Diagnosis and staging of multiple myeloma using serum-based laser-induced breakdown spectroscopy combined with machine learning methods

XUE CHEN,1,4 YAO ZHANG2,3,4 XIAOHUI LI,2,3,* ZIHENG YANG2,3 AICHUN LIU,1 AND XIN YU2,3

1Department of Hematology, Harbin Medical University Cancer Hospital, 150 Haping Road, 150081 Harbin, China
2Institute of Opto-electronics, Harbin Institute of Technology, 2 Yikuang Street, 150080 Harbin, China
3National Key Laboratory on Tunable Laser, Harbin Institute of Technology, 2 Yikuang Street, 150080 Harbin, China
4Contributed equally as co-first authors
*lixiaohui@hit.edu.cn

Abstract: Diagnosis and staging of multiple myeloma (MM) have been achieved using serum-based laser-induced breakdown spectroscopy (LIBS) in combination with machine learning methods. 130 cases of serum samples collected from registered MM patients in different progressive stages and healthy controls were deposited onto standard quantitative filter papers and ablated with a Q-switched Nd:YAG laser. Emissions of Ca, Na, K, Mg, C, H, O, N and CN were selected for malignancy diagnosis and staging. Multivariate statistics and machine learning methods, including principal component analysis (PCA), k-nearest neighbor (kNN), support vector machine (SVM) and artificial neural network (ANN) classifiers, were used to build the discrimination models. The performances of the classifiers were optimized via 10-fold cross-validation and evaluated in terms of accuracy, sensitivity, specificity, and area under the receiver operating characteristic curves (AUC). The kNN, SVM and ANN classifiers achieved comparable discrimination performances with accuracies of over 90% for both diagnosis and staging of MM. For diagnosis of MM, the classifiers achieved performances with AUC of ∼0.970, sensitivity of ∼0.930 and specificity of ∼0.910; for staging of MM, the corresponding values were AUC of ∼0.970, sensitivity of ∼0.910 and specificity of ∼0.930. These results show that the serum-based LIBS in combination with machine learning methods can serve as a fast, less invasive, cost-effective, and robust technique for diagnosis and staging of human malignancies.

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

1. Introduction

Laser-induced breakdown spectroscopy (LIBS) is a spectroscopic technique that analyzes the fingerprint atomic or molecular emissions of plasmas generated by focusing a pulsed laser on the samples. LIBS has been widely recognized in the community with many technical merits, including simplicity in experimental apparatus, no or limited sample preparation, ability of simultaneous multiple elements detection, standoff and on-site operation, etc. [1]. LIBS can also be easily integrated with the widely-used microscopy and endoscopy in the biomedical community. These characteristics make LIBS a potential solution for biomedical applications.

In recent years, LIBS has been applied on different biomedical samples, including bio-aerosols [2–5], soft tissues [6–8], and tumor tissues [9–13]. Discrimination and diagnosis of malignancies using blood-sample-based LIBS in combination with machine learning methods [12,14–17] is a new direction in the biomedical application community. Melikechi et al. realized age-specific discrimination of blood plasma samples of healthy and ovarian cancer prone mice using LIBS.
assisted with linear discriminant analysis (LDA) and random forest (RF) [14]. Gaudioso et al. demonstrated diagnosis of melanoma by analyzing serum and tissue homogenates of lungs, lymph nodes and spleens harvested from diseased mice and healthy controls, using LIBS in combination with LDA, support vector machine (SVM), Fisher discriminant analysis (FDA) and gradient boosting [12]. Chen et al. presented the first diagnosis of human lymphoma and multiple myeloma using serum-based and whole-blood-based LIBS in combination with LDA, quadratic discriminant analysis (QDA), and k-nearest neighbors classification (kNN) [15,16]. Chu et al. reported a follow-up work which realized discrimination of nasopharyngeal carcinoma using serum-based LIBS assisted by an extreme learning machine and RF method [17]. These Refs. [12,14–17] demonstrate that blood-sample-based LIBS supported by machine learning methods can realize quick and robust diagnosis of human malignancies.

Multiple myeloma (MM) is a plasma-cell neoplasm characterized with uncontrolled clonal of monoclonal plasma cells in the bone marrow [18,19]. MM is the second most common hematologic malignancy. International Agency for Research on Cancer estimated that MM would account for about 0.9% of new cancer cases and about 1.1% of cancer deaths around the world [20]. In China, over 14600 people die from MM each year [21]. Conventional diagnostic procedures for MM consist of laboratory studies, urine studies, bone marrow biopsy, and radiologic evaluation [18]. To complete the whole set of tests, it is time-consuming and expensive. Therefore, new methods need to be developed to realize fast, cost-effective, and accurate diagnosis of MM.

Our group has reported preliminary diagnosis of MM using serum-based LIBS in a recent work [16], with limited number of cases. In this work, with a much-expanded collection of cases, the serum-based LIBS technique is not only applied for diagnosis but also for staging of MM. The serum samples of registered MM patients in different progressive stages and healthy controls were deposited onto standard quantitative filter papers and ablated with a Q-switched Nd:YAG laser. The serum-LIBS spectra were compared between the MM patients and healthy controls, and among MM patients in different progressive stages. Three machine learning methods, including kNN, SVM, and artificial neural networks (ANN) classifiers, were used to build the models to diagnose and stage the disease. Cross-validation was used to evaluate and optimize the performances of the discrimination models, in terms of accuracy, sensitivity, specificity, and area under the receiver operating characteristic curves (AUC). Very good discrimination performances have been achieved, indicating that the serum-based LIBS assisted by machine learning methods can be a fast, cost-effective, and robust diagnosis and staging technique for MM.

2. Materials and methods

2.1. Serum sample preparation

Serum samples were collected from the MM patients registered in Department of Hematology, Harbin Medical University Cancer Hospital (HMUCH) and from healthy donators. All the subjects have signed informed consent in compliance with the Declaration of Helsinki. The protocol has been approved by the Clinical Research Ethics Committee of HMUCH. For each subject, about 2 mL blood sample was drawn from the vein on the inner portion of the arm, and collected using an EDTA-treated tube to prevent blood clotting. The blood samples were centrifugalized for 5 min to compact the cells. The serum samples were taken from the upper part and transferred to sterile micro centrifuge tubes. In total, serum samples of 130 subjects were collected, of which 55 were healthy volunteers and 75 were MM patients. The MM patients were separated into three progressive stages, including 30 Stage I, 16 Stage II and 29 Stage III cases. Staging of the patients were performed by two independent pathologists in Department of Pathology, HMUCH. All the serum samples were stored in a refrigerator at −40 degree Celsius until the LIBS measurements.
Prior to the LIBS analysis, pretreatment was performed to the serum samples. A piece of the micro centrifuge tube was taken for each subject. The serum sample was naturally thawed. After stirring in an ultrasonic tank for 1 min, 100 microliter of serum sample was taken from the micro centrifuge tube using a pipette, and then uniformly deposited onto a piece of quantitative filter paper of 25 mm by 25 mm in size. The thickness of the serum layer was estimated to be about 3–5 µm. The quantitative filter paper was made following China national standard GB/T1914-2007. The main content of the filter paper is purified cellulose. The contribution of the ash content to the spectra was thought to be neglectable due to its very low (less than 0.01%) concentration in the filter paper [16]. The filter paper with the serum sample was naturally dried for 20 min in an air-filtered laminar flow cabinet, then sent for LIBS measurements.

2.2. LIBS measurement

The LIBS system for serum sample analysis has been reported elsewhere [15,16]. In short, a Q-switched Nd: YAG laser operating at 1064 nm with a pulse width of ∼8 ns was used to generate the plasma (see Fig. 1). The laser was focused on the surface of the serum-deposited filter paper using a 75 mm focal length quartz lens. The laser pulse energy was set to ∼25 mJ. The sample was translated relative to the laser pulse using a three-dimensional translation stage platform (SHOT-302GS, OptoSigma) in a zig-zag mode to guarantee access to new point on the filter paper for each ablation. A four-channel spectrometer (AvaSpec-ULS2048-4, Avantes) covering the spectral range of 210–850 nm with resolution of 0.09–0.22 nm (Δλ/λ=2.5–4.8×10^{-4}) is used to collect the emission from the plasma. The plasma emission was coupled to a 4-in-1 fiber-bundle using a combination of two quartz lenses with focal length of 50 mm, and then transmitted to the spectrometer. The whole system was synchronized using a delay generator (9214+, Quantum). After receiving a trigger from the translation stage controller, the delay generator outputs triggers for the laser and the spectrometer. The delay time for spectrum collection was optimized to 0.2 µs following firing of the laser and the integration time was fixed to 1.05 ms (limited by the specification of the spectrometer).

The list of the serum samples was randomly ordered for LIBS measurements, with one (or two) cancer sample(s) followed by one normal sample, such that the interference caused by the fluctuation of experimental conditions would be balanced between the cancer and the normal categories. The LIBS measurements were conducted in the ambient air environment under atmospheric pressure. For each sample, a total of 1960 ablations were conducted. To minimize the impact of laser energy fluctuation and external environment on the measurements, the spectra were averaged with a batch size of 20, therefore 98 average spectra were obtained for each sample. The typical intensity variation among the average spectra was about 10–12%. The average spectra were then used for subsequent diagnosis and staging analysis.
2.3. **Data processing and machine learning methods**

The spectral data processing was performed using Python with scipy, and machine learning was processed using the scikit-learn library. The original average spectra were preprocessed to remove baseline using a home-made algorithm. In short, spectral peaks and valleys were detected first through the whole spectrum; then, sub-baseline between two adjacent valleys was determined by a least squares fit; finally, the sub-baselines were connected to obtain the whole baseline. The baseline-removed spectra were normalized to the Na I 589.99 nm line because of its good signal-to-noise ratio (SNR), to reduce the impact of experimental fluctuations [8]. Then, 52 representative emission lines with high intensities and good SNRs were selected for diagnosis and staging analysis. The detailed information of the selected lines was listed in Table 1 [22].

The intensity of the selected lines was retrieved for each sample. Two spectral data matrices (denoted as “original data matrices”) for diagnosis and staging were obtained, with dimension of 12740 × 52 and 7350 × 52, respectively. An outlier data removal processing was performed on the original data matrices. The original data matrices were standardized to have zero mean and unit variance along each column (i.e., spectral feature) and then subjected to principal component analysis (PCA). Here, standardization is used to normalize the features of different numerical scales to the same range, such that the features with large numerical values won’t dominate the following classification. The outlier data points were identified in the principal component (PC) space, using the Euclidean distance of the data points relative to the centroid of these points as the criteria [16]. The data points that were at least two standard deviations away from the centroid were treated as outliers. The final dimension of the outlier-free spectral data matrices was 12156 × 52 and 7008 × 52, respectively. The two outlier-free spectral data matrices were standardized again to have zero mean and unit variance along each column, and then used to train the classifiers for diagnosis and staging of MM, respectively.

| Emitting species | Wavelength (nm) |
|------------------|-----------------|
| C I              | 247.87          |
| Ca I             | 422.67, 430.25, 443.50, 445.48, 527.03, 558.87, 559.44, 612.22, 616.21, 643.91 |
| Ca II            | 317.93, 393.36, 396.85 |
| CN B^3Σ^+ - X^3Σ^- | 358.59, 359.04, 385.09, 385.47, 386.19, 387.13, 388.34, 415.81, 416.78, 418.10, 419.72, 421.60 |
| Hα               | 656.28          |
| K I              | 766.49, 769.90  |
| Mg II            | 279.55, 280.27  |
| Mg I             | 285.21          |
| Na I             | 568.26, 568.82, 588.99, 589.59, 818.32, 819.48, 819.91 |
| N I              | 742.36, 744.23, 746.83, 818.80, 821.07, 821.63, 824.24 |
| O I              | 715.67, 777.19, 777.42, 777.54, 822.18, 844.64 |

Three machine learning methods, i.e., kNN, SVM, and ANN classifiers, were used to diagnose and stage the MM. These classifiers have been used for classification or discrimination of complex samples in the community. kNN classifier is an instance-based classification method. It determines the membership of an unknown data point by the majority vote of its $k$ nearest neighbors following specific distance metrics [8,15,16]. It exhibited high robustness in discrimination of complex substances such as polymers [23,24], explosives [25] and soft tissues [8]. The SVM classifier distinguishes two classes of data by finding the best hyperplane that separates data points of one class from those of the other class [8]. It can deal with nonlinear problems using the kernel functions. It has been used for classification of pharmaceutical samples [26], suspect power [27] and soft tissues [8]. For kNN and SVM classifiers, the standardized outlier-free
spectral data matrices were subjected to PCA first, then the scores (i.e., the expression of the spectral data in the PC space) of a specific number of PCs were used as the input for the models, to reduce the dimension of data and avoid overfitting.

ANN mimics human brain with connections between different neuronal layers. By adjusting the weights for the connections, it reduces the errors between the output and the target values. It has been successfully used for discrimination of animal tissues [28] and bacterial strains [3,29]. A three-layer neural network was applied in this work, including an input layer, one hidden layer, and an output layer. The sigmoid function was used as the activation function. The standardized outlier-free spectral data matrices were directly used as the input for the network. The target values of samples were assigned as numeric values: for the diagnosis application, the MM and healthy control classes were assigned as 1 and 2, respectively; for staging application, the MM in Stage I, Stage II and Stage III were assigned as 1, 2 and 3, respectively [28,30]. The classifier was trained to minimize the mean square error (MSE) between the output and the target values. Here, the backward propagation algorithm was used. Early stopping was used to avoid overfitting. After training, the membership of an unknown sample was assigned to the class with the closest target value respective to its output.

The performances of the classifiers were optimized and evaluated using the 10-fold cross-validation. The samples were randomly divided into 10 disjoint subsamples (or folds). Ten independent sub-classifiers were trained. For each sub-classifier, 9 folds were used to train the model, and the rest one fold was used to validate the classifier. The mean results of the 10 sub-classifiers were reported as the final performances. It should be emphasized that, when doing the data partition, each fold includes data from all the 130 subjects. Thus, the 10-fold cross-validation can not only compensate the intra-patient variations, but also compensate the inter-patient variations. Although the leave-one-out cross-validation (LOOCV) may lead to more accurate evaluation of the model, with large number (98) of replica spectra for each subject, the 10-fold cross-validation would report similar results to the LOOCV, and thus can accurately evaluate the performances of the classifiers (see discussions in section 3.2). The accuracy, sensitivity and specificity of the classifiers were determined with margins of error calculated as the standard variance among different sub-classifiers. The receiver operating characteristic (ROC) curves were also obtained for sub-classifiers and the AUC values were determined and compared.

3. Results and discussion

3.1. Spectral analysis of the serum samples

Figure 2 shows the normalized average LIBS spectra of serum samples of MM patients and healthy controls. The spectra of the two categories are visually similar. Atomic and ionic emissions from C, H, O, N, Ca, Na, K, Mg, and molecular emissions from CN B-X system can be observed.

To make a more comprehensive comparison, the intensities of several representative lines were statistically analyzed. As shown in Fig. 3 (a), with the large number of subjects (75 MM patients vs. 55 healthy controls), the spectral intensities of the MM patients and healthy controls are generally comparable, yet marginal difference could be observed in intensities of C and Ca. Shown in Fig. 3 (b) is the comparison of spectral intensities of selected lines of the serum-LIBS spectra of MM patients in different progressive stages. The serum samples in Stage II show the highest intensities on Ca, Mg, C, O and N. The intensities of Ca in Stage I are higher than those in Stage III, while the intensities of other elements are comparable to those in Stage III.

The differences in spectral intensities indicate that relative concentrations of the elements are different between normal and MM patients, and among patients in different stages. These special patterns in elemental concentration may be used for diagnosis and staging of MM. However, it should be noted that the variations of spectral intensities between different categories are
**Fig. 2.** Average LIBS spectra of serum samples of MM patients and healthy controls.

**Fig. 3.** Comparison of intensities of representative lines of serum-LIBS spectra: (a) between MM patients and healthy controls, (b) among MM patients in different progressive stages.
complex, such that it is not possible to discriminate the samples by directly comparing the raw intensity of the emission lines. This indicates the importance of machine learning methods to achieve robust diagnosis and staging of the malignancy.

3.2. Diagnosis and staging of MM using machine learning methods

To diagnose and stage MM, three machine learning methods, i.e., kNN, SVM and ANN classifiers were applied. For kNN and SVM classifiers, the two standardized outlier-free spectral data matrices were subjected to PCA first to reduce the dimension of data. The PCA results show that, most total variances can be explained by a small number of PCs. For both data matrices, more than 95% of total variance is explained by the first 9 PCs. Shown in Fig. 4 is the three dimensional scatter plot of the samples in PC3-PC10-PC12 space. Significant overlapping can be observed between the MM and healthy control classes, indicating that linear model such as LDA may not achieve good discrimination performances. However, the two classes do show different distribution patterns, which indicates the possibility of discrimination using more advanced classifiers, such as kNN, SVM and ANN.

![Scatter plot in PC3-PC10-PC12 space for diagnosis of MM.](image)

The major contributors to the PCs are atomic emissions, especially the emissions from Ca, Mg, K, and Na. For example, for PC3, the Na I and K I lines show strong positive contributions, while the Ca I and Ca II lines contribute negatively. For PC10, the contributions from Na I and K I lines decrease greatly, while the Ca I and Ca II lines still show strong contributions; however, unlike in PC3, the Ca contributions reverse from negative to positive values. For PC12, the negative contributions of Mg I and Mg II lines increase. While the Ca II lines still contribute positively, the Ca I line contributions reverse to negative values.

3.2.1. kNN classifier

kNN classifier was used first for diagnosis and staging of MM. The score vectors of a specific number of PCs were used as the predictor features. The number of selected PCs were determined by evaluating the variation of discrimination losses with the number of PCs, internally to the cross-validation. Generally, the discrimination loss will decrease with the increasing number of
A preliminary comparison of cross-validation methods has been performed for the kNN classifier and the SVM classifier discussed in the following subsection, for diagnosis of MM. For the kNN classifier, with the number of nearest neighbors ($k$ value) as $k=10$, cityblock distance function and 19 input PCs, the 10-fold cross-validation reports accuracy of $91.9 \pm 0.5\%$, and the leave-one-out cross-validation (LOOCV) reports accuracy of $93.1 \pm 0.1\%$; for the SVM classifier, with the Gaussian kernel function and 19 input PCs, the 10-fold cross-validation reports accuracy of $91.6 \pm 0.7\%$, and the LOOCV reports accuracy of $90.4 \pm 0.2\%$. It can be seen the 10-fold cross-validation generally reports similar results to the LOOCV, thus can be used for evaluation of the performances of the classifiers. Therefore, the 10-fold CV was used throughout this work to evaluate the classifiers.

The $k$ value and distance metric functions were optimized internally using the 10-fold cross-validation to achieve the best discrimination performances. It was found that, for both diagnosis and staging, the best performances were obtained with $k=10$ using the cityblock distance function. For diagnosis of MM, the accuracy was achieved as $91.9 \pm 0.5\%$, with sensitivity of $0.932 \pm 0.007$ and specificity of $0.903 \pm 0.011$, see Table 2. For staging of MM, the accuracy was achieved as $91.8 \pm 0.1\%$. The sensitivity values for Stage I, Stage II, Stage III were $0.909 \pm 0.010$, $0.881 \pm 0.027$ and $0.914 \pm 0.020$, respectively, and the corresponding specificity values were $0.953 \pm 0.015$, $0.981 \pm 0.008$ and $0.962 \pm 0.006$, respectively.

### 3.2.2. SVM classifier

The SVM classifier was then used for diagnosis and staging of MM. The optimum number of input PCs for the classifier was also determined as 19 via the 10-fold cross-validation. Thus, the score vectors of the first 19 PCs were also used as the predictor features. The SVM classifier distinguishes two classes of data by finding the best hyperplane that separates data points of one
Table 2. Discrimination performances of kNN classifiers for diagnosis and staging of MM

| Discrimination performances | Diagnosis | Staging |
|-----------------------------|-----------|---------|
|                             | MM        | Normal  | Stage I | Stage II | Stage III |
| Sensitivity                 | 0.932 ± 0.007 | 0.903 ± 0.011 | 0.909 ± 0.010 | 0.881 ± 0.027 | 0.914 ± 0.020 |
| Specificity                 | 0.903 ± 0.011 | 0.932 ± 0.007 | 0.953 ± 0.015 | 0.981 ± 0.008 | 0.962 ± 0.006 |
| AUC                         | 0.972 ± 0.002 | 0.972 ± 0.002 | 0.977 ± 0.004 | 0.987 ± 0.004 | 0.983 ± 0.003 |
| Accuracy                    | 91.9 ± 0.5%  | 91.8 ± 0.1%  | 91.6 ± 0.7%  | 91.3 ± 0.8%  | 91.2 ± 0.7%  |

class from those of the other class. In the cases when the binary classification problems do not have a simple hyperplane as a useful separating criterion, nonlinear transformation with kernel functions can be used [8]. After a comparison among linear, polynomial and Gaussian kernel functions, performed internally to the cross-validation, we found that the classifiers achieved best performances with the Gaussian kernel function.

The discrimination performances of SVM classifiers for diagnosis and staging of MM using Gaussian kernel function are shown in Table 3. The classifiers achieved the accuracies of 91.6 ± 0.7% and 91.3 ± 0.8% for diagnosis and staging, respectively. The sensitivity for diagnosis of MM was 0.924 ± 0.011, and the corresponding specificity was 0.909 ± 0.020. For staging of MM, the sensitivity values for Stage I, Stage II, Stage III were 0.917 ± 0.013, 0.881 ± 0.030 and 0.917 ± 0.013, respectively, and the corresponding specificity values were 0.932 ± 0.010, 0.972 ± 0.009 and 0.947 ± 0.006, respectively.

Table 3. Discrimination performances of SVM classifiers for diagnosis and staging of MM using Gaussian kernel function

| Discrimination performances | Diagnosis | Staging |
|-----------------------------|-----------|---------|
|                             | MM        | Normal  | Stage I | Stage II | Stage III |
| Sensitivity                 | 0.924 ± 0.011 | 0.909 ± 0.020 | 0.917 ± 0.013 | 0.881 ± 0.030 | 0.917 ± 0.013 |
| Specificity                 | 0.909 ± 0.020 | 0.924 ± 0.011 | 0.932 ± 0.010 | 0.972 ± 0.009 | 0.947 ± 0.006 |
| AUC                         | 0.970 ± 0.004 | 0.970 ± 0.004 | 0.974 ± 0.005 | 0.984 ± 0.004 | 0.979 ± 0.005 |
| Accuracy                    | 91.6 ± 0.7%  | 91.3 ± 0.8%  | 91.6 ± 0.7%  | 91.3 ± 0.8%  | 91.3 ± 0.8%  |

3.2.3. ANN classifier

Finally, the ANN classifier was used for diagnosis and staging of MM. The number of hidden layer neurons is an important parameter to optimize the network. If the number is too small, the network lacks flexibility; on the other hand, if the number is too large, it will increase the training cost and may be prone to overfitting. The number of hidden layer neurons were varied between 1 to 40. As shown in Fig. 6, the optimum value was determined as 21 and 19 for diagnosis and staging of MM, respectively, via the 10-fold cross-validation.

After optimization, the ANN classifiers achieved best accuracies of 93.3 ± 0.4% and 93.7 ± 0.8% for diagnosis and staging, respectively. The sensitivity for diagnosis of MM was 0.935 ± 0.009, and the specificity was 0.930 ± 0.009. For staging of MM, the sensitivity values for Stage I, Stage II, Stage III were 0.944 ± 0.015, 0.944 ± 0.012 and 0.926 ± 0.018, respectively, and the corresponding specificity values were 0.933 ± 0.011, 0.935 ± 0.006 and 0.944 ± 0.007, respectively. The ROC curve can also be used to compare the performances of the classifiers. It plots sensitivity versus (1-specificity) with different discrimination thresholds. It shows the tradeoff between true positive rate and false positive rate of a classifier. For an ideal classifier, the working point should locate at the upper left corner of the curve. The area under the ROC curve (AUC) is another parameter to evaluate the performances of the classifier. An ideal classifier would have an...
Fig. 6. Optimization of hidden layer neurons of ANN classifiers (a) diagnosis (b) staging.

AUC = 1. Shown in Fig. 7 are the ROC curves of the kNN, SVM and ANN classifiers for diagnosis and staging of MM. For each classifier, the ROC curves of the 10 training folds are presented. All three classifiers show good discrimination characteristics. For each classifier, variations of the ROC curves between different training folds can be observed. Meanwhile, the ROC curves of different types of classifiers overlap with each other, indicating that the performances of the three classifiers are generally comparable. The corresponding AUC values are shown in Tables 2–4. When considering the error margins, the AUC values of different types of classifiers are also generally comparable.

Table 4. Discrimination performances of ANN classifiers for diagnosis and staging of MM

| Discrimination performances | Diagnosis | Staging |
|-----------------------------|-----------|---------|
|                            | MM        | Normal  | Stage I | Stage II | Stage III |
| Sensitivity                | 0.935 ± 0.009 | 0.930 ± 0.009 | 0.944 ± 0.015 | 0.944 ± 0.012 | 0.926 ± 0.018 |
| Specificity                | 0.930 ± 0.009 | 0.935 ± 0.009 | 0.933 ± 0.011 | 0.935 ± 0.006 | 0.944 ± 0.007 |
| AUC                         | 0.979 ± 0.002 | 0.979 ± 0.002 | 0.973 ± 0.006 | 0.984 ± 0.004 | 0.974 ± 0.004 |
| Accuracy                    | 93.3 ± 0.4% | 93.7 ± 0.8% |

3.3. Discussion

Diagnosis and staging of malignancies are challenging issues in the clinical practices. Here, we investigated the possibility of diagnosis and staging of MM using serum-based LIBS, using a large case number (130) of samples. Spectral analysis demonstrates differences in spectral intensities between the serum samples of MM patients and healthy controls, and among the MM patients in different stages. Yet, the complex trend of variation of spectral intensities hiders discrimination by direct comparison of raw spectral intensities. Therefore, machine learning methods should be introduced for robust diagnosis and stage of the malignancy.

The kNN, SVM and ANN classifiers all achieved good discrimination performances, with accuracies of over 90\% for both diagnosis and staging of MM. The ANN classifier showed slightly higher accuracies than kNN and SVM classifiers. However, it should be noted that, the classifiers may prone to bias from the true generalization performances with only in-model cross-validation. When considering the potential bias and the error margins, the three classifiers are considered to have generally comparable performances. For diagnosis of MM, the classifiers achieved discrimination performances with AUC of ~0.970, sensitivity of ~0.930 and specificity
The results show that, the serum-based LIBS in combination machine learning methods, can achieve accurate and robust diagnosis and staging of MM. Using the serum samples as the analyte, which is routine in clinical practices, this technique is faster, less invasive, and more cost-effective than conventional pathology. Besides malignancy diagnosis, the technique can also stage human malignancies. This could help to evaluate the severity of a certain cancer in the early period, and provide valuable information for medical treatment. Furthermore, with the merits of simplicity in sample preparation, compact apparatus, and ambient working condition, it is possible to adopt the LIBS system in various medical branches to carry out vast malignancy screening, which can help to reduce the morbidity and mortality of malignancies.

It is fair to point out that, although the fitting of the classifiers, including the optimization of their settings (i.e. the number of PCs to retain for the kNN and SVM classification, and the size of the hidden layer in the ANN classification, as well as the choices of the distance and kernel functions in the kNN and SVM classifications, respectively, and k and the kernel hyper-parameter in the kNN and SVM classifications, respectively) was cross-validated to minimize over-fitting in model selection, the reported performance measures, being computed on the same data as used for the model fitting, should not be used for an objective evaluation of the generalization performance of these methods. An independently collected dataset would be ideal for providing objective estimates of the expected generalization performance of the reported methods and of any of their fixed representatives (such as those corresponding to the settings optimized on the given dataset).
4. Conclusions

In this work, serum-based LIBS in combination with machine learning methods were applied for diagnosis and staging of multiple myeloma (MM). Serum samples of MM patients in different progressive stages and healthy controls were analyzed using LIBS. Multivariate statistics and machine learning methods, including PCA, kNN, SVM and ANN classifiers, were used to discriminate the samples of different categories. The classifiers were optimized via 10-fold cross-validation and evaluated in terms of accuracy, sensitivity, specificity, and ROC curves.

The kNN, SVM and ANN classifiers achieved comparable discrimination performances with accuracies of over 90% for both diagnosis and staging of MM. For diagnosis of MM, the classifiers achieved performances with AUC of ~0.970, sensitivity of ~0.930 and specificity of ~0.910; for staging of MM, the corresponding values were AUC of ~0.970, sensitivity of ~0.910 and specificity of ~0.930. The results show that serum-based LIBS in combination with machine learning methods can be a fast, less invasive, cost-effective, and robust technique for diagnosis and staging of MM. It can help to reduce the morbidity and mortality of malignancies. Further expansion of the serum library and improvement of the classifiers are in progress. Applications to other types of malignancies are under investigation.

Funding. National Natural Science Foundation of China (61975042); Heilongjiang Provincial Postdoctoral Science Foundation (LBH-Q19016).

Acknowledgements. The authors thank all cancer patients and healthy volunteers participating in this protocol.

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

References
1. 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), 335A–336A (2010).
2. M. Baudelet, J. Yu, M. Bosu, 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, 163903 (2006).
3. S. Manzoor, S. Moncayo, F. Navarro-Villoslada, J. A. Ayala, R. Izquierdo-Hornillos, F. J. M. de Villena, and J. O. Caceres, “Rapid identification and discrimination of bacterial strains by laser induced breakdown spectroscopy and neural networks,” Talanta 121, 65–70 (2014).
4. S. J. Rehse, J. Diedrich, and S. Palchaudhuri, “Identification and discrimination of Pseudomonas aeruginosa bacteria grown in blood and bile by laser-induced breakdown spectroscopy,” Spectrochim. Acta, Part B 62(10), 1169–1176 (2007).
5. S. J. Rehse, H. Salimnia, and A. W. Miziolek, “Laser-induced breakdown spectroscopy (LIBS): an overview of recent progress and future potential for biomedical applications,” J. Med. Eng. Technol. 36(2), 77–89 (2012).
6. R. Kanawade, F. Mehari, C. Knipfer, M. Rohde, K. Tangermann-Gerk, M. Schmidt, and F. Stelzle, “Pilot study of laser induced breakdown spectroscopy for tissue differentiation by monitoring the plume created during laser surgery — An approach on a feedback Laser control mechanism,” Spectrochim. Acta, Part B 87, 175–181 (2013).
7. F. Mehari, M. Rohde, R. Kanawade, C. Knipfer, W. Adler, F. Klämpfl, F. Stelzle, and M. Schmidt, “Investigation of the differentiation of ex vivo nerve and fat tissues using laser-induced breakdown spectroscopy (LIBS): Prospects for tissue-specific laser surgery,” J. Biophotonics 9(10), 1021–1032 (2016).
8. X. Li, S. Yang, R. Fan, X. Yu, and D. Chen, “Discrimination of soft tissues using laser-induced breakdown spectroscopy in combination with k nearest neighbors (kNN) and support vector machine (SVM) classifiers,” Opt. Laser Technol. 102, 233–239 (2018).
9. A. El-Hussein, A. K. Kassem, H. Ismail, and M. A. Harith, “Exploiting LIBS as a spectrochemical analytical technique in diagnosis of some types of human malignancies,” Talanta 82(2), 495–501 (2010).
10. 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,” Biomed. Opt. Express 7(1), 57–66 (2016).
11. A. Kumar, F.-Y. Yueh, J. P. Singh, and S. Burgess, “Characterization of malignant tissue cells by laser-induced breakdown spectroscopy,” Appl. Opt. 43(28), 5399–5403 (2004).
12. R. Gaudioso, E. Ewasi-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).
13. G. Teng, Q. Wang, H. Zhang, W. Xiaogl, H. Yang, X. Qi, C. Cui, B. S. Idrees, K. Wei, and M. N. Khan, “Discrimination of infiltrative glioma boundary based on laser-induced breakdown spectroscopy,” Spectrochim. Acta, Part B 165, 105787 (2020).
14. N. Melikechi, Y. Markushin, D. C. Connolly, J. Lasue, E. Ewusi-Annan, and S. Makrogiannis, “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).

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

16. 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).

17. Y. Chu, T. Chen, F. Chen, Y. Tang, S. Tang, H. Jin, L. Guo, Y. F. Lu, and X. 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).

18. K. Brigle and B. Rogers, “Pathobiology and diagnosis of multiple myeloma,” Seminars in Oncology Nursing 33(3), 225–236 (2017).

19. A. Mahindra, T. Hideshima, and K. C. Anderson, “Multiple myeloma: biology of the disease,” Blood Reviews 24, S5–S11 (2010).

20. 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: A Cancer Journal for Clinicians 68(6), 394–424 (2018).

21. J. Moros, J. Serrano, C. Sánchez, J. Macías, and J. J. Laserna, “New chemometrics in laser-induced breakdown spectroscopy for recognizing explosive residues,” J. Anal. At. Spectrom. 27(12), 2111–2122 (2012).

22. N. C. Dingari, I. Barman, A. K. Myakalwar, S. P. Tewari, and M. Kumar Gundawar, “Incorporation of Support Vector Machines in the LIBS toolbox for sensitive and robust classification amidst unexpected sample and system variability,” Anal. Chem. 84(6), 2686–2694 (2012).

23. J. Cisewski, E. Snyder, J. Hannig, and L. Oudejans, “Support vector machine classification of suspect powders using laser-induced breakdown spectroscopy: Chemometric applications for better data interpretation,” Talanta 134, 65–73 (2015).

24. F.-Y. Yueh, H. Zheng, J. P. Singh, and S. Burgess, “Preliminary evaluation of laser-induced breakdown spectroscopy for tissue classification,” Spectrochim. Acta, Part B 64(10), 1059–1067 (2009).

25. S. Manzoor, L. Ugena, J. Tornero-López, H. Martin, M. Molina, J. J. Camacho, and J. O. Cáceres, “Laser induced breakdown spectroscopy for the discrimination of Candida strains,” Talanta 155, 101–106 (2016).

26. S. Moncayo, S. Manzoor, J. D. Rosales, J. Anzano, and J. O. Caceres, “Qualitative and quantitative analysis of milk for the detection of adulteration by Laser Induced Breakdown Spectroscopy (LIBS),” Food Chem. 232, 322–328 (2017).