Check for updates

TrCSL: a transferred CNN-SE-LSTM model for high-accuracy quantitative analysis of laser-induced breakdown spectroscopy with small samples†

Shengjie Ma, [iD] [abc] Shilong Xu,*[abc] Congyuan Pan,[d] Jiajie Fang,[abc] Fei Han,[abc] Yuhao Xia,[abc] Wanying Ding,[abc] Youlong Chen [iD] [abc] and Yihua Hu*[abc]

When utilizing the laser-induced breakdown spectroscopy (LIBS) technology for high-precision quantitative analysis, a substantial number of samples are typically required to construct an accurate prediction model. However, in many practical applications, obtaining sufficient samples often faces challenges. The scarcity of samples not only reduces the reliability of experiments but also limits the potential and flexibility of LIBS technology in a broader range of applications. In this study, we introduced a transferred convolutional neural network-squeeze and excitation-long short-term memory (TrCSL) model, aimed at achieving high-precision quantitative analysis even with small samples. The TrCSL model combines the strengths of transfer learning, convolutional neural networks (CNN), squeeze and excitation (SE) block mechanisms, and long short-term memory (LSTM) networks to enhance feature extraction and learning capabilities. We trained on 100 sets of steel slag samples to obtain the pre-training model, which was then transferred to the small samples and underwent fine-tuning of its parameters. Compared to the traditional partial least squares regression (PLSR) and support vector regression (SVR) algorithms, the TrCSL model shows an improvement of about 0.4 in the R^2 value for quantitative analysis results on 20 carbon steel samples. In addition, the experimental results also show that the quantitative analysis accuracy of the TrCSL model on only 20 samples is close to that of traditional PLSR and SVR algorithms on 80 samples. The TrCSL model proposed in this paper possesses enhanced universality and superior prediction accuracy, offering a novel approach to improving LIBS quantitative analysis precision with small samples.

1. Introduction

Laser-induced breakdown spectroscopy (LIBS) is an advanced analytical technique that uses spectral emissions from plasma generated by a laser to determine the element composition and detailed information of materials.[1–4] Known for its rapidity, sensitivity, and uncomplex sample preparation, LIBS is widely acknowledged as an analytical technique with significant potential for a broad range of applications. As the technology advances, LIBS has been successfully applied across various critical fields, including materials science,[5,6] environmental monitoring,[7,8] and archaeological research.[9,10]

[a]State Key Laboratory of Pulsed Power Laser Technology, National University of Defense Technology, Hefei 230037, People's Republic of China. E-mail: xushi1988@nudt.edu.cn; skl_hyh@163.com

[b]Key Laboratory of Electronic Restriction of Anhui Province, National University of Defense Technology, Hefei 230037, People's Republic of China

[c]Advanced Laser Technology Laboratory of Anhui Province, National University of Defense Technology, Hefei 230037, People's Republic of China

[d]Hefei GStar Intelligent Control Technical Co., Ltd, Hefei 230037, People's Republic of China

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d4ja00459k

LIBS quantitative analysis includes univariate and multivariate analysis,[11] with the purpose of establishing a linear correlation between sample element concentration and the intensity of characteristic spectral lines. Although univariate analysis is simple to operate, it is susceptible to matrix effects, which will severely influence the accuracy. To enhance the quantitative analysis accuracy, multivariate analysis methods have been widely applied, where principal component regression (PCR) and partial least squares regression (PLSR) are the two most commonly used techniques.[12–15] However, when the relationship between element concentration and spectral intensity is nonlinear, the effectiveness of these models may be limited. Therefore, some nonlinear models, such as support vector regression (SVR),[16,17] random forest (RF),[18,19] decision tree (DT),[20] extreme learning machine (ELM),[21,22] and deep learning,[23–27] have been developed and applied to the quantitative analysis of LIBS spectra.

In recent years, some deep learning and neural network models have been widely applied in quantitative analysis for LIBS spectra. Li *et al.* proposed a multi-component quantitative analysis method for LIBS based on a deep convolutional neural network (CNN).[25] They trained the model using more than 1400 LIBS spectra. The results were compared with the back-propagation neural network (BPNN) and PLSR, and they

1810 | *J. Anal. At. Spectrom.*, 2025, **40**, 1810–1820

This journal is © The Royal Society of Chemistry 2025

concluded that the CNN-based method is a promising LIBS quantitative analysis tool with good accuracy and high efficiency. Song *et al.* introduced a spectral knowledge-based regression (SKR) model, which integrates the advantages of knowledge-driven linear models and data-driven nonlinear models to enhance the quantitative analysis accuracy.[28] The results demonstrate that the SKR model inherits the high precision of nonlinear models as well as the simplicity and interpretability of linear models. Consequently, it can significantly improve the accuracy and reliability of LIBS quantitative analysis. Ding *et al.* used the variable importance-based long short-term memory (LSTM) model to conduct quantitative analysis of Ca in infant formula powder,[27] which could obtain higher quantitative analysis accuracy with fewer samples.

Numerous studies have indicated that the quantitative analysis accuracy of data-driven models depends on the number of samples. In general, to improve the precision and robustness of the model, researchers tend to select a large number of samples for training. Some models with outstanding performance even require hundreds to thousands of samples. For instance, in ref. 29, Zhang selected 550 samples for carbon, ash and caloric value prediction. Hou *et al.* established a PLSR prediction model using 189 coal samples.[30] Li *et al.* constructed a principal component analysis-partial least squares (PCA-PLS) regression prediction model based on 100 sets of standard air-dried base coal samples.[31] Hou *et al.* built a prediction model for coal using 77 samples, and the error is less than 1%.[32] Zhang *et al.* conducted a partial least squares regression model with 101 coal samples, and the root mean square error (RMSE) of the volatiles prediction results was reduced to less than 1%.[33] The aforementioned results indicate that when the number of samples is sufficient, the accuracy and reliability of LIBS quantitative analysis can be significantly enhanced. However, when faced with small samples (less than 20 samples[31]), although it can reduce the calculation burden, a smaller number of samples also means that we can obtain less information, and it may even be insufficient to support model training, leading to poor interpretability and quantitative analysis accuracy. Therefore, improving the accuracy of LIBS quantitative analysis under small sample conditions is an important challenge in this field, which urgently requires in-depth investigation.

For LIBS quantitative analysis with small samples, Ma *et al.* proposed a three-layer stacking model for Al–Cu–Mg–Fe–Ni alloys.[34] For 15 alloy samples, the recognition accuracy of the stacking model reached 96.47%, which greatly exceeded the 71.76% accuracy of the Random Forest algorithm. The model has achieved good results in qualitative analysis, but its applicability to LIBS quantitative analysis has not been discussed. Liu *et al.* developed an artificial neural network (ANN) model trained using the Gaussian negative log-likelihood (GLL) function.[35] This method achieved quantitative accuracies of 0.9877, 0.9939, 0.9876, and 0.9899 for four elements (Mn, Mo, Cr, and Cu) in only six stainless steel samples. However, the problem of model overfitting that may arise from such a small number of samples is not discussed. Li *et al.* proposed a small-sample model for high-precision quantitative analysis of coal.[31] They only extract partial spectra of the samples for model training instead of calculating the average spectra, which allows for the

repeated extraction of random data from the samples. In this way they can form a rich batch of training data to prevent model overfitting. When the number of samples was 15, the average relative error (ARE) of carbon element prediction results using the new model was less than 4%. The aforementioned methods can effectively prevent model overfitting, but their applicability for samples with different matrices has not been discussed.

In recent years, with the rapid development of deep learning, transfer learning has shown tremendous potential and advantages in dealing with small-sample problems.[36] By referring to models that have been pre-trained on large-scale datasets, it is possible to construct high-performance models even in situations where data are scarce. Compared to traditional machine learning methods, the transfer learning method can significantly reduce the dependence on large amounts of data. This method not only improves the generalization ability of the model but also accelerates the training process and significantly reduces the risk of overfitting. This advantage is particularly evident in fields where data acquisition is challenging, ensuring that the model can be effectively applied in these demanding environments. Transfer learning is currently being successfully applied in various domains, including text sentiment analysis,[37] image classification,[38] human activity recognition,[39] and so on. Those successful applications provide a potential possibility for the use of transfer learning in high-precision quantitative analysis of LIBS with small samples.

In this paper, we combine the multiple advantages of transfer learning, CNN, the squeeze-and-excitation (SE) attention mechanism and LSTM and propose a transferred CNN-SE-LSTM (TrCSL) model to achieve high-precision quantitative analysis of LIBS with small samples. We initially trained on 100 sets of steel slag samples to construct the source model. During this process, the previous layers of the source model were able to capture more universal features, which is crucial for enhancing the overall performance of the transfer learning model. Subsequently, we transferred the source model to the target dataset and fine-tuned the parameters. Finally, we analysed the performance of the proposed TrCSL model and discussed the impact of model hyperparameters, spectra baseline, and spectral fluctuations on the quantitative analysis results. In addition, we compared it with traditional methods, and the experimental results showed that the TrCSL model, with only 20 sets of carbon steel samples, can achieve quantitative analysis results comparable to the traditional PLSR and PSO-SVR algorithms on 80 sets of samples. This comparison not only highlights the superior performance of the TrCSL model in handling small-sample data but also demonstrates its great potential in the field of quantitative analysis. The significant advantage of the TrCSL model undoubtedly provides a new solution for quantitative analysis under small-sample conditions.

2. Experimental setup and model establishment

2.1 The TrCSL model for small samples

In this paper, we concentrate on enhancing the LIBS quantitative analysis accuracy with small samples. The main challenge is the potential insufficiency of sample numbers to meet the

This journal is © The Royal Society of Chemistry 2025

J. Anal. At. Spectrom., 2025, **40**, 1810–1820 | **1811**

training requirements, which may affect the robustness and reliability of the model. To overcome this challenge, we proposed the TrCSL model to enhance the quantitative analysis accuracy with small samples. We selected 100 sets of steel slag samples as the source data to build the CNN-SE-LSTM pre-training model. Subsequently, we transferred the model to the target data (carbon steel samples) and fine-tuned the model parameters using the target data. Finally, we conducted the performance evaluation of the TrCSL model. As shown in Fig. 1, the process of using the TrCSL model for LIBS quantitative analysis with small samples mainly includes four steps: data preparation, pre-training of the CNN-SE-LSTM model, model transfer, and performance evaluation.

Step 1: data preparation. Initially, we collect LIBS spectral data for the steel slag (the source data) and carbon steel (the target data) samples. Considering the possible influence of the matrix effect between these two types of samples, we perform baseline correction on the LIBS spectra to minimize the influence of baseline on model performance. Subsequently, the baseline corrected spectral data for both steel slag and carbon steel samples are randomly divided into the training and test sets with a ratio of 7 : 3.

Step 2: construction of the CNN-SE-LSTM pre-training model. CNN-SE-LSTM is a deep learning model that integrates the CNN, the SE block, and LSTM. In the hybrid model, the CNN extracts the feature from the input data, the SE block enhances the expressiveness of these features, and the LSTM receives the serialized features to ultimately predict the element concentration. The detailed information about the CNN-SE-LSTM model is provided in Section 2.3.

Step 3: model transfer. When the structure of the pre-trained model is determined, we inherit the network structure and parameters of the source model into the target model and fine-tune the model using the training set of the target data. In this way, the target model can not only inherit the rich knowledge and experience of the source model in feature extraction but

also adapt to the specific requirements of the target task through fine-tuning.

Step 4: performance evaluation. In the final step, it is necessary to evaluate the quantitative analysis accuracy of the TrCSL model and compare it with traditional machine learning methods to demonstrate its superiority in small-sample quantitative analysis.

2.2 Sample preparation and spectra collection

The powder samples we used in this paper are shown in Fig. S1(a).† Laser irradiation on powder samples tends to cause sample splashing, which is not conducive for testing. Therefore, it is necessary to first pelletize the samples. The pellet press is shown in Fig. S1(b),† which is produced by Shanghai Xinnuo Instrument Equipment Co. Ltd, and the instrument number is ZYP-40TS. During the pelletizing process, we first took 5 g of the sample and 9 g of boric acid and placed them into the mold. Then, we pressed them under a pressure of 30 tons for one minute and finally demolded them to obtain circular samples with a diameter of about 40 mm and a thickness of about 5 mm, as shown in Fig. S1(c).† It should be noted that each time the samples are prepared, the mold must be cleaned with alcohol to reduce the residual interference between different samples.

Subsequently, we conducted tests on the pelletized samples. All LIBS spectral data used in this paper were provided by Hefei GStar Intelligent Control Technical Co., Ltd. In our experiments, a three-channel spectrometer was used, and a total of 16 375 spectral data points were collected for each sample in the range from 216 nm to 942 nm. To reduce the potential impact of sample surface heterogeneity on the experimental results, we placed each sample on an electric rotating 3D stage to allow the laser to irradiate different areas of the sample and generate plasma. In this way, we collected 300 sets of independent spectral data for each sample. More information about the experimental equipment can be found on the website at **https://www.goldstar-china.com/**.

2.3 The CNN-SE-LSTM pre-training model

As shown in Fig. 2, the CNN-SE-LSTM model is primarily composed of four components: the input part, the data processing part, the feature extraction part, and the output part. We take the LIBS spectral data as input of the model, which then undergoes preliminary data processing before feature extraction. In the data processing part, different operations are performed based on the input dimensions. For one-dimensional LIBS spectral data, we conduct variable selection to reduce its dimensionality. For two-dimensional LIBS spectral data, we perform a conversion from 1D to 2D. The feature extraction part is constituted by a hybrid model of CNN, SE, and LSTM, and here we only display the feature extraction part of 2D CNN-SE-LSTM. The output part is composed of a dropout layer and two fully connected (FC) layers.

2.3.1 The data processing part. First, the original LIBS spectra should be normalized so that the spectral intensity is in the range of [0,1]. Considering the concentration differences among samples and the influence of matrix effects, we chose
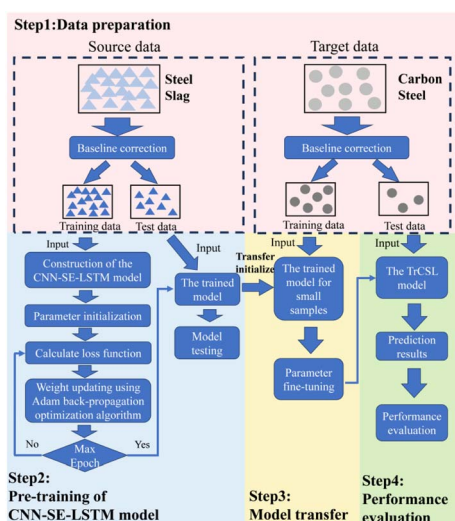

Fig. 1 Procedure of the TrCSL model for LIBS quantitative analysis with a small sample.
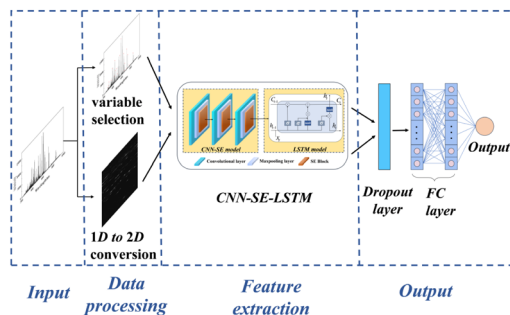
1812 | J. Anal. At. Spectrom., 2025, 40, 1810–1820

This journal is © The Royal Society of Chemistry 2025

Fig. 2 The structure of the CNN-SE-LSTM pre-training model.

Fig. 4 The feature extraction part of the CNN-SE-LSTM model.

the min–max normalization method[40] to normalize each spectrum separately, and the principle of this method can be expressed by eqn (1)

$$I_n = \frac{I_{raw} - I_{min}}{I_{max} - I_{min}},\tag{1}$$

where I_n is the normalized spectra, I_{raw} is the raw spectra, and I_{max} and I_{min} are the maximum and minimum intensity on the spectra. After the normalization of raw LIBS data, we perform different preprocessing operations according to the input dimension.

For 1D LIBS spectra, we collected a total of 16 375 sampling points in the wavelength range from 216.589 nm to 942.568 nm. Utilizing the entire unprocessed LIBS spectral data as input poses a significant challenge for both model training and computational performance. Therefore, to reduce the complexity and expedite the training process for 1D LIBS data, it is necessary to perform variable selection and dimensionality reduction on the original data before feature extraction. The common dimensionality reduction method is the PCA method. In this study, by integrating the elemental and concentration information shown in Tables S1 and S2† and referencing the atomic spectrum database of the National Institute of Standards and Technology (NIST), we have selected the characteristic spectral lines shown in Table S3† along with 5 data points surrounding each feature spectral line as the input of the 1D CNN-SE-LSTM model.

For 2D LIBS spectra, we perform a conversion from 1D to 2D. As depicted in Fig. 3, we reorganize the 16 375 sampling points into a new 125×131 matrix, forming a new LIBS spectra matrix, which serves as the input for the CNN-SE-LSTM model.

2.3.2 The feature extraction part. Taking 2D CNN-SE-LSTM as an example, as shown in Fig. 4, the feature extraction part mainly consists of three components: the CNN, the SE block,
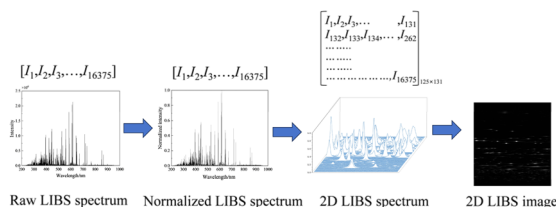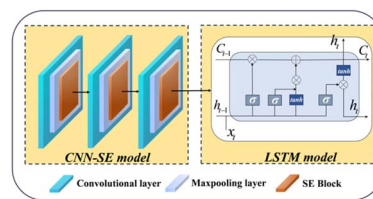
and the LSTM, where the CNN is responsible for extracting features from the input data, the SE module further enhances the expressiveness of these features, and the LSTM receives these serialized features to make accurate predictions of the element concentration.

2.3.2.1 CNN. CNN was proposed by Yann LeCun from New York University in 1998 and is essentially a type of multi-layer perceptron.[41] By employing local connections and weight sharing, CNN significantly reduces the number of weights, which in turn lowers the complexity of the model. This approach not only mitigates the risk of overfitting but also makes the network more amenable to optimization. A notable feature and advantage of CNN is their ability to deeply extract features through convolutional kernels, while pooling layers further assist the neural network in quickly extracting feature values, effectively reducing computational requirements. As a result, CNN has a significant advantage in the field of image processing. Here, we take the 2D CNN as an example to further introduce its principle and application.

2.3.2.1.1 Convolutional layer. The convolutional layer is the cornerstone of CNN, where feature extraction is achieved through convolutional operations. These operations involve the interaction between two matrices: one representing the input data and the other being the filter matrix for feature extraction, known as the convolutional kernel. These small-sized convolution kernels slide sequentially over the input data, extracting local features through dot product operations. The principle of convolution calculation is shown in Fig. 5(a). If the input size of CNN is $W_{in} \times H_{in} \times D_{in}$, the convolution kernel size is $w \times h$, the number is k, the stride is s, and the padding is p, then the output size of CNN is
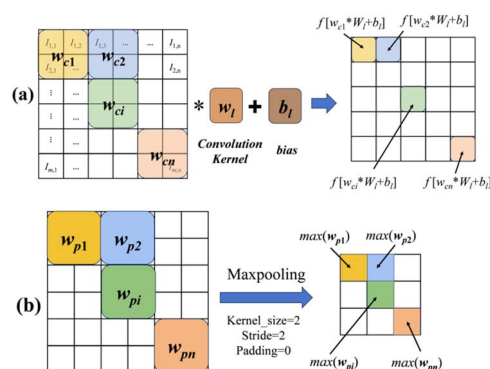

Fig. 3 The conversion of 1D LIBS spectra to 2D.

Fig. 5 (a) The convolution layer and (b) the max pooling layer.

This journal is © The Royal Society of Chemistry 2025

J. Anal. At. Spectrom., 2025, **40**, 1810–1820 | 1813

$$W_{out} = (W_{in} + 2p - w)/s + 1$$
$$H_{out} = (H_{in} + 2p - h)/s + 1, \quad (2)$$
$$W_{out} = k$$

Thus, the outcome of the convolutional operation is related to the size and the number of convolutional kernels, as well as the stride and padding. More information about these parameters can be found in the ESI.†

2.3.2.1.2 Pooling layer. The convolutional layer is followed by a pooling layer, whose function is to further reduce the size of the feature matrix and the number of parameters, while retaining the main information to reduce the risk of overfitting. Common pooling methods include max pooling and average pooling. Taking the max pooling as an example, as shown in Fig. 5(b), when performing max pooling on a 2×2 feature matrix, only the maximum value within the matrix is retained. The size of the result is reduced to a quarter of its original dimensions. On the one hand, the pooling operation can filter out unnecessary information; on the other hand, it can effectively reduce the dimension of the feature matrix and speed up the training of the model. However, we should also note that for a larger feature matrix, pooling operations can result in a sharp reduction in size and sometimes even the loss of some important information.

2.3.2.2 The SE block. To further extract features from LIBS spectra and enhance the accuracy of quantitative analysis, we have introduced the SE block after each "convolution + pooling" structure. The introduction of this mechanism adds a channel attention function, which assigns differentiated weights to different feature channels to enhance or suppress the features, thereby more efficiently extracting useful information and optimizing the performance of the model. The process of the SE block includes two main steps: squeeze and excitation. The structure of the SE block is shown in Fig. 6.

2.3.2.2.1 Squeeze. During the operation of the SE block, we first employ global average pooling to compress the spatial dimensions of each channel's feature map, thereby generating a global feature vector. This vector is capable of accurately extracting the global information across channels. The construction of such a global feature vector provides crucial global awareness for the subsequent optimization of the network, thereby enhancing the understanding and representation of overall features.

2.3.2.2.2 Excitation. We utilize two FC layers to calculate the weights for each channel. The first FC layer is dedicated to

dimensionality reduction of the feature vector to extract the main features. Subsequently, the second FC layer maps these features back to the original number of channels. During this process, the rectified linear unit (ReLU) activation function is first employed to introduce nonlinearity, followed by the sigmoid function, which generates weight values between 0 and 1. These weights quantify the contribution of each channel to the final task.

Through these two steps, the SE block is able to adaptively adjust the contribution of each channel in the network, thereby enhancing the model's ability to extract the main information.

2.3.2.3 LSTM. LSTM has been innovatively improved upon the foundation of recurrent neural networks (RNN), effectively addressing the common issue of gradient explosion in RNN. By incorporating a special type of memory cell, LSTM is capable of processing sequential data more efficiently without losing long-term dependencies. As shown in Fig. S2,† the structure of LSTM is relatively simple and straightforward compared to the complex architecture of CNN. A detailed description of the LSTM module can be found in the ESI.†

2.3.3 The output module. The output module is composed of a dropout layer and two fully connected layers. As shown in Fig. 2, we have added a dropout layer after the feature extraction part, which is a regularization technique proposed to prevent overfitting. During the forward propagation of the network's training stage, the dropout layer randomly sets the output of certain neurons to 0 with a certain probability p. This mechanism reduces the interdependencies between neurons and introduces regularization mechanisms, thereby reducing the risk of overfitting.

The fully connected layer typically resides at the end of a neural network model and is responsible for transforming all feature matrices from the dropout layer into a 1D feature vector. In this layer, each neuron is connected to all neurons, achieving comprehensive information integration. However, in cases where the amount of data is small, the fully connected layer may also lead to overfitting. To enhance the model's generalization ability, we combine the fully connected layer with the dropout technique in this paper.

Finally, according to the previous introduction, there are numerous parameters that need to be set in advance in the CNN-SE-LSTM model, which will largely determine the prediction performance. The parameters related to the neural network itself include the convolutional kernel size and number, the activation function, the type of pooling, the layer structure and number, the number of hidden layers, and the Dropout rate. The parameters related to neural network training are batch size, learning rate, epoch, *etc.* Details of these parameters can be found in the ESI,† and the effects of different parameters on the model performance are analysed in detail in Section 3.2.

2.4 Model transfer

Subsequently, we transferred the pre-trained CNN-SE-LSTM model to the target dataset and fine-tuned it. The overall strategy for model transfer is shown in Fig. 7. We first froze the parameters and weights of the feature extraction layers and
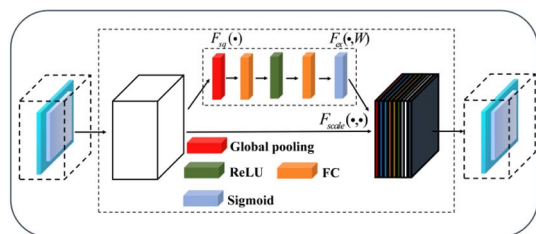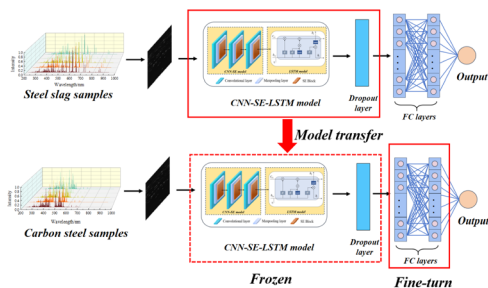

Fig. 6 The structure of the SE block.

1814 | *J. Anal. At. Spectrom.*, 2025, **40**, 1810–1820

This journal is © The Royal Society of Chemistry 2025

Fig. 7 The process of model transfer.

Fig. 8 Normalized spectra intensity of the first 5 steel slag and carbon steel samples. (a) The raw LIBS spectra of steel slag samples; (b) the raw LIBS spectra of carbon steel samples; (c) the baseline corrected LIBS spectra of steel slag samples; (d) the baseline corrected LIBS spectra of steel slag samples.

Dropout layers of the pre-trained CNN-SE-LSTM model and transferred them to the target samples. Then, we trained the transferred model with 20 sets of carbon steel samples, during which only the parameters in the fully connected layers were fine-tuned, and ultimately obtained the structure and parameters of the TrCSL model.

2.5   Performance evaluation

In order to better evaluate the prediction performance of the TrCSL model, the RMSE and the correlation coefficient (R^2) are used to evaluate the prediction accuracy. R^2 represents the linear relationship between the actual and prediction values. RMSE is used to measure the deviation between the actual value and the predicted value and is to evaluate the performance of the model from the perspective of statistics.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y}_i)^2}, \qquad (3)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}}, \qquad (4)$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, and $\overline{y}_i$ is the mean of y_i. N is the number of samples.

The relative standard deviation (RSD) is used to evaluate the stability of the LIBS spectra.

$$\text{RSD}(\%) = \frac{S}{\overline{y}} \times 100\% = \frac{\sqrt{\frac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{n-1}}}{\overline{y}} \times 100\%, \qquad (5)$$

where S is the standard deviation and $n = 300$ is the number of measurements.

3.   Results and discussion

3.1   LIBS spectra and preprocessing results

In this study, we utilized the experiment system described in Section 2.1 to collect LIBS spectral data from 100 steel slag samples and 20 carbon steel samples. Fig. 8(a) and (b) display the normalized spectra intensity of the first 5 sets of steel slag and carbon steel samples. It was observed that the LIBS lines of
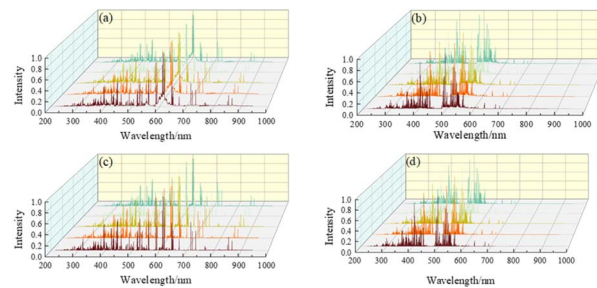
slag samples were essentially uniformly distributed across the 300–700 nm wavelength band. In contrast, the LIBS lines of the carbon steel samples were more concentrated in the 300–450 nm and 500–600 nm bands. Furthermore, the slag samples exhibited a distinct "arch-shaped" feature in the 600–650 nm band, whereas the carbon steel samples had continuous background interference in the 500–600 nm band. The presence of these baselines may lead to an overestimation of the intensity for the characteristic lines for some elements, thereby affecting the accuracy of model training. To address this issue, we employed the adaptive iteratively reweighted penalty least squares algorithm[42] for baseline correction, with a balance parameter λ set to 10^4 and the number of iterations set to 50. The baseline corrected spectra are shown in Fig. 8(c) and (d). After baseline correction, the base of the spectra is much flatter, and the "arch-shaped" feature present in the spectra of the two types of samples mentioned above has also been essentially eliminated. The above results show that the baseline correction can significantly improve the background stability of the spectra for both slag and carbon steel samples, effectively mitigating the adverse effects of matrix effects.

3.2   Training of the TrCSL model

We initially conducted a thorough analysis of the parameters for the CNN. In the SE block, the parameters for the two FC layers were fixed at 16 and 64, respectively, to optimize the extraction and transformation of features. For the LSTM, we configured a two-layer network structure with 128 hidden units in each layer to ensure that the model can extract the complex dynamics within time-series data. In terms of the optimization algorithm, we chose the adaptive moment estimation (Adam), a gradient descent method based on adaptive estimates, widely used in deep learning for its high efficiency.

First, we compared the influence of the network structure on the prediction performance of 1D and 2D CNN-SE-LSTM models. Here, we define a combination of one convolutional layer, one max pooling layer, and one SE block as one 'conv' unit, and the convolution kernel size is uniformly set to 5 × 5. Taking the Mn element as an example, Fig. 9(a) illustrates the influence of different numbers of conv units on model
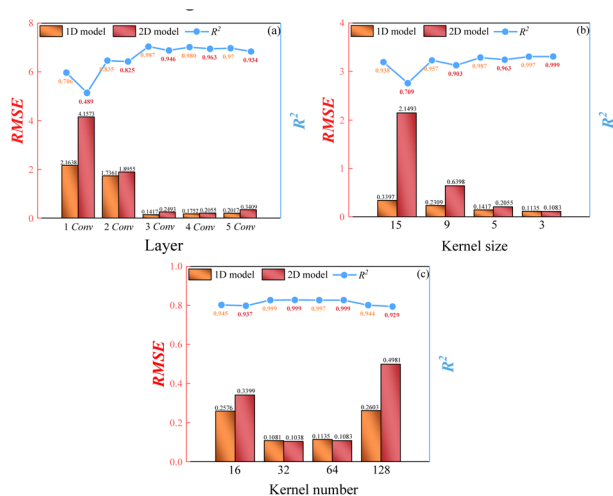
Fig. 9 Influence of different parameters on the performance of the CNN–SE–LSTM model. (a) The network structure, (b) the kernel size, and (c) the kernel number.

performance. It is clear the 1D CNN-SE-LSTM model has an advantage in prediction accuracy over the 2D CNN-SE-LSTM model. As the number of conv units increased from 1 to 5, the RMSE of the 1D CNN-SE-LSTM model decreased by 1.9935, 0.1594, 0.1076, 0.0303, and 0.1392 compared to 2D CNN-SE-LSTM, while the R^2 value increased by 0.2169, 0.0104, 0.0413, 0.0176, and 0.0360, respectively. Additionally, we found that when the number of conv units is 3, the 1D CNN-SE-LSTM model can achieve optimal performance, and for the 2D CNN-SE-LSTM model, 4 conv units are required. However, as the number of conv units further increases, the prediction performance declines. This may be because as the number of network layers increases, so do the model parameters, making the optimization process more complex and more prone to overfitting.

Next, we analysed the impact of the convolutional kernel size and number on the performance. As shown in Fig. 9(b), the effect of different kernel sizes on model performance is clearly demonstrated. When the kernel size is 3, both the 1D and 2D CNN-SE-LSTM models can achieve the best predictive performance, with the R^2 reaching 0.9974 and 0.9985, respectively. Fig. 9(c) illustrates the impact of the convolutional kernel number on the model's prediction performance. We find that the model performs best when the kernel number is set to 32. As the number continues to increase, the prediction performance of the model shows a declining trend. The reason may be that an excessive number of convolutional kernels may lead to parameter redundancy, thereby affecting the model's convergence.

Based on the aforementioned result, the 1D CNN-SE-LSTM model appears to outperform the 2D CNN-SE-LSTM model, and this superiority may stem from the following aspects:

(1) The input for the 2D CNN-SE-LSTM model is not an image in the traditional sense. As shown in Fig. 3, when constructing 2D LIBS images, we simply spliced the spectral intensities of different bands together. This approach may lead to some

discontinuities in the 2D LIBS images, affecting the overall image quality. Moreover, in the resulting LIBS images, the wavelength information of the original data is reflected in the spatial positions of the LIBS images.

(2) Before inputting the LIBS spectra into the 1D CNN-SE-LSTM model, we manually selected characteristic spectral lines for different elements according to the NIST database, effectively filtering out a large amount of irrelevant information. In contrast, the input of the 2D model is only to convert the 1D data into a 2D form, and the data itself are not filtered. Compared to the 2D CNN-SE-LSTM model, the 1D model significantly reduces the scale of the input data, and it is easier to train. The 1D model only takes 12 s, while the 2D model takes 10 min. Besides, the results of the 2D CNN-SE-LSTM model are greatly influenced by the model parameters, thus requiring more effort to adjust model parameters.

However, it is worth noting that although there is little difference in performance between the 1D and 2D models, the former is much more complex in data preprocessing. If opting for the 1D CNN-SE-LSTM model, prior variable selection must be conducted for different samples, and the characteristic spectral lines selected for each type of sample may be different. As shown in Fig. 10(b) and (e), the spectral line of the Mn element at 441.488 nm in both carbon steel and slag samples exhibits a distinct peak and is essentially unaffected by the continuous background and noise, with no interference from the characteristic spectral lines of other elements around it, thus making it suitable as input for the 1D CNN-SE-LSTM
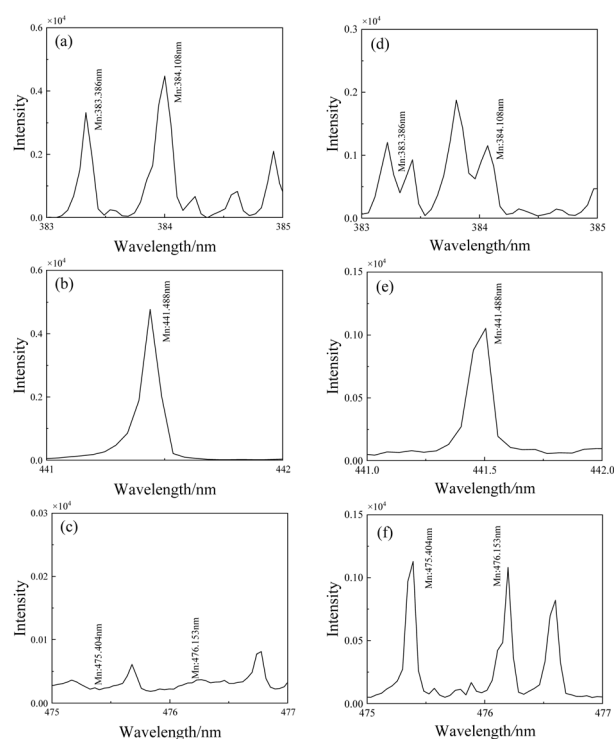

Fig. 10 The distribution of spectral lines of Mn element in different wavebands. (a–c) The carbon steel samples and (d–f) the steel slag samples.

model. For carbon steel samples, the characteristic spectral lines of the Mn element at 383.386 nm and 384.108 nm, along with their surrounding five sampling points, can serve as input for the CNN-SE-LSTM model (Fig. 10(a)), whereas for the slag samples, these lines are significantly influenced by the surrounding characteristic spectral lines of other elements and are not suitable as model input (Fig. 10(d)). Similarly, slag samples have prominent characteristic spectral lines at 475.404 nm and 476.153 nm (Fig. 10(f)), while the corresponding lines in carbon steel samples are obscured by the continuous background (Fig. 10(c)). Considering the model's versatility, we ultimately chose the 2D CNN-SE-LSTM model as the pre-training model.

Subsequently, we determined the parameters for the SE block and the LSTM. Fig. S3† illustrates the impact of the parameters of FC in the SE block on the prediction performance. As the number of hidden layer parameters gradually increases, the model's RMSE first decreases and then increases, achieving the best solution when the hidden layer parameters are set to 16. Similar observations occurred in the LSTM module. As shown in Fig. S4,† the model performed best when the number of hidden layer parameters was set to 32.

Additionally, some other hyperparameters of the model may also affect the results of quantitative analysis, such as the Dropout rate, the activation function and so on. We have conducted a detailed analysis of the impact of these parameters on model performance in the ESI file (Fig. S5 and S6),† and based on these results, we set the Dropout rate to 0.2 and the activation function to the ReLU function.

Finally, following the method illustrated in Fig. 7, we transferred the CNN-SE-LSTM pre-training model to the target samples and fine-tuned the parameters of the fully connected layers. The final parameters of the TrCSL model are shown in Table 1.

3.3 Performance of the TrCSL model for small samples

First, we analyzed the influence of baseline correction on model performance. Based on the results shown in Fig. 11, it is not
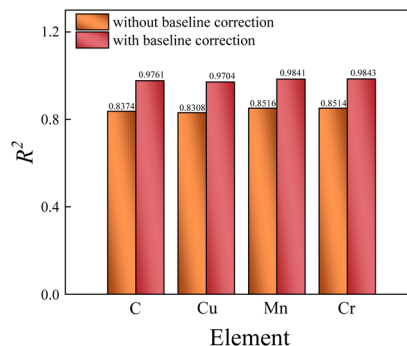

Fig. 11 Performance of the TrCSL model without/with baseline correction.

difficult to find that after baseline correction, the R^2 value of the prediction results for the TrCSL model increased by about 0.14. The reason may be that the matrix effect between different samples may seriously affect the training effectiveness. After baseline correction, it can effectively reduce the impact of the matrix effect, thereby enhancing the model performance.

Subsequently, we utilized the TrCSL model to conduct quantitative analysis on 20 sets of carbon steel samples and carried out ablation experiments to explore the importance of each component in the TrCSL model. Fig. 12 presents the prediction results of the four models on the test set, where the TrCNN model refers to the CNN model after transfer learning, which does not include the SE and LSTM modules. Similarly, the TrCNN-SE and TrCNN-LSTM models denote the absence of the LSTM and SE modules, respectively. Taking the prediction results of the Mn element as an example, the R^2 of the TrCNN model is only 0.7370. When the SE and LSTM modules are added separately, the R^2 increases by 0.1921 and 0.1745, respectively. When both the SE and LSTM modules were added simultaneously, the R^2 increased by 0.2471. The results validate the necessity of the SE and LSTM modules in the TrCSL model, which can effectively enhance model performance. In addition, we also find that the SE block slightly outperforms the LSTM module in enhancing model performance, with a difference in

Table 1 Parameters of the TrCSL model

|	Detail parameters
Input layer	125 × 131
CNN module	Number of layers: 4
	Size of kernels: 3 × 3
	Number of kernels: 32
Pooling layer	Average pooling
Dropout rate	0.2
Activation function	ReLU
SE block	Hidden layer parameter: 16
LSTM module	Hidden layer parameter: 32
FC layer	128
Optimizer	Adam optimizer
Loss function	MSE loss
Initial learning rate	0.01
Learning rate attenuation rate	0.1 for every 100 epochs
Epoch	800
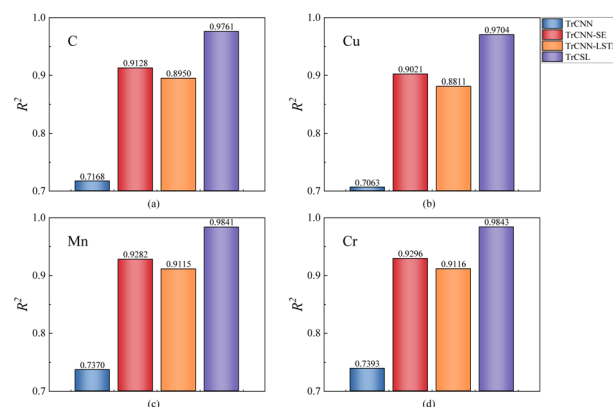
Fig. 12 The results of the ablation experiment. (a) The C element, (b) the Cu element, (c) the Mn element, and (d) the Cr element.

R^2 of approximately 2% when either is added alone. As we discussed in Section 2.2, each component in the TrCSL model is indispensable. The CNN extracts features from the input data, the SE block enhances the expressiveness of these features, and the LSTM processes the serialized features to ultimately predict the element concentration.

According to Fig. 12, we further observed that the TrCSL model's prediction results for Mn and Cr elements have the highest values of R^2, while the R^2 value of Cu element is the lowest. This discrepancy may be associated with spectral fluctuations, which has been further confirmed and thoroughly discussed in ref. 43 and 44. Taking sample no. 1 as an example, we selected a total of 8 characteristic spectral lines: C: 414.626 nm, C: 426.902 nm, Cu: 316.420 nm, Cu: 386.046 nm, Mn: 404.136 nm, Mn: 407.028 nm, Cr: 363.146 nm, and Cr: 399.112 nm. Fig. S7† shows the intensity fluctuation degree, and it is not difficult to find that the two characteristic spectral lines of Cu element exhibit the most significant fluctuations, while those of Cr and Mn elements show the least fluctuations. Furthermore, we calculated the RSD for these 8 characteristic spectral lines in Fig. 13. Among them, the Mn and Cr elements have the lowest RSD, which is close to 6%, and Cu element has the highest RSD. In conjunction with the results, we can conclude that spectral fluctuations may reduce the prediction accuracy of the TrCSL model.

Then, we contrasted the TrCSL model with traditional methods. We randomly selected different numbers of samples from 100 sets of steel slag samples and performed quantitative analysis using the PLSR and PSO-SVR algorithms. The number of samples we selected increased from 10 to 100 in increments of 10, and the samples were divided into a 70% training set and a 30% test set. Taking the Mn element as an example, the relationship between the LIBS quantitative analysis results and the sample numbers is shown in Fig. 14. It is not difficult to find that when the number of samples is small (number <20), the R^2 values of both PLSR and PSO-SVR algorithms are less than 0.6, mainly due to the insufficient model training. In comparison, the TrCSL model can achieve an R^2 of 0.98 on 20 sets of carbon steel samples, which represents a performance improvement close to 0.4. As the number of samples gradually increases, the
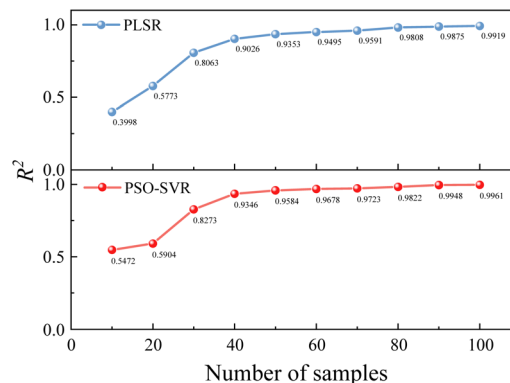

Fig. 14 The influence of sample numbers on quantitative analysis results.

quantitative analysis result significantly improves. When the number of samples reaches 40, the R^2 value has risen to 0.9. When the number of samples further increases to 90, the R^2 reaches 0.99, where the model is already able to achieve satisfactory quantitative analysis results. Furthermore, by comparing the data in Fig. 12 and 14, we found that our proposed TrCSL model, with only 20 sets of carbon steel samples, can achieve the quantitative analysis results comparable to the PLSR and SVR algorithms on 80 sets of samples. This comparison not only highlights the excellent performance of the TrCSL model in handling small-sample data but also demonstrates its great potential in the field of quantitative analysis.

Finally, we compared the TrCSL model's performance with the PCA-PLS and GLL methods in ref. 31 and 35. Fig. 15 shows the R^2 of the quantitative analysis results for the three methods. The TrCSL model demonstrated superior performance, with R^2 values exceeding those of both PCA-PLS and GLL. Moreover, the R^2 values of TrCSL and GLL were significantly higher than PCA-PLS, underscoring deep learning's effectiveness in feature extraction from small samples. When the sample size was 20, the performance of the TrCSL and GLL models was comparable. However, when the sample size was reduced to 10, the TrCSL model exhibited a significantly greater improvement in R^2, which highlights the TrCSL model's superior adaptability and robustness when dealing with small-sample data.
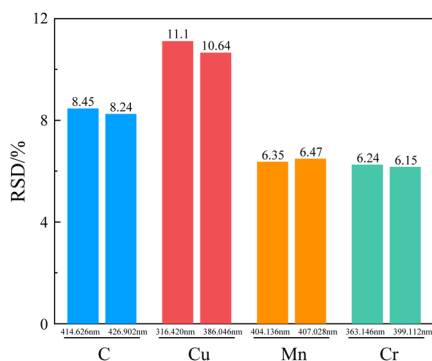

Fig. 13 The RSD values of Cr: 399.12 nm, Cr: 363.146 nm, Cu: 386.046 nm, Cu: 316.420 nm, Mn: 407.028 nm, Mn: 404.136 nm, C: 426.902 nm, and C: 414.626 nm for sample no. 1.
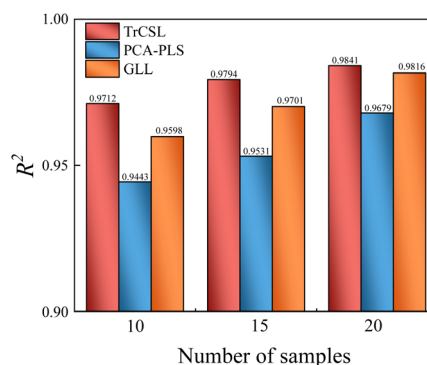

Fig. 15 Comparison of three models on small samples.

1818 | *J. Anal. At. Spectrom.*, 2025, **40**, 1810–1820

This journal is © The Royal Society of Chemistry 2025

The TrCSL model has achieved satisfactory results, mainly due to two reasons: (1) the introduction of the SE block mechanism into the model has significantly enhanced its ability to extract features from input LIBS images, which allows the model to understand the data more deeply; (2) when dealing with small samples, any unusual fluctuation in the data can have a significant impact on the overall analysis results, and the TrCSL model can effectively alleviate this interference through transfer learning, which can ensure the stability and reliability of prediction results. Furthermore, the TrCSL model has a broad applicability. For the input LIBS spectral data, there is no need for complex spectral line selection; it is only necessary to convert one-dimensional LIBS spectral data into a two-dimensional form. This simplified data processing procedure not only avoids cumbersome steps but also significantly improves the efficiency of data analysis.

4. Conclusion

In this paper, we proposed the TrCSL model, aimed at improving the accuracy of LIBS quantitative analysis with small samples. The model integrates the advantages of transfer learning, CNN, attention mechanisms, and LSTM. We transferred the pre-trained model obtained from 100 sets of steel slag data to 20 sets of carbon steel data and fine-tuned the model. At the same time, we also analyze the influence of model hyperparameters, spectra baseline correction results, spectral fluctuations and other factors on the performance of the model. The results demonstrate that the TrCSL model, with only 20 sets of carbon steel samples, can achieve the performance of the traditional PLSR and PSO-SVR algorithms when applied to a larger dataset of 80 samples. The TrCSL model we proposed has higher quantitative analysis accuracy and better universality ability for small samples, and this method is expected to improve the accuracy of LIBS quantitative analysis when the samples are not sufficient.

Data availability

The relevant experimental data can be downloaded at **https://github.com/msj-msj/msj-msj-TrCSL_0508_1.git**.

Author contributions

Shengjie Ma: investigation, conceptualization, software, formal analysis, visualization, writing – original draft. Shilong Xu: methodology and writing – review and editing. Congyuan Pan: software and resources. Jiajie Fang: investigation, resources, and visualization. Fei Han: investigation and resources. Yuhao Xia: data curation. Wanying Ding: visualization. Youlong Chen: resources. Yihua Hu: writing – review and editing, supervision, conceptualization, and writing – review and editing.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Notes and references

1 V. Gardette, V. Motto-Ros, C. Alvarez-Llamas, *et al.*, *Anal. Chem.*, 2023, **95**, 49–69.

2 Z. Wang, M. S. Afgan, W. Gu, *et al.*, *TrAC, Trends Anal. Chem.*, 2021, **143**, 116385.

3 R. S. Harmon and G. S. Senesi, *Appl. Geochem.*, 2021, **128**, 104929.

4 K. Liu, C. He, C. Zhu, *et al.*, *TrAC, Trends Anal. Chem.*, 2021, **143**, 116357.

5 A. H. Galmed, M. Maaza, B. M. Mothudi, *et al.*, *Mater. Today: Proc.*, 2021, **36**, 600–603.

6 J. D. Pedarnig, S. Trautner, S. Grünberger, *et al.*, *Appl. Sci.*, 2021, **11**, 9274.

7 X. Y. Jin, T. Gu, Y. Zhan, *et al.*, *Appl. Spectrosc. Rev.*, 2024, 1–44.

8 Y. Zhang, T. Zhang and H. Li, *Spectrochim. Acta, Part B*, 2021, **181**, 106218.

9 P. Siozos, N. Hausmann, M. Holst, *et al.*, *J. Archaeol. Sci. Rep.*, 2021, **35**, 102769.

10 N. Hausmann, D. Theodoraki, V. Piñon, *et al.*, *Sci. Rep.*, 2023, **13**, 19812.

11 J. Cui, W. Song, Z. Hou, *et al.*, *J. Anal. At. Spectrom.*, 2022, **37**, 2059–2068.

12 Y. Huang and A. Bais, *Spectrochim. Acta, Part B*, 2022, **193**, 106451.

13 D. Diaz, A. Molina and D. W. Hahn, *Appl. Spectrosc.*, 2020, **74**, 42–54.

14 A. Erler, D. Riebe, T. Beitz, *et al.*, *Sensors*, 2020, **20**, 418.

15 S. Kashiwakura and K. Wagatsuma, *ISIJ Int.*, 2020, **60**, 1245–1253.

16 E. Képeš, J. Vrábel, O. Adamovsky, *et al.*, *Anal. Chim. Acta*, 2022, **1192**, 339352.

17 M. Yao, G. Fu, T. Chen, *et al.*, *J. Anal. At. Spectrom.*, 2021, **36**, 361–367.

18 Y. Ding, W. Zhang, X. Zhao, *et al.*, *J. Anal. At. Spectrom.*, 2020, **35**, 1131–1138.

19 W. Yang, L. I. Mao-Gang, F. Ting, *et al.*, *Chin. J. Anal. Chem.*, 2022, **50**, 100057.

20 J. Moros, L. M. Cabalín and J. J. Laserna, *Anal. Chim. Acta*, 2022, **1191**, 339294.

21 C. Yan, J. Qi, J. Liang, *et al.*, *J. Anal. At. Spectrom.*, 2018, **33**, 2089–2097.

This journal is © The Royal Society of Chemistry 2025

J. Anal. At. Spectrom., 2025, **40**, 1810–1820 | 1819

22 J. Liang, C. Yan, Y. Zhang, *et al.*, *Chemom. Intell. Lab. Syst.*, 2020, **197**, 103930.

23 X. Xu, F. Ma, J. Zhou, *et al.*, *Comput. Electron. Agric.*, 2022, **199**, 107171.

24 C. X. LU, B. Wang, X. P. Jiang, *et al.*, *Plasma Sci. Technol.*, 2018, **21**, 034014.

25 L. N. Li, X. F. Liu, W. M. Xu, *et al.*, *Spectrochim. Acta, Part B*, 2020, **169**, 105850.

26 F. Poggialini, B. Campanella, S. Legnaioli, *et al.*, *Appl. Spectrosc.*, 2022, **76**, 959–966.

27 Y. Ding, L. Yang, W. Chen, *et al.*, *Appl. Opt.*, 2023, **62**, 2188–2194.

28 W. Song, M. S. Afgan, Y. H. Yun, *et al.*, *Expert Syst. Appl.*, 2022, **205**, 117756.

29 L. Zhang, Y. Gong, Y. Li, *et al.*, *Spectrochim. Acta, Part B*, 2015, **113**, 167–173.

30 Z. Hou, Z. Wang, L. Li, *et al.*, *Spectrochim. Acta, Part B*, 2022, **191**, 106406.

31 A. Li, X. Zhang, X. Wang, *et al.*, *J. Anal. At. Spectrom.*, 2022, **37**, 2022–2032.

32 Z. Hou, Z. Wang and T. Yuan, *et. al.*, *J. Anal. At. Spectrom.*, 2016, **31**, 722–736.

33 W. Zhang, Z. Zhuo, P. Lu, *et al.*, *J. Anal. At. Spectrom.*, 2020, **35**, 1621–1631.

34 Q. Ma, Z. Liu, T. Sun, *et al.*, *Opt. Express*, 2023, **31**, 27633–27653.

35 Z. Liu, Q. Ma, T. Zhang, *et al.*, *Opt Laser. Technol.*, 2025, **181**, 111720.

36 K. Weiss, T. M. Khoshgoftaar and D. D. Wang, *J. Big Data*, 2016, **3**, 1–40.

37 I. Ameer, N. Bölücü, M. H. F. Siddiqui, *et al.*, *Expert Syst. Appl.*, 2023, **213**, 118534.

38 H. E. Kim, A. Cosa-Linan, N. Santhanam, *et al.*, *BMC Med. Imaging*, 2022, **22**, 69.

39 Y. Zhang, Y. Shao, R. Luo, *et al.*, *IEEE Internet Things J.*, 2023, 8637–8646.

40 Y. Rao, W. Ren, W. Kong, *et al.*, *J. Anal. At. Spectrom.*, 2024, **39**, 925–934.

41 Y. LeCun, L. Bottou, Y. Bengio, *et al.*, *Proc. IEEE*, 1998, **86**, 2278–2324.

42 Z. M. Zhang, S. Chen and Y. Z. Liang, *Analyst*, 2010, **135**, 1138–1146.

43 A. Gong, W. Huang, Y. Xiao, *et al.*, *Knowl. Base Syst.*, 2024, **304**, 112450.

44 D. Zhang, H. Ma, J. Nie, *et al.*, *Talanta*, 2025, **281**, 126872.