DOI 10.1007/s10812-023-01487-w
Journal of Applied Spectroscopy, Vol. 89, No. 6, January, 2023 (Russian Original Vol. 89, No. 6, November–December, 2022)

CLASSIFICATION OF CORONAVIRUS SPIKE PROTEINS BY DEEP-LEARNING-BASED RAMAN SPECTROSCOPY AND ITS INTERPRETATIVE ANALYSIS

Wenbo Mo,[a,b] Jiaxing Wen,[a,b] Jinglin Huang,[a] Yue Yang,[a] Minjie Zhou,[a] Shuang Ni,[a] Wei Le,[a] Lai Wei,[a] Daojian Qi,[a] Shaoyi Wang,[a] Jingqin Su,[a] Yuchi Wu,[a] Weimin Zhou,[a] Kai Du,[a] Xuewu Wang,[b] and Zongqing Zhao[a,*]

UDC 535.375.5:616-022.6

The outbreak of COVID-19 has spread worldwide, causing great damage to the global economy. Raman spectroscopy is expected to become a rapid and accurate method for the detection of coronavirus. A classification method of coronavirus spike proteins by Raman spectroscopy based on deep learning was implemented. A Raman spectra dataset of the spike proteins of five coronaviruses (including MERS-CoV, SARS-CoV, SARS-CoV-2, HCoV-HKU1, and HCoV-OC43) was generated to establish the neural network model for classification. Even for rapidly acquired spectra with a low signal-to-noise ratio, the average accuracy exceeded 97%. An interpretive analysis of the classification results of the neural network was performed, which indicated that the differences in spectral characteristics captured by the neural network were consistent with the experimental analysis. The interpretative analysis method provided a valuable reference for identifying complex Raman spectra using deep-learning techniques. Our approach exhibited the potential to be applied in clinical practice to identify COVID-19 and other coronaviruses, and it can also be applied to other identification problems such as the identification of viruses or chemical agents, as well as in industrial areas such as oil and gas exploration.

Keywords: *coronavirus, spike protein, Raman spectroscopy, deep learning, interpretative analysis.*

Introduction. Since December 2019, coronavirus disease 2019 (COVID-19) has spread rapidly around the world. As of 5 March 2022, the global death toll from COVID-19 is approaching six million (https://www.who.int/emergencies/diseases/novel-coronavirus-2019). The world is at a critical juncture of a pandemic, and there is an urgent need for convenient, precise, rapid, and large-scale deployable assay detection methods for coronavirus detection, especially for the outbreak control and prevention of COVID-19. As a highly specific detection technology, Raman spectroscopy has simple pretreatment requirements and a short detection time, providing a new method for the rapid and accurate classification of coronaviruses. Related studies confirmed that the Raman spectra of spike proteins can be used for the classification and clinical detection of coronaviruses [1–3]. The whole detection process, including pretreatment, can be controlled within 20 min in a rapid Raman mapping mode [2]. Compared with the mainstream detection methods (such as nucleic acid detection [4–8] and antibody detection [9–11]), Raman methods combine their advantages of high accuracy and speed.

The potential of Raman spectroscopy to identify the species of biomacromolecules goes beyond that, including the classification of bacteria, pathogens, and cancerous tumors [2, 12, 13], and more importantly, whether Raman spectroscopy can distinguish COVID-19 from other viruses. In addition to SARS-CoV-2, which has given rise to this outbreak of viral pneumonia, there is a total of six coronaviruses (HCoV-229E, HCoV-OC43, HCoV-NL63, HCoV-HKU1, SARS-CoV, and MERS-CoV) that can infect humans [14–20]. Common coronaviruses, including HCoV-229E, HCoV-OC43, HCoV-NL63, and HCoV-HKU1, usually cause mild to moderate upper respiratory illness, such as a runny nose, headache,

*To whom correspondence should be addressed.

[a]Laser Fusion Research Center at China Academy of Engineering Physics, Mianyang, China; email: 498427431@qq.com; [b]Department of Engineering Physics, Tsinghua University, Beijing, China. Abstract of article is published in Zhurnal Prikladnoi Spektroskopii, Vol. 89, No. 6, p. 903, November–December, 2022.

cough, sore throat, and fever [21]. MERS-CoV, SARS-CoV, and SARS-CoV-2 often lead to severe symptoms and even develop into pneumonia with a relatively high case fatality rate [22]. The coexistence of various coronaviruses is becoming the norm in the context of COVID-19. As the virus with the highest similarity to COVID-19, the accurate classification of coronavirus will lay the foundation for the future application of Raman spectroscopy in virus identification.

Owing to the small Raman scattering cross-section of the molecules and the impurity components of the biological samples, the typical signal-to-noise ratio (SNR) of Raman spectroscopy in a short measurement time is relatively low, which gives traditional classification methods (such as principal component analysis) a low identification accuracy for distinguishing samples with high similarity. The development of deep-learning technology makes it possible to extract effective feature signals from complex spectra, and it has been widely used in the research of various spectral classification problems [2, 12, 13].

A neural network model was trained to classify the Raman spectra of spike proteins of five coronaviruses, including MERS-CoV, SARS-CoV, SARS-CoV-2, HCoV-HKU1, and HCoV-OC43. The average accuracy of the spectra obtained in the measurement time of 0.5 s exceeded 97%. The spectral features affecting the judgment of the neural network are determined in the Discussion section. An interpretive analysis of the classification results of the neural network is presented, which indicates that the spectral differences captured by the neural network are consistent with structural differences in the spike proteins of these coronaviruses. Despite being a preliminary result, the experimental results show that Raman spectroscopy can be used to classify different kinds of viruses. The process of interpretative analysis proposed in this paper will provide inspiration for the classification approach of Raman spectra based on neural networks in the biological, chemical, and industrial fields.

Materials and Methods. The spike proteins of five different coronaviruses were purchased from Sino Biological Inc. (Beijing, China). These proteins were buffered by PBS or 20 mM PB, 300 mM NaCl, and 10% glycerol with a concentration of ~1 mg/mL. An ultrafiltration centrifuge process was performed to remove the buffer and make a pre-concentration for these proteins before the Raman measurement. The ultrafiltration tubes were purchased from Merck Millipore, with the nominal molecular weight limit of 30 kDa. The initial 50-μL protein solutions were centrifuged five times to a final volume of ~15 μL for each spike protein. The centrifugal process was implemented by a refrigerated centrifuge with a centrifugal force of 14,000×g at 4^{o}C for 25 min each time.

The normal Raman spectra of spike proteins were collected on Ag-coated Si substrates. The Ag coatings were deposited by direct current magnetron sputtering equipment. The Si slices used in this work were successively cleaned with acetone, ethanol, and deionized water (DI, 18.25 MΩ) before Ag deposition. The deposition processes were implemented in pure Ar at a working pressure of 0.5 Pa at room temperature under a base pressure of ~3.5×10^{-4} Pa. The Ag target power was set constant to 120 W with a deposition time of 333 s. The thickness of the deposited Ag coating was ~300 nm.

The Raman spectra were measured by a portable Raman spectrometer (Renishaw Virsa) equipped with a 532-nm laser. A wavenumber calibration was carried out using a Si sample before the measurement. The Raman signal was collected by a ×50 objective long focal lens. The mapping data were recorded by a rapid Raman mapping mode with an integration time of 0.5 s for a single spot and a laser power of 20 mW. Five 20 × 20 discrete spot maps were taken for the spike proteins of each coronavirus with 2-μm spacing between spots to avoid overlap between spectra. The spectral range was selected as 400–1750 cm^{-1}, which covers the majority of information on proteins. All the spectra were preprocessed with convolution filtering and background correction and normalized between 0 and 1.

The dataset consists of 10,000 normal Raman spectra from the spike proteins of five coronaviruses, including MERS-CoV, SARS-CoV, SARS-CoV-2, HCoV-HKU1, and HCoV-OC43. Then, the dataset was divided into training, validation, and testing datasets, with a ratio of 6:2:2, that is, the training set, validation set, and test set contain 6000, 2000, and 2000 Raman spectra, respectively. The number of spectra from each coronavirus that make up each subset is the same.

The neural network model used in this paper was a multilayer perceptron (MLP), in which several fully connected hidden layers were added between the output layer and the input layer. The size of the input layer was equal to the spectral length of 895. There were three fully connected hidden layers after the input layer, and the number of hidden neurons in each layer was 512, 256, and 32. The model ended with a fully connected classification layer, where the number of neurons was equal to the number of classes of the coronaviruses. The number of hidden layers and the number of hidden neurons in each layer were selected via a grid search using the training set and validation set. The test set was only used to report final accuracy and was not used in the model selection process. The more complex convolutional neural network (CNN) [23] and residual neural network (Resnet) [24] architectures were experimented, but found that their performances were almost the same, whereas the model complexity of MLP and the computational resources required for training were much smaller. The

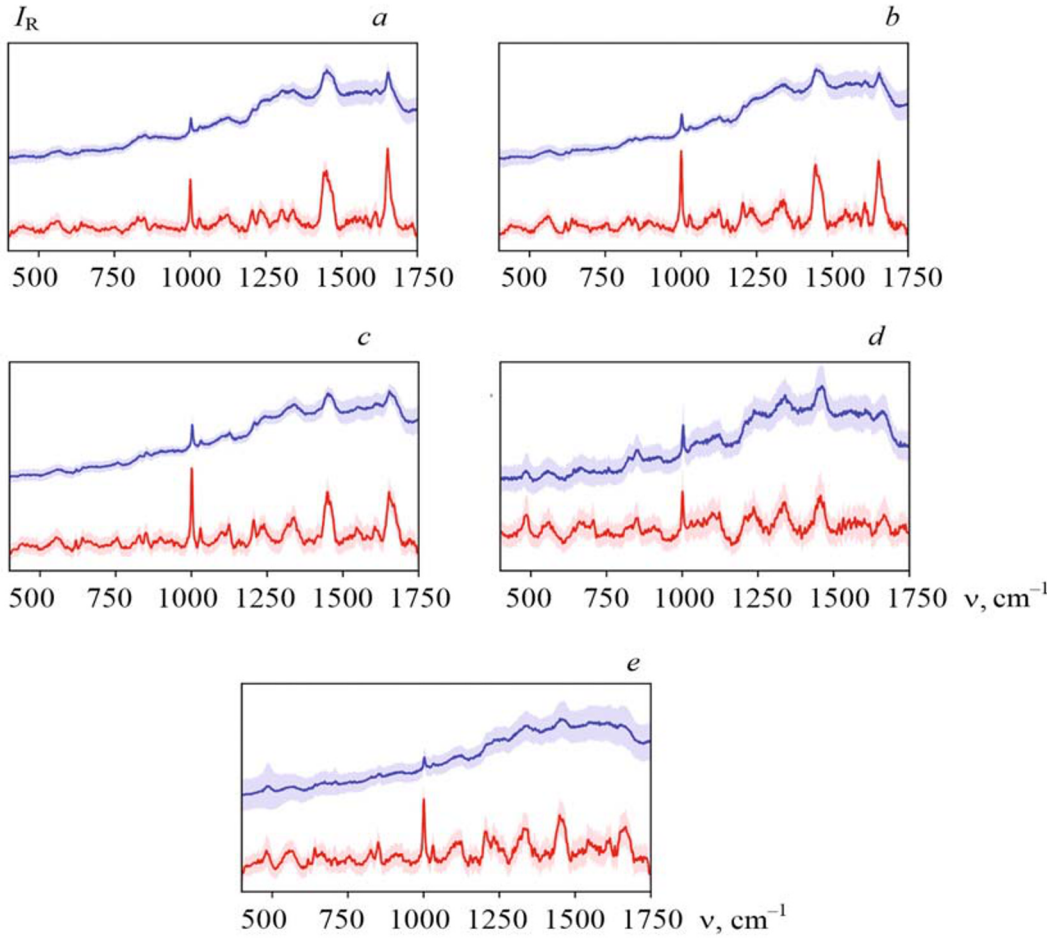

Fig. 1. Spectra of spike proteins of five coronaviruses: (a) MERS-CoV, (b) SARS-CoV, (c) SARS-CoV-2, (d) HCoV-HKU1, and (e) HCoV-OC43. In each subgraph, the blue line represents raw data, and the red line shows preprocessed data. The solid lines show the spectra after averaging all Raman spectra for each coronavirus spike protein. The shaded area shows the standard deviation of the measurement. All the spectra are normalized between 0 and 1.

stochastic gradient descent (SGD) optimizer was used in the training process. The learning rate was set to 0.01, the batch size was 100, and the epoch was set as 200.

Results and Discussion. According to the similarity of the genomic organization of coronaviruses, these seven coronaviruses can be divided into four groups. HCoV-229E and HCoV-NL63 genomes belong to the genus *Alphacoronavirus*, which show a higher sequence identity (63%) than other coronaviruses, and they were divided into one group. The other five coronaviruses belong to another genus, *Betacoronavirus*. HCoV-HKU1 and HCoV-OC43 both belong to the subgenus *Embecovirus* and were classified into one group. The SARS-CoV-2 genome shows 79.6% nucleotide sequence similarity with SARS-CoV, belonging to the same subgenus *Sarbecovirus*, and they were put in one group. Although MERS-CoV belongs to a single subgenus *Merbecovirus*, its genome is more analogous to SARS-CoV and SARS-CoV-2 than to other coronaviruses [22, 25, 26].

Coronavirus virions are spherical or pleomorphic, with club-like projections of the spike protein decorating the surface. The spike protein is distributed widely on the surface of the virus and serves as the key to attacking the human body, which can be considered one of the markers for coronavirus detection [27, 28]. Spike proteins of various coronaviruses have unique molecular compositional features that result in subtle differences in their corresponding Raman spectra. In clinical studies [1–3], special lysates were added to the collected samples to destroy the envelope of coronavirus and inactivate and release spike proteins. After that, the spike proteins can be detected by Raman spectroscopy for classification.

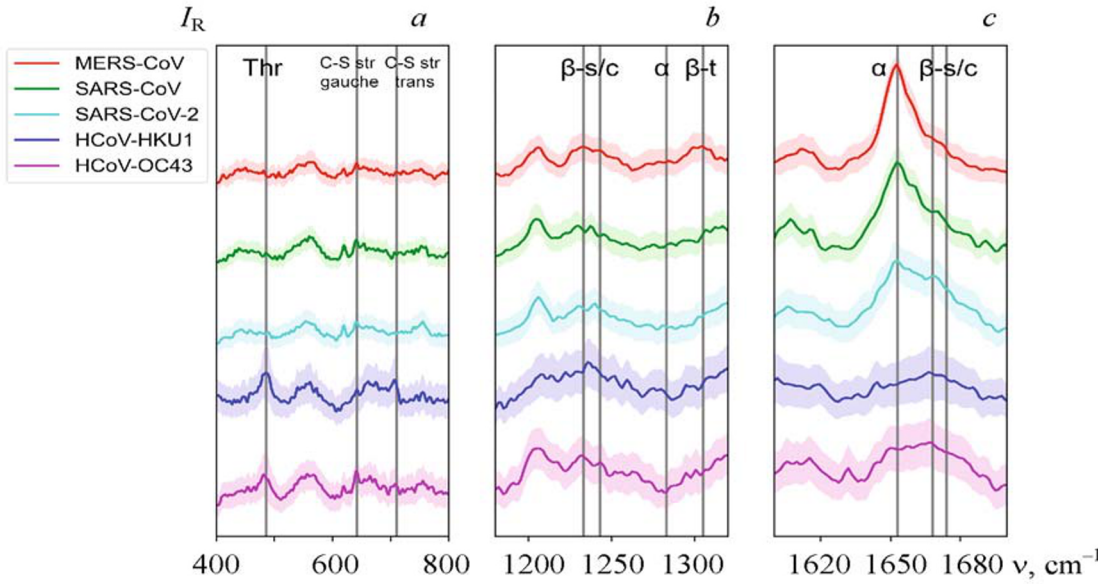

Fig. 2. Comparisons of the (a) lower wavenumber region and (b) the amide III and (c) amide I bands for five coronaviruses spike proteins. Thr: Threonine, str: stretch vibration, β-s/c: β-sheet/coil, β-t: β-turn, α: α-helix.

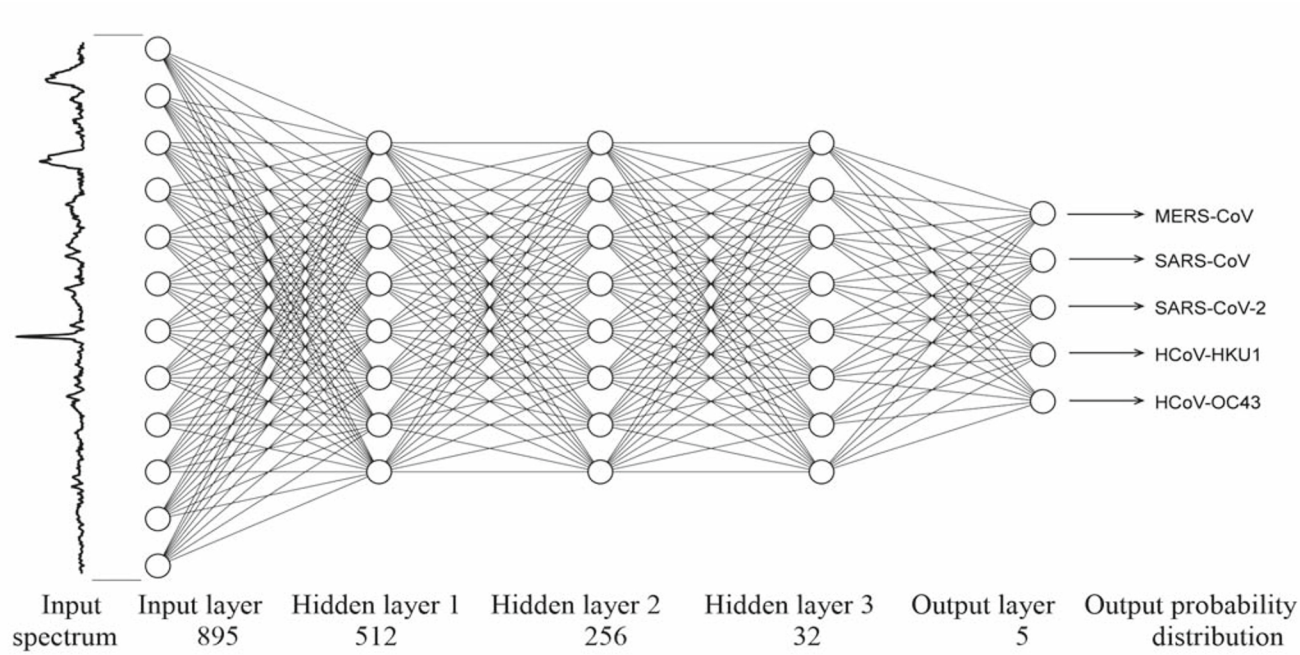

Fig. 3. Schematic diagram of the architecture of the multilayer perceptron model for the five-class identification task. As there are too many neurons in the actual model, the number of neurons shown in the figure is only schematic.

To demonstrate the effectiveness of our approach while simplifying the experimental cost, the five coronaviruses belonging to *Betacoronavirus* were selected as the objects of classification. Therefore, the dataset in this study contains the normal Raman spectra of the spike proteins of five coronaviruses, namely MERS-CoV, SARS-CoV, SARS-CoV-2, HCoV-HKU1, and HCoV-OC43. The mean spectra of the spike proteins of five coronaviruses are shown in Fig. 1.

There were some differences in Raman bands among the spike proteins of five coronaviruses, especially in a lower wavenumber region (400–800 cm^{-1}), amide I (1600–1700 cm^{-1}) band, and amide III (1200–1300 cm^{-1}) band, as shown in Fig. 2. In the lower wavenumber regions, the spike proteins of HCoV-HKU1 and HCoV-OC43 showed more intense Raman bands than those of MERS-CoV, SARS-CoV, and SARS-CoV-2, which reflected the difference in the Raman bands of the side chains. This may be due to the different amino acid sequences of the latter three coronavirus spike proteins, as they have higher amino acid homology. According to the Raman band assignment presented in Huang et al. [2], the amide I and amide III bands of MERS-CoV, SARS-CoV, and SARS-CoV-2 showed more alpha-helical components than those of beta-sheets/coils, whereas HCoV-HKU1 and HCoV-OC43 exhibited the opposite trend [2]. These differences in the Raman bands were caused by the different conformations of spinous process proteins in the gene sequence. The above results showed that the spike proteins of different coronaviruses had diverse Raman bands, indicating that Raman spectroscopy is a valid technology for identifying coronaviruses. In the next part, deep learning methods were applied to automatically capture these differences in spectra and implement the classification of coronaviruses.

The process of classifying coronavirus spike proteins from the Raman spectra of the spike proteins using an MLP model is illustrated in Fig. 3. The 895-dimensional spectra were input into MLP, and the probability distribution of five coronaviruses was finally output in the output layer after the nonlinear operation of three hidden layers.

The neural network model was trained to complete the five-class identification task. A one-hot encoded label was used when labeling the training data of five coronaviruses, which meant that the character label was expressed as a vector. MERS-CoV was expressed as [1 0 0 0 0], SARS-CoV as [0 1 0 0 0], SARS-CoV-2 as [0 0 1 0 0], HCoV-HKU1 as [0 0 0 1 0], and HCoV-OC43 as [0 0 0 0 1]. The MLP output a probability distribution across the five coronaviruses, and the maximum was taken as the predicted class. For instance, when the MLP output of a testing spectrum is [0.1 0.1 0.1 0.6 0.1], the spectrum will be considered to come from the fourth coronavirus, namely HCoV-HKU1.

The performance breakdown for individual classes is displayed in the confusion matrix. All values equal to 0 are not show.

True \ Predicted	MERS-CoV	SARS-CoV	SARS-CoV-2	HCoV-HKU1	HCoV-OC43
MERS-CoV	98.52		1.48		
SARS-CoV		99.39	0.61		
SARS-CoV-2			98.3		1.7
HCoV-HKU1				96.21	3.79
HCoV-OC43		0.42	2.64	2.78	94.17

The average accuracies of MERS-CoV, SARS-CoV, SARS-CoV-2, HCoV-HKU1, and HCoV-OC43 were 98.52 ± 0.74, 99.39 ± 0.49, 98.3 ± 1.63, 96.21 ± 0.69, and 94.17 ± 0.24%, respectively. The overall average accuracy reached 97.32 ± 0.76%. These five coronaviruses can be divided into two groups: MERS-CoV, SARS-CoV, and SARS-CoV-2 in one group, with HCoV-HKU1 and HCoV-OC43 in the other group. There were few misclassifications among different groups of coronaviruses but more misclassification within groups, confirming their higher spectral similarity.

The more common classification technique of principal component analysis (PCA) was also implemented. Only 67.16% accuracy was achieved when two principal components were retained. The accuracy reached 93% when the number of retained principal components exceeded 20. However, the improvement in accuracy was not obvious when the number of principal components continued to increase, which indicated that the spectra of the five coronaviruses were highly similar and could not be distinguished by a few features. In contrast, our neural network comprehensively utilized all the characteristic information in the full wavenumber range of the spectra, which is helpful to greatly improve the classification accuracy.

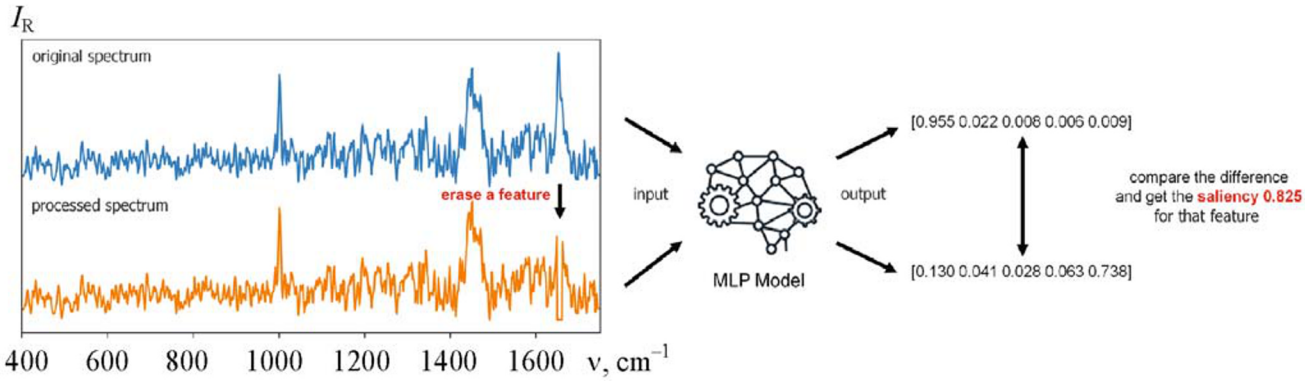

Fig. 4. An example of the process of calculating the saliency of a spectral feature.

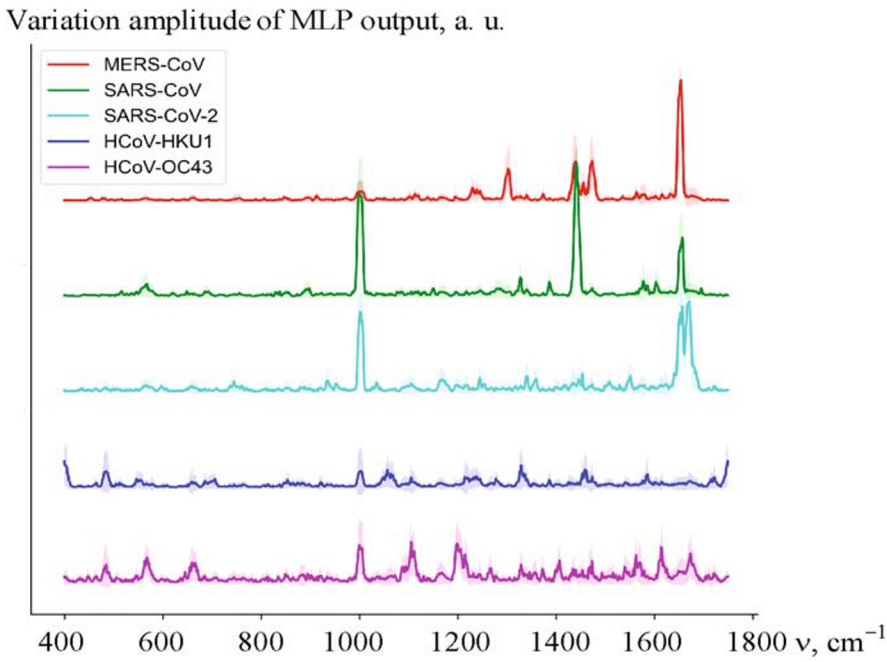

Fig. 5. The mean saliency map of the spectra of the five coronaviruses.

The interpretation of the output of neural networks has always been the focus of attention. Therefore, it is necessary to analyze which characteristics of these spectra of coronavirus spike proteins affect the discrimination results of neural network output. The normalized intensity of each wavenumber and the four wavenumbers before and after that (a total of nine wavenumbers) was set to zero, and then it was inputted into the MLP model again for prediction. When a certain feature in the spectrum disappears, the influence on the judgment of the neural network was analyzed by the process. The variation amplitude of the output was compared to reveal the influence of the spectral features in this wavenumber range on the neural network prediction. An example of this process is shown in Fig. 4.

Despite some differences, the term saliency map [29] in the field of deep learning was employed to call such results. In our saliency map, the abscissa is the wavenumber, and the ordinate is the variation amplitude of the MLP output, that is, the saliency. The mean saliency map of the spectra of MERS-CoV, SARS-CoV, SARS-CoV-2, HCoV-HKU1, and HCoV-OC43 are shown in Fig. 5.

The shaded area shows the standard deviation of all spectra in the test dataset.

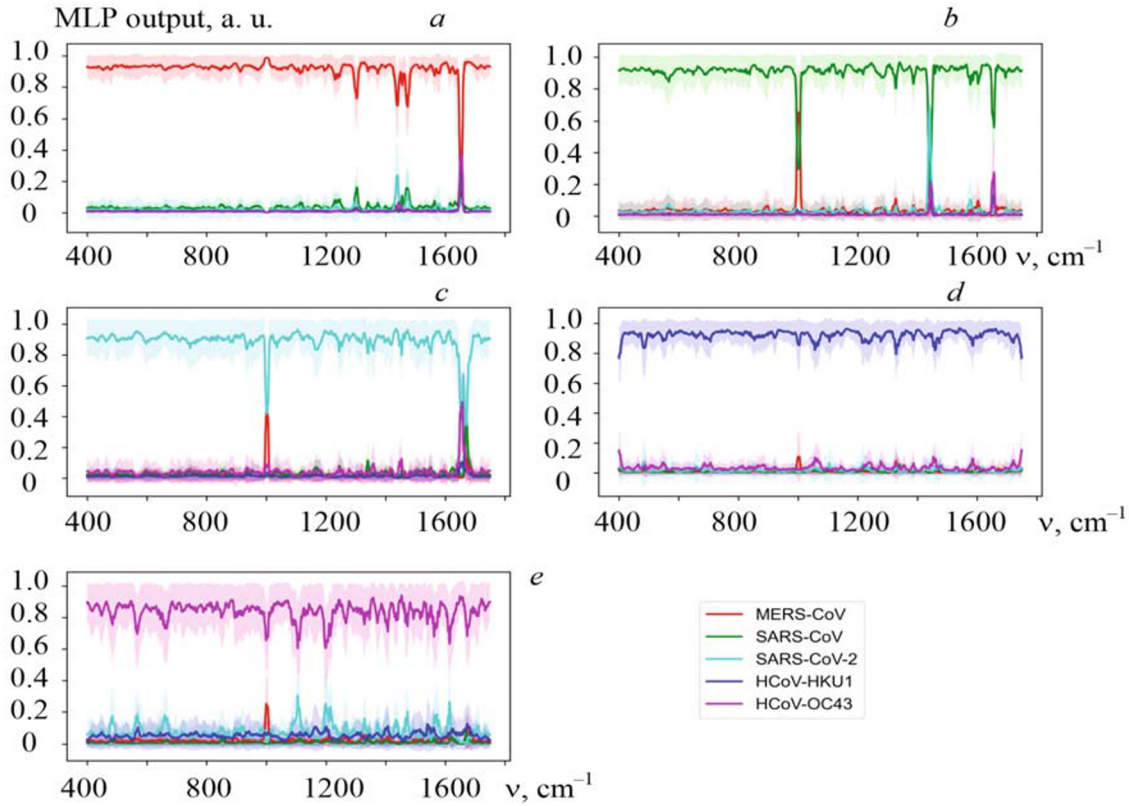

Fig. 6. The mean multilayer perceptron output of the spectra processed as shown in Fig. 4 from (a) MERS-CoV, (b) SARS-CoV, (c) SARS-CoV-2, (d) HCoV-HKU1, and (e) HCoV-OC43. The shaded area shows the standard deviation of all spectra in the test dataset.

From the characteristics of the saliency distribution, MERS-CoV, SARS-CoV, and SARS-CoV-2 were relatively similar, whereas HCoV-HKU1 and HCoV-OC43 were relatively similar. In the saliency map of MERS-CoV, SARS-CoV, and SARS-CoV-2, only two or three positions showed very high saliency, and the saliency in most wavenumber ranges was almost zero. This illustrates the relatively high degree of similarity among MERS-CoV, SARS-CoV, and SARS-CoV-2, with only a few differences in the spectra. HCoV-HKU1 and HCoV-OC43 show certain saliency in many positions, but were not particularly high, which showed that the spectra of HCoV-HKU1 and HCoV-OC43 were more different from those of the other three coronaviruses. Therefore, a single feature alone cannot determine the MLP output, which was consistent with the results of the classification accuracy of the neural network above.

This interpretative analysis method is also helpful to explore the specific position of spectral differences between different coronaviruses. The MLP output of the spectra processed as shown in Fig. 4 was directly plotted for each coronavirus, as shown in Fig. 6. As mentioned earlier, the output is a five-dimensional vector; hence, there are five curves in each subgraph. To some extent, these five components represented the probabilities that the input spectra were derived from the five coronaviruses. Therefore, these five components were directly called MERS-CoV, SARS-CoV, SARS-CoV-2, HCoV-HKU1, and HCoV-OC43 in the legends. The corresponding output component of coronavirus was always the highest in each subgraph. In most wavenumber ranges, the output changed little when the features were erased. However, the elimination of some features led to a sharp decrease in the component and an increase in the component corresponding to another coronavirus. These features are the main difference between the two coronaviruses.

The differences in spectral characteristics between the spike proteins of different coronaviruses are demonstrated in Table 1. The difference at approximately 1003 cm^{-1} arose from the phenylalanine component in the side chains of the five coronavirus spike proteins. Therefore, it will not be discussed separately in the following comparisons.

The differences in spectral characteristics between MERS-CoV and SARS-CoV were mainly distributed in the regions of wavenumbers 1300, 1460, and 1650 cm^{-1}, which was due to the difference in beta-turn components of

TABLE 1. Differences in Spectral Characteristics between the Spike Proteins of Different Coronaviruses

Wavenumber of characteristic differences, cm^{-1}	MERS-CoV	SARS-CoV	SARS-CoV-2	HCoV-HKU1	HCoV-OC43
MERS-CoV	–	1300, 1460, 1650	1440	1650	1650
SARS-CoV	1003	–	1450	1650	1450, 1650
SARS-CoV-2	1003, 1660	1660	–	1650	1650
HCoV-HKU1	1003	1340	490	–	560, 710, 1240, 1450, 1580
HCoV-OC43	1003, 1660	1660	490, 560, 1100, 1210, 1580, 1610	560, 1170 1210, 1390, 1560, 1670	–

amide I, alanine components in side chains, and alpha-helical components of amide III [2]. The spectral characteristics near wavenumber 1650 cm^{-1} were also the difference between MERS-CoV and SARS-CoV-2. Additionally, the spectral characteristics near 1440 cm^{-1} were also one of the differences between MERS-CoV and SARS-CoV-2, and was attributed to CH_2 bending vibrations in the backbone [30]. For SARS-CoV and SARS-CoV-2, their differences in spectral characteristics were located near the two positions of wavenumbers 1440 and 1650 cm^{-1}.

There were many differences in spectral features between HCoV-HKU1 and HCoV-OC43. At a lower wavenumber region, the differences were mainly concentrated at 490, 560, and 710 cm^{-1}, which was caused by the threonine, glycine, amide IV, isoleucine, and C–S stretch vibrations [22]. In the middle and high wavenumber regions, the possible assignments of the Raman bands that present differences were C–N stretch vibrations, tyrosine, histidine, phenylalanine, amide III, tryptophan, CH_2 bending vibration, and amide I [30].

The results were in good agreement with the difference observed directly in Fig. 2. The peaks at 1003, 1450, 1460, and 1650 cm^{-1} were the main characteristics of the spectra of the spike proteins of these coronaviruses. The intensity and shape of these peaks in different coronaviruses intuitively showed some differences. Moreover, our approach also identified more abundant differences that were so small that they drowned in the noise and were difficult to observe directly. The results showed that deep-learning methods had a strong ability to extract subtle features in Raman spectroscopy, which improved the detection sensitivity. The sample composition is more complex in future clinical applications, and only with the help of the method can the effective characteristic signal be extracted accurately and the correct classification be realized.

It should be noted that the amount of spectral data in this study is relatively limited. However, as a preliminary result, the feasibility of this approach was proved to a considerable extent. In further research, it is necessary to obtain more Raman spectra of various coronaviruses in different states and environments and build a comprehensive and reliable database to make better use of the advantages of deep learning in processing big data.

Conclusions. A deep-learning method for Raman spectra was applied to identify the spike protein of the coronavirus. As shown in our dataset of five coronaviruses, the average classification accuracy exceeded 97%. The classification results of our model were consistent with the spectral analysis through the interpretive analysis of saliency maps, which showed that this method and the classification results were reliable in biochemical foundations. This interpretative analysis method will provide a valuable reference for the future use of deep learning to identify complex Raman spectra.

The approach can easily be transplanted to the detection of COVID-19 by collecting real positive and negative samples of throat swabs or saliva to train the neural network model. In clinical applications, surface-enhanced Raman scattering can be applied to reduce the measurement time and further accelerate the detection speed [30–33]. Moreover, our model can be applied to other identification problems with high requirements for detection speed and accuracy, such as the identification of viruses or chemical agents, as well as in industrial areas such as oil and gas exploration. Combined with our proposed explanatory analysis method, it is helpful to evaluate the reliability of deep-learning classification results.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (No. 11975214) and the Science Challenge Program (Project No. TZ2018005).

REFERENCES

1. F. Cui and H. S. Zhou, *Biosens. Bioelectron.*, **165**, Article ID 112349 (2020).
2. J. Huang, J. Wen, M. Zhou, S. Ni, W. Le, G. Chen, L. Wei, Y. Zeng, D. Qi, M. Pan, *Anal. Chem.*, **93**, No. 26, 9174–9182 (2021).
3. M. Asif, M. Ajmal, G. Ashraf, N. Muhammad, A. Aziz, T. Iftikhar, J. Wang, and H. Liu, *Curr. Opin. Electrochem.*, **23**, 174–184 (2020).
4. M. Shen, Y. Zhou, J. Ye, A. A. A. Al-Maskri, Y. Kang, S. Zeng, and S. Cai, *J. Pharm. Anal.*, **10**, 97–101 (2020).
5. V. M. Corman, O. Landt, M. Kaiser, R. Molenkamp, A. Meijer, D. K. W. Chu, T. Bleicker, S. Brünink, J. Schneider, M. L. Schmidt, et al., *Eurosurveillance*, **25**, No. 3, Article ID 2000045 (2020).
6. M. N. Esbin, O. N. Whitney, S. Chong, A. Maurer, X. Darzacq, and R. Tjian, *RNA*, **26**, No. 7, 771–783 (2020).
7. J. Huggett, K. Dheda, S. Bustin, and A. Zumla, *Genes Immun.*, **6**, 279–284 (2005).
8. M. L. Wong and J. F. Medrano, *Biotechniques*, **39**, 75–85 (2005).
9. D. Jacofsky, E. M. Jacofsky, and M. Jacofsky, *J. Arthroplasty*, **35**, S74–S81 (2020).
10. Q. Lin, D. Wen, J. Wu, L. Liu, W. Wu, X. Fang, and J. Kong, *Anal. Chem.*, **92**, 9454–9458 (2020).
11. G. Liu and, J. F. Rusling, *ACS Sens.*, **6**, 593–612 (2021).
12. C.-S. Ho, N. Jean, C. A. Hogan, L. Blackmon, S. S. Jeffrey, M. Holodniy, N. Banaei, A. A. Saleh, S. Ermon, and J. Dionne, *Nat. Commun*, **10**, 1–8 (2019).
13. W. Lee, A. T. Lenferink, C. Otto, and H. L. Offerhaus, *J. Raman Spectrosc.*, **51**, 293–300 (2020).
14. D. Wrapp, N. Wang, K. S. Corbett, J. A. Goldsmith, C.-L. Hsieh, O. Abiona, B. S. Graham, and J. S. McLellan, *Science*, **367**, 1260–1263 (2020).
15. Z. Li, A. C. Tomlinson, A. H. Wong, D. Zhou, M. Desforges, P. J. Talbot, S. Benlekbir, J. L. Rubinstein, and J. M. Rini, *eLife*, **8**, Article ID e51230 (2019).
16. S. Abdul-Rasool and B. C. Fielding, *Open Virol. J.*, **4**, 76 (2010).
17. J. R. St-Jean, H. Jacomy, M. Desforges, A. Vabret, F. Freymuth, and P. J. Talbot, *J. Virol.*, **78**, 8824–8834 (2004).
18. P. C. Woo, S. K. Lau, C.-M. Chu, K.-H. Chan, H.-W. Tsoi, Y. Huang, B. H. Wong, R. W. Poon, J. J. Cai, and W.-K. Luk, *J. Virol.*, **79**, 884–895 (2005).
19. C. Drosten, S. Günther, W. Preiser, S. Van Der Werf, H.-R. Brodt, S. Becker, H. Rabenau, M. Panning, L. Kolesnikova, and R. A. Fouchier, *N. Engl. J. Med.*, **348**, 1967–1976 (2003).
20. R. J. De Groot, S. C. Baker, R. S. Baric, C. S. Brown, C. Drosten, L. Enjuanes, R. A. Fouchier, M. Galiano, A. E. Gorbalenya, and Z. A. Memish, *J. Virol.*, **87**, 7790–7792 (2013).
21. S. Su, G. Wong, W. Shi, J. Liu, A. C. Lai, J. Zhou, W. Liu, Y. Bi, and G. F. Gao, *Trends Microbiol.*, **24**, 490–502 (2016).
22. M. M. Kesheh, P. Hosseini, S. Soltani, and M. Zandi, *Rev. Med. Virol.*, Article ID e2282 (2021).
23. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, *Proc. IEEE*, **86**, 2278–2324 (1998).
24. K. He, X. Zhang, S. Ren, and J. Sun, *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 770–778 (2016).
25. D. X. Liu, J. Q. Liang, and T. S. Fung, *Encycl. Virol.*, 428 (**2021**).
26. A. B. Gussow, N. Auslander, G. Faure, Y. I. Wolf, F. Zhang, and E. V. Koonin, *Proc. Natl. Acad. Sci. USA*, **117**, 15193–15199 (2020).
27. H. Yao, Y. Song, Y. Chen, N. Wu, J. Xu, C. Sun, J. Zhang, T. Weng, Z. Zhang, and Z. Wu, *Cell*, **183**, 730–738, Article ID e713 (2020).
28. C. Liu, Y. Yang, Y. Gao, C. Shen, B. Ju, C. Liu, X. Tang, J. Wei, X. Ma, and W. Liu, *bioRxiv* (**2020**).
29. K. Simonyan, A. Vedaldi, and A. Zisserman, *arXiv preprint arXiv*, Article ID 1312.6034 (2013).
30. C. Carlomagno, D. Bertazioli, A. Gualerzi, S. Picciolini, P. Banfi, A. Lax, E. Messina, J. Navarro, L. Bianchi, and A. Caronni, *Sci. Rep.*, **11**, 1–13 (2021).
31. H. Jinglin, Z. Minjie, L. Wei, C. Guo, N. Shuang, N. Gao, L. Zeyu, Z. Zongqing, H. Zhibing, and L. Bo, *Qiangjiguang Yu Lizishu*, **32**, 069001–069002 (2020).
32. J. E. Sanchez, S. A. Jaramillo, E. Settles, J. J. V. Salazar, A. Lehr, J. Gonzalez, C. R. Aranda, H. R. Navarro-Contreras, M. O. Raniere, and M. Harvey, *RSC Adv.*, **11**, 25788–25794 (2021).
33. G. Yin, L. Li, S. Lu, Y. Yin, Y. Su, Y. Zeng, M. Luo, M. Ma, H. Zhou, and L. Orlandini, *J. Raman Spectrosc.*, **52**, 949–958 (2021).