Machine learning-based LIBS spectrum analysis of human blood plasma allows ovarian cancer diagnosis

ZENGQI YUE,1 CHEN SUN,1 FENGYE CHEN,1 YUQING ZHANG,1 WEIJIE XU,1 SAHAR SHABBIR,1 LONG ZOU,1 WEIGUO LU,2 WEI WANG,3 ZHENWEI XIE,2 LANYUN ZHOU,2 YAN LU,2,4 AND JIN YU1,5

1School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai 200240, China
2Center for Uterine Cancer Diagnosis & Therapy Research of Zhejiang Province, Women’s Reproductive Health Key Laboratory of Zhejiang Province, and Department of Gynecologic Oncology, Women’s Hospital and Institute of Translational Medicine, Zhejiang University School of Medicine, Hangzhou 310011, China
3Department of Clinical Laboratory, Tongde Hospital of Zhejiang Province, Hangzhou 310012, China
4yanlu76@zju.edu.cn
5jin.yu@sjtu.edu.cn

Abstract: Early-stage screening and diagnosis of ovarian cancer represent an urgent need in medicine. Usual ultrasound imaging and cancer antigen CA-125 test when prescribed to a suspicious population still require reconfirmations. Spectroscopic analyses of blood, at the molecular and atomic levels, provide useful supplementary tests when coupled with effective information extraction methods. Laser-induced breakdown spectroscopy (LIBS) was employed in this work to record the elemental fingerprint of human blood plasma. A machine learning data treatment process was developed combining feature selection and regression with a back-propagation neural network, resulting in classification models for cancer detection among 176 blood plasma samples collected from patients, including also ovarian cyst and normal cases. Cancer diagnosis sensitivity and specificity of respectively 71.4% and 86.5% were obtained for randomly selected validation samples.

© 2021 Optical Society of America under the terms of the OSA Open Access Publishing Agreement

1. Introduction

Ovarian cancer as one of the common gynecologic cancers [1] presents a high mortality rate when a patient is diagnosed in an advanced stage [2]. The absence of specific clinical symptom combined with the lack of performant diagnosis method would delay effective treatments [3]. Early screening method is therefore currently expected by clinical medicine [4]. At the same time, the occurrence of ovarian cancer [1] does not yet justify a systematic screening over a large population with a high precision but invasive diagnosis technique, such as tissue biopsy. Noninvasive techniques such as ultrasound imaging often requires a high expertise to distinguish between cancer and a benign abnormal case such as cyst, while blood cancer antigen CA-125 test generally presents insufficient robust levels of diagnostic accuracy [5]. Supplementary tests are thus extremely important for an accurate diagnosis of ovarian cancer. The improvement idea is to develop an intermediate noninvasive technique with higher diagnostic performances than ultrasound imaging and CA-125 test to help practitioners in their decision of further diagnosis with biopsy for example. Blood analysis in the molecular or atomic levels could be an efficient way to satisfy the above need in condition that it is coupled with a suitable and effective information extraction method.

Optical spectroscopy as an analytical technique is able to acquire the fingerprint of a blood sample. The obtained information can be often complex in nature and implicit in expression, as
for analysis and diagnosis in biology and medicine in general. Modern data mining methods [6] developed within artificial intelligence (AI), such as machine learning and deep learning [7] are at the same time, required and powerful for extracting suitable characteristic features of a sample. Recent progresses in medical image treatment by the AI approach fully demonstrate the capability of the related algorithms for classification and identification of medical images [8,9].

Combined with machine learning algorithms, spectroscopic techniques have been implemented for cancer diagnosis, including attenuated total reflection Fourier transform infrared (ATR-FTIR) spectroscopy [10], Raman [11] and surface-enhanced Raman spectroscopy (SERS) [12]. Laser-induced breakdown spectroscopy (LIBS) as a multi-elemental detection technique [13,14] has demonstrated its potential in biomedical applications [15], especially for bacteria detection and identification [16–18], biological tissue mapping [19], neurodegenerative disease diagnosis [20] and cancer screening [21]. Combined with chemometrics or more recently machine learning-based regression models [22,23], LIBS is able to classify and identify various biological samples according to their LIBS spectra [24]. For cancer diagnosis with LIBS more specifically, early works often involved samples harvested from laboratory model animals and was designed to study spectral markers pertinent for identification and detection of the targeted diseases [25,26]. More recently, tests using human blood samples were reported with a quite limited number of samples in the order of several tens [27–29], which may influence the significance of the study for clinical applications due to a large variability of the patients. It is also worth to point out that the above mentioned works of cancer detection with LIBS paid attention to the differentiation between normal and cancer cases, other intermediate or evolving abnormal cases, cyst for example, were not included in the collection of studied samples. Furthermore, among various operation modes with different types of biological samples, direct analysis of blood-related liquids is preferred for a clinical approach because it corresponds to an easy, cost-effective and commonly applied implementation of biomedical test which is suitable for a wide coverage screening and can be incorporated in a routine physical examination.

The present work was designed according to a clinical application scenario where an ensemble of blood plasma samples is collected from a population of female patients after their initial medical imaging examination intended to a screening of ovarian cancer, and their health situation needs to be further diagnosed to decide among the three cases of normal, ovarian cyst and ovarian cancer. LIBS analysis of blood plasma was thus chosen to provide supplementary diagnostic information. Although the sample collection represented a time-consuming task, a significantly large set of 176 blood plasma samples was collected from a population of female patients examined by the hospitals, including the three finally diagnosed case-types of normal, ovarian cyst and ovarian cancer. The measurements were performed using the liquid sample preparation method of surface-assisted LIBS introduced in our previous works [30,31]. The recorded LIBS spectra were used to study classification and identification models. One third of each case-type of samples was randomly selected as the model validation samples, while the rest of the samples was used as the model training samples. Basically, the study of the classification and identification models included the steps of data pretreatment, feature selection, neural network training and model validation. In the following, we first presented the used samples, the experimental setup and the measurement protocol. The spectral data treatment method was then presented in detail. In the section “results and discussions”, we presented the way that the classification model was optimized step by step together with the principle of each optimization process and the role of several common elements in blood plasma, such as K, Na and Mg before delivering and discussing about the performances of the finally optimized model.

2. Samples, experimental setup, and measurement protocol

Blood samples were collected from 176 patients examined by Women’s Hospital of the School of Medicine, Zhejiang University and Tongde Hospital of Zhejiang Province in China, the resulted
plasmas were stored in a fridge at −80°C before being prepared for the experiment. Plasma as the liquid component of blood is known as playing a vital role in an intravascular osmotic effect that keeps electrolyte concentration balanced and protecting the body from infection and other blood disorders [32]. A disease in the body can influence its chemical composition, the analysis of which can therefore provide indications of the health state of the patient. The advantage of the use of plasma instead of whole blood for LIBS is to avoid the spectral interference from lines emitted by iron contained in red blood cells at a relative high concentration. For the purpose of the present study, all the samples were labeled with the help of usual medical diagnosis methods practiced in the hospitals. Thereby, 79 samples belonged to the healthy normal case (45%), 34 were diagnosed to be the cyst cases (19%) and 63 to be the ovarian cancer cases (36%). For an effective fingerprint measurement, the liquid sample preparation protocol of surface-assisted LIBS [30,31] was applied. As shown in Fig. 1(a) and (b), 150 µℓ of each liquid sample at the room temperature was picked with a pipette and dropped on a high purity graphite plate (purity ≥ 99.92% according to the provider) with a polished and cleaned surface of 20 × 20 mm² (thickness 5 mm). The liquid was then spread out uniformly over the whole surface of the plate with the tip of the pipette which was changed for each sample to avoid cross contamination. The obtained liquid-covered graphite surface was put under an infrared lamp and heated up for about 10 minutes for drying. The sample was then left for cooling down under the ambient temperature for about 10 minutes. The result was a thin and semi-transparent layer of residual of the blood plasma on the surface of a graphite plate.

The used experimental setup with its central part illustrated in Fig. 1(c) was the same as in our previous work [33] and the detailed description can be found elsewhere [34]. Briefly, a Q-switched Nd:YAG laser operating at its fundamental of 1064 nm delivered 7 ns and 30 mJ laser pulses to the samples after being focused by a lens of 50 mm focal length for the LIBS measurements. A sample was placed on a motorized 3-D displacement stage allowing its translation synchronized with the laser pulses in order to perform single-shot LIBS spectrum recordings. During the experiment with a sample, replicate measurements were performed on its surface and distributed in the form of a matrix of ablation craters. A center-to-center distance of 0.8 mm was left between neighboring craters to avoid their overlapping. Emission from laser-induced plasma was collected by a two-lens system and captured by an optical fiber connected to an echelle spectrometer with a wide spectral range from 230 nm to 900 nm and a spectral resolution power of λ/Δλ = 5000, equipped with an intensified charge coupled device (ICCD) camera (Mechelle 5000 and iStar from Andor Technology). The ICCD camera was triggered by laser pulses with a detection delay of 0.8 µs and gate width of 2.5 µs. Measurements were randomly performed with samples of different case-types in order to avoid systematic drift of the setup. For each sample, a measurement matrix of about 400 to 500 single-shot ablations together with the corresponding replicate LIBS spectrum recordings was performed on the substrate surface covered by a residue of blood plasma, yielding a total number of 85176 single-shot spectra for the 176 samples. In Fig. 1(d), typical replicate-averaged spectra are shown for the three case-types of blood plasma samples and for a substrate graphite. We can see that the prominent spectral features on an intensity scale of 10^5 counts correspond to the major metal and nonmetal elements in a biological material: K, Na, Ca, Mg and C, H, O, N. Notice that the lines from the last 4 nonmetal elements can also be contributed by the substrate for C, and by the ambient gas for H, O, N, as shown in Fig. 1(d). On a much smaller intensity scale of 10^3 counts, some minor elements, Fe, Si, P, Cu, can be identified in the blood plasma spectra, as shown in the insets in Fig. 1(d). A first glance on the spectra does not evidence obvious difference between those of the blood plasma samples, indicating the need of a more sophisticated data processing method to reveal their specific characteristics.
Fig. 1. (a) and (b): sample preparation procedure; (c) the central part of the experimental setup; (d) typical LIBS spectra of blood plasmas corresponding to the normal, cyst and cancer cases and of a substrate graphite. The designations of the prominent lines on an intensity scale of $10^5$ counts are given on the spectrum of a normal sample, while some much less important lines are shown in the insets on the spectra of a cyst and a cancer samples on an intensity scale of $10^3$ counts and with an enlarged spectral range.

3. Data treatment procedure and methods

The data treatment flowchart is shown in Fig. 2, it consisted of several steps, including spectral pretreatment, organization of training and validation sample and data sets, feature selection, standardization, model training and validation.

3.1. Data pretreatment

Data pretreatment included the following operations: i) Spectrum averaging in order to reduce the fluctuations due to laser pulse shot-to-shot energy jitter and sample inhomogeneity. For a given sample, the raw single-shot replicate spectra were randomly arranged in a sequence, a first average spectrum was generated as the result of averaging over the first 30 spectra from the n°1 to the n°30. A second average spectrum was then generated in shifting the averaging range by 10 spectra to include the spectra from the n°11 to the n°40. The operation was repeated until all the raw single-shot spectra of the sample were involved in the generation of average spectra. A number from 37 to 47 average spectra were generated for each of the samples, resulting 7887 average spectra for the 176 samples of the three case-types. ii) Normalization: each average spectrum was normalized with its total spectral intensity calculated by integrating the spectral intensity over the whole spectral range. The above operations generated the pretreated spectra.
3.2. Dataset organization

For the further steps in the study of identification and classification models, we needed to isolate a part of the samples as the model validation ones in order to assess the prediction performance of the trained models. In a machine learning data treatment procedure, the validation samples do not take part in the model training process that is exclusively contributed by the model training samples. It is however required that the training samples and the validation samples share a same feature space with similar distributions [35]. In the case of a regression model trained and validated for quantitative analysis, such requirement can be satisfied by displaying the standard samples as a function of their known concentrations of the element to be analyzed, and selecting validation samples in such way that their elemental concentrations being randomly and uniformly distributed within the concentration range covered by the concentrations of the training samples. Similarly, for our task in this work of identifying and classifying among the 3 case-types of normal, cyst and cancer, the selection of the validation samples consisted in taking respectively from the 3 case-types, one third of the samples in a random way so that the selected samples are statistically equivalent to the remaining ones. Different from the case of regression model for quantitative analysis, our task of identification and classification did not rely on a clearly identified parameter, such as the concentration of an element to be determined. The fact that a sample is classified into a given case-type depends on a rather undefined ensemble of parameters which can be extracted from the LIBS spectrum. It was why we decided to select the validation samples according to principal component analysis (PCA) scores, although such method did not correspond to the unique one that could be used, and the actual selection of the validation samples could slightly influence the final performance of the trained models.

The pretreated spectra of each sample were therefore further averaged generating a mean spectrum to represent the sample in a PCA plot. The 2-D PCA plots of the samples respectively belonging to the 3 case-types are presented in Fig. 3(a), (b) and (c). For each case-type, we manually selected one third of the samples as validation ones in such way that these samples were distributed uniformly in different areas in the PCA plot occupied by the ensemble of the
samples as shown in Fig. 3(a), (b) and (c), where for each case-type of the samples, the selected validation samples are represented in crosses, while the rest of the samples used as the model training samples are represented in cycles. As a convention for visual recognition in this paper, we used green, orange and red respectively for the normal, cyst and cancer cases. The detailed composition of the validation and training sets of samples is given in Table 1. The training samples with their pretreated spectra provided the training data set, and the validation samples with their pretreated spectra provided the validation data set.

![Fig. 3. PCA plots of normal samples in green (a), cyst samples in orange (b), and cancer samples in red (c) together with the indication of the validation samples selected for each case-type represented by crosses to be distinguished from the training samples represented by cycles.](image)

| Table 1. Organization of the samples into a training and a validation sets. |
|-----------------------------|-------------|-------------|-------------|-----------|
|                             | Normal      | Cyst        | Cancer      | Total     |
| Training sample set         | 53          | 23          | 42          | 118       |
| Validation sample set       | 26          | 11          | 21          | 58        |
| Total                       | 79          | 34          | 63          | 176       |

3.3. Feature selection

A feature selection process based on SelectKBest algorithm [36] with a chi-squared test [37] was applied to the ensemble of pretreated spectra of the 118 training samples. In statistics, the chi-squared test is used to determine whether there are statistically significant differences among two or more distributions of data by calculating the distances separating the distributions. In our case, for each spectral channel, a chi-squared value was calculated for the intensities of the individual pretreated spectra with respect to the mean channel intensity over all the pretreated spectra. The result value represented the distances among the 3 ensembles of channel intensities associated to the 3 case-types. The calculation was carried out for the 23826 channels of the spectrum, resulting in the corresponding channel scores for ranking them from the highest (large distances among the populations) to the lowest scores. The top 100 spectral channels were retained to determine the selected features in all the pretreated spectra of the training data set. These spectral features were used as the input variables to train the classification model. At the same time, the retained channels were applied to the pretreated spectra of the 58 validation samples in order to identify the 100 spectral features used for the assessment of the trained model in the step of model validation.

The results of feature selection are shown in Fig. 4. The score obtained by the 100 highest ranked spectral channels are shown in Fig. 4(a) in red cycles, and in Fig. 4(b) the corresponding spectral features are indicated by red cycles in an average spectrum of a normal sample to show their respective intensities. We can see that the most important spectral features for the
identification and classification according to the 3 targeted case-types of normal, cyst and cancer belong to K I 766.5 nm, K I 769.9 nm, Na I 589.0 nm, Na I 589.6 nm, Na I 819.5 nm, Mg II 279.6 nm, Mg II 280.3 nm, and C I 247.9 nm lines. Such result can be understood by the fact that certain major metal elements like sodium, potassium and to a lesser extent magnesium, are the most important electrolytes in living systems, their concentrations in blood plasms play a vital role in maintaining homeostasis in the body [38]. Concentration imbalance of the electrolytes can be the cause of abnormalities in the human body and should play an important role in the identification and classification of the case-types of normal, cyst and cancer in our study. The case of carbon remains less straightforward for an explication because of the contribution also from the substrate as we mentioned above. We can also remark the absence of minor elements among the 100 highest ranked spectral features. A detailed look in the scores obtained by the detected minor elements reveals a ranking of 464th for Si and 4923rd, 5151st, 6465th respectively for Cu, P and Fe, far behind the above discussed major elements. This observation shows the marginal roles of the minor elements detected in blood plasma in diagnosis of ovarian cancer, due to certainly their very low line intensities.

3.4. Standardization

As a usual operation in a machine learning data treatment facilitating the gradient descent type model optimization, the standardization was implemented in our study for the selected and identified features respectively of the training and validation samples. The selected features of the training data set were first scaled with a linear transformation which brought their values into the range of [0, 1]. For a given selected spectral channel, the maximal ($I_{max}$) and minimal ($I_{min}$) values of the channel intensities were identified over all the pretreated spectra of all the samples of the training set. The standardized channel intensity of an actual spectrum was then calculated by $(I - I_{min})/(I_{max} - I_{min})$. The pair of values $I_{max}$ and $I_{min}$ were then applied to the same spectral channel of the validation data set. The operation generated respectively for the training and the validation data sets the standardized selected features and the standardized identified features.
3.5. Neural network training by cross-validations

The classification model training process was implemented according to our previous work initially devoted to quantitative analysis with LIBS spectra from soil samples with a regression model based on back-propagation neural network (BPNN) [23]. Since its introduction into LIBS data analysis, this method has been applied to various scenarios of LIBS analysis including laser pulse energy variation correction [33], chemical matrix effect correction in rock analysis [39], determination of carbon concentration in steel [40], and simultaneous determination of concentrations of water and potassium in potash online analysis [41]. In the present work, the method was adapted to the case of identification and classification of a collection of samples. The used neural network had 3 layers, with an input layer of 100 neurons corresponding to the 100 standardized selected features of each pretreated training spectrum, a hidden layer of 50 neurons, and an output layer of 3 neurons corresponding to the 3 output case-types. A 5-fold cross-validation optimization procedure was employed for neural network training with the pretreated spectra of the training samples. For each fold of the cross-validation, the identification of a sample among the 3 case-types of normal, cyst and cancer, was decided according to the majority of the individual identifications for the test spectra of the sample. In the end of the cross-validation, an ensemble of definitive identifications was assigned to all the training samples according to the majority of the 5 cross-validation identifications. The calibration performance of the trained models was then assessed by a comparison between the models-assigned case-types of the training sample and their label values, and presented in a confusion matrix for the training samples, together with the associated figures of merit. A more detailed description of the model training process can be found in the Supplement 1 associated to the paper.

3.6. Model validation by the validation data set

The prediction performance of the trained models was assessed in this step by the validation data set which was excluded from the training process. The validation process was similar to the cross-validation tests in the model training step. The pretreated spectra of a given validation sample with 100 standardized identified features each, were used as an ensemble of data to successively test the 5 trained models, generating 5 ensemble of identifications for the validation sample. The majority of identifications in each ensemble determined the prediction by the corresponding individual model. The majority of the 5 individual predictions determined final model-predicted case-type of the sample. The model-predicted case-types of the validation samples were compared to their label values, resulting in a confusion matrix for the validation samples, which led to the calculation of the figures of merit detailing the performances of the models for prediction with independent samples.

The figures of merit used in this work for the assessment of the calibration as well as the prediction performances of the models correspond to sensitivity and specificity with their usual definitions: sensitivity \(=\frac{TP}{TP+FN}\), specificity \(=\frac{TN}{TN+FP}\), where \(TP\), \(TN\), \(FP\) and \(FN\) stand for respectively true positives, true negatives, false positives and false negatives.

4. Results and discussions

4.1. Initial models leading to classification according to 3 case-types

The classification results with the above discussed initial models leading to classification according to 3 case-types are shown in Table 2 for the training samples and in Table 3 for the validation samples, together with the confusion matrix and the figures of merit. For calibration with the training samples as shown in Table 2, we can see that the identification of normal samples is satisfactory with a very low rate of wrong classification of 1.9% (1 over 53). At the same time, for the cyst and cancer samples, if they were considered as an ensemble, their wrong classification to the normal case remains limited with a rate of 4.6% (3 over 65). However, misclassification
within the ensemble of cyst and cancer samples becomes quite important. The model therefore effectively explores the pertinent information to distinguish between the normal and the ensemble of cyst and cancer samples, while within the ensemble of cyst and cancer samples its effectiveness greatly decreased. For prediction with the validation samples as shown in Table 3, we can see that the performances are globally degraded as shown by a comparison between the figures of merit for the training data set in Table 2 and for the validation data set in Table 3. The robustness of the model is therefore not sufficient. This remains understandable because of the limited number of the training samples, which appears quite small with respect to the large variability of human plasma samples. The representability of the validation samples by the training samples cannot thus be ensured in an optimized way despite the precaution taken in the data organization into the training and validation sample sets. Besides the weak robustness of the models, a better performance for classification between the normal samples and the ensemble of cysts and cancer samples can still be observed for the validation samples. In addition, the important misclassification within the ensemble of cyst and cancer samples leads to, for an application of cancer screening, a large false positive rate of 36.4% (8 over 22) mainly due to misclassifications from cyst samples, and a false negative rate of 33.3% (7 over 21) also due to misclassifications to cyst samples, corresponding to cancer diagnosis sensitivity and specificity of respectively 66.7% and 78.4%. These results show rooms for improvement for the initial classification models.

In order to figure out the way to improve the performances of the classification models, we investigated in detail the reasons for the mediocre performances with the ensemble of cyst and cancer samples. We first looked at the mean positions of the samples in a PCA plot (PC1 and PC2) determined by the respective mean values of the 100 standardized selected or identified features calculated over the pretreated spectra of a sample. Such plot is shown in Fig. 5(a), where the mean positions of the normal, cyst and cancer samples are respectively represented in green, orange and red, the training samples with crosses and the validation samples with circles.
We can see first that globally the normal samples are clustered together in an area separated from the cyst and cancer samples. The PC1 expresses the most discriminative character for such separation. This can explain the satisfactory classification results for normal samples with respect to the ensemble of cyst and cancer samples and vice versa. There are however 2 normal validation samples wrongly classified as cancer ones in Table 3. In the PCA plot, these 2 samples correspond to the points in Fig. 5(a) surrounded by a green circle, the positions of which are actually located outside of the area occupied by the major part of the normal samples and merged into the zone occupied by the ensemble of cyst and cancer samples. Table 3 also indicates 3 cancer validation samples wrongly identified as normal ones. We can find them in Fig. 5(a) surrounded by a red circle and merged into the zone of normal samples. Concerning the general situation of the cyst and cancer samples, we can see in Fig. 5(a) that their positions are mutually merged in a same zone without distinct areas from each other. This can explain the unsatisfactory classification results for the training samples as well as for the validation samples between the cyst and cancer samples. These observations seem to tell us that the features selected for 3 case-types classifications would not express effective characteristics of the spectra for the distinction between the cases of cyst and cancer, although they provide a quite satisfactory distinction between the normal and the ensemble of cyst and cancer samples.

![Fig. 5.](image)

A look at the composition and the coefficients of the PC1 reveals the main contributions from K lines (5.44), Na lines (−2.48) and Mg lines (0.17), while the PC2 are mainly contributed by Na lines (5.97), K lines (1.97) and Mg lines (0.40). This means that the separation between the normal samples and the ensemble of cyst and cancer samples are mainly due to the K lines which express different behaviors for these 2 types of samples. A plot of the relative intensity of the K I 766.5 nm line for all the samples in Fig. 5(b) clearly shows such difference. We can see also that the several wrongly classified samples show their K I line intensity differentiating from that of the other samples of the same type, as indicated in Fig. 5(b) by the data points surrounded by a circle. At the same time, for the cyst and the cancer samples, the intensities of the K I line exhibit similar behaviors in accordance with the PCA plot in Fig. 5(a), where the PC1 scores...
do not allow their separation. The PC2 which is mainly contributed by the Na lines does not allow clear separation of the samples, confirmed by the similar behaviors of the Na I 589.0 nm line for the 3 types of the samples shown in Fig. 5(c). These observations would tell us that the features selected for classification according to the 3 case-types of normal, cyst and cancer are dominant by K lines as indicated by the scores shown in Fig. 4(a), which offers a satisfactory separation between the normal samples and the ensemble of cyst and cancer samples. Meanwhile however, such domination prevents other spectral characteristics from a sufficient expression, which otherwise, may help for a better differentiation between cyst and cancer samples.

The idea was thus to proceed the classification task into 2 successive steps of one first step separating the normal samples and the ensemble of cyst and cancer ones, followed by a second step for a further separation between the cyst and cancer samples with new spectral features selected without dominance of the K lines.

4.2. Improved models of classification in 2 steps of 2 case-types

An improved model training process with a schema of classification in 2 steps of 2 case-types was implemented according to the flowchart shown in Fig. 6. In the end of the first step which was identical to the initial one-step model training, the resulted model 1 were validated by the validation samples, resulting in the separation of the “normal” samples and the ensemble of “cyst-cancer” samples. Here, the use of quotation marks expresses the fact that misclassification can happen with the model 1 in such way that the resulted 2 classes of identified samples can mutually contain individuals from the other type.

![Flowchart of model training with a schema of classification in 2 steps of 2 case-types.](image)

The ensemble of validation samples identified as “cyst-cancer” was further processed in the second step as the new validation samples. The cyst and cancer samples in the training sample set of the first classification model were used as the new training sample set for the second classification model. The same feature selection process was applied to the new training data set for the purpose of classification according to 2 case-types. The 100 highest ranked spectral channels are shown in Fig. 7 for comparison with the results shown in Fig. 4. The obtained scores of these channels are shown in Fig. 7(a) in red cycles, and in Fig. 7(b) the corresponding
spectral features are indicated by red cycles in an average spectrum of a cancer sample to show their respective intensities. We can see that the most relevant spectral channels for classification according to the 2 case-types of cyst and cancer belong to Mg II 279.6 nm, Mg II 280.3 nm, Mg I 285.2 nm lines, complemented by Na I 589.0 nm, Na I 589.6 nm, Na I 819.5 nm lines, C I 247.9 nm line, as well as Ca II 393.4 nm and 396.8 nm lines and H I 656.3 nm line. The K lines are not selected among the 100 channels as pertinent for differentiating cyst and cancer, confirming our expectation discussed in the above section. Minor elements were again absent among the 100 top features. A detailed look in the scores obtained by the detected minor elements reveals a ranking of 2497th, 2634th, 3703rd, and 3850th respectively for P, Fe, Cu and Si, showing their negligible contributions for differentiation between cyst and cancer.

A comparison between the scales of the vertical coordinate in Fig. 7(a) and Fig. 4(a) shows an important diminution of the scores of the 100 highest ranked spectral channels, which means a significant reduction of the difference between the 2 populations of data in the new training data set, in accordance with the results shown in Fig. 5. In other words, the number of the features effective to distinguish the 2 populations of data decreases as a consequence of a reduced distance between them. It was therefore justified to use less features for model training in order to avoid overfitting. After testing several options, the 30 highest ranked channels (including emission lines from Mg and Na) were used to selected 30 features in the pretreated spectra of the new training samples and to identify the same number of features in the pretreated spectra of the new validation samples. The same model training process was performed using the new training spectra with the new selected features, to optimize a neural network with 30 neurons in the input layer, 20 neurons in the hidden layer and 2 neurons in the output layer, resulting in the model 2 and the confusion matrix 2 together with the corresponding figures of merit for classification of the training samples as shown in Table 4. The trained model 2 was then validated using the new validation spectra with the new identified features. The obtained result was combined with the that obtained in the first step of classification, leading to the final confusion matrix together with the corresponding figures of merit for classification of the validation samples as shown in Table 5.
Table 4. Confusion matrix and figures of merit for classification of the training samples with the improved models in 2 steps of classification according to 2 the case-types of cyst and cancer.

| Sample label | Model-predicted class | Cyst | Cancer |
|--------------|-----------------------|------|--------|
| Cyst         |                       | 17   | 6      |
| Cancer       |                       | 4    | 38     |
| Average      |                       | 8.2  | 8.2    |

Classification performance in cross-validation

|          | Sensitivity (%) | Specificity (%) |
|----------|-----------------|-----------------|
| Cyst     | 73.9            | 90.5            |
| Cancer   | 90.5            | 73.9            |

Table 5. Confusion matrix and figures of merit for classification of the validation samples with the improved models in 2 steps of classification according to the 2 case-types of cyst and cancer.

| Sample label | Model-predicted class | Normal | Cyst | Cancer |
|--------------|-----------------------|-------|------|--------|
| Normal       |                       | 24    | 0    | 2      |
| Cyst         |                       | 0     | 8    | 3      |
| Cancer       |                       | 3     | 3    | 15     |
| Average      |                       | 78.8  | 90.2 |

Classification performance in model validation

|          | Sensitivity (%) | Specificity (%) |
|----------|-----------------|-----------------|
| Normal   | 92.3            | 90.6            |
| Cyst     | 72.7            | 93.6            |
| Cancer   | 71.4            | 86.5            |

For calibration with the training samples, a comparison between Table 4 and Table 2 shows that the performance for cyst identification is unchanged with a wrong classification rate of 26.1% (6 over 23). At the same time, the performance for cancer identification is slightly degraded. The wrong classification rate is increased from 7.1% (3 over 42) to 9.5% (4 over 42). For prediction with the validation samples, a comparison between Table 5 and Table 3 shows that the performance of identification is improved for the both cases of cyst and cancer. The wrong classification rate for cyst is greatly reduced from 54.5% (6 over 11) to 27.3% (3 over 11). The wrong classification rate for cancer is reduced from 33.3% (7 over 21) to 28.6% (6 over 21). These results show that the improved models of 2-step classifications offer a better prediction performance for validation samples, even though the improvement for calibration remains unclear. For an application of cancer screening, a false positive rate of 25.0% (5 over 20) and a false negative rate of 28.6% (6 over 21) were obtained, which are clearly improved comparing to the initial one-step classification models, allowing cancer diagnosis sensitivity and specificity of respectively 71.4% and 86.5%.

5. Conclusion

In this work, we have developed a method of identification and classification of blood plasma samples collected from the patients and including, among the 176 samples, the 3 case-types of normal, ovarian cyst and ovarian cancer. The method is based on LIBS spectrum recording coupled with spectral data treatment using a machine learning approach. A first classification model allowed a satisfactory classification between the normal samples and the ensemble of cyst and cancer samples, whereas numerous misclassifications happened between the cyst and cancer samples, leading to mediocre sensitivity and specificity for cancer identification of respectively 66.7% and 78.4% when the models were tested with independent validation samples. A detailed investigation on the spectral features selected for the model training revealed the domination of K lines in LIBS spectrum, which was effective for separating the normal samples and the ensemble of cyst and cancer samples. Such domination inhibited the expression of other features.
more suitable for the discrimination between cyst and cancer. A second ensemble of models was trained, where the normal samples were separated from the ensemble of cyst and cancer ones in the first step of classification, while the second step focused on the discrimination between the cyst and cancer cases. A new feature selection disgraced the K lines and put forwards other features, Mg and Ca lines for instance. The new models exhibited a better performance of differentiation between cyst and cancer samples, leading to improved cancer identification sensitivity and specificity of respectively 71.4% and 86.5% when the models were tested with independent validation samples. Emission lines from some minor elements in blood plasma, Fe, Si, P, Cu, were identified in our experiment. Their contribution to the classification of the samples has been observed clearly negligible as compared to the major metal elements, K, Na, Mg and Ca, considered as the most important electrolytes in blood and playing a vital role in maintaining homeostasis in the body. An imbalance of their concentrations therefore indicates a state of abnormality in a patient [38].

**Funding.** National Natural Science Foundation of China (11574209, 11805126, 61975190); The Key R&D Program of Zhejiang Province (2021C03126C); Startup funding for young scholars at Shanghai Jiao Tong University.

**Acknowledgments.** This study was reviewed and approved by the Ethics Committees of Women’s Hospital of Zhejiang University, School of Medicine (Hangzhou, China, ID 2019063). The study was conducted in accordance with the International Ethical Guidelines for Biomedical Research Involving Human Subjects. All samples have been collected and utilized following strict human subject protection guidelines, written informed consent and IRB review of protocols.

**Disclosures.** The authors declare on conflicts of interest.

**Supplemental document.** See Supplement 1 for supporting content.

**References**

1. F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, “Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries,” CA Cancer J. Clin. 68(6), 394–424 (2018).
2. L. A. Torre, B. Trabert, C. E. DeSantis, K. D. Miller, G. Samimi, C. D. Runowicz, M. M. Gaudet, A. Jemal, and R. L. Siegel, “Ovarian cancer statistics, 2018,” CA Cancer J. Clin. 68(4), 284–296 (2018).
3. K. E. Brain, S. Smits, A. E. Simon, L. J. Forbes, C. Roberts, I. J. Robbé, J. Steward, C. White, R. D. Neal, and J. Hanson, “Ovarian cancer symptom awareness and anticipated delayed presentation in a population sample,” BMC Cancer 14(1), 171 (2014).
4. S. Brandner, J. Müller-Nordhorn, W. Stritter, C. Fotopoulou, J. Sehouli, and C. Holmberg, “Symptomization and triggering processes: Ovarian cancer patients’ narratives on pre-diagnostic sensation experiences and the initiation of healthcare seeking,” Soc. Sci. Med. 119, 123–130 (2014).
5. M. Paraskevaidi, K. M. Ashton, H. F. Stringfellow, N. J. Wood, P. J. Keating, A. W. Rowbottom, P. L. Martin-Hirsch, and F. L. Martin, “Raman spectroscopic techniques to detect ovarian cancer biomarkers in blood plasma,” Talanta 189, 281–288 (2018).
6. A. Esteve, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, and J. Dean, “A guide to deep learning in healthcare,” Nat. Med. 25(1), 24–29 (2019).
7. Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” Nature 521(7553), 436–444 (2015).
8. J. N. Kather, A. T. Pearson, N. Halama, D. Jager, J. Krause, S. H. Loosen, A. Marx, P. Boor, F. Tacke, U. P. Neumann, H. I. Grabsch, T. Yoshikawa, H. Brenner, J. Chang-Claude, M. Hoffmeister, C. Trautwein, and T. Luedde, “Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer,” Nat. Med. 25(7), 1054–1056 (2019).
9. X. Mei, H. C. Lee, K. Y. Diao, M. Huang, B. Lin, C. Liu, Z. Xie, Y. Ma, P. M. Robson, M. Chung, A. Bernheim, V. Mani, C. Calcagno, K. Li, S. Li, H. Shan, J. Lv, T. Zhao, J. Xia, Q. Long, S. Steinberger, A. Jacobi, T. Deyer, M. Lukhsa, F. Liu, B. P. Little, Z. A. Fayad, and Y. Yang, “Artificial intelligence-enabled rapid diagnosis of patients with COVID-19,” Nat. Med. 26(8), 1224–1228 (2020).
10. H. J. Butler, P. M. Brennam, J. M. Cameron, D. Finlayson, M. G. Hegarty, M. D. Jenkins, D. S. Palmer, B. R. Smith, and M. J. Baker, “Development of high-throughput ATR-FTIR technology for rapid triage of brain cancer,” Nat. Commun. 10(1), 1–9 (2019).
11. P. Giamougiannis, C. L. M. Morais, R. Grabowska, K. M. Ashton, N. J. Wood, P. L. Martin-Hirsch, and F. L. Martin, “A comparative analysis of different biofluids towards ovarian cancer diagnosis using Raman microspectroscopy,” Anal. Bioanal. Chem. 413(3), 911–922 (2021).
12. H. Shin, S. Oh, S. Hong, M. Kang, D. Kang, Y.-G. Ji, B. H. Choi, K.-W. Kang, H. Jeong, Y. Park, S. Hong, H. K. Kim, and Y. Choi, “Early-stage lung cancer diagnosis by deep learning-based spectroscopic analysis of circulating exosomes,” ACS Nano 14(5), 5435–5444 (2020).
13. D. W. Hahn and N. Omenetto, “Laser-Induced Breakdown Spectroscopy (LIBS), part I: review of basic diagnostics and plasma–particle interactions: still-challenging issues within the analytical plasma community,” Appl. Spectrosc. 64(12), 3358–3366 (2010).

14. D. W. Hahn and N. Omenetto, “Laser-induced breakdown spectroscopy (LIBS), Part II: review of instrumental and methodological approaches to material analysis and applications to different fields,” Appl. Spectrosc. 66(4), 347–419 (2012).

15. R. Gaudioso, N. Melikechi, Z. A. Abdel-Salam, M. A. Harith, V. Palleschi, V. Motto-Ros, and B. Busser, “Laser-induced breakdown spectroscopy for human and animal health: a review,” Spectrochim. Acta, Part B 152, 123–148 (2019).

16. M. Baudelet, L. Guyon, J. Yu, J.-P. Wolf, T. Amodeo, E. Fréjafon, and P. Laloi, “Spectral signature of native CN bonds for bacterium detection and identification using femtosecond laser-induced breakdown spectroscopy,” Appl. Phys. Lett. 88(6), 063901 (2006).

17. M. Baudelet, J. Yu, M. Bossa, J. Jovelet, J. P. Wolf, T. Amodeo, E. Fréjafon, and P. Laloi, “Discrimination of microbiological samples using femtosecond laser-induced breakdown spectroscopy,” Appl. Phys. Lett. 89(16), 163903 (2006).

18. S. J. Rehse, “A review of the use of laser-induced breakdown spectroscopy for bacterial classification, quantification, and identification,” Spectrochim. Acta, Part B 154, 50–69 (2019).

19. L. Sancey, V. Motto-Ros, B. Busser, S. Koth, J. M. Benoist, A. Piednoir, F. Lux, O. Tillement, G. Panczer, and J. Yu, “Laser spectrometry for multi-elemental imaging of biological tissues,” Sci. Rep. 4(1), 6065 (2014).

20. R. Gaudioso, E. Ewusi-Annan, W. Xia, and N. Melikechi, “Diagnosis of alzheimer’s disease using laser-induced breakdown spectroscopy and machine learning,” Spectrochim. Acta, Part B 171, 105931 (2020).

21. X. Chen, X. H. Li, S. B. Yang, X. Yu, and A. C. Liu, “Discrimination of lymphoma using laser-induced breakdown spectroscopy conducted on whole blood samples,” Biomed. Opt. Express 9(3), 1057–1068 (2018).

22. T. F. Boucher, M. V. Ozanne, M. L. Carmosino, M. D. Dyar, S. Mahadevan, E. A. Breves, K. H. Lepore, and S. M. Clegg, “A study of machine learning regression methods for major elemental analysis of rocks using laser-induced breakdown spectroscopy,” Spectrochim. Acta, Part B 107, 1–10 (2015).

23. C. Sun, Y. Tian, L. Gao, Y. S. Niu, T. L. Zhang, H. Li, Y. Q. Zhang, Z. Q. Yue, N. D. Gilon, and J. Yu, “Machine learning allows calibration models to predict trace element concentration in soils with generalized LIBS spectra,” Sci. Rep. 9(1), 11363 (2019).

24. R. Gaudioso, E. Ewusi-Annan, N. Melikechi, X. Sun, B. Liu, L. F. Campesato, and T. Merghoub, “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).

25. J. H. Han, Y. Moon, J. J. Lee, S. Choi, Y.-C. Kim, and S. Jeong, “Differentiation of cutaneous melanoma from surrounding skin using laser-induced breakdown spectroscopy,” Spectrochim. Acta, Part B 107, 57–66 (2016).

26. N. Melikechi, Y. Markushin, D. C. Connolly, J. Lasue, E. Ewusi-Annan, and S. Makrogianis, “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).

27. X. Chen, X. Li, X. Yu, D. Chen, and A. Liu, “Diagnosis of human malignancies using laser-induced breakdown spectroscopy in combination with chemometric methods,” Spectrochim. Acta, Part B 139, 63–69 (2018).

28. Y. W. Chu, T. Chen, F. Chen, Y. Tang, S. S. Tang, H. L. Jin, L. B. Guo, Y. F. Lu, and X. Y. Zeng, “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).

29. Y. W. Chu, F. Chen, Z. Q. Sheng, D. Zhang, S. Y. Zhang, W. L. Wang, H. L. Jin, J. W. Qi, and L. B. Guo, “Blood cancer diagnosis using ensemble learning based on a random subspace method in laser-induced breakdown spectroscopy,” Biomed. Opt. Express 7(1), 57–66 (2016).

30. J. S. Xiu, X. S. Bai, E. Negre, V. Motto-Ros, and J. Yu, “Indirect laser-induced breakdown of transparent thin gel layer for sensitive trace element detection,” Appl. Phys. Lett. 102(24), 244101 (2013).

31. Y. Tian, C. H. Yan, T. L. Zhang, H. S. Tang, H. Li, J. L. Yu, J. Bernard, L. Chen, S. Martin, N. Delepine-Gilon, J. Bocková, P. Veis, Y. Chen, and J. Yu, “Classification of wines according to their production regions with the contained trace elements using laser-induced breakdown spectroscopy,” Spectrochim. Acta, Part B 135, 91–101 (2017).

32. https://en.wikipedia.org/wiki/Blood_plasma

33. Z. Yue, C. Sun, L. Gao, Y. Zhang, S. Shabbir, W. Xu, M. Wu, L. Zou, Y. Tan, F. Chen, and J. Yu, “Machine learning efficiently corrects LIBS spectrum variation due to change of laser Fluence,” Opt. Express 28(10), 14345 (2020).

34. Y. Zhang, C. Sun, L. Gao, Z. Yue, S. Shabbir, W. Xu, M. Wu, and J. Yu, “Determination of minor metal elements in steel using laser-induced breakdown spectroscopy combined with machine learning algorithms,” Spectrochim. Acta, Part B 166, 105802 (2020).

35. W. Dai, G. R. Xue, Q. Yang, and Y. Yu, “Co-clustering based classification for out-of-domain documents,” Proc. 13th ACM SIGKDD Int’l Conf. Knowledge Discovery and Data Mining, 210–219 (2007).

36. Introduction to Algorithms, 2nd Edition, T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, eds. (MIT Press and McGraw-Hill 2001).

37. https://en.wikipedia.org/wiki/Chi-squared_test

38. https://en.wikipedia.org/wiki/Electrolyte_imbalance
39. W. Xu, C. Sun, Y. Tan, L. Gao, Y. Zhang, Z. Yue, S. Shabbir, M. Wu, L. Zou, F. Chen, S. Liu, and J. Yu, “Total alkali silica classification of rocks with LIBS: influences of the chemical and physical matrix effects,” J. Anal. At. Spectrom. 35(8), 1641–1653 (2020).

40. Y. Zhang, C. Sun, Z. Yue, S. Shabbir, W. Xu, M. Wu, L. Zou, Y. Tan, F. Chen, and J. Yu. “Correlation-based carbon determination in steel without explicitly involving carbon-related emission lines in a LIBS spectrum,” Opt. Express 28(21), 32019 (2020).

41. L. Zou, S. Chen, M. Wu, Y. Zhang, Z. Yue, W. Xu, S. Shabbir, F. Chen, B. Liu, W. Liu, and J. Yu, “Online simultaneous determination of H2O and KCl in potash with LIBS coupled to convolutional and back-propagation neural networks,” J. Anal. At. Spectrom. 36, 303–313 (2021).