

## Fast Screening and Primary Diagnosis of COVID-19 by ATR–FT-IR

Liyang Zhang,<sup>||</sup> Meng Xiao,<sup>||</sup> Yao Wang,<sup>||</sup> Siqi Peng, Yu Chen, Dong Zhang, Dongheyu Zhang, Yuntao Guo, Xinxin Wang, Haiyun Luo,<sup>\*</sup> Qun Zhou,<sup>\*</sup> and Yingchun Xu<sup>\*</sup>Cite This: <https://dx.doi.org/10.1021/acs.analchem.0c04049>

Read Online

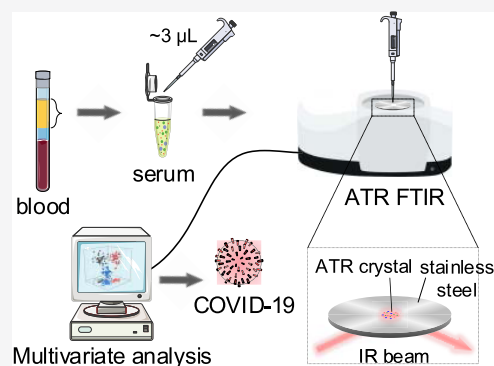
ACCESS |

Metrics &amp; More

Article Recommendations

Supporting Information

**ABSTRACT:** The outbreak of coronavirus disease 2019 (COVID-19) has led to substantial infections and mortality around the world. Fast screening and diagnosis are thus crucial for quick isolation and clinical intervention. In this work, we showed that attenuated total reflection–Fourier transform infrared spectroscopy (ATR–FT-IR) can be a primary diagnostic tool for COVID-19 as a supplement to in-use techniques. It requires only a small volume ( $\sim 3 \mu\text{L}$ ) of the serum sample and a shorter detection time (several minutes). The distinct spectral differences and the separability between normal control and COVID-19 were investigated using multivariate and statistical analysis. Results showed that ATR–FT-IR coupled with partial least squares discriminant analysis was effective to differentiate COVID-19 from normal controls and some common respiratory viral infections or inflammation, with the area under the receiver operating characteristic curve (AUROC) of 0.9561 (95% CI: 0.9071–0.9774). Several serum constituents including, but not just, antibodies and serum phospholipids could be reflected on the infrared spectra, serving as “chemical fingerprints” and accounting for good model performances.



## INTRODUCTION

Coronavirus disease 2019 (COVID-19) is a pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a newly appearing coronavirus which has spread over the world and led to substantial infections and mortality.<sup>1</sup>

Reverse transcription polymerase chain reaction (RT-PCR) is a conventional and standard assay for viral diagnosis and has been widely used for SARS-CoV-2 RNA detection. SARS-CoV-2 RNA can be detected in both upper and lower respiratory specimens including nasal swab, oropharyngeal swab, sputum, and bronchoalveolar lavage fluid (BALF).<sup>2,3</sup> Despite BALF, which is not a requisite for COVID-19 diagnosis because of the harder sampling, the sputum was reported to have the highest positive rate (74.4–88.9%), followed by nasal swabs (53.6–73.3%) during the first 14 days after onset (d.a.o.).<sup>3</sup> The positive rate for throat swab was reported to be around 60%.<sup>3,4</sup> Viral RNA is also detected in serum samples with a percentage of 0% (0/31),<sup>5</sup> 8% (1/12),<sup>6</sup> or 15% (6/41).<sup>7</sup> Notably, several factors may influence the performance of RT-PCR such as improper sample preparation or varied qualities of detection kits and thus lead to high false-positive rates. In addition, viral replication is inhibited in the late stage of infection, accounting for the high false-negative rates in this stage. Also, it is time-consuming to perform the whole test procedures.

Serological assay based on immunoglobulin-G (IgG) and IgM levels can serve as a complement to nucleic acid detection.<sup>8,9</sup> The median time of IgM and IgG seroconversion was reported to be 5 ( $n = 41$ ) and 14 ( $n = 208$ ) days after

onset, respectively.<sup>10</sup> The combination of IgM and IgG tests yielded a higher detection sensitivity of 88.66% and specificity of 90.63% (397 PCR confirmed patients and 128 negative patients in total) than a single IgG or IgM test.<sup>11</sup> Additionally, a higher positive detection rate of 99.4% ( $n = 173$ , 95% CI 96.8–100%) was achieved when applying both antibody and nucleic acid tests, compared to a single RNA test of 67.1% (95% CI 59.4–74.1%).<sup>12</sup> Nevertheless, there remain some problems unclear such as the antibody responses of COVID-19 patients, the potential false positive caused by immunological cross reactivity, and the varied performances of commercially available detection kits.

Rapid and reliable diagnosis of COVID-19 is of great significance to help screen the COVID-19 patients and deliver more appropriate treatment. In the last decade, transmission or attenuated total reflection (ATR) Fourier transform infrared spectroscopy (FT-IR) and Raman spectroscopy have been utilized to identify viral infections or predict viral load in blood,<sup>13</sup> sera,<sup>13–15</sup> plasma,<sup>16</sup> or infected cells,<sup>17,18</sup> differentiate different viral infections,<sup>19</sup> and verify the infectious agent type (as bacterial or viral) based on the white blood cell data.<sup>20</sup> Subtle molecular and chemical changes in blood components

Received: September 25, 2020

Accepted: December 18, 2020

in response to bacterial or viral infections can be recorded and reflected by the infrared spectra. For example, the strong band at  $1631\text{ cm}^{-1}$  attributed to the  $\beta$ -pleated sheet protein marker of Ig is unique to the positive serum spectra induced by hepatitis B and C virus.<sup>14</sup> In comparison with other assays, infrared spectroscopy enables us to inspect almost all biological components at once, which may be beneficial to COVID-19 diagnosis. Additionally, it is easier to perform and takes less operation time (typically for several minutes).

In this work, we showed the feasibility of ATR–FT-IR in COVID-19 screening and primary diagnosis. The spectral differences between COVID-19 and healthy controls and the potential spectral markers were identified by multivariate and statistical analysis. For the purpose of the performance test, especially the specificity of the proposed model, healthy controls and some common respiratory viral infections or inflammation were considered.

## MATERIALS AND METHODS

**Participants.** We collected a total of 115 blood samples from 20 healthy donors and 76 patients, of which 41 were confirmed with COVID-19, 15 had respiratory viral infections caused by influenza A/B or respiratory syncytial virus (RSV), and 20 were with inflammation-related diseases (Table 1).

**Table 1. Number of Participants, Specimens, and the Measured Spectra in This Study**

	participants ( <i>n</i> = 96)	specimens ( <i>n</i> = 115)	spectra ( <i>n</i> = 289)
COVID-19 (cohort 1)	35	35	59
COVID-19 (cohort 2)	6	6	18
Influenza A	4	8	23
Influenza B	2	2	5
RSV	9	24	73
inflammation <sup>a</sup>	20	20	60
control	20	20	51

<sup>a</sup>Details are presented in Table S1.

Influenza A/B and RSV are chosen because they are common in respiratory infections and share similar flu-like symptoms<sup>21,22</sup> to COVID-19. Based on a report regarding 1099 patients, the most common symptoms of COVID-19 are fever (43.8% on admission and 88.7% during hospitalization) and cough (67.8%).<sup>23</sup> We know that inflammation-related diseases can also contribute to alterations in serum components (see Discussion). Here, respiratory bacterial infections, pulmonary infection, intra-abdominal infection, bacteremia, and some other diseases were enrolled in the study (Table S1). We aim to investigate whether and how the serum infrared spectra are specific for COVID-19.

The COVID-19 patients were from two cohorts, all of whom were diagnosed by reverse transcriptase PCR (RT-PCR) following *China national guidelines for diagnosis and treatment of Corona Virus Disease 2019 (COVID-19)* (trial version 5, revised).<sup>24</sup> The first cohort included 35 critically ill COVID-19 patients who were admitted from Feb 15 to March 30 at the Sino-French New City Branch of Tongji Hospital in Wuhan. Of the 35 patients, one was with <7 days postsymptom onset and one was within 7–14 days, while 33 were with >14 days. In the second cohort collected in the Peking Union Medical College Hospital, five confirmed cases were with >14 days after

onset, while one was with 2 days. All of them were with mild symptoms. Clinical residual blood samples (stored at  $-80\text{ }^{\circ}\text{C}$ ) from influenza A, influenza B, or respiratory syncytial virus (RSV)-infected patients admitted from Feb 16 to March 1, 2020, were revived. The viral nucleic acids were determined by X'Pert Xpress Flu/RSV (Cepheid AB, Sweden). Of note, for each patient with influenza A or RSV infections, two to three blood specimens from different infection stages before recovery were collected to enrich and generalize the data set. Patients with other diseases were diagnosed using standard clinical methods and proved to be without any SARS-CoV-2 infection (Table S1). This study was approved by the Ethics Committees from Tsinghua University and Peking Union Medical College Hospital.

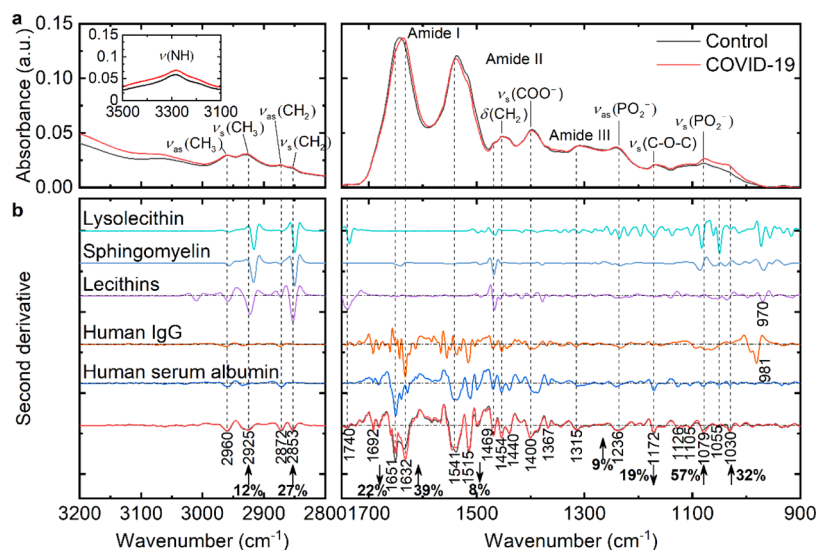
### Sample Preparation and ATR–FTIR Spectroscopy.

The serum samples were obtained by blood centrifugation under 5000 rpm for 5 min. Prior to measurements, the serum specimens were incubated at  $56\text{ }^{\circ}\text{C}$  for 30 min to inactivate the potential pathogens. Each specimen was measured one to three times on a PerkinElmer infrared spectrometer coupled with a diamond ATR accessory at a resolution of  $4\text{ cm}^{-1}$ . Sixteen scans were accumulated per spectrum. For each measurement, an aliquot of 3 or  $4\text{ }\mu\text{L}$  serum sample was transferred onto the ATR crystal and allowed to dry under mild airflow at room temperature. Water absorption was monitored by OH stretching at around  $3300\text{ cm}^{-1}$  and bending at around  $1635\text{ cm}^{-1}$ . It took about 3–5 min for samples to be sufficiently dried. Afterward, the spectra between 4000 and  $600\text{ cm}^{-1}$  were collected. The spectral background was recorded separately for each sample to achieve a higher signal-to-noise ratio. Prior to data analysis, the raw spectral data were preprocessed with baseline calibration using the built-in rubber-band algorithm on the PerkinElmer spectrum (version 10.5.4). The corresponding constant baseline drift at  $4000\text{ cm}^{-1}$  was subtracted for each sample spectrum, and then, all the spectra were normalized to amide I by peak absorbance. The second-derivative infrared (SD-IR) spectra were calculated from the normalized spectra.

**Multivariate and Statistical Analysis.** All data analyses were performed on MATLAB (version R2020b, Mathworks, Natick, U.S.A). The second-derivative infrared (SD-IR) spectra were calculated by the 13-point Savitzky–Golay algorithm using in-house protocols.

Analysis of variance (ANOVA) is a well-established method to evaluate the difference of the observed statistic(s) among groups by calculating the F-statistic (*i.e.*, the ratio of between-group variance to within-group variance).<sup>25</sup> The larger F-statistic yields a lower *p* value, indicating that the groups are more likely to be different in this statistic. In this work, the *p* value was used to evaluate the statistical spectral differences of major bands in absorbances or band locations of healthy controls and COVID-19 patients. Hence, the potential spectral markers can be primarily identified.

Hierarchical cluster analysis (HCA) and principal component analysis (PCA) are widely used unsupervised approaches in microorganism differentiation.<sup>26–28</sup> In HCA, the spectral distance (usually Euclidian distance) is calculated to construct the hierarchical structure, in which root nodes are formed by spectra with the shortest distances, while leaf nodes are formed by the longest distances. PCA aims to reduce the dimensions of the original matrix data in terms of maximum variance.<sup>26,29</sup> In the new vector space formed by principal components (PC), data distribution is more apparent. Here, the natural



**Figure 1.** Spectral profiles of COVID-19 and control serum samples and some abundant constituents in human serum. (a) Averaged original spectra of COVID-19 and control samples. Spectra were normalized to amide I. The principal absorbance bands were annotated. (b) Averaged SD-IR spectra of COVID-19 and control samples and the most abundant constituents in human serum. Bands with notable variations were annotated (see also Table S3). Lysolecithin, sphingomyelin, lecithin, and human IgG (Bioss, Beijing, China, purified by Protein-A affinity chromatography) were in dried powder form, whereas human serum albumin was buffered with water. The band located at  $981\text{ cm}^{-1}$  in the human IgG spectrum may arise from impurities.

**Table 2. Normalized Spectra: Band Assignments and Statistical Comparisons between COVID-19 and Control Spectra in Band Locations and Relative Absorbance<sup>a,d,e,f</sup>**

assignments <sup>19,27</sup>	band locations/( $\text{cm}^{-1}$ )			relative absorbance (a.u.)			
	control	COVID-19	<i>p</i>	control	COVID-19	<i>p</i>	changes (%) <sup>b</sup>
amide A	3285.65 (1.04)	3282.75 (1.19)	***	0.430 (0.028)	0.500 (0.049)	***	+16.3
$\nu_{\text{as}}(\text{CH}_3)$	2958.31 (0.55)	2958.99 (0.83)	***	0.215 (0.007)	0.217 (0.008)		
$\nu_{\text{as}}(\text{CH}_2)$	2930.76 (1.05)	2929.71 (1.55)	***	0.220 (0.010)	0.225 (0.015)	*	+2.3
$\nu_{\text{s}}(\text{CH}_3)$	2872.61 (0.49)	2873.14 (0.76)	***	0.150 (0.006)	0.151 (0.008)		
amide I	1642.27 (2.28)	1636.84 (2.24)	***	1 (0) <sup>c</sup>	1 (0) <sup>c</sup>		
amide II	1537.18 (1.05)	1538.19 (0.95)	***	0.878 (0.03)	0.857 (0.025)	***	-2.4
$\delta(\text{CH}_2)$	1453.33 (0.62)	1452.9 (0.7)	***	0.342 (0.021)	0.339 (0.014)		
$\nu_{\text{s}}(\text{COO}^-)$	1397.53 (1.05)	1398.22 (0.68)	***	0.386 (0.024)	0.380 (0.016)		
amide III	1308.06 (1.7)	1311.19 (1.14)	***	0.280 (0.021)	0.279 (0.022)		
$\nu_{\text{as}}(\text{PO}_2^-)$	1242.29 (0.81)	1241.62 (1.01)	***	0.264 (0.023)	0.271 (0.016)		
$\nu_{\text{s}}(\text{C-O-C})$	1170.12 (0.33)	1167.6 (2.85)	***	0.154 (0.016)	0.151 (0.010)		
$\nu_{\text{s}}(\text{PO}_2^-)$	1079.63 (0.77)	1077.36 (1.02)	***	0.162 (0.017)	0.193 (0.025)	***	+19.1

<sup>a</sup> $\nu_{\text{s}}$ : symmetric stretching vibrations;  $\nu_{\text{as}}$ : asymmetric stretching vibrations;  $\delta$ : bending vibrations. Data are in mean (standard deviation) or %.

<sup>b</sup>Increases in average band absorbance compared to controls. <sup>c</sup>All the spectra were normalized to amide I. <sup>d</sup>Statistical differences were compared between COVID-19 and the control group using one-way ANOVA. <sup>e</sup>\* $p < 0.05$ . <sup>f</sup>\*\* $p < 0.01$ . <sup>f</sup>\*\*\* $p < 0.001$ . <sup>f</sup>See more in Figure S2.

separability of normal controls and nonsevere and severe COVID-19 patients was tested by the two methods.

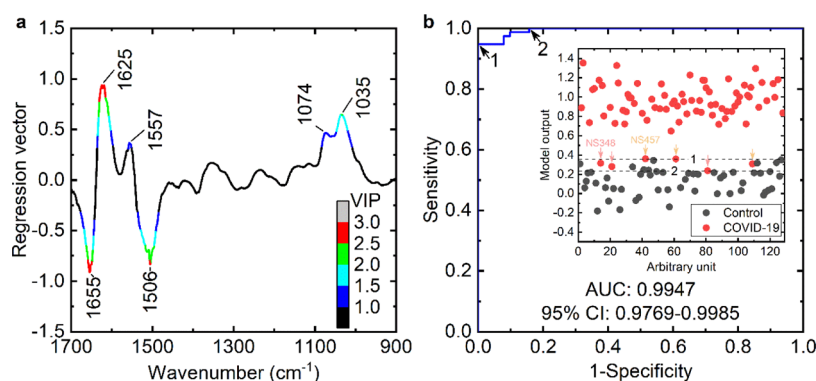
Partial least squares discriminant analysis (PLS-DA) is a supervised chemometric technique whereby the so-called latent variables are successively extracted to find the maximum correlation between the X-matrix and Y-matrix.<sup>14,30</sup> In the present study, the data set was divided into three groups: normal controls, COVID-19, and infections caused by other diseases. PLS1 and PLS2 models<sup>31</sup> were applied for two-group and three-group classification, respectively. Specifically, in the case of PLS1, the two classes were coded as [0] and [1], whereas the three classes were coded as [1 0 0], [0 1 0], and [1 0 0] in the PLS2 model. In comparison with PCA, PLS-DA is more suitable to identify the significant wavenumbers for discrimination by inspecting the regression vector<sup>14</sup> or the variable importance in projection (VIP).<sup>32</sup>

## RESULTS

**COVID-19 Vs Controls: Band Assignments, Spectral Differences, and Spectral Interpretation.** It is known that serum is composed of proteins, cholesterol, glucose, urea, triglycerides, and other more dilute compounds, all of which can be recorded in the spectra, whereas only components with higher abundance could be identified in the spectra and provide insightful information. Reports show that a wide range of abundant biomolecules in plasma can be quantified using FTIR.<sup>33</sup> To better illustrate the spectral features, we summarized the normal levels of the major serum components (Table S2).

Averaged spectra minimize the influence of individual differences and thus are more representative. As for the original spectra, all bands had significant frequency shifts in the COVID-19 group ( $p < 0.001$ , Figure 1a and Table 2),





**Figure 3.** Results of the PLS-DA model for classification between COVID-19 and control groups. (a) Regression vector with respect to the spectral region of 900–1700  $\text{cm}^{-1}$ . VIP values in different ranges are shown in different colors. VIP = variable influence on projection. (b) ROC plot of COVID-19 samples (outer) and the model output (inner). Two threshold values corresponding to two points in the ROC curve were selected. AUC = area under curve. CI = confidence interval.

were observed. Nevertheless, we found that this band together with the bands centered at 2925 and 2853  $\text{cm}^{-1}$  in the SD-IR spectra had a significant blue shift in the absorption frequency (Figure 1b and Table S3), suggesting the potential conformational changes in lipid components.

**Differentiate COVID-19 Patients from Normal Controls.** PCA was first performed on the SD-IR spectra to reduce the dimension of infrared spectral space. It requires the first five and ten components to account for 90% of the total variance in the data ranging from 1600 to 1700  $\text{cm}^{-1}$  and 1000 to 1700  $\text{cm}^{-1}$ , respectively. Nevertheless, the score plot with respect to the first two PCs provides meaningful information (Figure 2). Obvious transitions from normal controls to nonsevere COVID-19 patients and to the severe ones were observed (Figure 2a). Overlapping was mitigated when applying the spectral range of 1000–1700  $\text{cm}^{-1}$  (Figure 2b). This implies that other serum components in addition to proteins may play a role in differentiating between healthy controls and COVID-19 patients and evaluating the illness severities. Two patients (NS348 and NS457) in the nonsevere group and one patient (S55) in the severe group were found to be closer to normal controls. It is noteworthy that patient NS348 was in the early stage of illness with 2 days after onset, while NS457 was an asymptomatic patient who was validated with at least >14 d.a.o. S55 was a critically ill patient but with time after onset of <7 days when sampled. Note that two components explain about 60% and 55% of the total variance when applying a spectral window of 1600–1700 or 1000–1700  $\text{cm}^{-1}$ , respectively. For comparison with the 2-PC results, the score plot regarding three PCs is shown in Figure S4.

We further applied HCA to classify the groups using the shortest distance method and Euclidian distance based on the spectral region of 1000–1700  $\text{cm}^{-1}$ . Most of the severe patients were clustered together except six spectra from patients S57, S14, S24, and S55 (Figure 2c). Among the spectra from nonsevere patients, six of them from patients NS348 and NS457 were mixed with normal controls, whereas the remaining was separated well with others (Figure 2c). Results of the two unsupervised methods show the potential feasibility of FT-IR to discern COVID-19 patients in different severities, yet supervised learning methods with the labeled data set may provide more insightful information.

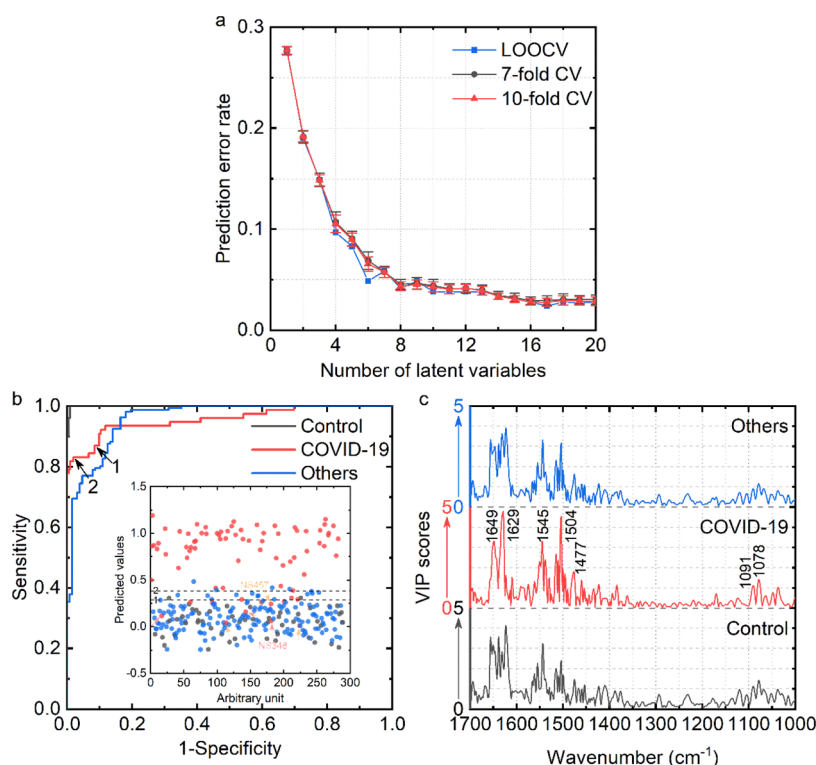
PLS-DA is commonly used in chemometrics for its ability to handle multivariate data and high model interpretability.<sup>31</sup> We first performed a PLS-DA analysis using SD-IR spectra in the

region between 1000 and 1700  $\text{cm}^{-1}$ . The first latent variable (LV, *i.e.*, score) was extracted under leave-one-out cross validation with an  $R^2Y$  value of 0.793 and a  $Q^2$  value of 0.790. Regularly,  $Q^2 > 0.5$  is admitted for good model predictability.<sup>43</sup> The regression vector of variables (loadings for LV1) and the corresponding variable importance in the projection (VIP) values are presented in Figure 3a. VIP is a parameter to evaluate the influence of individual X-variables on the model and is preferred if the value is larger than one.<sup>32,44</sup> Clearly, the most discriminatory bands are located at 1655, 1625, 1557, 1506, 1074, and 1035  $\text{cm}^{-1}$ . This finding is consistent with the SD-IR spectra shown in Figure 1b.

In the PLS-DA model, normal controls and COVID-19 samples were labeled “0” and “1”, respectively. The receiver operating characteristic (ROC) curve was generated by disturbing the decision threshold. The area under the curve (AUC) value of the model was 0.9947 (95% CI 0.9769–0.9985). From the model output shown in Figure 3b, we observed a distinct separation between the two classes, despite six sample spectra which were from patient NS348 and NS457 mixed with controls, that is, when screening COVID-19 patients, especially the ones with moderate symptoms, such a simple model based on a single latent variable may not be effective enough. Nevertheless, SD-IR spectra coupled with PLS-DA analysis achieve high sensitivity and specificity when choosing proper decision thresholds, at least for the training data in this work.

**Differentiate COVID-19 from Normal Controls, Respiratory Viral Infections, and Inflammation-Related Diseases.** Patients with influenza A/influenza B/RSV infections or common inflammatory diseases were enrolled in as interferences to further assess the performance of the proposed method. We first applied a database searching algorithm based on Pearson’s correlation coefficient.<sup>45</sup> Despite the high sensitivity for the controls, the overall performance is limited (Table S4). PCA was not satisfactory either (not shown), three PCs explaining about 50–55% of the data variance when using different spectral windows. The above-mentioned results highlight the risks of misclassification using unsupervised methods.

Then, a triple-class PLS-DA model was established to differentiate the following three groups: normal controls, COVID-19, and other diseases. The spectral range between 900 and 1700  $\text{cm}^{-1}$  achieved the best results. Three most frequent cross-validation (CV) methods, namely, leave-one-out



**Figure 4.** PLS-DA model performances for the triple-class classification. (a) Prediction error rates as a function of the number of latent variables. For 7-fold and 10-fold cross-validation, the error bars were presented. (b) ROC graphs for each group. The inner graph shows the model predicted output of the COVID-19 class. The decision threshold values of 1 and 2 are 0.288 and 0.383, respectively. (c) VIP scores for each class. Significant peaks were labeled.

(LOO), 7-fold, and 10-fold were used to determine the optimum number of latent variables by means of the prediction error rate.<sup>30,46</sup> When applying 7-fold or 10-fold CV, the data were randomly rearranged, which was repeated 100 times to obtain statistical results. Clearly, three methods achieved consistent prediction error rates, indicating good model predictability (Figure 4a). In view of the model complexity, finally, five LVs was selected. Fewer variables are not adequate, whereas more variables contribute little to the prediction capability.

The PLS-DA model with five LVs using the whole dataset was then applied to produce the ROC curves (Figure 4b). Here, the model output is a matrix of  $n \times 3$ , each column corresponding to one class. By disturbing the decision threshold, we got one ROC curve for each class. This helps illustrate how one class is differentiated from others. As shown in Figure 4b, the control group achieves the highest AUC value of 0.9994 (95% CI: 0.9970–1.0000), which suggests that normal controls are nearly not likely to be identified as patients by the model and vice versa. The AUC values for COVID-19 and other diseases are 0.9561 (95% CI: 0.9071–0.9774) and 0.9588 (95% CI: 0.9327–0.9752), respectively. The sensitivity and specificity of COVID-19 identification can be adjusted by modifying the decision thresholds. For instance, when it was set as 0.288 (threshold 1, Figure 4b), a high sensitivity of 87% was achieved, whereas a high specificity of 98% was fulfilled when it was set as 0.383 (threshold 2, Figure 4b, sensitivity: 83.1%). Overall, threshold 1 is more acceptable. We found that nonsevere patients NS348 and NS457 were still mixed with others (Figure 4b, inner graph). Finally, VIP scores were calculated to inspect the most significant peaks for classification. As indicated from Figure 4c, bands in spectral

ranges of 1450–1650 and 1050–1100  $\text{cm}^{-1}$  should be paid attention to. Of a simple note, inflammatory markers may not be likely to be distinguished by infrared spectroscopy because of their relatively low levels (<1 mg/mL, or even lower than ng/mL, Table S2). We may conclude that the discriminatory bands may arise from subtle changes in proteins or lipids or other constituents in serum.

## DISCUSSION

The selection of biospecimens is necessary to be discussed. In this study, we chose serum because of its relatively stable components, thus minimizing the individual differences caused by disease-irrelevant factors. However, other types of specimens also deserve trying, for example, white blood cells. COVID-19 patients, especially for severe cases, are reported with significantly lower total lymphocytes, CD4<sup>+</sup> T cells, CD8<sup>+</sup> T cells, B cells, and NK cells ( $P < 0.001$ ),<sup>47</sup> as well as reduced percentages of monocytes, eosinophils, and basophils.<sup>48,49</sup> Virus-containing samples such as nasal and throat swabs may be feasible too.

As for ATR-FT-IR measurements, several factors may influence the qualities of the spectra. For instance, hemolysis caused by improper sampling might lead to spectral distortion and impede correct identification (data not shown). It is known that unfolding, conformational changes, and denaturation may happen in proteins when exposed to elevated temperatures.<sup>50,51</sup> For the samples collected, preheating at 56 °C for 30 min before measurement contributes to minor spectral differences in most cases, whereas it may lead to relatively large variations in very few cases. The influence of the thermal factor on serum components and the corresponding infrared spectra is worthwhile to further investigate. Never-

theless, the IgG secondary structure seems to be stable within temperatures between 20 and 55 °C.<sup>51</sup> Of another note, when considering the practical operation, protein coagulation induced by heat may make it harder for uniform sampling.

Immune responses to infection can contribute to alterations in serum constituents, and several biomarkers have been suggested to identify certain infections. For example, procalcitonin (PCT), circulating cytokines (interferon [IL]-1 $\beta$ , IL-6, IL-18, *etc.*), and acute-phase proteins (C-reactive protein [CRP], ferritin, *etc.*) have been used to differentiate between bacterial and viral infection.<sup>52,53</sup> However, it is not always an easy task to find specific markers for a given viral infection. Over the past several months, proteomic and metabolomic profiles of peripheral blood samples have been investigated to correlate disease severity of COVID-19 with certain proteins and lipids.<sup>54,55</sup> Details can be found elsewhere.

We know that FT-IR also provides hints toward proteins, lipids, and other constituents in sera, but in a more comprehensive and macroscopic view. Provided that human serum is dominated by proteins and proteins are dominated by albumin and Ig (Table S2, see also Figure 1b), we can postulate that alterations in the amide I are closely related to the two types of proteins. The increases in Ig levels and decreases in albumin levels in the sera of patients with COVID-19, revealed by the SD-IR spectra in Figure 1b, have been proved by a lot of reports. The IgG level in the patient sera elevates several times after the onset of COVID-19.<sup>56</sup> A report by Long *et al.* showed that 19.5% (8/41) of the patients with COVID-19 have a fourfold increase in the IgG titers.<sup>9</sup> The decreased albumin level was also reported by several reports.<sup>54,57</sup> The reduction in the albumin level is related to the acute phase response (APR) of patients to viral infection. The hepatic synthesis of proteins is drastically regulated during the acute phase of illnesses such as infection, tissue injury, neoplastic growth, or immunological disorders, usually with increased C-reactive protein (CRP) and serum amyloid A and decreased transferrin and albumin.<sup>58</sup> The dysregulation of phospholipids in patients with COVID-19 has also been indicated by the SD-IR spectra shown in Figure 1b, which has been discussed earlier (Results, part one).

More detailed interpretation of the infrared spectra and the corresponding spectral differences still lack a lot of knowledge. The multivariate and statistical analysis provided chances for the evaluation of COVID-19 diagnosis by infrared spectroscopy. In summary, the second-derivative spectrum can increase the separation of the overlapping bands and thus is more powerful than the original spectrum. We show that both unsupervised and supervised methods are promising to differentiate COVID-19 patients from healthy controls. Nonsevere and severe patients can be separated by unsupervised methods such as PCA and HCA. However, as indicated by this work, asymptomatic patients or newly diagnosed patients with low IgG levels may not be correctly identified. In this case, the decision threshold of PLS-DA should be decided with caution. Lower threshold values yield higher sensitivity but lower specificity.

For the purpose of diagnosis, which is no doubt tougher than the only discrimination between COVID-19 and controls, common respiratory viral infections or inflammation were considered as interferences. Unsupervised methods were not so effective, whereas the PLS-DA model yielded good performances. For a given threshold, the model achieved a sensitivity of 83.1% and specificity of 98%. Bands in spectral

ranges of 1450–1650 and 1050–1100 cm<sup>-1</sup> were the most responsible for the model discrimination capability. Further illustration is required in future work.

## CONCLUSIONS

Taking the time cost, operation complexity, and detection performance into consideration, FTIR coupled with multivariate analysis is a feasible tool for screening and primary diagnosis of COVID-19. Several serum constituents including, but not just, antibodies and serum lipids could be reflected on the infrared spectra, serving as “chemical fingerprints” and accounting for good model performances. It was also showed that this assay exhibited high identification specificity for COVID-19 under the intervention from common respiratory viruses and inflammation-related diseases.

In terms of practical use, more clinical data are required to generalize the model, and special attention should be paid to the asymptomatic patients. Nevertheless, infrared spectroscopy can serve as an assistant diagnosis tool as a supplement to in-use techniques. The feasibility of other common specimens such as oropharyngeal swabs, identification and interpretation of the spectral markers, the potentials of FTIR spectroscopy on antibody response monitoring, and COVID-19 severity prediction require further investigations.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.analchem.0c04049>.

Patients with inflammatory diseases included in this study; molar and mass concentrations of major constituents in normal human serum/plasma; second-derivative spectra: band assignments and statistical comparisons between COVID-19 and control serum in band locations and relative absorbance; performance of the database searching algorithm using different spectral regions; original and the corresponding second-derivative infrared spectra from healthy controls (number of spectra = 51) and nonsevere ( $n = 18$ ) and severe ( $n = 59$ ) COVID-19 patients; box plot of relative absorbance (to amide I) of the existing bands in the original spectra; box plot of absorbance of the existing bands in the SD-IR spectra; and three-principal-component score plot of PCA results (PDF)

## AUTHOR INFORMATION

### Corresponding Authors

**Haiyun Luo** – Department of Electrical Engineering, Tsinghua University, Beijing 100084, China; Email: [lhy@tsinghua.edu.cn](mailto:lhy@tsinghua.edu.cn)

**Qun Zhou** – Department of Chemistry, Key Laboratory of Bioorganic Phosphorus Chemistry & Chemical Biology (Ministry of Education), Tsinghua University, Beijing 100084, China; Email: [zhouqun@tsinghua.edu.cn](mailto:zhouqun@tsinghua.edu.cn)

**Yingchun Xu** – Department of Laboratory Medicine, Peking Union Medical College Hospital, Peking Union Medical College, Chinese Academy of Medical Sciences, Beijing 100730, China; Email: [xycpumch@139.com](mailto:xycpumch@139.com)

## Authors

Liyang Zhang – Department of Electrical Engineering, Tsinghua University, Beijing 100084, China; [orcid.org/0000-0001-9909-9263](https://orcid.org/0000-0001-9909-9263)

Meng Xiao – Department of Laboratory Medicine, Peking Union Medical College Hospital, Peking Union Medical College, Chinese Academy of Medical Sciences, Beijing 100730, China

Yao Wang – Department of Laboratory Medicine, Peking Union Medical College Hospital, Peking Union Medical College, Chinese Academy of Medical Sciences, Beijing 100730, China

Siqi Peng – Department of Electrical Engineering, Tsinghua University, Beijing 100084, China; [orcid.org/0000-0001-5418-3656](https://orcid.org/0000-0001-5418-3656)

Yu Chen – Department of Laboratory Medicine, Peking Union Medical College Hospital, Peking Union Medical College, Chinese Academy of Medical Sciences, Beijing 100730, China

Dong Zhang – Department of Laboratory Medicine, Peking Union Medical College Hospital, Peking Union Medical College, Chinese Academy of Medical Sciences, Beijing 100730, China

Dongheyu Zhang – Department of Electrical Engineering, Tsinghua University, Beijing 100084, China

Yuntao Guo – Department of Electrical Engineering, Tsinghua University, Beijing 100084, China

Xinxin Wang – Department of Electrical Engineering, Tsinghua University, Beijing 100084, China

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.analchem.0c04049>

## Author Contributions

<sup>†</sup>L.Z., M.X., and Y.W. contributed equally.

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work was funded by the National Natural Science Foundation of China (52041701), the Tsinghua University-Peking Union Medical College Hospital Initiative Scientific Research Program (20191080604), Tsinghua University Spring Breeze Fund (2020Z99CFG007), and Beijing Nova Program (Z201100006820127).

## REFERENCES

- (1) WHO, Coronavirus disease (COVID-2019) situation reports. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/> (accessed June 1, 2020), June 1, 2020.
- (2) Zhou, P.; Yang, X.-L.; Wang, X.-G.; et al. *Nature* **2020**, *579*, 270–273.
- (3) Yang, Y.; Yang, M.; Yuan, J.; Wang, F.; Wang, Z.; Li, J.; Zhang, M.; Xing, L.; Wei, J.; Peng, L.; Wong, G.; Zheng, H.; Wu, W.; Shen, C.; Liao, M.; Feng, K.; Li, J.; Yang, Q.; Zhao, J.; Liu, L.; et al. *The Innovation* **2020**, *1*, No. 100061.
- (4) Ai, T.; Yang, Z.; Hou, H.; et al. *Radiology* **2020**, *296*, E32–E40.
- (5) Wölfel, R.; Corman, V. M.; Guggemos, W.; et al. *Nature* **2020**, *581*, 465–469.
- (6) Young, B. E.; Ong, S. W. X.; Kalimuddin, S.; et al. *JAMA, J. Am. Med. Assoc.* **2020**, *323*, 1488–1494.
- (7) Huang, C.; Wang, Y.; Li, X.; et al. *Lancet* **2020**, *395*, 497–506.
- (8) Li, X.; Geng, M.; Peng, Y.; Meng, L.; Lu, S. *J. Pharm. Anal.* **2020**, *10*, 102–108.
- (9) Long, Q.-X.; Liu, B.-Z.; Deng, H.-J.; et al. *Nat. Med.* **2020**, *26*, 845–848.

- (10) Guo, L.; Ren, L.; Yang, S.; et al. *Clin. Infect. Dis.* **2020**, *71*, 778–785.
- (11) Li, Z.; Yi, Y.; Luo, X.; et al. *J. Med. Virol.* **2020**, *92*, 1518–1524.
- (12) Zhao, J.; Yuan, Q.; Wang, H.; et al. *Clin. Infect. Dis.* **2020**, *71*, 2027.
- (13) Santos, M. C. D.; Nascimento, Y. M.; Araújo, J. M. G.; Lima, K. M. G. *RSC Adv.* **2017**, *7*, 25640–25649.
- (14) Roy, S.; Perez-Guaita, D.; Bowden, S.; Heraud, P.; Wood, B. R. *Clinical Spectroscopy* **2019**, *1*, 100001.
- (15) Sitole, L.; Steffens, F.; Krüger, T. P. J.; Meyer, D. *OMICS* **2014**, *18*, 513–523.
- (16) Sakudo, A.; Tsenkova, R.; Onozuka, T.; et al. *Microbiol. Immunol.* **2005**, *49*, 695–701.
- (17) Salman, A.; Erukhimovitch, V.; Talyshinsky, M.; Huleihel, M.; Huleihel, M. *Biopolymers* **2002**, *67*, 406–412.
- (18) Erukhimovitch, V.; Talyshinsky, M.; Souprun, Y.; Huleihel, M. FTIR Microscopy Detection of Cells Infected With Viruses. In *DNA Viruses: Methods and Protocols*; Lieberman, P. M., Ed.; Humana Press: Totowa, NJ, 2005; Vol. 292, p 161.
- (19) Santos, M. C. D.; Nascimento, Y. M.; Monteiro, J. D.; et al. *Anal. Methods* **2018**, *10*, 1280–1285.
- (20) Agbaria, A. H.; Beck Rosen, G.; Lapidot, I.; et al. *Anal. Chem.* **2018**, *90*, 7888–7895.
- (21) Eccles, R. *Lancet Infect. Dis.* **2005**, *5*, 718–725.
- (22) Falsey, A. R.; Walsh, E. E. *Clin. Microbiol. Rev.* **2000**, *13*, 371–384.
- (23) Guan, W. J.; Ni, Z. Y.; Hu, Y.; et al. *N. Engl. J. Med.* **2020**, *382*, 1708–1720.
- (24) China National Guidelines for Diagnosis and Treatment of Corona Virus Disease 2019 (COVID-19). <http://www.nhc.gov.cn/yzygj/s7653p/202002/d4b895337e19445f8d728fcfa1e3e13a.shtml> (accessed Feb 15, 2020), Feb 8, 2020.
- (25) Zhang, X.; Guo, Y.; Song, Y.; et al. *Proteomics* **2006**, *6*, 5260–5268.
- (26) Agbaria, A. H.; Beck Rosen, G.; Lapidot, I.; et al. *Anal. Chem.* **2018**, *90*, 7888–7895.
- (27) Wenning, M.; Scherer, S. *Appl. Microbiol. Biotechnol.* **2013**, *97*, 7111–7120.
- (28) Zarnowicz, P.; Lechowicz, L.; Czerwonka, G.; Kaca, W. *Curr. Med. Chem.* **2015**, *22*, 1710–1718.
- (29) Carneiro, C. R.; Silva, C. S.; de Carvalho, M. A.; Pimentel, M. F.; Talhavin, M.; Weber, I. T. *Anal. Chem.* **2019**, *91*, 12444–12452.
- (30) Lee, L. C.; Liong, C.-Y.; Jemain, A. A. *Analyst* **2018**, *143*, 3526–3539.
- (31) Gromski, P. S.; Muhamadali, H.; Ellis, D. I.; et al. *Anal. Chim. Acta* **2015**, *879*, 10–23.
- (32) Mehmood, T.; Liland, K. H.; Snipen, L.; Sæbø, S. *Chemom. Intell. Lab. Syst.* **2012**, *118*, 62–69.
- (33) Déléris, G.; Petitbois, C. *Vib. Spectrosc.* **2003**, *32*, 129–136.
- (34) Naumann, D. *Appl. Spectrosc. Rev.* **2001**, *36*, 239–298.
- (35) Anderson, N. L.; Anderson, N. G. *Mol. Cell. Proteomics* **2002**, *1*, 845–867.
- (36) He, X. M.; Carter, D. C. *Nature* **1992**, *358*, 209–215.
- (37) Buijs, J.; Norde, W.; Lichtenbelt, J. W. T. *Langmuir* **1996**, *12*, 1605–1613.
- (38) Schroeder, H. W.; Cavacini, L. *J. Allergy Clin. Immunol.* **2010**, *125*, S41–S52.
- (39) Williams, J. H.; Kuchmak, M.; Witter, R. F. *Lipids* **1966**, *1*, 89–97.
- (40) Hidaka, H.; Yamauchi, K.; Ohta, H.; Akamatsu, T.; Honda, T.; Katsuyama, T. *Clin. Biochem.* **2008**, *41*, 1211–1217.
- (41) Wu, D.; Shu, T.; Yang, X.; et al. *Natl. Sci. Rev.* **2020**, *7*, 1157–1168.
- (42) Liu, K.-Z.; Shaw, R. A.; Man, A.; Dembinski, T. C.; Mantsch, H. H. *Clin. Chem.* **2002**, *48*, 499–506.
- (43) Triba, M. N.; Le Moyec, L.; Amathieu, R.; et al. *Mol. Biosyst.* **2015**, *11*, 13–19.
- (44) Galindo-Prieto, B.; Eriksson, L.; Trygg, J. *J. Chemom.* **2014**, *28*, 623–632.



- (45) Primpke, S.; Lorenz, C.; Rascher-Friesenhausen, R.; Gerdts, G. *Anal. Methods* **2017**, *9*, 1499–1511.
- (46) Liu, W.; Sun, Z.; Chen, J.; Jing, C. *J. Spectrosc.* **2016**, *2016*, 1603609.
- (47) Wang, F.; Nie, J.; Wang, H.; et al. *J. Infect. Dis.* **2020**, *221*, 1762–1769.
- (48) Qin, C.; Zhou, L.; Hu, Z.; et al. *Clin. Infect. Dis.* **2020**, *71*, 762–768.
- (49) Li, Q.; Ding, X.; Xia, G.; et al. *EClinicalMedicine* **2020**, *23*, 100375.
- (50) Huggins, C.; Jensen, E. V. *J. Biol. Chem.* **1949**, *179*, 645–654.
- (51) Vermeer, A. W. P.; Norde, W. *Biophys. J.* **2000**, *78*, 394–404.
- (52) Chalupa, P.; Beran, O.; Herwald, H.; Kaspříková, N.; Holub, M. *Infection* **2011**, *39*, 411–417.
- (53) Slaats, J.; ten Oever, J.; van de Veerdonk, F. L.; Netea, M. G. *PLoS Pathog.* **2016**, *12*, e1005973.
- (54) Shen, B.; Yi, X.; Sun, Y.; et al. *Cell* **2020**, *182*, 59–72.
- (55) Whetton, A. D.; Preston, G. W.; Abubeker, S.; Geifman, N. *J. Proteome Res.* **2020**, *19*, 4219–4232.
- (56) Zhang, B.; Zhou, X.; Zhu, C.; et al. *Front. Mol. Biosci.* **2020**, *7*, 157.
- (57) Chen, N.; Zhou, M.; Dong, X.; et al. *Lancet* **2020**, *395*, 507–513.
- (58) Heinrich, P. C.; Castell, J. V.; Andus, T. *Biochem. J.* **1990**, *265*, 621–636.