Quan, F. Zhang, R. Sun, L. Qian, W. Ge, W. Liu, S. Liang, H. Chen, Y. Zhang, J. Li, J. Xu, Z. He, B. Chen, J. Wang, H. Yan, Y. Zheng, D. Wang, J. Zhu, Z. Kong, Z. Kang, X. Liang, X. Ding, G. Ruan, N. Xiang, X. Cai, H. Gao, L. Li, S. Li, Q. Xiao, T. Lu, Y. Zhu, H. Liu, H. Chen, T. Guo, *Cell* **2020**, 182, 59.
- [10] K. Naseer, A. Amin, M. Saleem, J. Qazi, *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2019**, 206, 197.
- [11] S. Khan, R. Ullah, A. Khan, R. Ashraf, H. Ali, M. Bilal, M. Saleem, *Photodiagnosis Photodyn. Ther.* **2018**, 23, 89.
- [12] J. H. Lee, B. C. Kim, B. K. Oh, J. W. Choi, *J. Biomed. Nanotechnol.* **2015**, 11, 2223.
- [13] N. M. Ralbovsky, L. Halamkova, K. Wall, C. Anderson-Hanley, I. K. Lednev, *J. Alzheimers Dis.* **2019**, 71, 1351.
- [14] U. Parlattan, M. T. Inanc, B. Y. Ozgor, E. Oral, E. Bastu, M. B. Unlu, G. Basar, *Sci. Rep.* **2019**, 9, 19795.
- [15] S. Feng, R. Chen, J. Lin, J. Pan, G. Chen, Y. Li, M. Cheng, Z. Huang, J. Chen, H. Zeng, *Biosens. Bioelectron.* **2010**, 25, 2414.

- [16] D. Lin, S. Feng, J. Pan, Y. Chen, J. Lin, G. Chen, S. Xie, H. Zeng, R. Chen, *Opt. Express* **2011**, *19*, 13565.
- [17] S. Desai, S. V. Mishra, A. Joshi, D. Sarkar, A. Hole, R. Mishra, S. Dutt, M. K. Chilakapati, S. Gupta, A. Dutt, *J. Biophotonics* **2020**, *13*, e202000189.
- [18] K. Mehta, A. Atak, A. Sahu, S. Srivastava, K. C. Murali, *Analyst* **2018**, *143*, 1916.
- [19] S. B. T. Guo, U. Neugebauer, J. Popp, *Anal. Methods* **2017**.
- [20] V. M. Corman, O. Landt, M. Kaiser, R. Molenkamp, A. Meijer, D. K. Chu, T. Bleicker, S. Brunink, J. Schneider, M. L. Schmidt, D. G. Mulders, B. L. Haagmans, B. van der Veer, S. van den Brink, L. Wijsman, G. Goderski, J. L. Romette, J. Ellis, M. Zambon, M. Peiris, H. Goossens, C. Reusken, M. P. Koopmans, C. Drosten, *Euro. Surveill.* **2020**, *25*.
- [21] L. M. Kucirka, S. A. Lauer, O. Laeyendecker, D. Boon, J. Lessler, *Ann. Intern. Med.* **2020**, *173*, 262.
- [22] Z. Li, Y. Yi, X. Luo, N. Xiong, Y. Liu, S. Li, R. Sun, Y. Wang, B. Hu, W. Chen, Y. Zhang, J. Wang, B. Huang, Y. Lin, J. Yang, W. Cai, X. Wang, J. Cheng, Z. Chen, K. Sun, W. Pan, Z. Zhan, L. Chen, F. J. Ye, *Med. Virol.* **2020**.
- [23] T. Nguyen, D. Duong Bang, A. Wolff, *Micromachines (Basel)* **2020**, *11*.
- [24] X. Xie, Z. Zhong, W. Zhao, C. Zheng, F. Wang, J. Liu, *Radiology* **2020**, 200343.
- [25] S. Li, W. Jiang, J. Huang, Y. Liu, L. Ren, L. Zhuang, Q. Zheng, M. Wang, R. Yang, Y. Zeng, Y. Wang, *Eur. Respir. J.* **2020**.
- [26] K. L. Lim, N. A. Johari, S. T. Wong, L. T. Khaw, B. K. Tan, K. K. Chan, S. F. Wong, W. L. E. Chan, N. H. Ramzi, P. K. C. Lim, S. L. Hakim, K. Voon, *PLoS One* **2020**, *15*, e0238417.
- [27] D. Tong, C. Chen, J. Zhang, G. Lv, X. Zheng, Z. Zhang, X. Lv, *Photodiagnosis Photodyn. Ther.* **2019**, *28*, 248.
- [28] S. Shanmukh, L. Jones, Y. P. Zhao, J. D. Driskell, R. A. Tripp, R. A. Dluhy, *Anal. Bioanal. Chem.* **2008**, *390*, 1551.
- [29] S. L. Hsu, T. Hahn, W. Suen, S. Kang, H. D. Stidham, A. R. Siedle, *Macromolecules* **2001**, *34*, 3376.
- [30] M. Furuta, T. Fujisawa, H. Urago, T. Eguchi, T. Shingae, S. Takahashi, E. W. Blanch, M. Unno, *Phys. Chem. Chem. Phys.* **2017**, *19*, 2078.
- [31] J. Depciuch, M. Sowa-Kucma, G. Nowak, D. Dudek, M. Siwek, K. Styczen, M. Parlinska-Wojtan, *J. Pharm. Biomed. Anal.* **2016**, *131*, 287.
- [32] L. Guo, L. Ren, S. Yang, M. Xiao, D. Chang, F. Yang, C. S. Dela Cruz, Y. Wang, C. Wu, Y. Xiao, L. Zhang, L. Han, S. Dang, Y. Xu, Q. W. Yang, S. Y. Xu, H. D. Zhu, Y. C. Xu, Q. Jin, L. Sharma, L. Wang, J. Wang, *Clin. Infect. Dis.* **2020**, *71*, 778.
- [33] A. C. Walls, Y. J. Park, M. A. Tortorici, A. Wall, A. T. McGuire, D. Veasler, *Cell* **2020**, *181*, 281.
- [34] D. Buitrago-Garcia, D. Egli-Gany, M. J. Counotte, S. Hossmann, H. Imeri, A. M. Ipekci, G. Salanti, N. Low, *PLoS Med.* **2020**, *17*, e1003346.
- [35] J. Liu, S. Li, J. Liu, B. Liang, X. Wang, H. Wang, W. Li, Q. Tong, J. Yi, L. Zhao, L. Xiong, C. Guo, J. Tian, J. Luo, J. Yao, R. Pang, H. Shen, C. Peng, T. Liu, Q. Zhang, J. Wu, L. Xu, S. Lu, B. Wang, Z. Weng, C. Han, H. Zhu, R. Zhou, H. Zhou, X. Chen, P. Ye, B. Zhu, L. Wang, W. Zhou, S. He, Y. He, S. Jie, P. Wei, J. Zhang, Y. Lu, W. Wang, L. Zhang, L. Li, F. Zhou, J. Wang, U. Dittmer, M. Lu, Y. Hu, D. Yang, X. Zheng, *EBioMedicine* **2020**, *55*, 102763.
- [36] X. Xie, Z. Zhong, W. Zhao, C. Zheng, F. Wang, J. Liu, *Radiology* **2020**, 296, E41.
- [37] Y. Fang, H. Zhang, J. Xie, M. Lin, L. Ying, P. Pang, W. Ji, *Radiology* **2020**, 296, E115.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Yin G, Li L, Lu S, et al. An efficient primary screening of COVID-19 by serum Raman spectroscopy. *J Raman Spectrosc.* 2021;1–10. <https://doi.org/10.1002/jrs.6080>