

Contents

1	Introduction	2
2	Machine learning methods	2
2.1	Unsupervised learning algorithms	3
2.2	Supervised learning algorithms	3
2.3	Semi-supervised learning algorithms	4
3	Data preprocessing with machine learning	4
3.1	Denoise and de-interference	4
3.2	Spectral data selection	5
3.3	Variable reconstruction	7
4	Qualitative and quantitative modeling with machine learning	8
4.1	Qualitative model	8
4.2	Quantitative model	10
5	Summary and prospects	16
	Declarations	19
	Acknowledgements	19
	References	19

1 Introduction

Laser-induced breakdown spectroscopy (LIBS) is a powerful atomic emission spectroscopy technology. Since its first report in 1962, LIBS technology has been regarded as the future superstar for chemical analysis due to its unique advantages, including simple sample pretreatment, fast detection speed, small sample damage, and less sample and environmental limitation. With the advent of high-performance lasers and spectrometers, the analytical performance of LIBS technology has been significantly improved, and it has been widely used in various fields, such as industry, agriculture, biology, medicine, and space exploration. However, the laser-induced plasma is spatially inhomogeneous and drastically temporal varying due to intense interaction among the laser, sample, plasma, and surrounding gases [1, 2]. Therefore, the quantitation results of the physical principle-based model are mostly unsatisfactory because these models apply only one or a few spectral line intensities and was not able to compensate for matrix effect as well as signal fluctuations. How to effectively improve the quantitative analysis performance of LIBS remains a critical challenging issue for all LIBS researchers.

LIBS is a complex spectrum composed of ionic, atomic, and molecular spectral lines, which include information from the sample and environmental gases. Meanwhile, the LIBS spectrum is also influenced by factors such as intensity fluctuations, matrix effects, self-absorption, and spectral line interference [3]. Therefore, signal acquisition and data processing methods have become important directions for LIBS research. Among them, the quantitative method established by manually extracting spectral feature information relies on personal

experience, which is subjective and one-sided, making it difficult to fully utilize the useful information in spectral data. Machine learning algorithms have been an emerging interdisciplinary method, which can select useful information to the maximum extent from LIBS data with minimal requirement for subjective explanations [4]. It is undeniable that various types of machine learning methods have emerged in LIBS in recent years, and it is necessary to summarize the progress and look forward to future development through a review. Zhang *et al.* [5, 6] summarized the research progress of chemometrics methods in LIBS from the spectral data pre-processing, qualitative and quantitative analysis a few years ago. In another review, the research situation and progress of the application of machine learning algorithms to LIBS were reviewed, and the problems and challenges such as over-fitting, under-fitting and spectra noise in the application of machine learning algorithms to LIBS still needs to overcome [7]. The above reviews give a good summary of the current application of machine learning algorithms in LIBS, but do not focus on the key problems encountered by machine learning in solving the quantitative analysis of LIBS, such as matrix effects problems and signal uncertainty. Wang *et al.* [1] provided a guideline for LIBS researchers with basic knowledge for further quantification improvement, including the introduction of LIBS uncertainty generation mechanism, plasma modulation methods and quantification methods. It is noted that machine learning algorithms based on physical principles should be introduced into LIBS quantifications and it could be the final solution for cases with large amount of data available.

In this review, the recent progress of machine learning was comprehensively summarized and discussed around its methodology and application for data preprocessing, qualitative and quantitative analysis in LIBS fields. In particular, the role of machine learning algorithms in improving analysis repeatability and suppressing matrix effects were emphasized. Furthermore, the application prospects and suggestions for machine learning in LIBS are proposed.

2 Machine learning methods

Machine learning is a subfield of artificial intelligence (AI) that uses algorithms trained on data sets to create self-learning models that can predict outcomes and classify information. According to the learning approaches, machine learning can be separated into unsupervised learning, supervised learning, and semi-supervised learning, among which supervised learning builds a model based on labeled data, unsupervised learning is based on unlabeled data, and semi-supervised learning is based on a mix of labeled and unlabeled data.

A comparison of supervised, unsupervised, and semi-supervised learning is listed in Table 1. To present a

**Table 1** Comparison of supervised, unsupervised, and semi-supervised.

Category	Supervised	Unsupervised	Semi-supervised
Input data	All data is labeled	All data is unlabeled	Partially labeled
Training	External supervision	No supervision	External/internal supervision
Example algorithms	Decision trees, Support vector machine, Linear discriminant analysis, Linear regression, K-nearest neighbor, Logistic regression, Naive Bayes	K-means, Principal component Analysis, Hierarchical clustering, Generative adversarial networks, Iterative self-organizing data	Generate semi-supervised models, Self-trained models (e.g., Self-trained Naive Bayes classifier), Co-trained models (e.g., Co-trained K-nearest neighbor), Semi-supervised support vector machines.
Advantages/disadvantages	High accuracy, time-consuming, a large number of known samples are needed for modeling, cannot give “unknown” information	Time-saving, less accurate predictions for high-dimensional datasets, sensitivity to noise and outliers	Fewer labeled samples are needed, complex iterative process, not as accurate as supervised learning, cannot handle more “complex tasks”

simple guideline for readers, these three types of machine learning were briefly introduced below, including algorithm principles, functions, advantages, and disadvantages.

2.1 Unsupervised learning algorithms

For unsupervised learning recognition, the distance between similar compounds is small in multidimensional space but it is larger between the different compounds, which leads to the analysis and cluster of unlabeled data sets. The most common methods of unsupervised learning recognition used in LIBS are K-means [8–10], principal component analysis (PCA) [11], hierarchical clustering [12], and iterative self-organizing data analysis technique (ISODATA) [12]. K-means is computationally efficient and can handle large datasets with high dimensionality. However, the algorithm is sensitive to the initial selection of centers and can converge to a suboptimal solution. Furthermore, K-means is sensitive to outliers, which can have a significant impact on the resulting clusters. ISODATA algorithm is a modification of the K-means clustering algorithm, the clustering process begins with an arbitrary clustering average and is not limited by the initial center selection. PCA provides a dimensionality reduction of the original data set by the generalization of the original variance, which is done by a transformation of the original high-dimensional space into a smaller set of independent variables called principal components (PCs). Hierarchical clustering is simple and easy to use, but it will face the problem of insufficient performance for large-scale datasets due to the corresponding high time complexity. These algorithms are “unsupervised” because they discover hidden correlations in the data without human intervention. Unsupervised learning models are used in LIBS for three main tasks including clustering, association, and dimensionality reduction.

2.2 Supervised learning algorithms

The basic idea of supervised pattern recognition is that the samples with a known class as a training set are used to construct a training model, and then the class or

grade of unknown samples can be predicted by the training model. The common supervised pattern recognition methods mainly include multiple linear regression (MLR) [13–15], partial least squares (PLS) [16, 17], soft independent modeling class analog (SIMCA) [18], K-nearest neighbor method (KNN) [19–21], support vector machines (SVM) [22–26], artificial neural networks (ANN) [27–31], random forest (RF) [32–36], kernel extreme learning machine (KELM) [37–39], and linear discriminant analysis (LDA) [40–42].

Each algorithm has its own characteristics, and it is necessary to choose the appropriate algorithm and avoid its shortcomings in practical application. MLR method has the advantages of being high-speed, simple, and easy to implement, especially for small data and simple relationships. If the data distribution presents complex curves or the characteristics are not independent, MLR can not be utilized to build a model. PLS method is a multivariate statistical method used to find the basic relationship between spectral intensity and elemental content (or clustering label), which combines features of PCA and multiple regression. PLS has been widely used in LIBS data processing and efficiently handles high dimensionality and collinearity problems, however, the PLS classification or regression model is easy to fall into over-fitting for a small number of training samples. SIMCA is a pattern recognition method based on PCA, the limitation of SIMCA is that a non-optimized differential model would be generated when the difference between the classes is close to the differences from the class itself. KNN algorithms uses proximity to make classifications or predictions about the grouping of an individual data point. KNN makes the most direct use of the relationships between samples, reducing the adverse effects of improper selection of class features on classification results, and can minimize the errors in the classification process to the greatest extent possible. KNN is not sensitive to outliers, but has high computational complexity and spatial complexity. SVM is a linear method in a very high dimensional feature space that is nonlinearly related to the input space. But it does not achieve satisfactory results when training high-dimensional data, which is due to the consumption of huge computer

memory and operation time. ANN is a processing system based on imitating the structure and function of the brain neural network, which have advantages in self-learning and processing nonlinear relationships. Of course, ANNs also have some disadvantages, such as high requirements for computing hardware, susceptible to overfitting, and difficulty interpreting the decision-making processes. RF is ease of use and flexibility have fueled its adoption in LIBS, as it handles both classification and regression problems. Compared to other methods, RF has higher accuracy, and higher speed, can balance the error data and evaluate the importance of each variable, especially avoiding over-fitting problems. KELM is an improved algorithm based on extreme learning machine (ELM) and combined with kernel functions. Some researchers applied a stable kernel function to replace the random feature space in ELM, which exhibits fast learning speed, better stability, and generalization performance.

2.3 Semi-supervised learning algorithms

Semi-supervised learning is a popular algorithm that serves as a bridge between the realms of supervised and unsupervised machine learning, the algorithm uses a small portion of labeled data and many unlabeled dates from which a model must learn and make predictions on new examples. the algorithms will cluster similar data using an unsupervised learning algorithm and then use the existing labeled data to label the rest of the unlabeled data. Semi-supervised learning is particularly useful when there is a large amount of unlabeled data available, but it is too expensive or difficult to label all of it. Nowadays, semi-supervised learning has been used in LIBS quantitative and identifying technology [43, 44]. Li *et al.* [43] proposed a novel semi-supervised LIBS quantitative analysis method for high-alloy steel samples based on a co-training regression model with a selection of effective unlabeled samples. But only when effective unlabeled samples are chosen for model training, the predicting accuracy and generalization ability of the model can be effectively improved. Wang *et al.* [92] used semi-supervised learning combined with LIBS for explosive identification, compared with KNN, PCA and SIMCA, and the semi-supervised learning algorithm generate better results. Therefore, the semi-supervised algorithm has a good application prospect in the classification of samples by using the labeled data to guide the learning procedure.

3 Data preprocessing with machine learning

LIBS spectral data is highly complex and redundant, the spurious data and redundant information in LIBS spectra

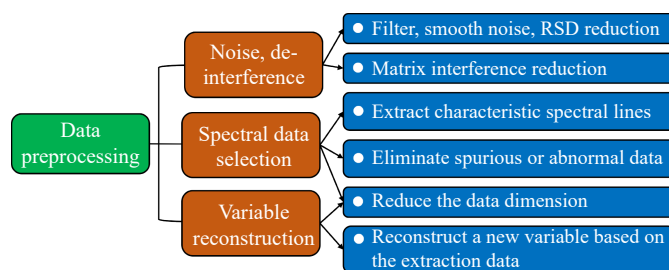


Fig. 1 The role of machine learning for data preprocessing in LIBS.

will worsen the model performance of classification or prediction. The data preprocessing can mine the useful and potential features in LIBS raw data, so as to improve LIBS analyzing performance by eliminating the redundant and irrelevant features. Figure 1 shows the role of machine learning for data preprocessing in LIBS. The application methods of data processing with machine learning in LIBS include denoising, de-interference, feature extraction, and variable reconstruction.

3.1 Denoise and de-interference

The noises and interference in LIBS have an important impact on the stability and reliability of spectral data, and the implementation of accurate quantitative LIBS analysis is always subject to noise interference. Machine learning algorithms have also been used to solve these problems, which can effectively expand the quantitative analysis performance and the application of LIBS.

Noise is an unavoidable and significant component of the LIBS signal, and different sources of noise seriously influence the prediction performance of the analysis model. To avoid using noise information as a feature, it is necessary to develop effective methods to filter noise [45–47]. Wavelet threshold denoising (WTD) is the most commonly used methods in noise filtering, which localize features in LIBS data to different scales and preserve important signal while removing noise [48, 49]. In the process of WTD, a threshold with an adjustment factor was implemented to overcome the over-smooth phenomenon and improve model performance. As shown in Fig. 2, wavelet analysis procedures can eliminate environmental denoising and background noise in LIBS spectrum [50]. Based on entropy analysis of noisy LIBS signal and noise, Zhang *et al.* [51] presented a method for selecting the optimal decomposition level, which reduced the limit of detection values by more than 50%. Before constructing the analytical model, WTD and Kalman filtering were used to preprocess coal ash spectra, and WTD showed better performance in filtering noise [52]. Yang *et al.* [53] proposed an empirical mode decomposition (EMD) based on the wavelet method to remove LIBS noise. Compared with other denoising methods, this method has a good denoising evaluation

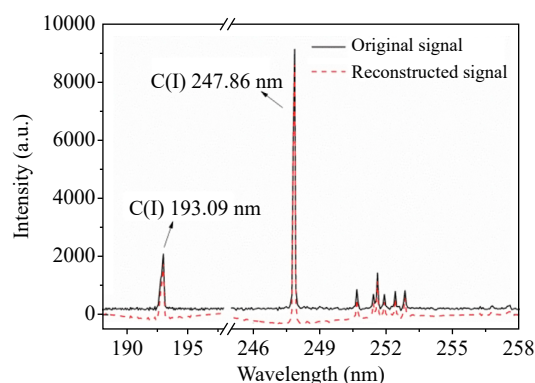


Fig. 2 Comparison of original and reconstructed spectra. Reproduced from Ref. [50].

index and stability.

According to the data characteristics, some researchers have improved the traditional noise filtering methods. Xie *et al.* [45] proposed an improved soft-hard trade-off threshold method, in which the RSD of each trace element was significantly improved after denoising. Duan *et al.* [54] proposed an improved wavelet double threshold function, by which the spectral denoising effect of Cu and Zn models were improved, and the corresponding models performance was improved for Cu and Zn. Based on the entropy analysis of noisy LIBS signal and noise, Zhang *et al.* [51] proposed a method to select the optimal decomposition level. Experimental data analysis showed that this method can reduce the fluctuation of noisy signals and improve the signal-to-noise ratio of LIBS.

Line broadening due to plasma processes or the instrument degrade spectral resolution leading to uncertainty in the elemental profile obtained by optical emission spectroscopy [55]. In the research of using machine learning algorithms to solve spectral line interferences, Liu *et al.* [56] developed an algorithm based on iterative discrete wavelet transform (IDWT) and Richardson-Lucy deconvolution (RLD) to reduce the impact of spectral interference and improve the accuracy of quantitative analysis. Tan *et al.* [57] studied an error compensation method based on the curve fitting method to realize the decomposition and correction of overlapping peaks. This improved the efficiency of quantitative analysis in LIBS by greatly reducing fitting residuals. Wang *et al.* [58] applied the Fourier self-deconvolution method to analyze overlapping peaks. This method effectively improved the detection of Pb concentration in polluted water by LIBS.

LIBS is strongly influenced by signal fluctuations, which are correlate to plasma properties, measurement conditions, and the physical properties of the samples. To reduce fluctuations of spectral intensity and increase the prediction ability of quantitative models, LIBS often requires to normalize spectra. Data standardization or

normalization is correction of the LIBS signal by dividing a factor, which including the spectral background, total area, internal standard, standard normal variate, plasma characteristic parameters, acoustic signal induced by the shock-wave, and ablated mass [59]. LIBS combined with data standardization based on machine learning methods has been reported. Wang *et al.* [60] designed a back propagation neural network (BPNN) model for standardizing the spectrum to a lower relative standard deviation (RSD) of emission line intensities, in which training spectrum, sample energy and image parameters were inputs. This data processing method not only provided a practical access to acquire stable spectrum information for both qualitative and quantitative LIBS analysis but also showed a bright future of combining LIBS data processing with machine learning methods.

3.2 Spectral data selection

LIBS is a hybrid spectrum that contains ions, atoms, and molecules of the elements in the sample and the ambient gas. Furthermore, due to plasma instability, operational errors, instrumentation abnormalities, or parameter changes, the collected LIBS spectra did not contain complete or correct plasma emissions, which were defined as spurious data or abnormal data. It is difficult to provide sufficient information for quantitative models with finite characteristic lines, and the use of full spectrum introduces a large amount of redundant data. To improve the robustness of the analytical model, machine learning algorithms are used to extract characteristic spectral lines strongly correlated with element content and to avoid spurious data or abnormal data.

Some researchers have used machine learning methods to identify and reject abnormal spectra. It is a popular method by combining prior knowledge with input selection algorithms to reduce the influence of redundant data on the analytical results and decrease the model complexity. Lu *et al.* [61] proposed a hybrid feature selection method combining with wavelet transform to analyze the heat value of coal using LIBS. The results of the study showed that feature selection method can effectively reduce the spectral dimension, remove irrelevant information, and select the relevant spectral data. As shown in Fig. 3, not only characteristic lines but also some molecular lines or background contribute to the quantitative model.

Data feature extraction is beneficial to improve the LIBS analytical performance by screening the important wavelength variables and eliminating the effect of plasma uncertainty on the spectrum. Therefore, feature extraction algorithms has been used in combination with qualitative identification and quantitative analysis models to improve the analytical performance of LIBS. Xie *et al.* [62] used wavelet packet transform to select effective feature spectral lines and combined them with

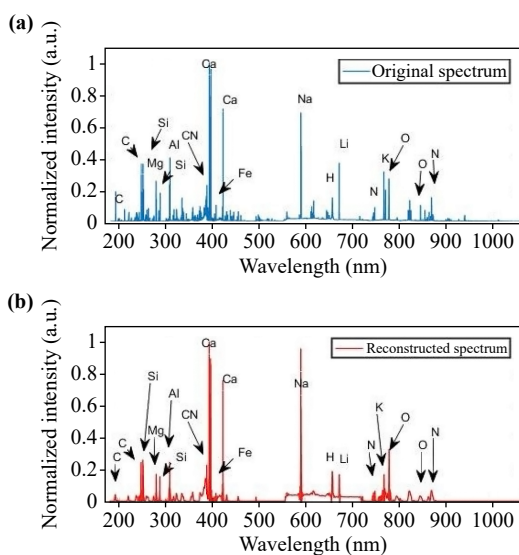


Fig. 3 The original spectrum and the reconstruction spectrum by the hybrid feature selection method. Reproduced from Ref. [61].

relevant vector machines to achieve accurate in situ component prediction. Chen *et al.* [63] proposed a weakly supervised method called spectral distance variable selection (SDVS), which utilizes prior information of samples to evaluate spectral feature weights. Compared with full-spectrum input and other feature selection methods, this method substantially improves prediction accuracy. Harefa *et al.* [64] used sequential forward selection (SFS) to eliminate the most irrelevant features, which required short computation time but improved the classification accuracy of four machine learning models (quadratic discriminant analysis, RF, Bernoulli naïve Bayes, and SVM). Chu *et al.* [65] proposed an approach using LIBS combined with the ensemble learning based on the random subspace method (RSM), which extracts important spectral lines (Na, K, Mg, Ca, H, O, N, C–N) from the LIBS spectrum of blood cancer samples, and the recognition ability of blood cancer types can be greatly improved. Therefore, a feature extraction algorithm can be used to judge the importance of each variable and maintain the most important variables [66]. Kong *et al.* [67] proposed an automatic method to select analytical and reference lines for internal standard method from the original spectra based on GA. The featured optimal analytical and reference lines can effectively improve the quantitative accuracy. Gan *et al.* [68] used the uninformed variable elimination (UVE) algorithm to remove the non-information noise variables, and then the competitive adaptive reweighted sampling (CARS) algorithm was used to screen the important wavelength variables related to Pythium, the extracted feature lines can effectively improve the accuracy of PLS model. Ma *et al.* [69] applied GA to screen 12 wavelength variables related to the characteristic spectral lines of Ca, Na, and

K elements in manure samples, and the method could significantly reduce the modeling variable information and improve the prediction accuracy of Ca content in manure using LIBS. He *et al.* [70] proposed a hybrid variable selection method mutual information-particle swarm optimization (MI-PSO) to realize precise screening of LIBS and Fourier transform infrared spectrometer (FTIR) spectral characteristic variables of coal samples. The MI was used to eliminate redundant variables in the spectral data, and the PSO was used to further filter the retained variables to find a set of variables with higher prediction accuracy. The algorithm mentioned above can more accurately predict the ash content and volatile matter of coal quality analysis. Duan *et al.* [71, 72] proposed an automatic variable selection method for quantitative analysis of soil samples using LIBS based on full spectrum correction (FSC) and modified iterative predictive weighted-partial least squares (mIPW-PLS), which automatically selects features without artificial processes. To illustrate the feasibility and effectiveness of the method, a comparison with GA and successive projections algorithm (SPA) for different elements (copper, barium and chromium) detection in soil was implemented. the method requires short computation time and improves prediction performance of quantification models. Recursive feature elimination can effectively reduce redundant variables and prevent overfitting. Lu *et al.* [73] extracted effective features from the de-noised LIBS spectrum using the recursive feature elimination with cross-validation (RFECV) method. According to the selected features, SVR model of coal was established. The performance of the models was significantly improved compared with the original model. Wang *et al.* [74] used an RFE method based on ridge for feature selection. The results showed that the root-mean-square error prediction (RMSEP) was significantly reduced compared with the PLS model with full spectrum as input. Ruan *et al.* [75] proposed to combine sequence reverse selection with RF for quantitative analysis of phosphorus and sulfur in steel, and the results showed that the RF model based on sequential backward selection (SBS) had a better prediction effect than the univariate method, the PLS model and the traditional RF model. Ruan *et al.* [76] also proposed an improved backward elimination feature selection method. Compared with the predicted results of the RF, VI-RF, and SBS-RF models, the improved SBS-RF model has higher sensitivity, specificity, and accuracy. To improve the accuracy of the model, Ding *et al.* [77] proposed a mean decrease accuracy (MDA) and mean decrease impurity (MDI) feature selection methods of RF to filter the LIBS data. Four models were constructed using convolutional neural networks (CNN) and RF: MDA-CNN, MDA-RF, MDI-CNN, and MDI-RF, and applied to predict soil sources. The experimental results indicate that this analysis method can effectively determine the soil source. You *et*

al. [78] used RF for variable selection to reduce the number of characteristic variables from 100 to 6, which significantly reduced the interference of irrelevant spectral lines. Lv *et al.* [79] proposed a feature extraction method that combines the linear regression (LR) and the sparse and under-complete autoencoder (SUAC) neural network. This method performs nonlinear feature extraction and dimension reduction on high-dimensional spectral data.

In LIBS, matrix effects are due to changes of emission line intensities of some elements when the physical properties and/or the chemical composition of the sample matrix varies. The matrix effects limit the performance of LIBS in absolute elemental analysis, since the spectral intensity of an emission line at a given concentration depends on the matrix. Many studies have focused on solving the influence of the matrix effects by data preprocessing. Wu *et al.* [80] used CARS method to select characteristic and related variables of Cr in LIBS spectra from edible vegetable oil and establish calibration model using LSSVM based on the selected variables. The number of variables was reduced from 132 to 10, and the CARS-LSSVM can reduce the influence of matrix effects on analytical element and improve prediction accuracy of LIBS analysis. Zhu *et al.* [81] proposed a multi-spectral line internal calibration method for quantitative analysis of Pb element in irregular lead-brass alloy samples. The linear fitting degree of the calibration curve reached 0.9846, indicating that this method can to some extent eliminate the influence of matrix effects and spectral interference, and significantly improve the measurement accuracy. Long *et al.* [82] proposed a data selection method based on plasma temperature matching (DSPTM) to reduce both matrix effects and signal uncertainty. By selecting spectra with smaller plasma temperature differences for all samples, the proposed method was used to build up the univariate and multiple linear regression (MLR) model to rely more on spectra with smaller matrix effects and signal uncertainty, therefore improving final quantification accuracy and precision.

3.3 Variable reconstruction

The purpose of variable reconstruction is to reconstruct a new variable based on the extraction data that are sensitive and useful to concentration of the element to be measured in LIBS spectrum. In some cases, a new set of derived features can provide better interpretability than the original LIBS data. In the method of variable reconstitution using machine learning algorithms, in addition to directly selecting the relevant spectral line information of the element to be measured from the LIBS spectrum, the spectral line information of the non-measured element is sometimes extracted, and the selected information is reconstituted into a new variable.

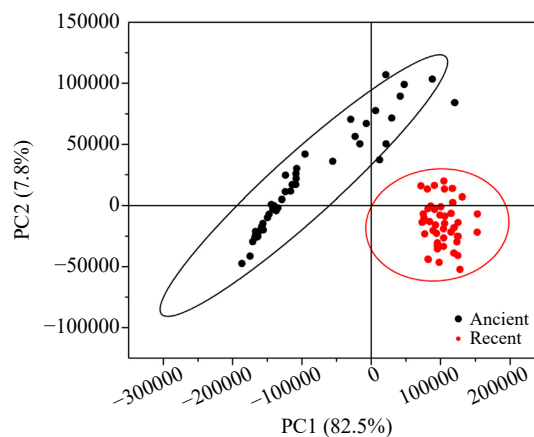


Fig. 4 PCA analysis for the LIBS spectra of ancient and recent bovine bone. Reproduced from Ref. [86].

PCA is one of the most used variable reconstitution method, which recombine many variables with certain correlations into a new set of unrelated comprehensive variables to replace the original variables. PCA can effectively reduce the data dimension and retain most of the information of the original LIBS spectrum [83]. In addition, PCA can also be used for data preprocessing, visualization, dimensionality reduction, model building, classification, quantification and non-conventional multi-variate mapping [84]. Sirven *et al.* [85] introduced PCA into the processing of LIBS data for outliers filtering, they used basic visual thresholding and omitted up to 30% of spectra prior to a rock classification in a preflight ChemCam testing. Abdel-Salam *et al.* [86] utilized PCA to extract features from the LIBS spectrum of recent and ancient bovine bone samples, and the first two principal components constituted 90.3% of the total variance were used for establishing identification model as shown in Fig. 4. Farhadian *et al.* [87] used the first three components which cover 96% information of LIBS data as the input variables of ANN model, and the accuracy reached 100% for the identification of energetic materials in the Ar atmosphere. Yuan *et al.* [88] extracted 13 principal components as the input variables of SVM model, and the classification accuracy reached 100% for the rapid classification of steel materials.

In addition, more and more new variable reconstruction algorithms have been proposed and used to improve the performance of LIBS quantitative analysis. Li *et al.* [89] used GA to select the intensity ratios of the spectral lines belonging to the target and domain matrix elements, and then these selected line-intensity ratios were taken as inputs to construct an analysis model based on an ANN to analyze the elements copper (Cu) and vanadium (V) in steel samples. The results showed that the GA combining ANN can excellently improve the prediction accuracy of Cu and V elements in steel samples compared with traditional internal calibration

methods. Zhong *et al.* [90] introduced the concept of standardized root mean square error of cross-validation (SRMSECV) to select the median area of all spectra of the same sample as the center and discard the spectra outside the spectral area interval. Under the optimized areal screening span, the average of determination coefficients (R^2) and the accuracy of multi-element analysis were improved to some extent. Neural networks are a subset of machine learning, and one of the most impressive forms of ANN architecture is that of the Convolutional Neural Network (CNN). Dong *et al.* [91] proposed a lightweight CNN model, which extract spectral low-level features through the first three convolutional layers, this model can improve the accuracy of quantitative analysis of flow slurry by solving problems such as matrix effect, self-absorption effect, and limited sample size with more dimensions. In addition, machine learning algorithms have also been used for parameter optimization in LIBS experiments. Prochazka *et al.* [92] developed an ANN algorithm to predict the signal-to-noise ratio (SNR) of selected spectral lines based on specific experimental parameters (laser pulse energy and gate delay) and on the sample's physical and mechanical properties. It has been concluded that the optimization process can be substituted or significantly shortened by means of the ANN.

In summary, to further improve the accuracy and precision of the model based on machine learning algorithms, various data preprocessing is utilized before the analysis model is established, including feature extraction, variable reconstruction, noise filtering, interference processing, matrix effects, and self-absorption correction. Among them, feature extraction from LIBS spectra and plasma images are the most essential data preprocessing methods, which are to maximize the effective information and avoid interference of redundant or invalid information (such as noise and baseline in LIBS spectra) in the LIBS spectrum to the accuracy/precision and efficiency of analytical models.

4 Qualitative and quantitative modeling with machine learning

Based on the data preprocessing method mentioned above, it is ready to be used as an input variable for the analysis model. LIBS-based analysis models can be divided into qualitative and quantitative categories, qualitative analysis is the identification, clustering, or classification of samples, while quantitative analysis is the analysis of the elemental composition of samples. In this section, the research progress of machine learning algorithms for improvements in the qualitative and quantitative performance of LIBS are reviewed in detail. Figure 5 shows the role of machine learning for qualitative and quantitative analysis in LIBS.

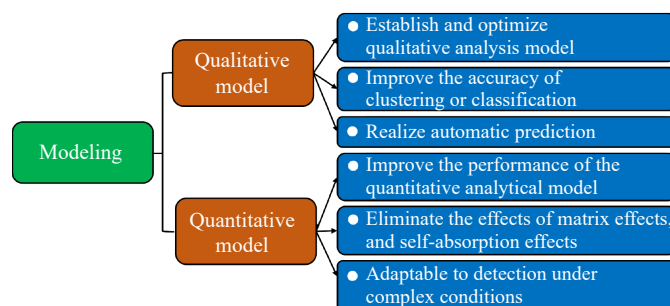


Fig. 5 The role of machine learning for qualitative and quantitative analysis in LIBS.

4.1 Qualitative model

Affected by matrix effects, the physical properties and composition of the sample will affect the element signal. LIBS analysis technology displays significant matrix effects which greatly hinder the application of this technology. However, the matrix effects are not always unhelpful for analytical results, which are beneficial for sample classification. The most basic kind of qualitative analysis is the clustering or identification of data points according to their mutual similarities. Qualitative models are often organized in a graph (dendrogram or scatter plot) and utilized for the discrimination of objects based on their characteristic spectra. The application of machine learning algorithms in the qualitative analysis of LIBS is manifested in three aspects: (i) establish and optimize qualitative analysis model; (ii) improve the accuracy of clustering or classification; (iii) realize automatic prediction.

Establish and optimize qualitative analysis model. To find suitable machine learning algorithms for clustering or classification various types of samples, researchers have conducted extensive algorithm comparison studies. Vítková *et al.* [40] used subset of PCA scores in conjunction with both ANN and LDA as variables for classification of 18 different biominerals, and the ANN model showed better performance than the LDA model. The method can create a database for simple and fast identification of archeological or paleontological materials in situ. Tang *et al.* [93] compared the predictive performance of four different machine learning methods (PLS-DA, SVM, RF and RF based on variable importance (VIRF)) for slag samples classification, and VIRF showed the highest classification accuracy. Yang *et al.* [94] studied the classification performances of six machine learning algorithms (PCA, DT, RF, PLS-DA, LDA, and SVM) for the geographic origins of 20 kinds of rice samples, and LDA was demonstrated to be the most efficient tool for rice geographic origin classification assisted by LIBS with high accuracy and analytical speed. LDA project the high-dimensional sample data into the optimal discriminant vector subspace (low dimension) in order to compress the dimension of the



feature space and extract classification information. Alarsan *et al.* [95] identified heart diseases using three algorithms (DT, RF, and Gradient-Boosted Trees), and RF shows the highest accuracy of 98.03%. Because RF can balance the error data and evaluate the importance of each variable, especially avoiding over-fitting problems. Yu *et al.* [96] used several methods (PCA, PLS-DA, LDA, SVM) to classify nephrite samples from five different locations, and SVM showed the highest accuracy of 100% for predicting training data and 99.3% for predicting testing data. Gyftokostas *et al.* [97] demonstrated that LIBS coupled with machine learning (LDA, ERTC, RFC, and XGBoost) is a powerful tool for live oil authenticity and geographic discrimination. Zhao *et al.* [98] traced the geographical origins of acacia honey and multi-floral honey using SVM and LDA, the accuracy of the SVM model was 99.7% which was superior to the LDA model. Luo *et al.* [99] performed three pattern recognition methods (discriminant analysis, RBF-ANN, and MLP) to identify rice species, and the MLR model showed higher accuracy of 100% and 97.9% in the training and test sets, respectively. Huang *et al.* [100] adopted traditional machine learning methods (CNN, LDA, KNN, RF, and SVM) to identify 25 adulterated milk powders mixed with four different types of exogenous proteins, and SVM model obtained the highest accuracy of 93.9%. Kiss *et al.* [101] used with a K-means algorithm to cluster various matrices within the tumorous tissue, Typical skin tumors were selected for LIBS analysis, the imaging of biotic elements (Mg, Ca, Na, and K) provides the elemental distribution within the tissue. The elemental images were correlated with the tumor progression and its margins, as well as with the difference between healthy and tumorous tissues. Ding *et al.* [102] applied LDA and SVM to identify three kinds of plant leaves (*Ligustrum lucidum*, *Viburnum odoratissimum*, and bamboo), the average classification accuracy rate of SVM for the test set was up to 98.89%, which is better than LDA. Li *et al.* [103] applied PCA, KNN, and SVM to classify fat, skin and muscle tissues with an accuracy of over 99.83%, a sensitivity of over 0.995 and a specificity of over 0.998. Babu *et al.* [104] applied PCA and ANN to classify the unaged, gamma-irradiated, and water-aged specimens, the ANN-adopted LIBS analysis was successful with good classification accuracy compared to PCA. From the above study, it can be seen that supervised algorithms, such as SVM, LDA, ANN, and RF, have higher classification accuracy than unsupervised algorithms. Singh *et al.* [105] compared the analytical merits (accuracy and precision) of applying the PCR and PLSR algorithms to identical LIBS data from a set of stainless steel samples. A few guidelines were proposed in this work for selecting PCR or PLSR depending on the analytical situation. However, it is difficult to summarize the fixed and feasible algorithm selection law here, because the characteristics of the spectra excited by

various substances are very different, and their advantages and disadvantages can only be compared under the same experimental equipment and the same experimental conditions.

Improve the accuracy of clustering or classification. The main research purpose of machine learning algorithms for LIBS qualitative analysis is to improve the accuracy of clustering or classification. It is necessary to choose the best one from different machine learning methods, or improve existing machine learning algorithms to establish an optimized qualitative analysis model. No data library is currently available commercially, and even if it was, the transfer between different LIBS systems is not possible yet. Therefore, each research group is building its database according to sample type, application requirements, experimental conditions, etc. [84]. For example, PCA is very suitable for data visualization, and SIMCA and PLS-DA can achieve automatic prediction. SIMCA has lower sensitivity than PLS-DA, but its model is more robust to unknown samples [84]. Different machine learning algorithms have their advantages and disadvantages, and in some cases, combining different methods can overcome the defects of a single algorithm and extract potential features from LIBS data.

In addition, the introducing of different preprocessing methods could also improve the classification accuracy and efficiency. Liu *et al.* [106] investigated the classification and identification of rice geographical origin using combined LIBS and hyperspectral imaging with machine learning method. PCA was utilized to realize data dimensionality and extract the data feat of LIBS, HSI and fusion data, and PLS-DA, SVM and ELM were used to achieve rapid and accurate rice quality and identity detection.

Realize automatic prediction. The above researches also show that the combination of LIBS and machine learning algorithms has been successfully used for automatic sorting of unknown samples. Several researchers have reported the recycling of metals use of LIBS and machine learning algorithms. Campanella *et al.* [107] developed a strategy based on LIBS and ANN for the sorting of aluminum scrap samples. The neural networks approach enables more reproducible results, which can accommodate the unavoidable signal variations due to the low intrinsic reproducibility of the LIBS systems. The results demonstrate the possibility of an efficient (> 75%) classification of non-ferrous metallic automotive scraps using LIBS and ANN method working in conditions simulating the industrial environment. Park *et al.* [108] studied a 3D sensing system for LIBS-based metal scrap identification, and PCA algorithms was used to reduce the dimensions of the wavelength data into principal components. The maximum classification accuracy of 95% was achieved when LIBS spectra were acquired from optimized rather than non-optimized

sample surfaces. Demir *et al.* [109] used LIBS and PCA for the classification of 700 °C molten aluminum alloys without sample preparation.

To show the performance of different machine learning algorithms in LIBS field clearly, Table 2 presents a summary of sample classification using LIBS combined with machine learning algorithms, including the types of machine learning methods, improvement ways, comparison methods, material types, optimal classification results, and references. As can be seen from Table 2, PCA is widely used in combination with other machine learning algorithms for qualitative analysis of LIBS, because PCA has excellent correlation data extraction and dimensionality reduction capabilities, which can effectively eliminate the influence of noisy data and redundant data on classification results and improve computing efficiency.

4.2 Quantitative model

Quantitative analysis is the most important goal of any analysis technique. However, due to the complexity of the interaction process between laser and material, the limitations of experimental instruments, matrix effects, self-absorption effects, and environmental conditions, the quantitative analysis ability has always been a bottleneck in the development of LIBS technology [133]. To capture the complex relationship between spectra and analyte information, machine learning methods often have a higher model complexity than traditional univariate calibration method based on physical principles. On the one hand, if there are too few samples used for training, the quantitative analysis model is prone to over-fitting. In practice, a data-driven model may require enough calibration samples to ensure that its applicability domain covers the validation sample. Of course, the large use of standard samples increases the computational time and the cost of LIBS quantitative analysis. Furthermore, most machine learning models are not designed with model interpretability. It is difficult to judge whether the decision-making process conforms to the physical principles behind LIBS. Such an issue may reduce the robustness of LIBS qualification and quantification based on machine learning methods [251]. The research of machine learning algorithms in the quantitative analysis of LIBS focuses on the following four aspects: (i) improve the performance of the quantitative analytical model; (ii) eliminate the effects of matrix effects and self-absorption effects; (iii) adaptable to detection under complex conditions.

Improve the performance of the quantitative analytical model. The machine learning algorithms used to construct the quantitative analysis model of LIBS include a single-algorithm model and a multi-algorithm model. The single machine learning model is a prediction model based on only one machine learning

algorithm, which is simple to implement and efficient. In most cases, the combination of preprocessing methods and a single machine learning algorithm can accurately determine the element content in the target sample, and the machine learning algorithm and corresponding parameters should be optimized according to the specific applications [134].

PLSR is one of the most widely used quantitative analysis method due to its excellent performance and simple calculation. Akhmetzhanov *et al.* [135] achieved quantitative detect rare-earth elements (REE) in ores using PLSR, which solved the problems of significant overlapping of REE lines in LIBS emission spectra and high pairwise correlation between REE contents in certified reference materials (CRMs). They confirmed that using PLSR can compensate for the low resolution of handheld LIBS instruments and achieve quantitative analysis of Ce and La in REE-rich ores [136]. Rao *et al.* [137] used PLSR, PCR, and ANN to detect trace elements in plutonium, PLSR is superior in determining iron and nickel contents in plutonium metal, with a limit of detections (LoD) of 15 and 20 ppm, respectively. In Rao's another work, they created a boosted regression ensemble model (boosted regression tree, BRT) to predict silicon content in ceria pellets doped with silicon [138]. Its predictive accuracy was higher than that of traditional PCA, PLS, and ANN regression models. Gu *et al.* [139] applied conditional univariate quantitative analysis, MLR and PLSR to the quantitative analysis of steel alloys. PLSR showed low relative errors for two unknown steel alloy samples with values below 6.62% and 1.49%, respectively. Yaroshchuk *et al.* [140] applied PCR, PLSR, multi-block PLS, and serial PLS to the quantitative analysis of Fe content in iron ore. In comparison with PCR and PLS, the performance of the multi-block PLS algorithm is poor. Erler *et al.* [141] evaluated multiple regression methods including PLSR, least absolute shrinkage and selection operator regression (LASSO), and Gaussian process regression (GPR), for predicting Ca, K, Mg and Fe in soil. LASSO and GPR yielded slightly better results than PLSR. The advantages of GPR are mainly reflected in dealing with nonlinear and small data problems, and the model may fail when encountering high-dimensional spaces. Rao *et al.* [142] quantified gallium in cerium matrices via ensemble regressions, SVR, Gaussian kernel regressions, and ANN. Gaussian kernel regression is the best prediction model with RMSEP of 0.33% and an LoD of 0.015%. Yuan *et al.* [143] applied BP and MLR to study the content of forsterite and fayalite in olivine, and the root-mean-square error (RMSE) value of the BP model was the lowest (28.64). Shi *et al.* [144] applied SVR and PLSR to determine the concentrations of five main elements (Si, Ca, Mg, Fe and Al) in sedimentary rock samples, and found that the SVR model performed better with satisfactory accuracy. Ding *et al.* [145] applied KELM and

Table 2 A summary of sample classification using LIBS combined with machine learning algorithms.

Methods	Improvement	Comparison	Materials	Best results	Ref.
LDA	IFALS	MLS-LDA, ASPI-LDA	Rocks	Accuracy: 98.54%	[110]
	–	PCA, DT, RF, PLS-DA, SVM	Rice	Accuracy: 99.20%	[94]
	RSM	KNN, LDA	Blood	Accuracy: 98.34%	[65]
	PCA	PCA-SVM	Olive oil	Accuracy: 100%	[111]
RF	–	–	Aluminium alloy	Accuracy: 98.45%	[112]
	SBS	RF, VIRF	Ceramics	Sensitivity: 0.9526 Specificity: 0.9910 Accuracy: 97.82%	[76]
	CWT	PCA-RF	Huanglongbing-infected navel oranges	Accuracy: 97.89%	[113]
	VI	PLS-DA, SVM, RF	Slag	Sensitivity: 0.9889 Specificity: 0.9944 Accuracy: 0.9926%	[93]
SVM	–	–	Steel	Accuracy: 95.3%	[114]
	LDA	–	Mineral	Recall: 0.53 Precision: 0.73	[115]
	–	KNN	Fat, skin and muscle tissues	Accuracy: 99.83% Sensitivity: 0.995 Specificity: 0.998	[103]
	–	PCA, PLS-DA, LDA	Nephrite	Accuracy: 99.3%	[96]
	PCA	–	Metal	Accuracy: 100%	[116]
	PCA	LDA	Plant leaves	Accuracy: 98.89%	[102]
	–	LDA	Honey	Accuracy: 99.7%	[98]
	SFS	SFS-RF, SVM	Soil	Accuracy: 97.88%	[64]
	PCA	–	Steel	Accuracy: 100%	[88]
	PCA	PCA-KNN, CA-ANN	Rocks	Accuracy: 98%	[117]
CNN	Restricted Boltzmann machines	PCA-SVM	Steel	Accuracy: 100%	[118]
	–	SVM	Rocks	Accuracy: 100%	[119]
	PCA	BP-ANN, SVM, KNN, RF	Soil	Accuracy: 99.60%	[120]
	–	SVM	Fish	Accuracy: 98.2%	[121]
KNN	–	–	Paint	Accuracy: 100%	[20]
	Multiplicative scatter correction (MSC)	–	Meat	Accuracy: 100%	[122]
	–	–	Explosive	Accuracy: 99.58%	[44]
	–	PLSR	Biochar	ARSDP: 8.13%	[123]
PLS-DA	–	SIMCA	E-waste	Accuracy: 98%	[124]
	PCA	BP-ANN	Pencil writing marks	Accuracy: 98.33%	[125]
	VI	PLS-DA, RF, VI-RF	Plastic	Accuracy: 99.55%	[83]
	SW	PLS-DA, SVM, RF	Plastic bottles	Accuracy: 93.93%	[126]
ANN	–	–	Organic materials	Accuracy: 95%	[17]
	WT	–	Copper and steel	Accuracy: 100%	[22]
	–	PCA	Aged epoxy micro-nanocomposites	Accuracy: 100%	[104]
	PCA	–	Nergetic materials	Accuracy: 100%	[87]
PCA	PCA	PCA-LDA	Mineralized tissues and bio-mineral	Accuracy: 87.5%	[40]
	–	–	Milk	Accuracy: 95.8%	[127]
	–	–	Rosewood	Accuracy: 100%	[128]
BP	–	–	Tooth	Accuracy: 100%	[129]
	PCA-GA	–	Plastic	Accuracy: 99.72%	[130]
KELM	PCA	PCA-SVM	Ginseng	Accuracy: 99.5%	[131]
	PSO	PSO-LSSVM, PSO-RF	Salvia miltiorrhiza	Accuracy: 94.87%	[37]
LSSVM	PSO	PSO-SVM, SVM, LSSVM	Aviation alloy	Accuracy: 99.56%	[49]
	–	PCA-LS-SVM	Mentha haplocalyx	Accuracy: 99.2%	[132]

PLSR for the quantitative analysis of total iron content and alkalinity of sinter, and the correlation KELM model shows better predictions for total iron and alkalinity with a correlation coefficient are above 0.9, which are all higher than that of PLSR. Wu *et al.* [146] applied RFR and PLSR to quantitative analysis of S and P elements in steel samples, RF calibration model made good predictions of S ($R^2=0.9974$) and P ($R^2=0.9981$). Xiang *et al.* [147] employed MLR, PLSR, LS-SVM and BP-ANN to quantitatively analyze heavy metals Pb and Cd elements in soil, and LS-SVM and BP-ANN offered promising results. Ye *et al.* [148] used LIBS combined with PLSR and RFR algorithms to measure chemical oxygen demand (COD) in river water samples, the results showed that RFR had a high R^2 value (0.9248) and low RMSE value (25.1215 mg/L). Labutin *et al.* [149] implemented a PCR method for the construction of a PCR model under spectral interferences using C I 833.51 nm line for carbon determination in low-alloy steels in air. The predicted carbon content in the rail templet was in an agreement with the reference value obtained by a combustion analyzer within the relative error of 6%.

Intelligent optimization algorithms have also been introduced into LIBS quantitative analysis. Sun *et al.* [150] implemented particle swarm optimization (PSO), GA and ant colony optimization algorithms to quantitatively analyze the Pb concentration in water, the values of mean relative error percentage and RSD of the test results obtained by PSO algorithm were the best among these algorithms. Intelligent optimization algorithms also have some shortcomings. The parameters selection seriously affects the quantitative analysis results of GA, and currently, the selection of these parameters mostly relies on experience. The PSO algorithm is simple and fast in solving problems, but it is prone to getting stuck in local optima. Ant colony algorithm has a high computational cost.

Machine learning algorithms can not only be used for quantitative analysis of elements in samples, but also quantitative analysis of other chemical properties of samples. Hao *et al.* [151] applied LIBS with PLSR to measuring the acidity of iron ore, the average relative error (ARE) and RMSE of the acidity achieved 3.65% and 0.0048, respectively. With the conventional internal standard calibration, it is difficult to establish the calibration models of the acidity for iron ore due to the serious matrix effects. PLSR is effective to address this problem due to its performance in compensating the matrix effects. This is because the interferential and nonlinear signals were eliminated by choosing the number of PCs during the model establishment. Wang *et al.* [152] applied VIRF, PLSR, and LS-SVM for the quantitative analysis of iron ore acidity, and the VIRF model showed excellent predictive performance. Similarly, Yang *et al.* [153] applied PLSR and RFR for measuring the basicity

of sintered ore, which can be defined by the concentration of oxides: CaO, SiO₂, Al₂O₃ and MgO. The RFR model showed better predictive capabilities with an RSD of 0.27% to 0.59%. Lu *et al.* [154] applied PLSR and LS-SVR to quantitative analysis of pH value in soil, and LS-SVR effectively improved the analysis accuracy with the values of R^2 , MAE, and RMSE of 0.987, 0.1 units (pH), and 0.079, respectively. Zhang *et al.* [155] applied PLSR, SVR, ANN, and PCR for the quantitative analysis of coal quality, and ANN model has the lowest average relative error (ARE) value, which was 0.69% (ash content), 0.87% (volatile matter content), and 0.56 MJ·kg⁻¹ (calorific), respectively. Képeš *et al.* [156] explored the application of ANN for predicting plasma temperatures, and they leveraged synthetic data to isolate temperature effects from other factors and studied the relationship between the LIBS spectra and temperature learnt by the ANN. Saeidfirozeh *et al.* [157] also developed an ANN method for characterising crucial physical plasma parameters (i.e., temperature, electron density, and abundance ratios of ionisation states) in a fast and precise manner that mitigates common issues arising in evaluation of laser-induced breakdown spectra. Thus, the introduction of machine learning algorithms has provided great convenience for the expansion of LIBS application fields, because machine learning algorithms can automatically obtain relevant information in spectra and establish robust quantization models.

Various machine learning algorithms have their unique strengths, some researchers have proposed some multi-algorithm models to improve the performance of LIBS quantitative analysis. The method combining two or more machine learning approaches to utilize their respective strengths to obtain the best analytical results. Meanwhile, the combined machine learning model presents the complex process of selecting suitable preprocessing methods and algorithms. However, the combined algorithms are computationally complex and time-consuming. For example, Ahmed *et al.* [158] adopted an ANN based on multi-line calibration (MLC-ANN) to improve the accuracy of LIBS quantitative analysis of aluminum alloys. Li *et al.* [159] proposed a multi-spectral line correction method based on ANN, which improves the accuracy and precision of LIBS analysis of steel compared to the traditional internal calibration method. Yang *et al.* [160] compared the prediction ability of the PLS, ANN, and PLS-ANN models in detecting nine essential element components in plant materials. The results show that the PLS-ANN model has the highest accuracy, the ANN model is the second, and the PLS model is the lowest. Li *et al.* [139] combined the standardization method and dominant factor based PLSR to improve the measurement accuracy of carbon content in coal with R^2 , RMSEP, and ARE were 0.99, 1.63 wt.%, and 1.82%, respectively. Shabbir *et al.* [161] combined feature selection with BPNN for the analysis

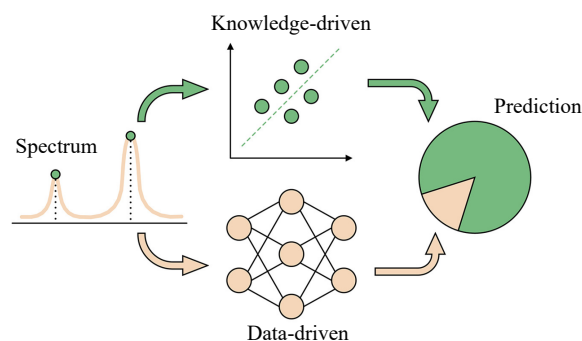


Fig. 6 Schematic description of combining spectral knowledge and machine learning for LIBS data analysis. Reproduced from Ref. [165].

of raw rocks with RMSEPs of 1.6, 18, 101, and 162 ppm for Li, Rb, Sr, and Ba elements, respectively. Zhang *et al.* [162] established PLSR model, SVR model and PLS-SVR model separately for the prediction of gelatin adulteration ratios, the result reveals that the PLS-SVM model can be employed as a preferred method for the accurate estimation of edible gelatin adulteration. Huang *et al.* [163] introduced PCA and canonical correlation analysis (CCA) into SVR for the analysis of T91 steel specimens with different degrees of microstructure aging, and the maximum values of mean relative error (MRE), RSDs and RMSEPs were 2.47%, 2.94% and 6.14, respectively. Yu *et al.* [164] adopted a multivariate multispectral correction method, combining DP-LIB with BPNN, to establish a multivariate GA-BP-ANN correction method, which effectively reduced the ARE of predicted samples and further improved the accuracy of LIBS quantitative analysis. The above research confirms that combined models have advantages in improving the quantitative analysis ability of LIBS, but the complexity of algorithms and data processing time also increase accordingly. But what is worrying is that some uninitiated researchers lack the fundamental knowledge of the capabilities and limitations of these complex algorithms when using them, to the extent that they overlook their physical principles. If physical principles are disregarded in the process of using machine learning algorithms, it will result in key variables being independent of the properties of the element, but rather related to pollutants and/or background features. To improve the accuracy and robustness of LIBS quantitative model without departing from physical mechanisms, a state-of-the-art strategy is to incorporate physical principles into machine learning algorithms, that is, a hybrid algorithm of machine learning and physical principles. Ideally, the intensity of the characteristic line of the element to be measured is linearly correlated with its concentration in the sample, these characteristic lines play a dominant role in the model decisions compared to the lines of the matrix elements. Song *et al.* [165] proposed a schematic description

of incorporating LIBS physical principles into machine learning as shown in Fig. 6. The method used knowledge-based lines, related to analyte compositions, to build a linear physical principle-based model and adopts KELM to account for the residuals of the linear model. The residual error of the linear model is corrected by machine learning and chemometrics models. As knowledge-driven and data-driven models are combined for the final prediction, how important spectral lines influence the result can be intuitively explained. The hybrid model inherits the advantage of physical principle-based methods, which is robust over a wider range of sample matrices. Furthermore, its good model complexity ensures that the complexity and nonlinearity of data can be handled efficiently.

A typical example of incorporating physical principles into machine learning is PLS model based on dominant factor (DF-PLS) proposed by Wang *et al.* [166–168]. The dominant factor is the major part of the concentration extracted from characteristic line intensity of the specific element based on physical principle: the linear relation between the line intensities and the elemental concentrations, non-linear self-absorption effects, and inter-element interference. The physical based dominant factor model increased the robustness and sample adaptivity of the final multivariate model. Combined with dominant factor, the PLS approach is further applied to minimize the residual errors by utilizing more spectral information to compensate for the fluctuations of plasma. It uses MLR to model the relationship between key emission lines and analyte concentration. The residual error of MLR is corrected by performing PLS on the full spectrum. The model combines advantages of both the univariate and PLS models. Li *et al.* [169] combined atomic and molecular emission spectra in the dominant factor to improve the quantification of coal. The method shows better performance than PLS in LIBS quantification tasks such as coal property analysis [167, 170] and content determination of brass alloy [171], and it has recently been combined with plasma images to correct for self-absorption [172]. Based on the DF-PLS, a hybrid model was developed to identify known calibration samples from a self-adaptive spectral database [173]. Firstly, the new spectra were standardized to reduce signal uncertainty, and the similarity between new and stored spectra was compared. Then, quantitative information of samples within and outside the database can be determined from the database directly and based on DF-PLS analysis, respectively. As the database was updated, the hybrid model can improve measurement reproducibility and reduce the measurement-to-measurement RSD. Further modifications to DF-PLS mainly include nonlinear extraction of the dominant factor and residual correction using non-linear models. The former extracts the dominant factor through a nonlinear transformation of line intensities [174], while the latter uses

machine learning methods such as SVR and KELM to increase the accuracy of residual correction [175, 176]. The accuracy of dominant factor-based methods exceeds that of their non-dominant factor-based counterparts in coal property analysis tasks in most cases [176]. In addition, the linear (physical principle-based) and nonlinear (data-driven) models in the hybrid model can be jointly optimized to improve the performance of LIBS quantitative analysis. Although machine learning is usually used as a black box, some study inspires that it essentially follows certain physical mechanisms, which need to be revealed and understood. Képeš *et al.* [177] applied various post-hoc interpretation techniques with the aim of interpreting the decision-making of CNN. They found synthetic spectra that yield perfect expected classification predictions and denoted the CNN can only learn meaningful spectroscopic features.

Eliminate the effects of matrix effects and self-absorption effects. If the test sample encounters severe matrix differences or there is self-absorption effect in the spectral lines of the analytical elements [178], it is difficult to obtain effective quantitative analysis from conventional univariate calibration models based on physical mechanisms. Some machine learning algorithms can automatically extract relevant information from LIBS spectra to construct multivariate quantitative models, so it can be used to solve the matrix effects and self-absorption effects encountered in the quantitative analysis of LIBS. PLSR and PCR are essentially multiple linear regression methods, which can take into account chemical and physical matrix effects by including peak information of matrix elements in the model while eliminating redundancy and non-linear response to the analyte concentrations. For example, PLSR method has been used to quantitative analysis the of concentrations of CaO, MgO, Al₂O₃, and SiO₂ in hematite and limonite ore samples [151]. The results showed that the PLSR models can compensate the matrix effects and obtained accurate quantitative analysis results. Amador-Hernandez *et al.* [179] applied PLSR and LIBS to quantify Au and Ag in precious metal, in their PLSR models, the spectral range where less strong resonance lines are observed is preferred since less self-absorption occurs. Death *et al.* [180] investigated the quantitative analysis of iron ore samples using PCR, the results confirmed that PCR can effectively reduce the effect of self-absorption on quantitative analysis. Zaytsev *et al.* [181] investigated the effectiveness of PCR in addressing matrix effects and spectral interference in quantitative analysis of LIBS. PCR provided good predictive capability in the spectral ranges where numerous matrix lines strongly interfered with analytical lines [182]. PCR is a linear regression model, so the non-linear response of some portions of the LIBS spectra due to self-absorption may be partitioned within principal components which attract lower regression scores, and thus make less of a contribution to the cali-

bration outcome than that of PCs containing non-self-absorbed spectral data. Huang *et al.* [183] reviewed the progress of LIBS application combined with machine learning methods for reducing matrix effect and self-absorption in soil analysis. Rethfeldt *et al.* [184] used univariate and multivariate regression method (interval PLS) to detect REE in minerals and soils by LIBS, iPLS method is better suited for the determination of REE contents in heterogeneous field samples. In the iPLS regression, only parts of the relevant element lines are included in the regression. Self-absorption and partial contamination of the flank of lines are excluded, resulting in improved regressions with higher coefficients of determination. Bhatt *et al.* [185] reviewed the performance of univariate and multivariate analyses methods in quantitative analysis of REE by LIBS. The review indicates that PLSR is one of the crucial multivariate analytical techniques for reducing the matrix effect. Kwapis *et al.* [186] reviewed the development of machine learning and LIBS measurements for nuclear applications. Multivariate techniques (PLSR and PCR) are used to address and mitigate the detrimental influence of matrix effects on predictions by including information from multiple emission lines up to the entire visible spectrum. PLS is very closely related to PCR, which is used to eliminate collinearity from LIBS spectra while simultaneously addressing overfitting through the reduction in dimensionality of the data set. Multiple separate PLS models have been developed to perform in situ online monitoring of elemental concentrations in molten salts. Above machine learning algorithms have been proven to be an effective correction of the chemical matrix effect, but physical matrix effect (surface roughness, hardness and heterogeneity) pose greater challenges to the model. Sun *et al.* [187] developed a transfer learning model training algorithm and demonstrate its effectiveness to overcome the physical matrix effect due to the change of sample physical state in LIBS analyses. This method will apply to analysis of rocks with LIBS in Mars explorations. They also found that samples with a same chemical composition but different physical forms are more appreciated for an efficient training of a transfer learning model.

It should be noted that PLSR or PCR method is difficult to establish a robust quantitative model when it cannot obtain sufficient linear correlation information from the spectrum with extremely strong variations in the degree of matrix or self-absorption effects. Therefore, some nonlinear machine learning techniques (ANN, CNN, BPNN, SVR, etc.) are now widely used to effectively solve these problems. These models do not account for causes and effects, and they automatically explain the correlation between the spectral intensity inputs and the elemental concentrations in the samples. An ANN is a non-linear machine learning technique, it can pose an advantage for modelling complex matrix effects and self-



absorption by including non-linearity with a high-degree of flexibility. It was reported that ANNs showed the potential to account for effects of chemical and physical matrices and overlapped lines when major elemental compositions of rock samples were measured [188]. As for self-absorption, ANNs can in principle account for these effects by modelling the non-linear relationship using a flexible statistical model. Sirven *et al.* [189] have confirmed that ANNs have advantages over conventional calibration curves and PLS especially in taking into account non-linearity between spectral intensities and the concentrations due to self-absorption in the plasma. The high variation of raw LIBS signal seriously reduces the accuracy and stability of the spectral analysis. To solve this problem, Xu *et al.* [190] applied CNN to predict soil type and soil properties based on the non-preprocessed LIBS spectra. The results confirmed that CNN models performed better in preventing overfitting than the conventional PLS combined with various spectral preprocessing approaches. Yang *et al.* [191] also proposed a robust least squares support vector machine (RLS-SVM) regression model to solve the data fluctuation in multiple measurements of LIBS. Through the improved segmented weighting function, the information on the spectral data in the normal distribution will be retained in the regression model while the information on the outliers will be restrained or removed.

Adaptable to detection under complex conditions. LIBS has the advantage of in-situ, fast, and remote monitoring, which allows it to adapt to complex environments such as high temperatures, space, deep sea, radiation, toxic and explosive. However, it is difficult for LIBS to obtain stable spectral signals in these complex environments, which makes the conventional calibration model useless, and the machine learning algorithm has good adaptability to extract relevant information from the complex spectrum to establish a robust calibration model. Yang *et al.* [192] proposed a method based on LIBS for measure transient surface temperatures, which holds great significance for fast sliding friction processes in linear electromagnetic propulsion, gun barrels, and high-speed trains. Three algorithms including (single-peak fitting) SPF, PLS, and BP-ANN were used to predict the surface temperature, and the results showed that BP-ANN performed best in the 1, 2, and 3 μ s exposure time. Since LIBS has a remote measurement function, the development of LIBS sensors/systems for determining the elemental composition of molten phases has been a hot topic of research. Sun *et al.* [193] developed a system that comprises a Cassegrain telescope in addition to a double-pulse LIBS to study the quantitative analysis of Si, Mn, Cr, Ni, and V in molten steel samples, and PLSR calibration method offered better repeatability and accuracy than that of univariate calibration. In addition, the system was used for estimation of C, Si, and Mn in molten steel samples in an industrial oven

[194]. Lee *et al.* [195] investigated LIBS as a possible option for remote online monitoring difficult-to-access nuclear reactors based on molten salts. The height of molten salts is easily fluctuated by vibration, in this study, the machine learning (PLS and ANN) models trained with both focusing and defocusing data were constructed, and the best RMSEP of 0.0210–0.0316 wt% were obtained for Sr and Mo elements using ANN models. This is because the training and test data sets considered the defocusing, which significantly affected the nonlinear pattern. In addition, the defocusing measurement results in self-absorption, which can show the saturated calibration curve or reversible tendency due to the thick plasma formation. The results suggested that the nonlinear model is more suitable to predict the compositions of the molten salt fluctuated by vibration. ChemCam is one of the sensor systems included on the Mars Science Laboratory rover Curiosity that landed on Mars in August 2012. Gasda *et al.* [196] reported a calibration model for manganese using the LIBS instrument that is part of the ChemCam instrument suite onboard the NASA Curiosity Rover. The optimal calibration model uses the PLS and Least Absolute Shrinkage and Selection Operator (LASSO) multivariate techniques. The double blended multivariate model shows a RMSEP accuracy of 1.39 wt% MnO. China's first Mars exploration mission, named Tianwen 1, landed on Mars on 15 May 2021. Yang *et al.* [197] investigated the performance of a designed deep CNN on datasets consisting of multi-distance spectra at eight different distances ranging from 2.0 to 5.0 m. More than 18 000 LIBS spectra were collected by a duplicate model of the mars surface composition detector (MarSCoDe) instrument for China's Tianwen-1 Mars mission.

Underwater LIBS of submerged solids has been suffering from serious spectral deformation and shot-to-shot fluctuation. Some multivariate analysis, such as PCR and PLSR models, has been applied to improve the quantitative performance of underwater LIBS. Takahashi *et al.* [198, 199] have achieved PCR and PLS based quantification of underwater LIBS data of submerged alloy samples. The non-linear effects of excitation temperature fluctuations on the signals are treated as systematic errors in the analysis. The effect of these errors on the analytical performance is evaluated by applying PCR and PLS with a temperature segmented database. The results demonstrated that the proposed database segmentation can improve quantitative accuracy of the PCR and PLS models. Zheng's group [200] developed an underwater LIBS system named LIBSea II, which was deployed on a remotely operated vehicle (ROV) of Haima for deep sea trial. To reduce the matrix effect and the instability of LIBS signals, the MLR calibration model of Zn and Cu elements were constructed. The correlation coefficients (R^2) of correlation relationships between the predicted concentration and reference concentration are 0.989 and

0.979 for Zn and Cu, respectively. These results indicate that LIBSea II for in-situ direct detection and quantitative analysis of submerged solids in real seawater environment.

The real-time in-line quantitative analysis instruments are highly demanded in many industrial sectors. LIBS is uniquely positioned in this regard, but the complexity of the field environment can have a serious impact on analytical performance, and fortunately, machine learning algorithms have the potential to compensate for this shortcoming. For example, Li *et al.* [201] designed a LIBS setup with optimized optical route and PCA-PLS algorithm for real time and high-precision online determination of total iron TFe, silica SiO₂, aluminum oxide Al₂O₃, and phosphorus P in iron ore. In this work, the spectral pretreatment algorithm was optimized for baseline removal and spectral normalization. The overlapped window slide algorithm avoids the deformation of emission peaks in spectral baseline removal, and two normalization steps by total back area and total spectral intensity within the sub-channel are applied to improve the spectral data stabilization.

In summary, due to the matrix effect and the fluctuation of LIBS signal, the conventional univariate models for LIBS quantitative analysis presents poorer accuracy and precision than other similar spectral analysis techniques. The introducing of a machine learning algorithm can solve these problems to a certain extent. However, due to changes in the experimental environment, the matrix differences of unknown samples and trained samples, the analytical model based on machine learning presents low generalization ability. Thus, the combination of physical mechanisms and machine learning models has attracted the interest and attention of researchers in the LIBS field. To show the quantitative performance of kinds of machine learning algorithms in the LIBS field comprehensively, Table 3 presents a summary of quantitative analysis using LIBS combined with machine learning algorithms, including the types of machine learning methods, improvement ways, methods for comparison, analyzed elements, quantitative results, and references. The aim of this list is to: (i) show which machine learning algorithms have been investigated in LIBS quantitative analysis; (ii) compare the ability of different algorithms to improve the performance of quantitative analysis, and provides a reference for future related research.

5 Summary and prospects

The research progress of machine learning in the LIBS field in recent decades was presented in this review, including data selection, variable selection, noise filtering, interference processing, and qualitative and quantitative analysis based on machine learning methods. For a successful analytical model, the selected input variables should contain or reflect the peak information of LIBS

original data while minimizing random noise, background interference, self-absorption, and matrix effects to the greatest extent possible. What's more, the machine learning algorithm should discriminate and reasonably utilize the input variables which are sensitive to analytical uncertainty, accuracy, precision, and generalization ability. To solve the application problems of machine learning algorithms in LIBS fields and simply the construction process of analytical models, a method chain including data preprocessing, physical principles, and intelligent modeling based on machine learning algorithms will be a promising way. It may replace traditional modeling methods, accelerate application field expansion and the improvement of analytical capabilities of LIBS technology. The application of machine learning in LIBS has the following characteristics:

1) Universality. Machine learning is versatile in the whole process of LIBS analysis, including data selection, feature selection and extraction, noise filtering, self-absorption correction, matrix effect suppression, sample identification, and quantitative analysis. It has played an important role in many applications filed, including geological exploration, industrial metallurgy, environmental pollution monitoring, food safety, and biomedicine. If there is enough data, selecting appropriate machine learning algorithms can greatly improve spectral stability and the accuracy of qualitative/quantitative analysis of the LIBS technique.

2) Specificity. All kinds of machine learning methods, such as unsupervised learning, supervised learning, and semi-supervised learning, can all be adapted to LIBS analysis. Although machine learning is effective for LIBS data processing and modeling, the same algorithm may have different analytical capabilities for samples with different matrix features. There is no general machine learning to solve all the LIBS problems. The selection of machine learning algorithms depends on specific application requirements. Therefore, selecting and optimizing machine learning algorithms is very important for LIBS analysis to achieve optimal analytical performance. When a single algorithm cannot meet these requirements, multiple algorithms can be combined, or a hybrid algorithm based on physical principles and machine learning can be established. When all above methods cannot meet these requirements, an improved or a new machine learning algorithm should be developed.

3) Selectivity. Although early machine learning was only used to solve the problem of multivariate simultaneous analysis, the application of the LIBS technique combined with machine learning methods gradually increased in terms of classification and regression analysis. An increasing number of studies indicate that combining different machine learning algorithms at different stages of LIBS data analysis is a new trend. Some methods may be very effective in preprocessing, but others may be beneficial for modeling, which includes classification and quantification. Therefore,



Table 3 A summary of works dealing with quantitative analysis using LIBS combined with machine learning algorithms.

Methods	Improvement	Comparison	Materials	Target	Best results	Ref.
SVR	WPT-DC-RFECV	WPT-RFECV	Coal	C, Si, Al, Ca, Na, H, N, O, Mg, K, Li, Fe	RMSEP: 0.5786 ARE: 2.27%	[202]
	WTD-RFECV	SVR, RFECV-SVR	Coal	Fe, Cu, Si, K	RMSEP: 0.010281% ARE: 2.16% R^2 : 0.9911	[73]
	–	–	Coal	C, H, N	RSD: 4.16% ARE: 3.28%	[66]
	–	GKR-ANN	Plutonium alloys	Ga	RMSEP: 0.33% LOD:0.015%	[142]
	–	PLSR	Sedimentary rock	Si, Ca, Mg, Fe, Al	RMSE: 0.40wt.% RSD: 1.86%	[144]
	CCA	PCA-SVR PCA-PLSR CCA-PLSR	T91 steel	Fe, Cr, Mo, Mn, V	MRE: 2.47% RSD: 2.94% RMSEP: 6.14%	[163]
	RBF	–	Liquid steel	Mn	MSE: 0.599% RSD: 8.26% R^2 : 0.997	[203]
	ALASSO	LASSO-SVR	Soil	Cr	R^2 : 0.998 RMSEV: 0.033wt.% RSD: 2.343%	[204]
	PLSR	RF-SVR, PCA-SVR, CARS-SVR, MC-UVE-SVR	Edible gelatin	C, H, O, N, Na, K, Ca, Mg, Cr	R_p^2 : 0.9708 RMSEP: 5.69wt.%	[162]
	–	PLSR	Soil	C, Ca, Na, O, H, Mg, Al, Fe	R^2 : 0.987 RMSE: 0.079%	[154]
	–	MLR, PLSR, BP-ANN	Soil	Pb, Cd	R^2 : 0.9877 RMSEP: 0.1521	[147]
	SPLSR	PLSR, LS-SVR	Iron ore	Fe, Si, Al, Ca and Mg, Ti	RMSEP: 0.6242% AREP: 0.85%	[205]
	CARS	LS-SVR	Edible vegetable oil	Cr	R^2 : 0.9926 RMSEP: 0.000586%	[80]
	SDVS	PLSR, iPLSR, NrVS-PLSR	Iron	Fe, Si, Ca, Na, H, O	RMSEP: 0.75% R^2 : 0.76	[63]
	Dominant Factor	PLSR	Brass	Cu, Pb, Zn, Fe, P, Sn, Sb	RMSEP: 5.25% R^2 : 0.999	[206]
	Dominant Factor	PLSR	water	Mn, Sr, Li	R^2 : 0.999	[172]
	Dominant Factor	PLSR	Coal	C, H, N, O, Si, Al, Fe, Ca, Mg, K, etc.	RSD: 0.3%	[173]
	PLSR	SVR, MLR	–	Coal	C	R^2 : 0.99 RMSEP: 1.43%
Backward interval		Interval-PLSR, PLSR	Soil	Cu, Ni	R^2 : 0.9449 RMSEP: 0.0363%	[16]
UVE-CARS		PLSR	Agricultural fungicide	Cl	RMSE: 0.66% R^2 : 0.9905	[68]
GA		–	Manure	Ca	R^2 : 0.90 RPD: 3.04	[69]
Double GA		–	Copper	Cu, Fe, S, Si, Ca, Mg, Al, As, Zn, etc.	RMSECV: 0.29% RMSEP: 0.21%	[207]
FSC-mIPW		GA, SPA	Soil	Cu, Ba, Cr	RMSEP: 0.2747%	[71]
mIPW		Interval-PLS	Soil	Cu, Ba, Cr, Mg, Ca	RMSEP: 0.4232wt.% R_p^2 : 0.9746	[72]
Ridge-RFE		PLSR, GA-PLSR,	Aluminum alloy	Fe, Si, Mg, Cu, Zn, Mn	R_{cv}^2 : 0.9566 RMSECV: 0.0601wt.% RMSEP: 0.0476wt.%	[74]
–		–	Alloy	Cu, Zn, Fe, Pb	RSD: 1.80% R^2 : 0.992 RMSEP: 1.30%	[208]
MSC		–	Pork	Pb	R^2 : 0.9908 RMSEP: 0.282 ARPE: 7.8%	[209]
MSC		–	Egg	Cu	R^2 : 0.9789 RMSEP: 50 ARPE: 7.14%	[210]
–		–	Navel orange	Cu	R^2 : 0.9928 ARE: 5.55%	[211]

Table 3 (continued)

Methods	Improvement	Comparison	Materials	Target	Best results	Ref.
PLSR	–	–	Navel orange	Pb	R^2 : 0.9633 RMSECV: 1.56 ARPE: 6.9%	[212]
	–	SVR	Alloy steel	Cr, Ni, Mn, Fe	R^2 : 0.9896 RMSE: 0.8230%	[213]
	–	PLSR, SVR	Alloy steel	Cr, Ni, Mn, Fe	R^2 : 0.9617 RMSE: 1.519	[214]
	–	–	Mineral	Fe, Si, Al, P	RMSE: 3.4wt.% R^2 : 0.95	[215]
	Savitzky-Golay	PCA	River	Al, Ca, Cd, Cr, Fe, K, Mg, Na, Ni, etc.	R^2 : 0.9836 RMSEC: 0.7120 RMSECV: 0.9430	[216]
	–	–	Liquid steel	Mn, Si	ARE: 6.125 R^2 : 0.997	[217]
	GA	–	Boron	B	R^2 : 0.9888 RMSEP: 0.8667wt.% MPE: 10.9685%	[218]
	–	PCA, ANN	Plutonium metal	Fe, Ni	RMSE: 0.00132% R^2 : 0.997	[137]
	–	Univariate Analysis, MLR	Steel alloys	Cr, Ni	R^2 : 0.995	[139]
	Savitzky-Golay	–	River	Ca, C	R^2 : 0.9753 RMSEP: 9.7339%	[219]
RFR	Multi-block	PCR, PLSR S-PLSR	Iron ore	Fe	R^2 : 0.94 RMSEP: 3.1%	[140]
	–	PLSR	River	C, Na, H	R^2 : 0.9248 RMSE: 25.1215%	[148]
	–	PLSR	Steel	S, P	R^2 : 0.9981	[146]
	–	PLSR	Sinter ore	Al, Fe, Si, Ca, Mg	RSD: 0.59%	[153]
	VI	PLSR, LS-SVM	Iron ore	Ca, Mg, Si, Al	RMSE: 0.0554wt.% R^2 : 0.9103	[152]
	MI-VIM	–	Oily sludge	Cu, Cr, Pb and Zn	R_p^2 : 0.9681 RMSEP: 0.6009	[220]
	SBS	PLSR, RFR	Steel	S, P	R^2 : 0.9991	[75]
	LR-SUAC	PLSR	Ceramic	Si, Al, Mg, Fe, Ti	R^2 : 0.93	[79]
	GA	–	Steel	C	RMSE: 0.0114%	[221]
	feature selection BPNN	–	Rock	Li, Rb, Sr, Ba	RMSEP: 0.0162wt.%	[161]
ANN	PLSR	PLSR, ANN	Plant	Mg, Fe, N, Al, B, Ca, K, Mn, P	RMSEC: 0.0028wt.% MPE: 4.22%	[160]
	–	PLSR, SVR, PCR	Coal	C, H, O, N, Fe, Mg, Al, Ca, Si, Li, etc.	AAE: 0.69%	[155]
	GA-BP	BP-ANN	Alloy steel	C, Fe, Cr, Mn, Si	R^2 : 0.9893	[164]
	MSLC	–	Alloy	Cu, Zn, Sn, Pb, Si	R^2 : 0.99	[158]
	MSLC	–	Steel	Cr, Ni	RMSE: 0.023wt.% RSD: 12.9%	[159]
	GA	–	Steel	Cu, V	RMSEP: 0.0040wt.%	[89]
	Rectified linear units	–	Coal	C, H, N, O, Li, Na, Mg, Al, K, Ca, etc.	R^2 : 0.7102 RMSE: 0.4786%	[222]
	LASSO	PCA-ANN, CARS-ANN, KBest-ANN	Rocks and mineral	Cu, Fe, As	RMSEP: 0.18% R^2 : 0.97	[223]
	Lightweight	PLSR, SVR	Phosphate ore	P	RMSEP: 0.89% R^2 : 0.9874	[91]
	–	BPNN, PLSR	Mineral	Si, Al, Fe, Ca, Mg, K, Na	RMSE: 0.022	[224]
KELM	–	SVR, LS-SVM, BP-ANN	Coal	C, S	RMSE: 0.7704% R^2 : 0.9832	[225]
	–	PLSR	Sinter ore	Al, Fe, Si, Ca, Mg	R^2 : 0.9	[145]
	MI-PSO	KELM	Coal	C, Si, Ca, Mg, Al, Fe, Mn, Na, Ti	RMSECV: 1.2886 R_{CV}^2 : 0.9868	[70]
	Dominant factor	MLR, PLSR, DF-PLSR, SVR, DF-SVR and KELM	Coal	C, H, O, I, N, Mg, Si, Al, Mn, Ca, Na,	RMSE: 1.075% R^2 : 0.965	[165]



selecting different methods at different stages is also very important and worth further research.

4) Limitation. Although machine learning is excellent in assisting LIBS analysis, it is not omnipotent. The insufficient number of samples and lack of physical principles make the model based on machine learning methods less robust, making it difficult to obtain accurate quantitative analysis of unknown samples outside the training set. Therefore, in the process of using machine learning algorithms, its limitations should also be fully recognized. If machine learning algorithms are abused for LIBS analysis regardless of physical mechanisms, it will result in the obtained material types or element contents not being related to the characteristic spectral information of their constituent elements, but may be related to information from noise, background, interference, or even pollutants.

5) Prospects. At present, although the machine learning has been proven to improve the analytical performance of LIBS, it depends on the specific applications. The reason is that the current training data is not enough for accurate prediction when the sample to be tested is unknown. The potential for overtraining is significant with LIBS spectral data, resulting in calibration or classification models that are less robust than even univariate models. Therefore, the future big model might provide a promising way to break the current limitations, the model will have better adaptability to uncertain factors such as matrix effect, self-absorption effect, noise interference, and equipment parameter drift. However, robust analytical models require a large number of training samples, and how to process massive spectral data is also an important problem to be studied in the application of machine learning algorithms in LIBS.

Declarations The authors declare that they have no competing interests and there are no conflicts.

Acknowledgements This research was financially supported by the project funded by the National Natural Science Foundation of China (Nos. 62305123, 32301694, and 12064029) and the National Key R&D Program of China (No. 2023YFE0204600).

References

- Z. Wang, M. S. Afgan, W. Gu, Y. Song, Y. Wang, Z. Hou, W. Song, and Z. Li, Recent advances in laser-induced breakdown spectroscopy quantification: From fundamental understanding to data processing, *Trends Analyt. Chem.* 143, 116385 (2021)
- Y. T. Fu, W. L. Gu, Z. Y. Hou, S. A. Muhammed, T. Q. Li, Y. Wang, and Z. Wang, Mechanism of signal uncertainty generation for laser-induced breakdown spectroscopy, *Front. Phys.* 16(2), 22502 (2021)
- S. Sheta, M. S. Afgan, Z. Y. Hou, S. C. Yao, L. Zhang, Z. Li, and Z. Wang, Coal analysis by laser-induced breakdown spectroscopy: A tutorial review, *J. Anal. At. Spectrom.* 34(6), 1047 (2019)
- N. C. Dingari, I. Barman, A. K. Myakalwar, S. P. Tewari, and M. K. 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 (2012)
- T. L. Zhang, W. U. Shan, H. S. Tang, K. Wang, Y. X. Duan, and L. I. Hua, Progress of chemometrics in laser-induced breakdown spectroscopy analysis, *Chin. J. Anal. Chem.* 43(6), 939 (2015)
- T. Zhang, H. Tang, and H. Li, Chemometrics in laser-induced breakdown spectroscopy, *J. Chemometr.* 32(11), e2983 (2018)
- D. Zhang, H. Zhang, Y. Zhao, Y. Chen, C. Ke, T. Xu, and Y. He, A brief review of new data analysis methods of laser-induced breakdown spectroscopy: Machine learning, *Appl. Spectrosc. Rev.* 57(2), 89 (2022)
- M. R. Wójcik, R. Zdunek, and A. J. Antończak, Unsupervised verification of laser-induced breakdown spectroscopy dataset clustering, *Spectrochim. Acta B* 126, 84 (2016)
- Y. Tang, Y. Guo, Q. Sun, S. Tang, J. Li, L. Guo, and J. Duan, Industrial polymers classification using laser-induced breakdown spectroscopy combined with self-organizing maps and K-means algorithm, *Optik (Stuttg.)* 165, 179 (2018)
- Y. Guo, Y. Tang, Y. Du, S. Tang, L. Guo, X. Li, Y. Lu, and X. Zeng, Cluster analysis of polymers using laser-induced breakdown spectroscopy with K-means, *Plasma Sci. Technol.* 20(6), 065505 (2018)
- P. Pořízka, J. Klus, E. Kepes, D. Prochazka, D. W. Hahn, and J. Kaiser, On the utilization of principal component analysis in laser-induced breakdown spectroscopy data analysis, a review, *Spectrochim. Acta B* 148, 65 (2018)
- L. A. He, X. Q. Wang, Y. Zhao, L. Liu, and Z. Peng, Study on cluster analysis used with laser-induced breakdown spectroscopy, *Plasma Sci. Technol.* 18(6), 647 (2016)
- T. Chen, L. Huang, M. Yao, H. Hu, C. Wang, and M. Liu, Quantitative analysis of chromium in potatoes by laser-induced breakdown spectroscopy coupled with linear multivariate calibration, *Appl. Opt.* 54(25), 7807 (2015)
- W. Sha, J. T. Li, C. P. Lu, and C. H. Zhen, Quantitative analysis of P in compound fertilizer by laser-induced breakdown spectroscopy coupled with linear multivariate calibration, *Spectroscopy & Spectral Anal.* 39(6), 1958 (2019) (in Chinese)
- H. Y. Li, L. Mazzei, C. D. Wallis, and A. S. Wexler, Improving quantitative analysis of spark-induced breakdown spectroscopy: Multivariate calibration of metal particles using machine learning, *J. Aerosol Sci.* 159, 105874 (2022)
- S. N. Zhu, Y. Ding, Y. J. Chen, F. Deng, F. F. Chen, and F. Yan, Quantitative analysis of Cu and Ni in oil-contaminated soil by LIBS combined with variable selection method and PLS, *Spectroscopy & Spectral Anal.* 40(12), 3812 (2020) (in Chinese)
- Jr De Lucia and J. L. Gottfried, Influence of variable selection on partial least squares discriminant analysis

- models for explosive residue classification, *Spectrochim. Acta B* 66(2), 122 (2011)
18. Y. Lee, S. H. Han, and S. H. Nam, Soft independent modeling of class analogy (SIMCA) modeling of laser-induced plasma emission spectra of edible salts for accurate classification, *Appl. Spectrosc.* 71(9), 2199 (2017)
 19. Z. Cao, J. Cheng, X. Han, L. Li, J. Wang, Q. Fan, and Q. Lin, Rapid classification of coal by laser-induced breakdown spectroscopy (LIBS) with K-nearest neighbor (KNN) chemometrics, *Instrum. Sci. Technol.* 51(1), 59 (2023)
 20. J. ShangGuan, Y. Tong, A. Yuan, X. Ren, J. Liu, H. Duan, Z. Lian, X. Hu, J. Ma, Z. Yang, and D. Wang, Online detection of laser paint removal based on laser-induced breakdown spectroscopy and the K-nearest neighbor method, *J. Laser Appl.* 34(2), 022009 (2022)
 21. X. Yan, X. Peng, Y. Qin, Z. Xu, B. Xu, C. Li, N. Zhao, J. Li, Q. Ma, and Q. Zhang, Classification of plastics using laser-induced breakdown spectroscopy combined with principal component analysis and K nearest neighbor algorithm, *Results in Optics* 4, 100093 (2021)
 22. L. Liang, T. Zhang, K. Wang, H. Tang, X. Yang, X. Zhu, Y. Duan, and H. Li, Classification of steel materials by laser-induced breakdown spectroscopy coupled with support vector machines, *Appl. Opt.* 53(4), 544 (2014)
 23. M. V. Dastjerdi, S. J. Mousavi, M. Soltanolkotabi, and A. N. Zadeh, Identification and sorting of PVC polymer in recycling process by laser-induced breakdown spectroscopy (LIBS) combined with support vector machine (SVM) model, *Iranian J. Sci. Technol. A* 42(2), 959 (2018)
 24. P. Yang, H. T. Liu, Z. L. Nie, and X. N. Qu, Accuracy improvement of geographical indication of rice by laser-induced breakdown spectroscopy using support vector machine with multi-spectral line, *J. Appl. Spectrosc.* 89(3), 579 (2022)
 25. J. Jia, H. Fu, Z. Hou, H. Wang, Z. Ni, and F. Dong, Calibration curve and support vector regression methods applied for quantification of cement raw meal using laser-induced breakdown spectroscopy, *Plasma Sci. Technol.* 21(3), 034003 (2019)
 26. J. Cisewski, E. Snyder, J. Hannig, and L. Oudejans, Support vector machine classification of suspect powders using laser-induced breakdown spectroscopy (LIBS) spectral data, *J. Chemometr.* 26(5), 143 (2012)
 27. E. D'Andrea, S. Pagnotta, E. Grifoni, S. Legnaioli, G. Lorenzetti, V. Palleschi, and B. Lazzerini, A hybrid calibration-free/artificial neural networks approach to the quantitative analysis of LIBS spectra, *Appl. Phys. B* 118(3), 353 (2015)
 28. F. G. Rendon-Sauz, T. Flores-Reyes, and A. Ponce-Flores, Rapid classification of bacteria using libs in multi-pulse laser regime and neural networks processing, *Revista Cubana De Fisica* 35(1), 10 (2018)
 29. X. Cui, Q. Wang, Y. Zhao, X. Qiao, and G. Teng, Laser-induced breakdown spectroscopy (LIBS) for classification of wood species integrated with artificial neural network (ANN), *Appl. Phys. B* 125(4), 56 (2019)
 30. J. El Haddad, M. Villot-Kadri, A. Ismael, G. Gallou, K. Michel, D. Bruyere, V. Laperche, L. Canioni, and B. Bousquet, Artificial neural network for on-site quantitative analysis of soils using laser induced breakdown spectroscopy, *Spectrochim. Acta B* 79–80, 51 (2013)
 31. Q. Q. Wang, Z. W. Huang, K. Liu, W. J. Li, and J. X. Yan, Classification of plastics with laser-induced breakdown spectroscopy based on principal component analysis and artificial neural network model, *Spectroscopy & Spectral Anal.* 32(12), 3179 (2012) (in Chinese)
 32. N. Li, J. Qi, P. Wang, X. Zhang, T. Zhang, and H. Li, Quantitative structure-activity relationship (QSAR) study of carcinogenicity of polycyclic aromatic hydrocarbons (PAHs) in atmospheric particulate matter by random forest (RF), *Anal. Methods* 11(13), 1816 (2019)
 33. L. Sheng, T. Zhang, G. Niu, K. Wang, H. Tang, Y. Duan, and H. Li, Classification of iron ores by laser-induced breakdown spectroscopy (LIBS) combined with random forest (RF), *J. Anal. At. Spectrom.* 30(2), 453 (2015)
 34. L. Zhan, X. Ma, W. Fang, R. Wang, Z. Liu, Y. Song, and H. Zhao, A rapid classification method of aluminum alloy based on laser-induced breakdown spectroscopy and random forest algorithm, *Plasma Sci. Technol.* 21(3), 034018 (2019)
 35. T. Feng, X. Zhang, M. Li, T. Chen, L. Jiao, Y. Xu, H. Tang, T. Zhang, and H. Li, Pollution risk estimation of the Cu element in atmospheric sedimentation samples by laser induced breakdown spectroscopy (LIBS) combined with random forest (RF), *Anal. Methods* 13(30), 3424 (2021)
 36. K. Liu, D. Tian, H. Xu, H. Wang, and G. Yang, Quantitative analysis of toxic elements in polypropylene (PP) via laser-induced breakdown spectroscopy (LIBS) coupled with random forest regression based on variable importance (VI-RFR), *Anal. Methods* 11(37), 4769 (2019)
 37. J. Liang, C. Yan, Y. Zhang, T. Zhang, X. Zheng, and H. Li, Rapid discrimination of *Salvia miltiorrhiza* according to their geographical regions by laser induced breakdown spectroscopy (LIBS) and particle swarm optimization-kernel extreme learning machine (PSO-KELM), *Chemom. Intell. Lab. Syst.* 197, 103930 (2020)
 38. Y. Mei, S. Cheng, Z. Hao, L. Guo, X. Li, X. Zeng, and J. Ge, Quantitative analysis of steel and iron by laser-induced breakdown spectroscopy using GA-KELM, *Plasma Sci. Technol.* 21(3), 034020 (2019)
 39. C. Yan, T. Zhang, Y. Sun, H. Tang, and H. Li, A hybrid variable selection method based on wavelet transform and mean impact value for calorific value determination of coal using laser-induced breakdown spectroscopy and kernel extreme learning machine, *Spectrochim. Acta B* 154, 75 (2019)
 40. G. Vítková, K. Novotny, L. Prokes, A. Hrdlicka, J. Kaiser, J. Novotny, R. Malina, and D. Prochazka, Fast identification of biominerals by means of stand-off laser-induced breakdown spectroscopy using linear discriminant analysis and artificial neural networks, *Spectrochim. Acta B* 73, 1 (2012)



41. P. Yang, R. Zhou, W. Zhang, S. S. Tang, Z. Q. Hao, X. Y. Li, Y. F. Lu, and X. Y. Zeng, Laser-induced breakdown spectroscopy assisted chemometric methods for rice geographic origin classification, *Appl. Opt.* 57(28), 8297 (2018)
42. Z. F. Zhao, L. Chen, F. Liu, F. Zhou, J. Y. Peng, and M. H. Sun, Fast classification of geographical origins of honey based on laser-induced breakdown spectroscopy and multivariate analysis, *Sensors (Basel)* 20(7), 1878 (2020)
43. X. M. Li, H. L. Lu, J. H. Yang, and F. Chang, Semi-supervised LIBS quantitative analysis method based on co-training regression model with selection of effective unlabeled samples, *Plasma Sci. Technol.* 21(3), 034015 (2018)
44. Q. Wang, G. Teng, C. Li, Y. Zhao, and Z. Peng, Identification and classification of explosives using semi-supervised learning and laser-induced breakdown spectroscopy, *J. Hazard. Mater.* 369, 423 (2019)
45. S. Xie, T. Xu, X. Han, Q. Lin, and Y. Duan, Accuracy improvement of quantitative LIBS analysis using wavelet threshold de-noising, *J. Anal. At. Spectrom.* 32(3), 629 (2017)
46. R. Wang, X. Ma, Q. Yu, Y. Song, H. Zhao, M. Zhang, and Y. Liao, Methods of data processing for trace elements analysis using laser induced breakdown spectroscopy, *Plasma Sci. Technol.* 17(11), 944 (2015)
47. H. Yang, L. Huang, T. B. Chen, G. F. Rao, M. H. Liu, J. Y. Chen, and M. Y. Yao, Spectral filtering method for improvement of detection accuracy of lead in vegetables by laser induced breakdown spectroscopy, *Chin. J. Anal. Chem.* 45(8), 1123 (2017)
48. N. Ali, Z. Huang, J. Zong, Y. Ma, Y. Xiao, L. Wang, P. Zhang, and D. Chen, Real-time analysis of mineral elements in oat using laser-induced breakdown spectroscopy, *J. Food Safety & Food Quality* 72(4), 131 (2021)
49. H. Guo, M. Cui, Z. Feng, D. Zhang, and D. Zhang, Classification of aviation alloys using laser-induced breakdown spectroscopy based on a WT-PSO-LSSVM model, *Chemosensors (Basel)* 10(6), 220 (2022)
50. T. Yuan, Z. Wang, Z. Li, W. Ni, and J. Liu, A partial least squares and wavelet-transform hybrid model to analyze carbon content in coal using laser-induced breakdown spectroscopy, *Anal. Chim. Acta* 807, 29 (2014)
51. B. Zhang, L. Sun, H. Yu, Y. Xin, and Z. Cong, A method for improving wavelet threshold denoising in laser-induced breakdown spectroscopy, *Spectrochim. Acta B* 107, 32 (2015)
52. J. Wei, J. Dong, T. Zhang, Z. Wang, and H. Li, Quantitative analysis of the major components of coal ash using laser induced breakdown spectroscopy coupled with a wavelet neural network (WNN), *Anal. Methods* 8(7), 1674 (2016)
53. L. Yang, Y. Zhang, J. Liu, Z. Zhang, M. Xu, F. Ji, J. Chen, T. Zhang, and R. Lu, Spectral preprocessing to improve accuracy of quantitative detection of elemental Cr in austenitic stainless steel by laser-induced breakdown spectroscopy, *Rev. Sci. Instrum.* 93(3), 033002 (2022)
54. H. Duan, S. Ma, L. Han, and G. Huang, A novel denoising method for laser-induced breakdown spectroscopy: Improved wavelet dual threshold function method and its application to quantitative modeling of Cu and Zn in Chinese animal manure composts, *Microchem. J.* 134, 262 (2017)
55. J. Chappell, M. Martinez, and M. Baudalet, Statistical evaluation of spectral interferences in laser-induced breakdown spectroscopy, *Spectrochim. Acta B* 149, 167 (2018)
56. K. Liu, R. Zhou, W. Zhang, Z. Tang, J. Yan, M. Lv, X. Li, Y. Lu, and X. Zeng, Interference correction for laser-induced breakdown spectroscopy using a deconvolution algorithm, *Anal. At. Spectrom.* 35, 762 (2020)
57. B. Tan, M. Huang, Q. Zhu, Y. Guo, and J. Qin, Decomposition and correction overlapping peaks of LIBS using an error compensation method combined with curve fitting, *Appl. Opt.* 56(25), 7116 (2017)
58. Y. Wang, Y. Bu, F. Wu, Y. Cao, Y. Yu, and X. Wang, Research on LIBS quantitative analysis of heavy metal concentration in polluted water-based on Fourier self-deconvolution method, in: *AOPC 2019: Optical Spectroscopy and Imaging*, SPIE, 2019, pp 167–172
59. J. Guezenoc, A. Gallet-Budynek, and B. Bousquet, Critical review and advices on spectral-based normalization methods for LIBS quantitative analysis, *Spectrochim. Acta B* 160, 105688 (2019)
60. R. Wang, and X. Ma, Study on LIBS Standard Method via Key Parameter Monitoring and Backpropagation Neural Network, *Chemosensors (Basel)* 10(8), 312 (2022)
61. P. Lu, Z. Zhuo, W. H. Zhang, J. Tang, Y. Wang, H. L. Zhou, X. L. Huang, T. F. Sun, and J. Q. Lu, A hybrid feature selection combining wavelet transform for quantitative analysis of heat value of coal using laser-induced breakdown spectroscopy, *Appl. Phys. B* 127(2), 19 (2021)
62. S. Xie, T. Xu, G. Niu, W. Liao, Q. Lin, and Y. Duan, Quantitative analysis of steel samples by laser-induced-breakdown spectroscopy with wavelet-packet-based relevance vector machines, *J. Anal. At. Spectrom.* 33(6), 975 (2018)
63. T. Chen, L. Sun, H. Yu, L. Qi, D. Shang, and Y. Xie, Efficient weakly supervised LIBS feature selection method in quantitative analysis of iron ore slurry, *Appl. Opt.* 61(7), D22 (2022)
64. E. Harefa and W. Zhou, Performing sequential forward selection and variational autoencoder techniques in soil classification based on laser-induced breakdown spectroscopy, *Anal. Methods* 13(41), 4926 (2021)
65. Y. Chu, F. Chen, Z. Sheng, D. Zhang, S. Zhang, W. Wang, H. Jin, J. Qi, and L. Guo, Blood cancer diagnosis using ensemble learning based on a random subspace method in laser-induced breakdown spectroscopy, *Biomed. Opt. Express* 11(8), 4191 (2020)
66. Y. Jiang, Z. Lu, X. Chen, Z. Yu, H. Qin, J. Chen, J. Lu, and S. Yao, Optimizing the quantitative analysis of solid biomass fuel properties using laser induced breakdown spectroscopy (LIBS) coupled with a kernel partial least squares (KPLS) model, *Anal. Methods* 13(45), 5467 (2021)

67. H. Y. Kong, L. X. Sun, J. T. Hu, and P. Zhang, Automatic method for selecting characteristic lines based on genetic algorithm to quantify laser-induced breakdown spectroscopy, *Spectroscopy & Spectral Anal.* 36(5), 1451 (2016) (in Chinese)
68. L. P. Gan, T. Sun, J. Liu, and M. H. Liu, Double pulse LIBS combined with variable screening to detect procymidone content, *Spectroscopy & Spectral Anal.* 39(02), 584 (2019) (in Chinese)
69. S. Ma, Q. Ma, L. Han, and G. Huang, Modelling of calcium content in manure using laser-induced breakdown spectroscopy and genetic algorithm combined with partial least squares, *Spectroscopy & Spectral Anal.* 37(5), 1532 (2017) (in Chinese)
70. T. He, J. Liang, H. Tang, T. Zhang, C. Yan, and H. Li, Quantitative analysis of coal quality by mutual information-particle swarm optimization (MI-PSO) hybrid variable selection method coupled with spectral fusion strategy of laser-induced breakdown spectroscopy (LIBS) and fourier transform infrared spectroscopy (FTIR), *Spectrochim. Acta B* 178, 106112 (2021)
71. F. Duan, X. Fu, J. Jiang, T. Huang, L. Ma, and C. Zhang, Automatic variable selection method and a comparison for quantitative analysis in laser-induced breakdown spectroscopy, *Spectrochim. Acta B* 143, 12 (2018)
72. X. Fu, F. J. Duan, T. T. Huang, L. Ma, J. J. Jiang, and Y. C. Li, A fast variable selection method for quantitative analysis of soils using laser-induced breakdown spectroscopy, *J. Anal. At. Spectrom.* 32(6), 1166 (2017)
73. P. Lu, Z. Zhuo, W. Zhang, J. Tang, H. Tang, and J. Lu, Accuracy improvement of quantitative LIBS analysis of coal properties using a hybrid model based on a wavelet threshold de-noising and feature selection method, *Appl. Opt.* 59(22), 6443 (2020)
74. G. Wang, L. Sun, W. Wang, T. Chen, M. Guo, and P. Zhang, A feature selection method combined with ridge regression and recursive feature elimination in quantitative analysis of laser induced breakdown spectroscopy, *Plasma Sci. Technol.* 22(7), 074002 (2020)
75. F. Ruan, J. Qi, C. Yan, H. Tang, T. Zhang, and H. Li, Quantitative detection of harmful elements in alloy steel by LIBS technique and sequential backward selection-random forest (SBS-RF), *J. Anal. At. Spectrom.* 32(11), 2194 (2017)
76. F. Ruan, L. Hou, T. Zhang, and H. Li, A modified backward elimination approach for the rapid classification of Chinese ceramics using laser-induced breakdown spectroscopy and chemometrics, *J. Anal. At. Spectrom.* 35(3), 518 (2020)
77. Y. Ding, Y. Shu, A. Hu, M. Zhao, J. Chen, L. Yang, W. Chen, and Y. Wang, Determination of soil source using laser induced breakdown spectroscopy combined with feature selection, *J. Anal. At. Spectrom.* 38(11), 2499 (2023)
78. W. You, Y. P. Xia, Y. T. Huang, J. J. Lin, and X. M. Lin, Research on selection method of LIBS feature variables based on CART regression tree, *Spectroscopy & Spectral Anal.* 41(10), 3240 (2021) (in Chinese)
79. Z. Lv, H. Yu, L. Sun, and P. Zhang, Composition analysis of ceramic raw materials using laser-induced breakdown spectroscopy and autoencoder neural network, *Anal. Methods* 14(13), 1320 (2022)
80. Y. Wu, T. Sun, J. Liu, L. Gan, and M. Liu, Detection of chromium content in edible vegetable oil with DP-LIBS combined with LSSVM and CARS methods, *Laser & Optoelectron. Prog.* 55(1), 013005-1 (2018) (in Chinese)
81. D. H. Zhu, M. C. Wang, L. J. Xu, X. J. Chen, B. T. Sun, J. Zhang, W. W. Liu, Y. Cao, L. M. Yuan, and Y. Cai, Detection of Pb element composition in irregular copper alloy samples based on multi-line internal standard method, *Spectroscopy & Spectral Anal.* 39(10), 3159 (2019) (in Chinese)
82. J. Long, W. Song, Z. Hou, and Z. Wang, A data selection method for matrix effects and uncertainty reduction for laser-induced breakdown spectroscopy, *Plasma Sci. Technol.* 25(7), 075501 (2023)
83. J. a. Liu, J. m. Li, N. Zhao, Q. x. Ma, L. Guo, and Q. m. Zhang, Rapid classification and identification of plastic using laser-induced breakdown spectroscopy with principal component analysis and support vector machine, *Spectroscopy & Spectral Anal.* 41(6), 1955 (2021) (in Chinese)
84. P. Pořízka, J. Klus, E. Képeš, D. Prochazka, D. W. Hahn, and J. Kaiser, On the utilization of principal component analysis in laser-induced breakdown spectroscopy data analysis – A review, *Spectrochim. Acta B* 148, 65 (2018)
85. J. B. Sirven, B. Salle, P. Mauchien, J. L. Lacour, S. Maurice, and G. Manhes, Feasibility study of rock identification at the surface of Mars by remote laser-induced breakdown spectroscopy and three chemometric methods, *J. Anal. At. Spectrom.* 22(12), 1471 (2007)
86. Z. A. Abdel-Salam, V. Palleschi, and M. A. Harith, Study of the feeding effect on recent and ancient bovine bones by nanoparticle-enhanced laser-induced breakdown spectroscopy and chemometrics, *J. Adv. Res.* 17, 65 (2019)
87. A. H. Farhadian, M. K. Tehrani, M. H. Keshavarz, and S. M. R. Darbani, Energetic materials identification by laser-induced breakdown spectroscopy combined with artificial neural network, *Appl. Opt.* 56(12), 3372 (2017)
88. M. Yuan, Q. Zeng, J. Wang, W. Li, G. Chen, Z. Li, Y. Liu, L. Guo, X. Li, and H. Yu, Rapid classification of steel via a modified support vector machine algorithm based on portable fiber-optic laser-induced breakdown spectroscopy, *Opt. Eng.* 60(12), 124114 (2021)
89. K. Li, L. Guo, J. Li, X. Yang, R. Yi, X. Li, Y. Lu, and X. Zeng, Quantitative analysis of steel samples using laser-induced breakdown spectroscopy with an artificial neural network incorporating a genetic algorithm, *Appl. Opt.* 56(4), 935 (2017)
90. Q. X. Zhong, T. Z. Zhao, X. Li, F. Q. Lian, H. Xiao, S. Z. Nie, S. N. Sun, and Z. W. Fan, Standardized cross-validation and its optimization for multi-element LIBS analysis, *Spectroscopy & Spectral Anal.* 40(2), 622 (2020) (in Chinese)
91. H. Dong, L. Sun, L. Qi, H. Yu, and P. Zeng, A lightweight convolutional neural network model for



- quantitative analysis of phosphate ore slurry based on laser-induced breakdown spectroscopy, *J. Anal. At. Spectrom.* 36(11), 2528 (2021)
92. D. Prochazka, P. Porizka, J. Hruska, K. Novotny, A. Hrdlicka, and J. Kaiser, Machine learning in laser-induced breakdown spectroscopy as a novel approach towards experimental parameter optimization, *J. Anal. At. Spectrom.* 37(3), 603 (2022)
 93. H. Tang, T. Zhang, X. Yang, and H. Li, Classification of different types of slag samples by laser-induced breakdown spectroscopy (LIBS) coupled with random forest based on variable importance (VIRF), *Anal. Methods* 7(21), 9171 (2015)
 94. P. Yang, R. Zhou, W. Zhang, S. Tang, Z. Hao, X. Li, Y. Lu, and X. Zeng, Laser-induced breakdown spectroscopy assisted chemometric methods for rice geographic origin classification, *Appl. Opt.* 57(28), 8297 (2018)
 95. F. I. Alarsan and M. Younes, Analysis and classification of heart diseases using heartbeat features and machine learning algorithms, *J. Big Data* 6(1), 81 (2019)
 96. J. Yu, Z. Hou, S. Sheta, J. Dong, W. Han, T. Lu, and Z. Wang, Provenance classification of nephrite jades using multivariate LIBS: A comparative study, *Anal. Methods* 10(3), 281 (2018)
 97. N. Gyftokostas, D. Stefanos, V. Kokkinos, C. Bouras, and S. Couris, Laser-induced breakdown spectroscopy coupled with machine learning as a tool for olive oil authenticity and geographic discrimination, *Sci. Rep.* 11(1), 5360 (2021)
 98. Z. Zhao, L. Chen, F. Liu, F. Zhou, J. Peng, and M. Sun, Fast classification of geographical origins of honey based on laser-induced breakdown spectroscopy and multivariate analysis, *Sensors (Basel)* 20(7), 1878 (2020)
 99. Z. Luo, L. Zhang, T. Chen, M. Liu, J. Chen, H. Zhou, and M. Yao, Rapid identification of rice species by laser-induced breakdown spectroscopy combined with pattern recognition, *Appl. Opt.* 58(7), 1631 (2019)
 100. W. Huang, L. Guo, W. Kou, D. Zhang, Z. Hu, F. Chen, Y. Chu, and W. Cheng, Identification of adulterated milk powder based on convolutional neural network and laser-induced breakdown spectroscopy, *Microchem. J.* 176, 107190 (2022)
 101. K. Kiss, A. Sindelárová, L. Krbal, V. Stejskal, K. Mrázová, J. Vrábek, M. Kaska, P. Modlitbová, P. Porizka, and J. Kaiser, Imaging margins of skin tumors using laser-induced breakdown spectroscopy and machine learning, *J. Anal. At. Spectrom.* 36(5), 909 (2021)
 102. J. Ding, D. C. Zhang, B. W. Wang, Z. Q. Feng, X. Y. Liu, and J. F. Zhu, The classification of plant leaves by applying chemometrics methods on laser-induced breakdown spectroscopy, *Spectroscopy & Spectral Anal.* 41(2), 606 (2021) (in Chinese)
 103. 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 (2018)
 104. M. S. Babu, T. Imai, and R. Sarathi, Classification of aged epoxy micro-nanocomposites through PCA- and ANN-adopted LIBS analysis, *IEEE Trans. Plasma Sci.* 49(3), 1088 (2021)
 105. M. Singh and A. Sarkar, Comparative study of the pls and pcr methods in laser-induced breakdown spectroscopic analysis, *J. Appl. Spectrosc.* 85(5), 962 (2018)
 106. Y. Liu, S. Zhao, X. Gao, S. Fu, Chao Song, Y. Dou, Shaozhong Song, C. Qi, and J. Lin, Combined laser-induced breakdown spectroscopy and hyperspectral imaging with machine learning for the classification and identification of rice geographical origin, *RSC Adv.* 12(53), 34520 (2022)
 107. B. Campanella, E. Grifoni, S. Legnaioli, G. Lorenzetti, S. Pagnotta, F. Sorrentino, and V. Palleschi, Classification of wrought aluminum alloys by artificial neural networks evaluation of laser induced breakdown spectroscopy spectra from aluminum scrap samples, *Spectrochim. Acta B* 134, 52 (2017)
 108. S. Park, J. Lee, E. Kwon, D. Kim, S. Shin, S. Jeong, and K. Park, 3D sensing system for laser-induced breakdown spectroscopy-based metal scrap identification, *Int. J. Precis. Eng. & Manuf. -Green Tech.* 9, 695 (2022)
 109. A. Demir, D. K. Ürk, K. Akben, M. Doğan, E. Pehlivan, Ö. Yalçın, M. A. Kıştan, G. Gökçe, and A. Obalı, Elemental Analysis and Classification of Molten Aluminum Alloys by LIBS, Springer Nature Switzerland, Cham, 2024, pp. 984-990.
 110. J. Yan, S. Li, K. Liu, R. Zhou, W. Zhang, Z. Hao, X. Li, D. Wang, Q. Li, and X. Zeng, An image features assisted line selection method in laser-induced breakdown spectroscopy, *Anal. Chim. Acta* 1111, 139 (2020)
 111. N. Gyftokostas, E. Nanou, D. Stefanos, V. Kokkinos, C. Bouras, and S. Couris, Classification of greek olive oils from different regions by machine learning-aided laser-induced breakdown spectroscopy and absorption spectroscopy, *Molecules* 26(5), 1241 (2021)
 112. T. Zhang, C. Yan, J. Qi, H. Tang, and H. Li, Classification and discrimination of coal ash by laser-induced breakdown spectroscopy (LIBS) coupled with advanced chemometric methods, *J. Anal. At. Spectrom.* 32(10), 1960 (2017)
 113. G. Rao, L. Huang, M. Liu, T. Chen, J. Chen, Z. Luo, F. Xu, X. Xu, and M. Yao, Identification of Huanglongbing-infected navel oranges based on laser-induced breakdown spectroscopy combined with different chemometric methods, *Appl. Opt.* 57(29), 8738 (2018)
 114. S. Lu, M. Dong, J. Huang, W. Li, J. Lu, and J. Li, Estimation of the aging grade of T91 steel by laser-induced breakdown spectroscopy coupled with support vector machines, *Spectrochim. Acta B* 140, 35 (2018)
 115. S. Müller and J. A. Meima, Mineral classification of lithium-bearing pegmatites based on laser-induced breakdown spectroscopy: Application of semi-supervised learning to detect known minerals and unknown material, *Spectrochim. Acta B* 189, 106370 (2022)
 116. L. M. Narla and S. V. Rao, Identification of metals and alloys using color CCD images of laser-induced breakdown emissions coupled with machine learning, *Appl. Phys. B* 126(6), 113 (2020)
 117. M. Yelameli, B. Thornton, T. Takahashi, T. Weer-

- akoon, and K. Ishii, Classification and statistical analysis of hydrothermal seafloor rocks measured underwater using laser-induced breakdown spectroscopy, *J. Chemometr.* 33(2), e3092 (2019)
118. Q. Zeng, G. Chen, W. Li, Z. Li, J. Tong, M. Yuan, B. Wang, H. Ma, Y. Liu, L. Guo, and H. Yu, Classification of steel based on laser-induced breakdown spectroscopy combined with restricted Boltzmann machine and support vector machine, *Plasma Sci. Technol.* 24(8), 084009 (2022)
 119. T. Chen, L. Sun, H. Yu, W. Wang, L. Qi, P. Zhang, and P. Zeng, Deep learning with laser-induced breakdown spectroscopy (LIBS) for the classification of rocks based on elemental imaging, *Appl. Geochem.* 136, 105135 (2022)
 120. W. Hao, X. Hao, Y. Yang, X. Liu, Y. Liu, P. Sun, and R. Sun, Rapid classification of soils from different mining areas by laser-induced breakdown spectroscopy (LIBS) coupled with a PCA-based convolutional neural network, *J. Anal. At. Spectrom.* 36(11), 2509 (2021)
 121. W. H. Yan, X. Y. Yang, X. Geng, L. S. Wang, L. Lu, Y. Tian, Y. Li, and H. Lin, Rapid identification of fish products using handheld laser induced breakdown spectroscopy combined with random forest, *Spectroscopy & Spectral Anal.* 42(12), 3714 (2022) (in Chinese)
 122. Y. W. Chu, S. S. Tang, S. X. Ma, Y. Y. Ma, Z. Q. Hao, Y. M. Guo, L. B. Guo, Y. F. Lu, and X. Y. Zeng, Accuracy and stability improvement for meat species identification using multiplicative scatter correction and laser-induced breakdown spectroscopy, *Opt. Express* 26(8), 10119 (2018)
 123. M. Guo, R. Zhu, L. Zhang, R. Zhang, G. Huang, and H. Duan, Quantitative detection of chromium pollution in biochar based on matrix effect classification regression model, *Molecules* 26(7), 2069 (2021)
 124. V. C. Costa, F. W. B. Aquino, C. M. Paranhos, and E. R. Pereira-Filho, Identification and classification of polymer e-waste using laser-induced breakdown spectroscopy (LIBS) and chemometric tools, *Polym. Test.* 59, 390 (2017)
 125. Y. Chen, Y. Liu, B. Han, W. Yu, and E. Wan, Identification of writing marks from pencil lead through machine learning based on laser-induced breakdown spectroscopy, *Optik (Stuttg.)* 259, 169008 (2022)
 126. K. Liu, D. Tian, X. Deng, H. Wang, and G. Yang, Rapid classification of plastic bottles by laser-induced breakdown spectroscopy (LIBS) coupled with partial least squares discrimination analysis based on spectral windows (SW-PLS-DA), *J. Anal. At. Spectrom.* 34(8), 1665 (2019)
 127. Z. A. Abdel-Salam, S. A. M. Abdel-Salam, I. I. Abdel-Mageed, and M. A. Harith, Evaluation of proteins in sheep colostrum via laser-induced breakdown spectroscopy and multivariate analysis, *J. Adv. Res.* 15, 19 (2019)
 128. W. Yu, Z. Sun, and Y. Liu, Rapid detection and identification of objects using a self-designed methodology based on LIBS and PCA-DVSM – taking rosewood for example, *Optik (Stuttg.)* 248, 168069 (2021)
 129. A. K. Pathak, A. Singh, R. Kumar, and A. K. Rai, Laser-induced breakdown spectroscopy coupled with PCA study of human tooth, *Natl. Acad. Sci. Lett.* 42(1), 87 (2019)
 130. H. Song, L. Ma, E. Zhu, Y. Wang, Y. Liu, W. Sun, P. Peng, and C. Li, Plastic classification and recognition by laser-induced breakdown spectroscopy and GA-BP neural network, *Laser & Optoelectron. Prog.* 57(15), 153002 (2020) (in Chinese)
 131. P. Dong, S. Zhao, K. Zheng, J. Wang, X. Gao, Z. Hao, and J. Lin, Rapid identification of ginseng origin by laser induced breakdown spectroscopy combined with neural network and support vector machine algorithm, *Acta Phys. Sin.* 70(4), 040201 (2021) (in Chinese)
 132. X. Liu, X. Che, K. Li, X. Wang, Z. Lin, Z. Wu, and Q. Zheng, Geographical authenticity evaluation of *Mentha haplocalyx* by LIBS coupled with multivariate analyses, *Plasma Sci. Technol.* 22(7), 074006 (2020)
 133. Q. Godoi, F. O. Leme, L. C. Trevisan, E. R. Pereira Filho, I. A. Rufini, Jr Santos, and F. J. Krug, Laser-induced breakdown spectroscopy and chemometrics for classification of toys relying on toxic elements, *Spectrochim. Acta B* 66(2), 138 (2011)
 134. P. M. Mukhono, K. H. Angeyo, A. Dehayem-Kamadjeu, and K. A. Kaduki, Laser induced breakdown spectroscopy and characterization of environmental matrices utilizing multivariate chemometrics, *Spectrochim. Acta B* 87, 81 (2013)
 135. T. F. Akhmetzhanov and A. M. Popov, Direct determination of lanthanides by LIBS in REE-rich ores: Comparison between univariate and DoE based multivariate calibrations with respect to spectral resolution, *J. Anal. At. Spectrom.* 37(11), 2330 (2022)
 136. T. F. Akhmetzhanov, T. A. Labutin, D. M. Korshunov, A. A. Samsonov, and A. M. Popov, Determination of Ce and La in REE-rich ores using handheld LIBS and PLS regression, *J. Anal. At. Spectrom.* 38(10), 2134 (2023)
 137. A. P. Rao, P. R. Jenkins, D. M. Vu, J. D. Auxier II, A. K. Patnaik, and M. B. Shattan, Rapid quantitative analysis of trace elements in plutonium alloys using a handheld laser-induced breakdown spectroscopy (LIBS) device coupled with chemometrics and machine learning, *Anal. Methods* 13(30), 3368 (2021)
 138. A. P. Rao, P. R. Jenkins, J. D. II Auxier, and M. B. Shattan, Comparison of machine learning techniques to optimize the analysis of plutonium surrogate material via a portable LIBS device, *J. Anal. At. Spectrom.* 36(2), 399 (2021)
 139. Y. H. Gu, Y. Li, Y. Tian, and Y. Lu, Study on the multivariate quantitative analysis method for steel alloy elements using LIBS, *Spectroscopy & Spectral Anal.* 34(8), 2244 (2014) (in Chinese)
 140. P. Yaroshchuk, D. L. Death, and S. J. Spencer, Comparison of principal components regression, partial least squares regression, multi-block partial least squares regression, and serial partial least squares regression algorithms for the analysis of Fe in iron ore using LIBS, *J. Anal. At. Spectrom.* 27(1), 92 (2012)
 141. A. Erler, D. Riebe, T. Beitz, H. G. Löhmannsröben, and R. Gebbers, Soil nutrient detection for precision agriculture using handheld laser-induced breakdown spectroscopy (LIBS) and multivariate regression methods

- (PLSR, Lasso and GPR), *Sensors (Basel)* 20(2), 418 (2020)
142. A. P. Rao, P. R. Jenkins, M. B. Auxier, Shattan, and A. K. Patnaik, Development of advanced machine learning models for analysis of plutonium surrogate optical emission spectra, *Appl. Opt.* 61(7), D30 (2022)
143. R. J. Yuan, X. Wan, Q. He, and H. P. Wang, Research on olivine component analysis using LIBS combined with back-propagation algorithm, *Spectroscopy & Spectral Anal.* 39(12), 3861 (2019) (in Chinese)
144. Q. Shi, G. Niu, Q. Lin, T. Xu, F. Li, and Y. Duan, Quantitative analysis of sedimentary rocks using laser-induced breakdown spectroscopy: Comparison of support vector regression and partial least squares regression chemometric methods, *J. Anal. At. Spectrom.* 30(12), 2384 (2015)
145. Y. Ding, F. Yan, G. Yang, H. Chen, and Z. Song, Quantitative analysis of sinters using laser-induced breakdown spectroscopy (LIBS) coupled with kernel-based extreme learning machine (K-ELM), *Anal. Methods* 10(9), 1074 (2018)
146. S. Wu, T. Zhang, H. Tang, K. Wang, X. Yang, and H. Li, Quantitative analysis of nonmetal elements in steel using laser-induced breakdown spectroscopy combined with random forest, *Anal. Methods* 7(6), 2425 (2015)
147. L. R. Xiang, Z. H. Ma, X. Y. Zhao, F. Liu, Y. He, and L. Feng, Comparative analysis of chemometrics method on heavy metal detection in soil with laser-induced breakdown spectroscopy, *Spectroscopy & Spectral Anal.* 37(12), 3871 (2017) (in Chinese)
148. S. Ye, X. Chen, D. Dong, J. Wang, X. Wang, and F. Wang, Rapid determination of water COD using laser-induced breakdown spectroscopy coupled with partial least-squares and random forest, *Anal. Methods* 10(40), 4879 (2018)
149. T. A. Labutin, S. M. Zaytsev, A. M. Popov, and N. B. Zorov, Carbon determination in carbon-manganese steels under atmospheric conditions by laser-induced breakdown spectroscopy, *Opt. Express* 22(19), 22382 (2014)
150. L. Sun, L. Liu, M. Zhu, M. Wang, Q. Wang, X. Peng, and J. Qu, Quantitative analysis of laser-induced breakdown spectroscopy of Pb in water using particle swarm optimization algorithm, in: 2015 Optoelectronics Global Conference (OGC), Shenzhen, China, 2015, pp 29–31
151. Z. Q. Hao, C. M. Li, M. Shen, X. Y. Yang, K. H. Li, L. B. Guo, X. Y. Li, Y. F. Lu, and X. Y. Zeng, Acidity measurement of iron ore powders using laser-induced breakdown spectroscopy with partial least squares regression, *Opt. Express* 23(6), 7795 (2015)
152. P. Wang, N. Li, C. Yan, Y. Feng, Y. Ding, T. Zhang, and H. Li, Rapid quantitative analysis of the acidity of iron ore by the laser-induced breakdown spectroscopy (LIBS) technique coupled with variable importance measures-random forests (VIM-RF), *Anal. Methods* 11(27), 3419 (2019)
153. G. Yang, X. Han, C. Wang, Y. Ding, K. Liu, D. Tian, and L. Yao, The basicity analysis of sintered ore using laser-induced breakdown spectroscopy (LIBS) combined with random forest regression (RFR), *Anal. Methods* 9(36), 5365 (2017)
154. C. Lu, G. Lv, C. Shi, D. Qiu, F. Jin, M. Gu, and W. Sha, Quantitative analysis of pH value in soil using laser-induced breakdown spectroscopy coupled with a multivariate regression method, *Appl. Opt.* 59(28), 8582 (2020)
155. Y. Zhang, Z. Xiong, Y. Ma, C. Zhu, R. Zhou, X. Li, Q. Li, and Q. Zeng, Quantitative analysis of coal quality by laser-induced breakdown spectroscopy assisted with different chemometric methods, *Anal. Methods* 12(27), 3530 (2020)
156. E. Képeš, H. Saeidirozeh, V. Laitl, J. Vrábek, P. Kubelík, P. Pořízka, M. Ferus, and J. Kaiser, Interpreting neural networks trained to predict plasma temperature from optical emission spectra, *J. Anal. At. Spectrom.* 39(4), 1160 (2024)
157. H. Saeidirozeh, A. K. Myakalwar, P. Kubelík, A. Ghaderi, V. Laitl, L. Petera, P. B. Rimmer, O. Shorttle, A. N. Heays, A. Krivková, M. Krus, S. Civis, J. Yáñez, E. Képes, P. Pořízka, and M. Ferus, ANN-LIBS analysis of mixture plasmas: Detection of Xenon, *J. Anal. At. Spectrom.* 37(9), 1815 (2022)
158. N. Ahmed, J. A. Awan, K. Fatima, S. M. Z. Iqbal, M. Rafique, S. A. Abbasi, and M. A. Baig, Machine learning-based calibration LIBS analysis of aluminium-based alloys, *Eur. Phys. J. Plus* 137(6), 671 (2022)
159. K. H. Li, L. B. Guo, C. M. Li, X. Y. Li, M. Shen, Z. Zheng, Y. Yu, R. F. Hao, Z. Q. Hao, Q. D. Zeng, Y. F. Lu, and X. Y. Zeng, Analytical-performance improvement of laser-induced breakdown spectroscopy for steel using multi-spectral-line calibration with an artificial neural network, *J. Anal. At. Spectrom.* 30(7), 1623 (2015)
160. P. Yang, X. Li, and Z. Nie, Determination of the nutrient profile in plant materials using laser-induced breakdown spectroscopy with partial least squares-artificial neural network hybrid models, *Opt. Express* 28(15), 23037 (2020)
161. S. Shabbir, W. Xu, Y. Zhang, C. Sun, Z. Yue, L. Zou, F. Chen, and J. Yu, Machine learning and transfer learning for correction of the chemical and physical matrix effects in the determination of alkali and alkaline earth metals with LIBS in rocks, *Spectrochim. Acta B* 194, 106478 (2022)
162. H. Zhang, S. Wang, D. Li, Y. Zhang, J. Hu, and L. Wang, Edible gelatin diagnosis using laser-induced breakdown spectroscopy and partial least square assisted support vector machine, *Sensors (Basel)* 19(19), 4225 (2019)
163. J. Huang, M. Dong, S. Lu, W. Li, J. Lu, C. Liu, and J. H. Yoo, Estimation of the mechanical properties of steel via LIBS combined with canonical correlation analysis (CCA) and support vector regression (SVR), *J. Anal. At. Spectrom.* 33(5), 720 (2018)
164. F. P. Yu, J. J. Lin, X. M. Lin, and L. Li, Detection of C element in alloy steel by double pulse laser induced breakdown spectroscopy with a multivariable GA-BP-ANN, *Spectroscopy & Spectral Anal.* 42(1), 197 (2022) (in Chinese)
165. W. Song, Z. Hou, W. Gu, M. S. Afgan, J. Cui, H. Wang, Y. Wang, and Z. Wang, Incorporating domain

- knowledge into machine learning for laser-induced breakdown spectroscopy quantification, *Spectrochim. Acta B* 195, 106490 (2022)
166. Z. Wang, J. Feng, L. Li, W. Ni, and Z. Li, A multivariate model based on dominant factor for laser-induced breakdown spectroscopy measurements, *J. Anal. At. Spectrom.* 26(11), 2289 (2011)
 167. J. Feng, Z. Wang, L. West, Z. Li, and W. Ni, A PLS model based on dominant factor for coal analysis using laser-induced breakdown spectroscopy, *Anal. Bioanal. Chem.* 400(10), 3261 (2011)
 168. Z. Wang, J. Feng, L. Li, W. Ni, and Z. Li, A multivariate model based on dominant factor for laser-induced breakdown spectroscopy measurements, *J. Anal. At. Spectrom.* 26(11), 2289 (2011)
 169. X. Li, Z. Wang, Y. Fu, Z. Li, and W. Ni, A model combining spectrum standardization and dominant factor based partial least square method for carbon analysis in coal using laser-induced breakdown spectroscopy, *Spectrochim. Acta B* 99, 82 (2014)
 170. Z. Hou, Z. Wang, L. Li, X. Yu, T. Li, H. Yao, G. Yan, Q. Ye, Z. Liu, and H. Zheng, Fast measurement of coking properties of coal using laser induced breakdown spectroscopy, *Spectrochim. Acta B* 191, 106406 (2022)
 171. Z. Wang, J. Feng, L. Li, W. Ni, and Z. Li, A nonlinearized PLS model based on multivariate dominant factor for laser-induced breakdown spectroscopy measurements, *J. Anal. At. Spectrom.* 26(11), 2175 (2011)
 172. Y. Zhang, Y. Lu, Y. Tian, Y. Li, W. Ye, J. Guo, and R. Zheng, Quantitation improvement of underwater laser induced breakdown spectroscopy by using self-absorption correction based on plasma images, *Anal. Chim. Acta* 1195, 339423 (2022)
 173. Z. Hou, Z. Wang, T. Yuan, J. Liu, Z. Li, and W. Ni, A hybrid quantification model and its application for coal analysis using laser induced breakdown spectroscopy, *J. Anal. At. Spectrom.* 31(3), 722 (2016)
 174. J. Feng, Z. Wang, L. Li, Z. Li, and W. Ni, A nonlinearized multivariate dominant factor-based partial least squares (PLS) model for coal analysis by using laser-induced breakdown spectroscopy, *Appl. Spectrosc.* 67(3), 291 (2013)
 175. M. Dong, L. Wei, J. Lu, W. Li, S. Lu, S. Li, C. Liu, and J. H. Yoo, A comparative model combining carbon atomic and molecular emissions based on partial least squares and support vector regression correction for carbon analysis in coal using LIBS, *J. Anal. At. Spectrom.* 34(3), 480 (2019)
 176. W. Song, M. S. Afgan, Y. H. Yun, H. Wang, J. Cui, W. Gu, Z. Hou, and Z. Wang, Spectral knowledge-based regression for laser-induced breakdown spectroscopy quantitative analysis, *Expert Syst. Appl.* 205, 117756 (2022)
 177. E. Kepes, J. Vrábel, T. Brázdil, P. Holub, P. Porízká, and J. Kaiser, Interpreting convolutional neural network classifiers applied to laser-induced breakdown optical emission spectra, *Talanta* 266, 124946 (2024)
 178. A. P. Rao, P. R. Jenkins, J. D. Auxier, M. B. Shattan, and A. K. Patnaik, Analytical comparisons of handheld LIBS and XRF devices for rapid quantification of gallium in a plutonium surrogate matrix, *J. Anal. At. Spectrom.* 37(5), 1090 (2022)
 179. J. Amador-Hernández, L. E. García-Ayuso, J. M. Fernández-Romeroa, and M. D. Luque de Castro, Partial least squares regression for problem solving in precious metal analysis by laser induced breakdown spectrometry, *J. Anal. At. Spectrom.* 15, 587 (2000)
 180. D. L. Deeth, A. P. Cunningham, and L. J. Pollard, Multi-element analysis of iron ore pellets by laser-induced breakdown spectroscopy and principal components regression, *Spectrochim. Acta Part B* 63(7), 763 (2008)
 181. S. M. Zaytsev, A. M. Popov, E. V. Chernykh, R. D. Voronina, N. B. Zorov, and T. A. Labutin, Comparison of single- and multivariate calibration for determination of Si, Mn, Cr and Ni in high-alloyed stainless steels by laser-induced breakdown spectrometry, *J. Anal. At. Spectrom.* 29(8), 1417 (2014)
 182. S. M. Zaytsev, I. N. Krylov, A. M. Popov, N. B. Zorov, and T. A. Labutin, Accuracy enhancement of a multivariate calibration for lead determination in soils by laser induced breakdown spectroscopy, *Spectrochim. Acta B* 140, 65 (2018)
 183. Y. C. Huang, S. S. Harilal, A. Bais, and A. E. Hussein, Progress toward machine learning methodologies for laser-induced breakdown spectroscopy with an emphasis on soil analysis, *IEEE Trans. Plasma Sci.* 51(7), 1729 (2023)
 184. N. Rethfeldt, P. Brinkmann, D. Riebe, T. Beitz, N. Köllner, U. Altenberger, and H. G. Löhmannsröben, Detection of rare earth elements in minerals and soils by laser-induced breakdown spectroscopy (LIBS) using interval PLS, *Minerals (Basel)* 11(12), 1379 (2021)
 185. C. R. Bhatt, F. Y. Yueh, and J. P. Singh, Univariate and multivariate analyses of rare earth elements by laser-induced breakdown spectroscopy, *Appl. Opt.* 56(8), 2280 (2017)
 186. E. H. Kwapis, J. Borrero, K. S. Latty, H. B. Andrews, S. Phongikaroon, and K. C. Hartig, Laser ablation plasmas and spectroscopy for nuclear applications, *Appl. Spectrosc.* 78(1), 9 (2024)
 187. C. Sun, W. J. Xu, Y. Q. Tan, Y. Q. Zhang, Z. Q. Yue, L. Zou, S. Shabbir, M. T. Wu, F. Y. Chen, and J. Yu, From machine learning to transfer learning in laser-induced breakdown spectroscopy analysis of rocks for Mars exploration, *Sci. Rep.* 11(1), 21379 (2021)
 188. V. Motto-Ros, A. S. Koujelev, G. R. Osinski, and A. E. Dudelzak, Quantitative multi-elemental laser-induced breakdown spectroscopy using artificial neural networks, *J. Eur. Opt. Soc. Rapid Publ.* 3, 08011 (2008)
 189. J. B. Sirven, B. Bousquet, L. Canioni, and L. Sarger, Laser-induced breakdown spectroscopy of composite Samples: Comparison of advanced chemometrics methods, *Anal. Chem.* 78(5), 1462 (2006)
 190. X. B. Xu, F. Ma, J. M. Zhou, and C. W. Du, Applying convolutional neural networks (CNN) for end-to-end soil analysis based on laser-induced breakdown spectroscopy (LIBS) with less spectral preprocessing, *Comput. Electron. Agric.* 199, 107171 (2022)
 191. J. H. Yang, C. C. Yi, J. W. Xu, and X. H. Ma, A laser



- induced breakdown spectroscopy quantitative analysis method based on the robust least squares support vector machine regression model, *J. Anal. At. Spectrom.* 30(7), 1541 (2015)
192. Z. Yang, B. Tang, and Y. Qiu, Measurement of transient temperature using laser induced breakdown spectroscopy (LIBS) with the surface temperature effect, *J. Anal. At. Spectrom.*, (10), 38 (2023)
 193. L. Sun, H. Yu, Z. Cong, Y. Xin, Y. Li, and L. Qi, Yang, In situ analysis of steel melt by double-pulse laser-induced breakdown spectroscopy with a Cassegrain telescope, *Spectrochim. Acta B* 112, 40 (2015)
 194. L. X. Sun, Y. Xin, Z. B. Cong, Y. Li, and L. F. Qi, Online compositional analysis of molten steel by laser-induced breakdown spectroscopy, *Adv. Mater. Res.* 694–697, 1260 (2013)
 195. Y. Lee, R. I. Foster, H. Kim, and S. Choi, Machine learning-assisted laser-induced breakdown spectroscopy for monitoring molten salt compositions of small modular reactor fuel under varying laser focus positions, *Anal. Chim. Acta* 1241, 340804 (2023)
 196. P. J. Gasda, R. B. Anderson, A. Cousin, O. Forni, S. M. Clegg, A. Ollila, N. Lanza, J. Frydenvang, S. Lamm, R. C. Wiens, S. Maurice, O. Gasnault, R. Beal, A. Reyes-Newell, and D. Delapp, Quantification of manganese for ChemCam Mars and laboratory spectra using a multivariate model, *Spectrochim. Acta B* 181, 106223 (2021)
 197. F. Yang, L. N. Li, W. M. Xu, X. F. Liu, Z. C. Cui, L. C. Jia, Y. Liu, J. H. Xu, Y. W. Chen, X. S. Xu, J. Y. Wang, H. Qi, and R. Shu, Laser-induced breakdown spectroscopy combined with a convolutional neural network: A promising methodology for geochemical sample identification in Tianwen-1 Mars mission, *Spectrochim. Acta B* 192, 106417 (2022)
 198. T. Takahashi, B. Thornton, T. Sato, T. Ohki, K. Ohki, and T. Sakka, Temperature based segmentation for spectral data of laser-induced plasmas for quantitative compositional analysis of brass alloys submerged in water, *Spectrochim. Acta B* 124, 87 (2016)
 199. T. Takahashi, B. Thornton, T. Sato, T. Ohki, K. Ohki, and T. Sakka, Partial least squares regression calculation for quantitative analysis of metals submerged in water measured using laser-induced breakdown spectroscopy, *Appl. Opt.* 57(20), 5872 (2018)
 200. C. Liu, J. Guo, Y. Tian, C. Zhang, K. Cheng, W. Ye, and R. Zheng, Development and field tests of a deep-sea laser-induced breakdown spectroscopy (LIBS) system for solid sample analysis in seawater, *Sensors (Basel)* 20(24), 7341 (2020)
 201. A. Li, X. Zhang, X. Liu, Y. He, Y. Shan, H. Sun, W. Yi, and R. Liu, Real time and high-precision online determination of main components in iron ore using spectral refinement algorithm based LIBS, *Opt. Express* 31(23), 38728 (2023)
 202. P. Lu, Z. Zhuo, W. Zhang, J. Tang, Y. Wang, H. Zhou, X. Huang, T. Sun, and J. Lu, A hybrid feature selection combining wavelet transform for quantitative analysis of heat value of coal using laser-induced breakdown spectroscopy, *Appl. Phys. B* 127(2), 19 (2021)
 203. Y. Yang, P. Wang, and C. Ma, Quantitative analysis of Mn element in liquid steel by LIBS based on particle swarm optimized support vector machine, *Laser & Optoelectron. Prog.* 52(7), 073004–1 (2015) (in Chinese)
 204. Y. Huang, J. Lin, X. Lin, and W. Zheng, Quantitative analysis of Cr in soil based on variable selection coupled with multivariate regression using laser-induced breakdown spectroscopy, *J. Anal. At. Spectrom.* 36(11), 2553 (2021)
 205. Y. M. Guo, L. B. Guo, Z. Q. Hao, Y. Tang, S. X. Ma, Q. D. Zeng, S. S. Tang, X. Y. Li, Y. F. Lu, and X. Y. Zeng, Accuracy improvement of iron ore analysis using laser-induced breakdown spectroscopy with a hybrid sparse partial least squares and least-squares support vector machine model, *J. Anal. At. Spectrom.* 33(8), 1330 (2018)
 206. Z. Wang, L. Li, L. West, Z. Li, and W. Ni, A spectrum standardization approach for laser-induced breakdown spectroscopy measurements, *Spectrochim. Acta B* 68, 58 (2012)
 207. H. Li, M. Huang, and H. Xu, High accuracy determination of copper in copper concentrate with double genetic algorithm and partial least square in laser-induced breakdown spectroscopy, *Opt. Express* 28(2), 2142 (2020)
 208. F. Chang, J. Yang, H. Lu, and H. Li, Repeatability enhancing method for one-shot LIBS analysis via spectral intensity correction based on probability distribution, *J. Anal. At. Spectrom.* 36(8), 1712 (2021)
 209. T. B. Chen, M. H. Liu, L. Huang, H. M. Zhou, C. H. Wang, H. Yang, H. Q. Hu, and M. Y. Yao, Effects of different pretreatment method on laser-induced breakdown spectroscopy measurement of Pb in pork, *Chin. J. Anal. Chem.* 44(7), 1029 (2016)
 210. H. Q. Hu, X. H. Xu, M. H. Liu, J. P. Tu, L. Huang, L. Huang, M. Y. Yao, T. B. Chen, and P. Yang, Determination of Cu in shell of preserved egg by LIBS coupled with PLS, *Spectroscopy & Spectral Anal.* 35(12), 3500 (2015) (in Chinese)
 211. W. B. Li, L. T. Yao, M. H. Liu, L. Huang, M. Y. Yao, T. B. Chen, X. W. He, P. Yang, H. Q. Hu, and J. H. Nie, Influence of spectral pre-processing on PLS quantitative model of detecting Cu in navel orange by LIBS, *Spectroscopy & Spectral Anal.* 35(5), 1392 (2015) (in Chinese)
 212. H. Yang, L. Huang, M. Liu, T. Chen, C. Wang, H. Hu, and M. Yao, Detection of Pb in navel orange by peel laser induced breakdown spectroscopy coupled with PLS, *Chin. J. Anal. Lab.* 35(7), 760 (2016)
 213. K. Ke, Y. Lu, and C. c. Yi, Improvement of convex optimization baseline correction in laser-induced breakdown spectral quantitative analysis, *Spectroscopy & Spectral Anal.* 38(7), 2256 (2018) (in Chinese)
 214. C. Yi, Y. Lv, H. Xiao, K. Ke, and X. Yu, A novel baseline correction method using convex optimization framework in laser-induced breakdown spectroscopy quantitative analysis, *Spectrochim. Acta B* 138, 72 (2017)
 215. P. Yaroshchuk, and J. E. Eberhardt, Automatic

- correction of continuum background in Laser-induced Breakdown Spectroscopy using a model-free algorithm, *Spectrochim. Acta B* 99, 138 (2014)
216. S. Yoon, J. Choi, S. J. Moon, and J. H. Choi, Determination and quantification of heavy metals in sediments through laser-induced breakdown spectroscopy and partial least squares regression, *Appl. Sci. (Basel)* 11(15), 7154 (2021)
 217. C. Ma and J. Cui, Quantitative analysis of composition in molten steel by LIBS based on improved partial least squares, *Laser Technology* 40(6), 876 (2016) (in Chinese)
 218. Z. Zhu, J. Li, Y. Guo, X. Cheng, Y. Tang, L. Guo, X. Li, Y. Lu, and X. Zeng, Accuracy improvement of boron by molecular emission with a genetic algorithm and partial least squares regression model in laser-induced breakdown spectroscopy, *J. Anal. At. Spectrom.* 33(2), 205 (2018)
 219. S. Ye, Y. H. Gu, X. F. Du, W. T. Zhang, J. J. Wang, X. Q. Wang, and D. M. Dong, Chemometrics method for real-time measurement of water COD based on laser-induced breakdown spectroscopy, *Spectroscopy & Spectral Anal.* 37(11), 3585 (2017) (in Chinese)
 220. M. Li, H. Fu, Y. Du, X. Huang, T. Zhang, H. Tang, and H. Li, Laser induced breakdown spectroscopy combined with hybrid variable selection for the prediction of the environmental risk Nemerow index of heavy metals in oily sludge, *J. Anal. At. Spectrom.* 37(5), 1099 (2022)
 221. J. He, C. Pan, Y. Liu, and X. Du, Quantitative analysis of carbon with laser-induced breakdown spectroscopy (LIBS) using genetic algorithm and back propagation neural network models, *Appl. Spectrosc.* 73(6), 678 (2019)
 222. J. Chen, Q. Li, K. Liu, X. Li, B. Lu, and G. Li, Correction of moisture interference in laser-induced breakdown spectroscopy detection of coal by combining neural networks and random spectral attenuation, *J. Anal. At. Spectrom.* 37(8), 1658 (2022)
 223. D. Luarte, A. K. Myakalwar, M. Velásquez, J. Álvarez, C. Sandoval, R. Fuentes, J. Yañez, and D. Sbarbaro, Combining prior knowledge with input selection algorithms for quantitative analysis using neural networks in laser induced breakdown spectroscopy, *Anal. Methods* 13(9), 1181 (2021)
 224. L. N. Li, X. F. Liu, W. M. Xu, J. Y. Wang, and R. Shu, A laser-induced breakdown spectroscopy multi-component quantitative analytical method based on a deep convolutional neural network, *Spectrochim. Acta B* 169, 105850 (2020)
 225. C. Yan, J. Qi, J. Ma, H. Tang, T. Zhang, and H. Li, Determination of carbon and sulfur content in coal by laser induced breakdown spectroscopy combined with kernel-based extreme learning machine, *Chemom. Intell. Lab. Syst.* 167, 226 (2017)