



香港城市大學
City University of Hong Kong

專業 創新 胸懷全球
Professional · Creative
For The World

CityU Scholars

Staging classification of omicron variant SARS-CoV-2 infection based on dual-spectrometer LIBS (DS-LIBS) combined with machine learning

WANG, Weiliang; SHI, Shengqun; LIU, Yuanchao; HOU, Zehai; QI, Jianwei; GUO, Lianbo

Published in:
Optics Express

Published: 04/12/2023

Document Version:
Final Published version, also known as Publisher's PDF, Publisher's Final version or Version of Record

Publication record in CityU Scholars:
[Go to record](#)

Published version (DOI):
[10.1364/OE.504640](https://doi.org/10.1364/OE.504640)

Publication details:
WANG, W., SHI, S., LIU, Y., HOU, Z., QI, J., & GUO, L. (2023). Staging classification of omicron variant SARS-CoV-2 infection based on dual-spectrometer LIBS (DS-LIBS) combined with machine learning. *Optics Express*, 31(25), 42413-42427. <https://doi.org/10.1364/OE.504640>

Citing this paper

Please note that where the full-text provided on CityU Scholars is the Post-print version (also known as Accepted Author Manuscript, Peer-reviewed or Author Final version), it may differ from the Final Published version. When citing, ensure that you check and use the publisher's definitive version for pagination and other details.

General rights

Copyright for the publications made accessible via the CityU Scholars portal is retained by the author(s) and/or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights. Users may not further distribute the material or use it for any profit-making activity or commercial gain.

Publisher permission

Permission for previously published items are in accordance with publisher's copyright policies sourced from the SHERPA RoMEO database. Links to full text versions (either Published or Post-print) are only available if corresponding publishers allow open access.

Take down policy

Contact lbscholars@cityu.edu.hk if you believe that this document breaches copyright and provide us with details. We will remove access to the work immediately and investigate your claim.



Staging classification of omicron variant SARS-CoV-2 infection based on dual-spectrometer LIBS (DS-LIBS) combined with machine learning

WEILIANG WANG,^{1,†} SHENGQUN SHI,^{1,†} YUANCHAO LIU,²
ZEHAU HOU,¹ JIANWEI QI,^{3,4} AND LIANBO GUO^{1,5} 

¹Wuhan National Laboratory for Optoelectronics (WNLO), Huazhong University of Science and Technology, Wuhan, Hubei 430074, China

²Department of Physics, City University of Hong Kong, Hong Kong, China

³State Key Laboratory of Experimental Hematology, National Clinical Research Center for Blood Diseases, Haihe Laboratory of Cell Ecosystem, Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Tianjin 300020, China

⁴qjianwei@ihcams.ac.cn

⁵lguo@hust.edu.cn

[†]These authors contributed equally to this work.

Abstract: Effective differentiation of the infection stages of omicron can provide significant assistance in transmission control and treatment strategies. The combination of LIBS serum detection and machine learning methods, as a novel disease auxiliary diagnostic approach, has a high potential for rapid and accurate staging classification of Omicron infection. However, conventional single-spectrometer LIBS serum detection methods focus on detecting the spectra of major elements, while trace elements are more closely related to the progression of COVID-19. Here, we proposed a rapid analytical method with dual-spectrometer LIBS (DS-LIBS) assisted with machine learning to classify different infection stages of omicron. The DS-LIBS, including a broadband spectrometer and a narrowband spectrometer, enables synchronous collection of major and trace elemental spectra in serum, respectively. By employing the RF machine learning models, the classification accuracy using the spectra data collected from DS-LIBS can reach 0.92, compared to 0.84 and 0.73 when using spectra data collected from single-spectrometer LIBS. This significant improvement in classification accuracy highlights the efficacy of the DS-LIBS approach. Then, the performance of four different models, SVM, RF, IGBT, and ETree, is compared. ETree demonstrates the best, with cross-validation and test set accuracies of 0.94 and 0.93, respectively. Additionally, it achieves classification accuracies of 1.00, 0.92, 0.92, and 0.89 for the four stages B1-acute, B1-post, B2, and B3. Overall, the results demonstrate that DS-LIBS combined with the ETree machine learning model enables effective staging classification of omicron infection.

© 2023 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

1. Introduction

Coronavirus disease 2019 (COVID-19) is one of the worst pandemics in human history, caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). In a matter of a few years, over 6 million lives have been lost for the devastating impact of COVID-19, leaving a profound impact on the global economy [1,2]. The emergence of the Omicron (B.1.1.529) variant of SARS-CoV-2 occurred in November 2021 and then became the predominant strain of the pandemic in 2022, resulting from stronger transmission ability and insidiously [3]. In this stage of the COVID-19 epidemic, it is not only crucial to detect the presence of the Omicron Variant of SARS-CoV-2, but also essential to determine the stage of infection. This diagnosis of infection stages is vital for

targeted treatment and managing the overall status of population infection, which could improve the treatment rate of patients and reduce the pressure of epidemic prevention and control [4]. Specifically, the emergence of novel and evolving variants of SARS-CoV-2 poses a serious threat to the COVID-19 pandemic control and disease treatment, which means rapid, sensitive, and powerful diagnostic technologies are needed, especially the methods for diagnosis of different infection periods [5].

Many diagnostic techniques have been developed for the detection of SARS-CoV-2 infection, including real-time reverse transcription polymerase chain reaction (rRT-PCR) and immunochromatographic assay (ICA). However, these methods are primarily employed to test throat or nasopharyngeal swab samples for determining the presence of infection, rather than accurately assessing the stage of infection [5,6]. In contrast, blood-based tests provide an effective assessment of the extent of SARS-CoV-2 infection such as detecting antibodies in the blood by ICA, which can indicate previous infection and immune response [7,8]. Furthermore, as the duration of SARS-CoV-2 infection progresses, there are dynamic fluctuations in the levels of certain blood elements [9]. Trace elements such as zinc, iron, and copper are crucial for immune response regulation. In COVID-19 infections, fluctuations in the levels of these trace elements are often observed, particularly during severe cases, varying with the progression of infection [10–13]. While no definitive theory directly links potassium, calcium, sodium, and magnesium to immune response or COVID-19 infection, it is worth noting that these major elements exhibit abnormal fluctuations as COVID-19 infection progresses [9,11,14]. Therefore, monitoring these elements in serum can reflect the evolving nature of the infection and provide valuable insights for monitoring the disease's course, assessing treatment efficacy, and understanding the patient's overall condition, thereby facilitating better prevention and control of the COVID-19 pandemic.

Currently, mass spectrometry, including inductively coupled plasma mass spectrometry, is highly regarded for accurate serum element detection, allowing simultaneous analysis of over 60 elements [15]. Additionally, atomic absorption spectroscopy and flame emission spectroscopy are also employed in the elemental detection of serum [9]. However, the above methods are difficult to widely use in clinical testing because of the disadvantages of complex, expensive, and rigorous environmental requirements. In contrast, laser-induced breakdown spectroscopy (LIBS) is an emerging analytical technique, which utilizes laser-induced plasma to generate emission spectra, enabling the identification and quantification of elements in a sample [16,17]. LIBS offers several advantages, including simplicity, rapid analysis, and minimal sample preparation, which benefits blood testing and related applications [18–20]. Analyzing the LIBS spectral information of blood allows for both qualitative and quantitative analysis of elements within the blood. Additionally, employing machine learning techniques to analyze the data can further enable the classification and identification of certain diseases [21–24]. For example, Chu et al. realized the diagnosis of nasopharyngeal carcinoma and a variety of blood cancers through the extreme learning machine and random subspace method, in which the detection accuracy rate is more than 95% [25,26]. Yue et al. developed a method for distinguishing normal individuals, ovarian cyst patients, and ovarian cancer patients, and the sensitivity and specificity of LIBS serum diagnosis were 71.4% and 86.5% [27]. The above works demonstrated that harnessing LIBS for serum examination, coupled with machine learning for analysis, presents a viable avenue for distinguishing different diseases. However, current studies focus on distinguishing individual diseases or multiple distinct diseases, and most studies solely detected major elements within the serum, the collected spectra rarely exhibit spectral lines of trace elements [28,29]. Staging classification of LIBS Omicron infection based on rapid detection has not been reported. In the context of LIBS detection for Omicron infection, if the spectra of trace elements can be detected as analytical features, it would undoubtedly enhance the accuracy of the analysis results. Therefore, how to collect the major and trace elemental spectra in serum is very important for diagnosing different infection stages of Omicron.

In this study, we proposed a rapid analytical method with dual-spectrometer LIBS (DS-LIBS) assisted with machine learning to classify different infection stages of Omicron. Compared to traditional single-spectrometer LIBS that are limited to focusing on either major or trace elemental spectra, DS-LIBS utilizes a broadband spectrometer to collect the constant elemental spectra and a narrowband spectrometer to collect trace elemental spectra separately, resulting in merged spectra with a greater amount of valuable information. The merged spectra were obtained from DS-LIBS that simultaneously encompass both major and trace elements within the serum. Then both the individual spectra and the merged spectra were subjected to machine learning models to demonstrate the classification accuracy improvement by the DS-LIBS. Finally, different machine learning models were employed to classify the new spectrum data for selecting the most suitable model for diagnosing different stages of Omicron infection.

2. Methods

2.1. Materials and sample pretreatment

Human serum samples from 172 patients were collected from donors drawn by the Institute of Hematology and Blood Diseases Hospital. The composition of samples and related class labels is shown in Table 1.

Table 1. The composition of samples

Class label	Sample category	Sample size
B1_Acute	Acute Infectious period	7
B1_Post	Post-acute Infectious period	60
B2	1 months Follow-up	62
B3	3 months Follow-up	43

B1: sixty-seven samples were collected after patients were diagnosed as Omicron-infected and admitted to the hospital. Of these, 7 were collected immediately after admission, known as the acute infectious period, and the remaining 60 were collected after treatment and recovery, known as the post-acute infectious period.

B2 and B3: the remaining 105 samples were collected from patients who were infected with Omicron and recovered, 62 samples of which were collected 1 month after discharge according to hospital standards, and 43 samples were collected 3 months after discharge.

The gender ratio of patients in each group is close to 1:1, as is the ratio of patients with or without underlying conditions. Additionally, the age distribution of patients within each group is consistent, mainly ranging from 20 to 60 years old. These settings aim to eliminate the influence of these variables on the results.

All samples were inactivated and stored in a -80 °C refrigerator until detection. Informed consent has been obtained for experiments on human subjects, and the study is approved by the ethics committee at the Institute of Hematology and Blood Diseases Hospital (ethical approval number: HHL2022005-EC-1).

2.2. DS-LIBS setup

Figure 1(a) shows the LIBS experimental setup: a laser beam from a Q-switched Nd:YAG laser (wavelength: 532 nm; pulse energy: 200 mJ; repetition rate: 10 Hz; pulse width: 8 ns; Nimma 400, Beamtech, China) passed the reflector and focusing mirror (focal length: 150 mm) onto the sample surface. The Motorized positioning systems (Y110TA150, Jiangyun, China) could move the sample on the horizontal plane, to make each laser pulse onto a fresh surface. The plasma emission was gathered by a light collector and then transmitted by fiber to the spectrometer. All instruments are controlled by Digital Delay Generator.

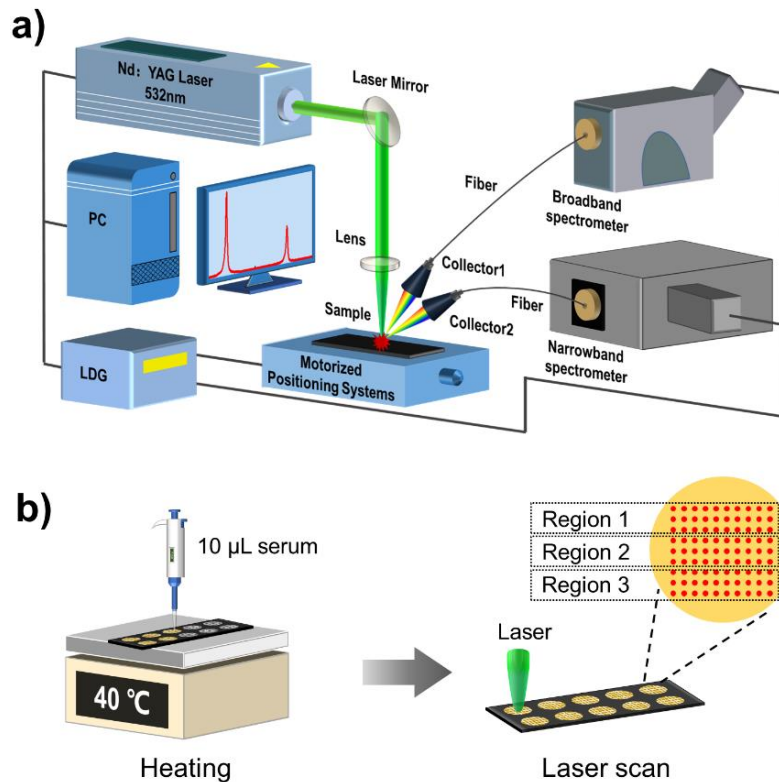


Fig. 1. a) The schematic diagram of the DS-LIBS setup, b) the serum liquid-solid conversion diagram and the laser scan mode

In most LIBS serum detection studies, a broadband spectrometer is used for LIBS spectral acquisition, allowing for the simultaneous collection of a larger amount of elemental information and facilitating faster qualitative and quantitative analysis [28,29]. Nevertheless, this collection strategy fails to capture the spectra of trace elements effectively. While the sodium content in serum can reach 3000 mg/L, the levels of trace elements such as zinc, iron, and copper are only around 1 mg/L. This implies that the spectral intensity of major elements will be much higher than that of trace elements. Consequently, it becomes challenging to simultaneously detect the spectra of major and trace elements without saturating the detector.

So the DS-LIBS utilizes two spectrometers, one designed for major element detection, utilizing a broadband spectrometer (Mechelle 5000, Andor Tech., United Kingdom) that covers characteristic spectral lines of all major elements in blood, and another narrowband spectrometer (Shamrock 500i, Andor Tech., United Kingdom) for trace element detection, featuring high resolution, sensitivity, adjustable slit size, and multiple interchangeable gratings to accommodate the varying signal strengths of trace elemental spectra. Because Mechelle 5000 offers a broad spectral coverage due to the characteristics of mid-range echelle gratings, and Shamrock 500i is a type of M-shaped Czerny-Turner spectrograph that can collect weak spectral signals in narrow wavelength regions. They are all coupled with intensified charge coupled device (ICCD) (iStar DH-334T-18F-E3, Andor Tech., United Kingdom). Through temporal control, the two spectrometers are synchronized to capture plasma emission spectra simultaneously. Subsequently, the spectral data was denoised using the wavelet transform function “wden” in MATLAB, and normalization was applied to both spectrometers’ data using MATLAB’s Min-Max technique to ensure measurement consistency for better fusion. The major elemental spectra and trace

elemental spectra are then extracted and concatenated to form new spectra containing both major and trace elements. After optimization, the parameters of the spectrometer and ICCD are set as shown in Table 2.

Table 2. The parameters of each spectrometer and ICCD

Option	ME5000 (Broadband spectrometer)	Shamrock (Narrowband spectrometer)
Exposure time	0.1 s	0.1 s
Gate width	30.0 μ s	20.0 μ s
Gate delay	1.5 μ s	2.0 μ s
Number of accumulations	15	15
Gain	1500	1500
Slit size	50 μ m \times 50 μ m	1 mm \times 100 μ m
Wavelength range	200 nm – 900 nm	Any 6 nm width in the 200 nm – 600 nm range

All serum samples underwent a liquid-solid transformation on the substrate, and then LIBS spectra were collected. The ordered microarray silicon substrate (OMSS) was utilized in this work, which has been described in detail in our previous work [21]. Figure 1(b) shows the process of the liquid-solid transformation involved, adding 10 μ L of serum drops to OMSS, and then placing it on the 40 $^{\circ}$ C heating table for 1 minute to dry to solid.

Figure 1(b) shows the laser scan mode for one sample, within a sample area, the laser scans a total of 90 points in a 9×10 grid, with every 3 rows forming a scanning region, and a total of three scanning regions. During data collection in these three regions, the acquisition parameters of the broadband spectrometer remain consistent, while the wavelength range of data collection for the narrowband spectrometer varies. Region 1 corresponds to the Zn collection area, where the narrowband spectrometer captures data from 213 nm to 219 nm, Region 2 is the Fe collection area with a wavelength range of 235 nm to 241 nm, and Region 3 is the Cu collection area, capturing data from 324 nm to 329 nm. Other parameters of the narrowband spectrometer remain unchanged, except for the differing wavelength ranges. In each region, using the “accumulate” function, compiling the spectra from 15 points into one spectrum. Each sample resulted in six constant element spectra and two spectra each for trace elements Zn, Fe, and Cu. Then the spectra from the three regions were averaged (for major elements spectra) or merged (for trace elements spectra). Using this collection method, a total of 344 spectra were obtained from the broadband spectrometer and 344 spectra from the narrowband spectrometer. Among them, each sample consists of 4 spectra, with 2 spectra from the broadband spectrometer and 2 spectra from the narrowband spectrometer.

2.3. Algorithm description

Random forest (RF) is an ensemble learning algorithm used for classification and regression problems. It builds multiple decision trees and introduces randomness in feature and data selection, enhancing stability and accuracy. Ultimately, the predictions from multiple decision trees are combined through voting or averaging to synthesize the final prediction result. RF is widely utilized in classification and recognition applications of LIBS, such as tumor detection, food adulteration identification, and bacterial detection [25].

Support vector machine (SVM) is a supervised learning algorithm used for classification and regression problems. It finds the optimal hyperplane in the feature space to maximize the margin between different classes, achieving efficient classification. SVM is suitable for both linear and nonlinear problems and performs well in high-dimensional data. It is also a commonly single classifier model used in LIBS classification applications [30].

Gradient boosting decision tree (GBDT) is also an ensemble learning model based on decision trees, primarily achieved through iterative training of multiple decision trees and correcting the prediction errors from the previous iteration to enhance the model's performance. This gradual improvement of predictive performance results in a final prediction that is a weighted combination of the weak learners [31].

Extremely randomized trees (ETree) is similar to Random Forest but with an increased level of randomness during the construction of decision trees and an ensemble learning algorithm. ETree not only randomizes feature selection but also the selection of split points, which reduces overfitting and performs well with datasets having many features. Furthermore, ETree parameter tuning is simple, and utilizing random features and split points also accelerates training speed [32].

Due to the different measurement dimensions of the two spectrometers, there is a challenge in the process of data fusion. In this work, we choose tree-based models (RF, GBDT, ETree) to study the influence of dimension. Different from the probability model, the tree-based model is less sensitive to the feature dimension, so it is not affected by the absolute scale of the feature when selecting nodes [33]. In addition, the ensemble learning model has been proven to have excellent performance in spectral classification tasks [34]. As one of the widely-used models in LIBS data classification, SVM was chosen to compare with other models [35]. The Fig. 2 outlines the entire process from data acquisition to classification through DS-LIBS.

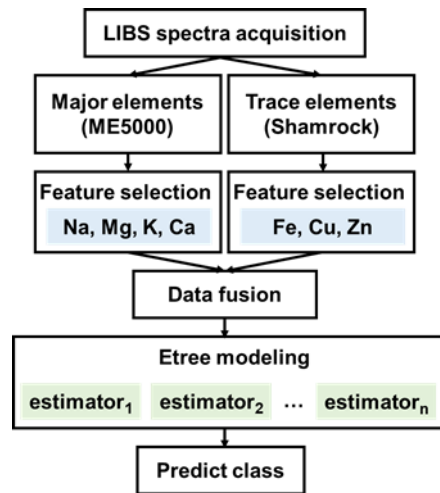


Fig. 2. The flowchart of the DS-LIBS data process

The samples from different infection stages were randomly divided into training and testing sets, following a ratio close to 8:2. For example, among the 60 samples in the B1 group, 48 samples were allocated to the training set and 12 samples to the testing set, and ten-fold cross-validation is applied within the training set for validation.

2.4. Classification evaluation index

The evaluation metrics used for classification effect evaluation are listed below.

Accuracy (Acc) measures the overall correctness of a classifier's predictions:

$$Acc = \frac{\sum_{i=1}^n (TP_i + TN_i)}{\sum_{i=1}^n (TP_i + FP_i + TN_i + FN_i)} \quad (1)$$

Precision indicates the proportion of correctly identified positive cases, which is crucial in avoiding false positive diagnoses:

$$Precision = \frac{1}{n} \sum_{i=1}^n \frac{TP_i}{(TP_i + FP_i)} \quad (2)$$

Recall measures the ability to correctly identify positive cases out of all the actual positive cases, which is important for minimizing false negative diagnoses:

$$Recall = \frac{1}{n} \sum_{i=1}^n \frac{TP_i}{(TP_i + FN_i)} \quad (3)$$

F1-Score balances precision and recall, providing a comprehensive measure of a classifier's performance:

$$F1_Score = \frac{2 * (Precision * Recall)}{(Precision + Recall)} \quad (4)$$

in which n represents the total number of categories, and i represents the specific category. TP_i (True Positive) refers to the number of instances that were correctly classified as positive by the model for Class i . TN_i (True Negative) represents the number of instances that were correctly classified as negative by the model for Class i . FP_i (False Positive) signifies the number of instances that were incorrectly classified as positive by the model for Class i . FN_i (False Negative) corresponds to the number of instances that were incorrectly classified as negative by the model for Class i .

3. Result and discussion

3.1. Spectra analysis and classification by the traditional LIBS

Figure 3 shows the serum spectra from 200 nm - 900 nm collected by the broadband spectrometer, including the spectral lines of metallic elements K, Ca, Na, Mg, and non-metallic elements C, H, O, and N. The spectral peaks contained in the spectra of each class are consistent, but there are certain differences in the intensity of each spectral peak. Figure 4 shows the serum's trace elemental spectra collected by the narrowband spectrometer, including the spectral lines of metal elements Zn, Fe, and Cu, which could not be collected by our broadband spectrometer. The narrowband spectrometer coverage range is only 6 nm, but after setting appropriate parameters, the sensitivity is very high, and the spectral lines of trace elements that are not detected by the broadband spectrometer can be measured. Metallic elements' spectra are derived from serum, as these elements are nearly absent in air and silicon substrates. On the other hand, non-metallic elements C, H, O, and N are both present in serum and air, resulting in non-metallic elemental spectra containing information from both serum and air.

The RF was used as the classification model, and the broadband spectrometer data and the narrowband spectrometer data were used as the data set. Before using the spectra from the two datasets, feature extraction was performed to extract the peak intensity of the spectra for both major elements and trace elements as features. The classification accuracy could reach 0.84 and 0.73, respectively. The specific classification evaluation index value and Confusion matrix are shown in Fig. 5. When using the broadband spectrometer data to classify patients from different periods, the evaluation metrics show a large difference between the training set and the test set, which indicates the insufficient generalization ability of the model. One possible reason is that in the spectrum collected by the broadband spectrometer, the elements that can be observed are major elements. These elements are more likely to change significantly in the blood during larger changes in the body, such as before and after Omicron infection. And during the infection process, their changes may not be so obvious. When using the narrowband spectrometer data for

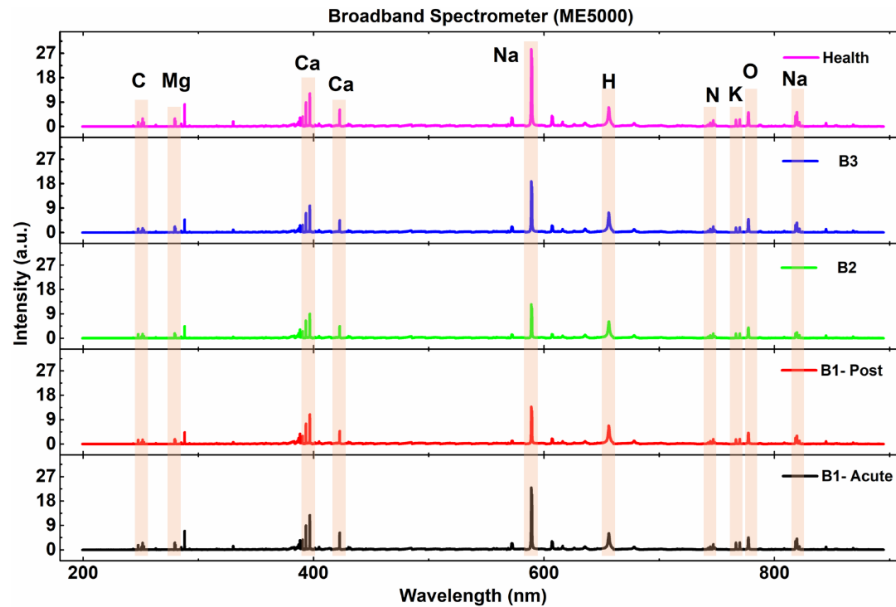


Fig. 3. The serum spectra of patients at different stages of Omicron infection collected by the broadband spectrometer

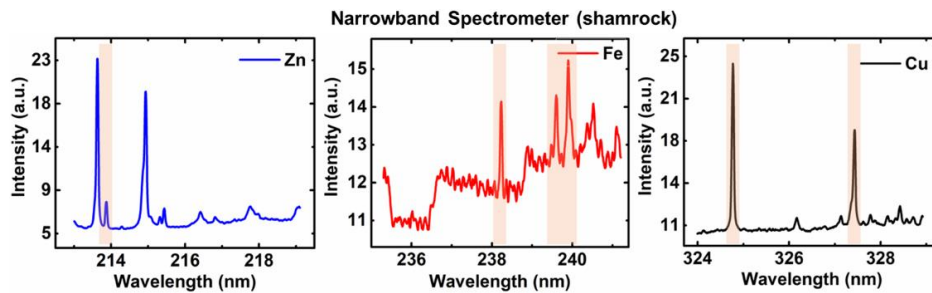


Fig. 4. The serum spectra of a patient in the Post-acute Infectious period collected by the narrowband spectrometer

classification, the evaluation metrics between the training and testing sets are relatively similar. However, the classification accuracy when using the narrowband spectrometer data is lower compared to using the broadband spectrometer data. This indicates that there is indeed some difference in the trace elemental spectra detected in the serum of patients with different infection stages, but the information contained is still insufficient to effectively distinguish patients from various infection stages.

The data of the whole spectrum without feature selection also had been used for classification. The accuracy using the whole spectrum data from the broadband spectrometer and the narrowband spectrometer was only 0.74 and 0.64, respectively, which indicates poor classification performance. The reason is that the full spectral data is used instead of data after feature selection, which may cause the curse of dimensionality, the data becomes sparse and the generalization ability is reduced.

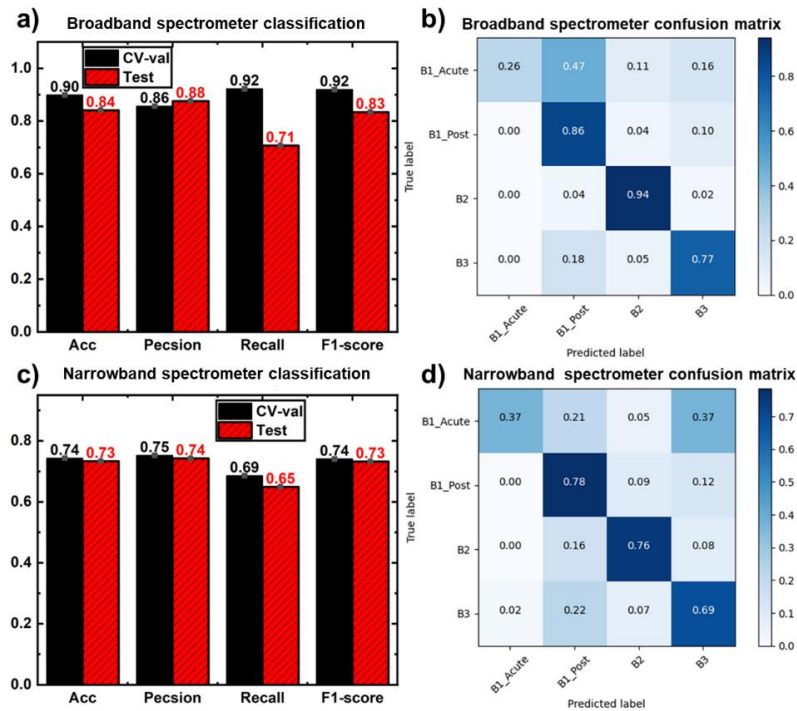


Fig. 5. a), c) Classification evaluation indexes and b), d) Confusion matrix for Classification using the broadband spectrometer and the narrowband spectrometer data

3.2. Spectrum detection and classification by DS-LIBS

To further improve the identification of different Omicron infection periods, feature selection was carried out on the previously used spectra. Considering that only the spectra of K, Ca, Na, and Mg could be determined to be completely derived from serum, we selected the spectra of these elements from the broadband spectrometer data as features to be used. The trace elements Zn, Fe, and Cu have certain links with Omicron infection and disease development, the narrowband spectrometer data also be selected. Figure 6 provides a concise overview of the whole data processing. By employing dual spectrometers, simultaneous acquisition of broadband spectral and narrowband spectral data is achieved. Subsequently, by comparing with the NIST database, major elemental spectra and trace elemental spectra are selectively chosen. After completing the feature selection process, data feature-level fusion is performed to obtain merged spectral data. Finally, machine learning models are employed to classify new data, ultimately achieving the discrimination of patients at different stages of Omicron infection.

Spectral fusion involves integrating spectral information from different bands into a single dataset to extract comprehensive and accurate information. There are three commonly employed methods of information fusion in the processes of information processing and decision-making: Data-level fusion, feature-level fusion, and decision-level fusion. Feature-level fusion offers the advantage of combining diverse and complementary information from multiple sources or methods, resulting in enhanced and robust feature representations. This leads to improved data quality, better handling of noise and missing data, clearer decision boundaries, and the ability to leverage prior knowledge, ultimately enhancing the effectiveness of analysis and decision-making processes. Therefore, by comparing the NIST database, we carried out feature selection on the spectral data from the two spectrograms and selected the bands after baseline subtraction where

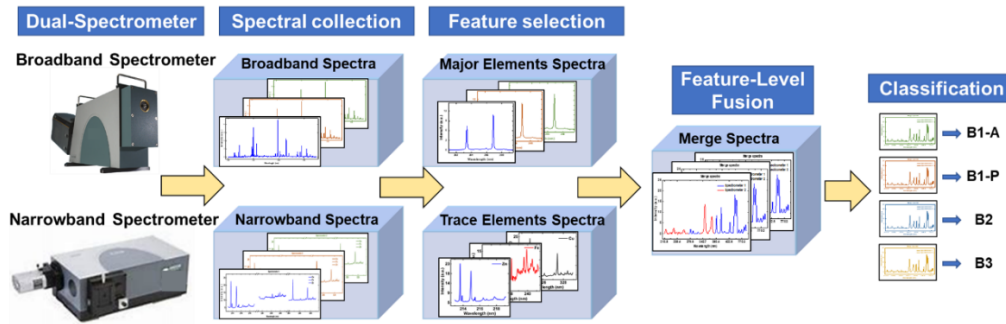


Fig. 6. The process of combining DS-LIBS with machine learning models for classification

the spectral lines of metal elements in K, Ca, Na, Mg, Zn, Fe, and Cu. Subsequently, the spectra from the two spectrometers are individually normalized, and finally, they are concatenated to obtain the new merged spectral data. The merged spectrum after Feature-Level Fusion is shown in Fig. 7.

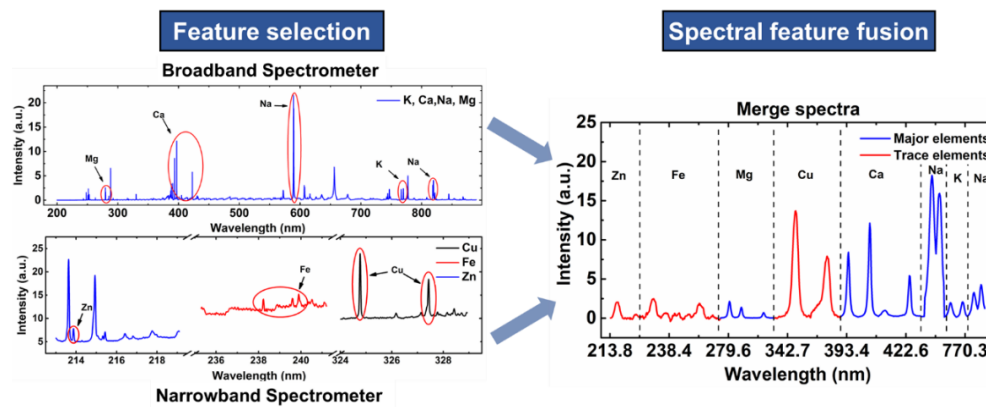


Fig. 7. The Feature-Level Fusion for spectra information

The random forest model is used as the classification model to classify the merged spectra data. Figure 8 shows the classification evaluation index and confusion matrix. The indicators from Fig. 8(a) suggest that the model performs similarly on the test set and training set, indicating good generalization capability. From Fig. 8(b), it is evident that the classification effect is best for the B1-Acute group, with an accuracy of 0.99, while the other three groups have accuracies around 0.90 and there is mutual misclassification among the three groups. This could be because, at the onset of infection, the composition of the human body's internal environment undergoes significant fluctuations, while during the recovery phase and post-recovery, most indicators gradually return to normal. In Table 3, the classification evaluation index values after using the random forest to classify the three data sets are uniformly displayed. The accuracy of the test set for classification after using merged spectra data has been changed from 0.84 to 0.92, while the precision, recall, and F1 index have all improved. The overall performance of the model is better. It can be seen from the classification results of spectra collected by the narrowband spectrometer that the spectra of trace elements also have the ability to distinguish different periods. In the merged spectral data, both major elemental spectra and trace elemental spectra are simultaneously included. This richer and more precise feature information can reflect a broader

range of information can reflect a broader range of human health condition data. As a result, it can better differentiate the serum of patients at different stages of Omicron infection.

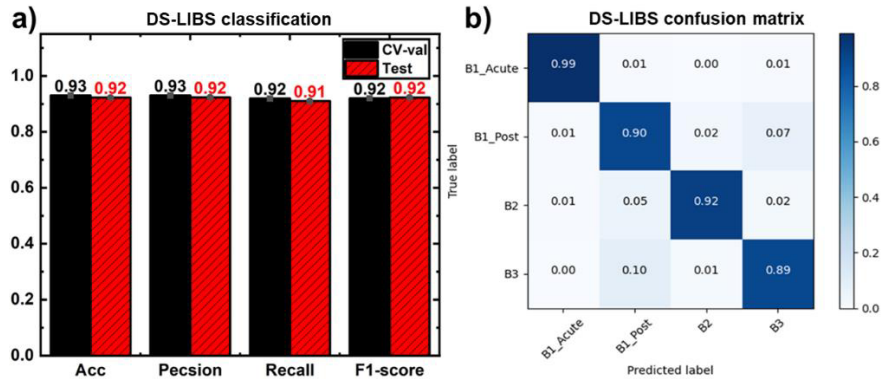


Fig. 8. a) Evaluation indexes and b) confusion matrix for classification using merged spectra.

Table 3. Evaluation index values for classification using various spectra^a

Source of classified data	ACC _{CV}	Precision _{CV}	Recall _{CV}	F1 _{CV}	ACC _T	Precision _T	Recall _T	F1 _T
Broadband spectrometer	0.90	0.86	0.92	0.92	0.84	0.88	0.71	0.83
Narrowband spectrometer	0.74	0.75	0.69	0.74	0.73	0.74	0.65	0.73
DS-LIBS	0.93	0.93	0.92	0.92	0.92	0.92	0.91	0.92

^aCV means Cross-Validation, T means test data.

3.3. Comparison of different machine learning models

To get a higher classification effect, SVM, GBDT, and ETree models are used to classify and recognize the merged spectra data. Figure 9 shows the confusion matrix and the ROC curves for each of the three models. The three confusion matrices indicate that the B1-Acute group continues to have the highest classification accuracy, with other groups achieving around 0.9 accuracy, though some misclassification among them exists. The B1-Acute group’s classification results show no missed diagnoses and a very low rate of false positives. Particularly under the ETree model, compared to the results in section 3.2, the primary improvement is that the samples belonging to the B-Acute class are mostly correctly classified. The three ROC curves show that all models perform well in classifying the various categories, with AUC values all above 0.99, especially the ETree model. The correlation diagram of RF is given in Fig. 8. Table 4 compares the evaluation indicators of the four models, including RF. The classification effect of the merged spectra data is very close under different models, which indicates that the data set itself is of high quality, and different models can obtain sufficient sample information from it, learn similar rules, and thus obtain similar classification performance. Among them, ETree is the better model with higher accuracy and better generalization ability. The results indicate that when dealing with DS-LIBS spectral data, employing ensemble learning models based on decision trees leads to better classification performance, because they are not affected by the absolute scale of the feature from the different measurement dimensions of the two spectrometers when selecting nodes.

The classification accuracy of the ETree model can reach 0.94, and Fig. 10 shows that the classification accuracies for the four stages B1-acute, B1-post, B2, and B3 are 1.0, 0.92, 0.92, and 0.89, respectively. The above results prove that by using DS-LIBS to obtain merged spectra

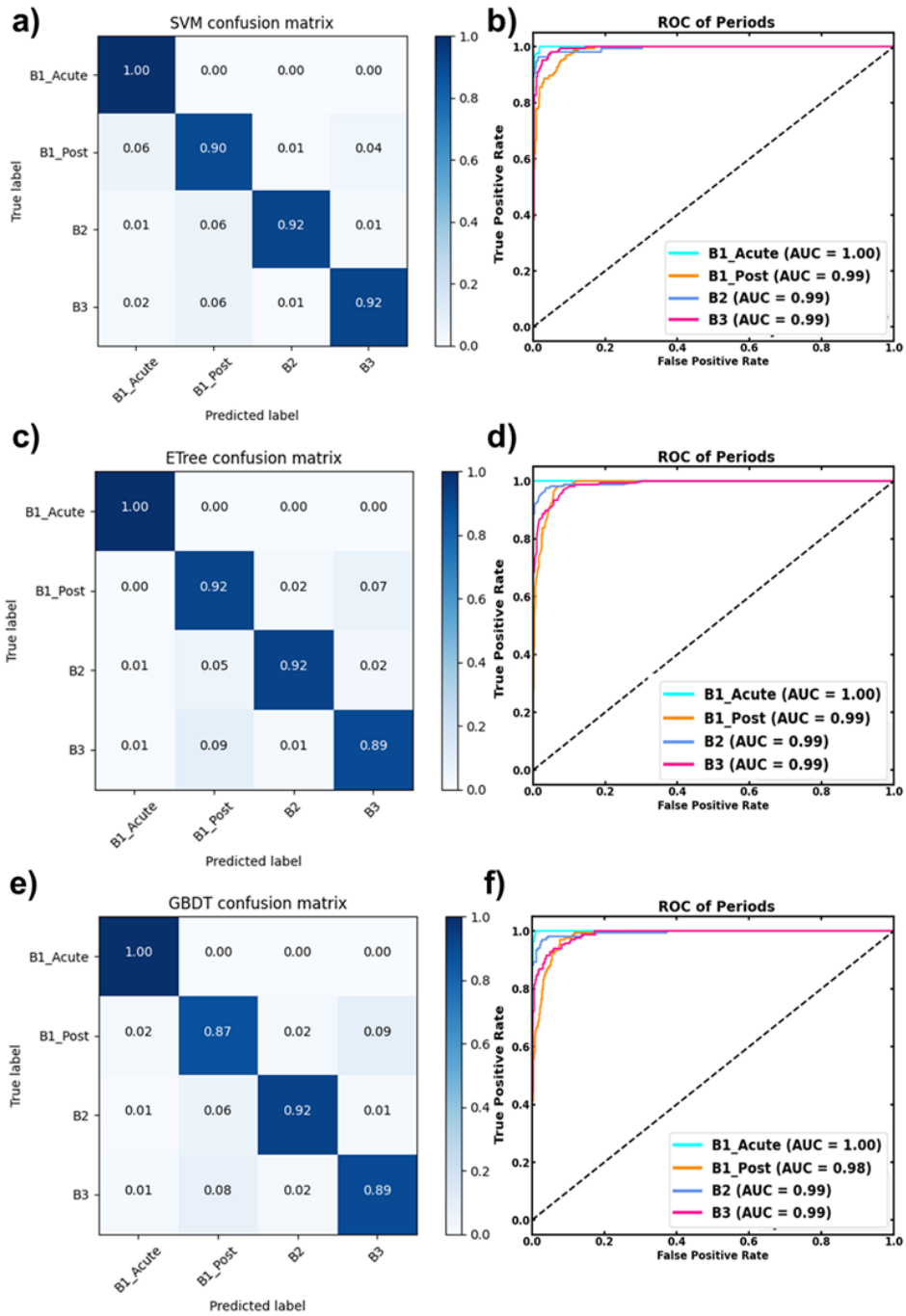


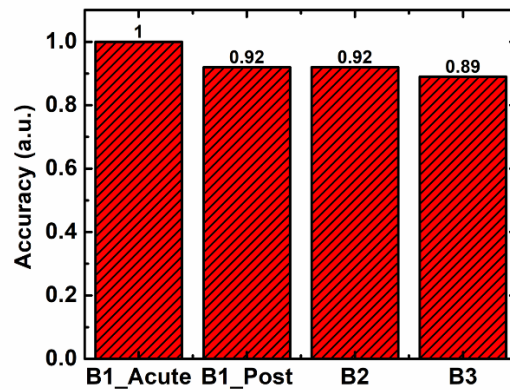
Fig. 9. a), c), e) Classification evaluation indexes and b), d), f) ROC curve of the different Classification model

Table 4. Evaluation index values for each classification model^a

	ACC _{cv}	Precision _{cv}	Recall _{cv}	F1 _{cv}	ACC _T	Precision _T	Recall _T	F1 _T
RF	0.93	0.93	0.92	0.92	0.92	0.92	0.91	0.92
SVM	0.91	0.92	0.91	0.91	0.93	0.93	0.93	0.93
GBDT	0.93	0.93	0.93	0.93	0.92	0.92	0.92	0.92
ETree	0.94	0.94	0.94	0.94	0.93	0.93	0.93	0.93

^aNote: _{cv} means Cross-Validation, _T means test data.

containing major elements and trace elements, combined with appropriate machine learning models, it is possible to effectively distinguish serum from patients at different periods of Omicron infection.

Classification accuracy across different infection stages**Fig. 10.** The discrimination accuracy of ETree for four infection stages

4. Conclusion

To diagnose patients in different stages of Omicron infection, we proposed a rapid serum analytical method by combining dual-spectrometer LIBS (DS-LIBS) with machine learning models. Unlike traditional single-spectrometer LIBS, which had difficulty in simultaneously detecting both major and trace elements in serum samples, DS-LIBS utilized a broadband spectrometer to obtain major elemental spectra and a narrowband spectrometer to collect trace elemental spectra, enabling the collection of more spectral information for staging classification.

Firstly, we separately used the broadband spectrometer and the narrowband spectrometer for staging the classification of Omicron infection, in which the classification accuracies by the RF model were 0.84 and 0.73, respectively, showing unsatisfactory results. Then, we used the DS-LIBS to collect and merge the spectra from the same serum samples and the classification accuracy was enhanced to 0.92 by using the RF model, demonstrating the accuracy enhancement by the DS-LIBS in staging classification of Omicron infection. Subsequently, four machine learning algorithms (RF, SVM, GBDT, and Etree) were performed for the staging classification using DS-LIBS. Etree exhibited accuracy rates of 0.94 and 0.93 on the training and testing sets, respectively. It achieved classification accuracies of 1.0, 0.92, 0.92, and 0.89 for the B1-acute, B1-post, B2, and B3 stages, respectively. These results surpassed the performance of RF, SVM, and GBDT. All in all, the DS-LIBS combined with the Etree model can effectively classify

the different stages of Omicron infection using serum, demonstrating the great potential of the DS-LIBS in auxiliary clinical diagnosis.

Funding. National Natural Science Foundation of China (62375101, 62075069).

Disclosures. The authors declare that there are no conflicts of interest related to this article.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. W. J. Wiersinga, A. Rhodes, A. C. Cheng, *et al.*, "Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19): A Review," *JAMA* **324**(8), 782–793 (2020).
2. C. Sohrabi, Z. Alsafi, N. O'Neill, *et al.*, "World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19)," *Int. J. Surg.* **76**, 71–76 (2020).
3. H. Shuai, J. F. Chan, B. Hu, *et al.*, "Attenuated replication and pathogenicity of SARS-CoV-2 B.1.1.529 Omicron," *Nature* **603**(7902), 693–699 (2022).
4. D. Focosi, M. Franchini, L. A. Pirofski, *et al.*, "COVID-19 Convalescent Plasma and Clinical Trials: Understanding Conflicting Outcomes," *Clin. Microbiol. Rev.* **35**(3), e0020021 (2022).
5. Q. Fernandes, V. P. Inchakalody, M. Merhi, *et al.*, "Emerging COVID-19 variants and their impact on SARS-CoV-2 diagnosis, therapeutics and vaccines," *Ann. Med.* **54**(1), 524–540 (2022).
6. M. Nagura-Ikeda, K. Imai, S. Tabata, *et al.*, "Clinical Evaluation of Self-Collected Saliva by Quantitative Reverse Transcription-PCR (RT-qPCR), Direct RT-qPCR, Reverse Transcription-Loop-Mediated Isothermal Amplification, and a Rapid Antigen Test To Diagnose COVID-19," *J. Clin. Microbiol.* **58**(9), e01438 (2020).
7. Z. He, L. Ren, J. Yang, *et al.*, "Seroprevalence and humoral immune durability of anti-SARS-CoV-2 antibodies in Wuhan, China: a longitudinal, population-level, cross-sectional study," *Lancet* **397**(10279), 1075–1084 (2021).
8. X. Xu, J. Sun, S. Nie, *et al.*, "Seroprevalence of immunoglobulin M and G antibodies against SARS-CoV-2 in China," *Nat. Med.* **26**(8), 1193–1195 (2020).
9. J. R. de Jesus, T. d. and A. Andrade, "Understanding the relationship between viral infections and trace elements from a metallomics perspective: implications for COVID-19," *Metallomics* **12**(12), 1912–1930 (2020).
10. M. Taheri, A. Bahrami, P. Habibi, *et al.*, "A Review on the Serum Electrolytes and Trace Elements Role in the Pathophysiology of COVID-19," *Biol. Trace Elem. Res.* **199**(7), 2475–2481 (2021).
11. K. K. Pvsn, S. Tomo, P. Purohit, *et al.*, "Comparative Analysis of Serum Zinc, Copper and Magnesium Level and Their Relations in Association with Severity and Mortality in SARS-CoV-2 Patients," *Biol. Trace Elem. Res.* **201**(1), 23–30 (2023).
12. D. Haschka, A. Hoffmann, and G. Weiss, "Iron in immune cell function and host defense," *Semin. Cell Dev. Biol.* **115**, 27–36 (2021).
13. T. Sonnweber, A. Boehm, S. Sahanic, *et al.*, "Persisting alterations of iron homeostasis in COVID-19 are associated with non-resolving lung pathologies and poor patients' performance: a prospective observational cohort study," *Respir Res.* **21**(1), 276 (2020).
14. K. Ozdemir, E. Saruhan, T. K. Benli, *et al.*, "Comparison of trace element (selenium, iron), electrolyte (calcium, sodium), and physical activity levels in COVID-19 patients before and after the treatment," *J. Trace Elem. Med. Biol.* **73**, 127015 (2022).
15. O. F. Kocak, F. B. Ozgeris, E. Parlak, *et al.*, "Evaluation of Serum Trace Element Levels and Biochemical Parameters of COVID-19 Patients According to Disease Severity," *Biol. Trace Elem. Res.* **200**(7), 3138–3146 (2022).
16. R. Gaudiuso, E. Ewusi-Annan, N. Melikechi, *et al.*, "Using LIBS to diagnose melanoma in biomedical fluids deposited on solid substrates: Limits of direct spectral analysis and capability of machine learning," *Spectrochim. Acta, Part B* **146**, 106–114 (2018).
17. N. Melikechi, Y. Markushin, D. C. Connolly, *et al.*, "Age-specific discrimination of blood plasma samples of healthy and ovarian cancer prone mice using laser-induced breakdown spectroscopy," *Spectrochim. Acta, Part B* **123**, 33–41 (2016).
18. L.-B. Guo, D. Zhang, L.-X. Sun, *et al.*, "Development in the application of laser-induced breakdown spectroscopy in recent years: A review," *Front. Phys.* **16**(2), 22500 (2021).
19. Y. Liu, B. Zhou, W. Wang, *et al.*, "Insertable, Scabbarded, and Nanoetched Silver Needle Sensor for Hazardous Element Depth Profiling by Laser-Induced Breakdown Spectroscopy," *ACS Sens.* **7**(5), 1381–1389 (2022).
20. X. Chen, X. Li, S. Yang, *et al.*, "Discrimination of lymphoma using laser-induced breakdown spectroscopy conducted on whole blood samples," *Biomed. Opt. Express* **9**(3), 1057–1068 (2018).
21. W. Wang, Y. Liu, Y. Chu, *et al.*, "Stable sensing platform for diagnosing electrolyte disturbance using laser-induced breakdown spectroscopy," *Biomed. Opt. Express* **13**(12), 6778–6790 (2022).
22. Z. Zhao, W. Ma, G. Teng, *et al.*, "Accurate identification of inflammation in blood based on laser-induced breakdown spectroscopy using chemometric methods," *Spectrochim. Acta, Part B* **202**, 106644 (2023).
23. X. Chen, Y. Zhang, X. Li, *et al.*, "Diagnosis and staging of multiple myeloma using serum-based laser-induced breakdown spectroscopy combined with machine learning methods," *Biomed. Opt. Express* **12**(6), 3584–3596 (2021).

24. X. Chen, X. Li, X. Yu, *et al.*, "Diagnosis of human malignancies using laser-induced breakdown spectroscopy in combination with chemometric methods," *Spectrochim. Acta, Part B* **139**, 63–69 (2018).
25. Y. Chu, T. Chen, F. Chen, *et al.*, "Discrimination of nasopharyngeal carcinoma serum using laser-induced breakdown spectroscopy combined with an extreme learning machine and random forest method," *J. Anal. At. Spectrom.* **33**(12), 2083–2088 (2018).
26. Y. Chu, F. Chen, Z. Sheng, *et al.*, "Blood cancer diagnosis using ensemble learning based on a random subspace method in laser-induced breakdown spectroscopy," *Biomed. Opt. Express* **11**(8), 4191–4202 (2020).
27. Z. Yue, C. Sun, F. Chen, *et al.*, "Machine learning-based LIBS spectrum analysis of human blood plasma allows ovarian cancer diagnosis," *Biomed. Opt. Express* **12**(5), 2559–2574 (2021).
28. E. M. Emara, H. Song, H. Imam, *et al.*, "Detection of hypokalemia disorder and its relation with hypercalcemia in blood serum using LIBS technique for patients of colorectal cancer grade I and grade II," *Lasers Med. Sci.* **37**(2), 1081–1093 (2022).
29. K. Bardarov, I. Buchvarov, T. Yordanova, *et al.*, "Laser-induced break down spectroscopy for quantitative analysis of electrolytes (Na, K, Ca, Mg) in human blood serum," in *International Conference on Quantum, Nonlinear, and Nanophotonics 2019 (ICQNN 2019)*, (2019).
30. G. Teng, Q. Wang, H. Zhang, *et al.*, "Discrimination of infiltrative glioma boundary based on laser-induced breakdown spectroscopy," *Spectrochim. Acta, Part B* **165**, 105787 (2020).
31. A. Allegra, A. Tonacci, R. Sciacotta, *et al.*, "Machine Learning and Deep Learning Applications in Multiple Myeloma Diagnosis, Prognosis, and Treatment Selection," *Cancers* **14**(3), 606 (2022).
32. A. Martini, S. A. Guda, A. A. Guda, *et al.*, "PyFitit: The software for quantitative analysis of XANES spectra using machine-learning algorithms," *Comput. Phys. Commun.* **250**, 107064 (2020).
33. T. Hastie, R. Tibshirani, and J. Friedman, "The elements of statistical learning. 2001," *J. Royal Stat. Soc.* **167**(1), 192 (2004).
34. J. Liang, M. Li, Y. Du, *et al.*, "Data fusion of laser induced breakdown spectroscopy (LIBS) and infrared spectroscopy (IR) coupled with random forest (RF) for the classification and discrimination of compound salvia miltiorrhiza," *Chemom. Intell. Lab. Syst.* **207**, 104179 (2020).
35. Y. Chu, Z. Zhang, Q. He, *et al.*, "Half-life determination of inorganic-organic hybrid nanomaterials in mice using laser-induced breakdown spectroscopy," *J. Adv. Res.* **24**, 353–361 (2020).