Journal of Network Intelligence
Taiwan Ubiquitous Information
 ISSN 2414-8105 (Online)
Volume **9**, Number **1**, February **2024**

A Comparative Analysis Study of Machine Learning Algorithms in the Quantitative Analysis of LIBS Steel

Ji-Shi Zheng

Department of Transportation
Fujian University of Technology
No.33 Xuefu South Road, Fuzhou, 350118, China
Zhengjishi@fjut.edu.can

Wei-Yu Hou*

Department of Transportation
Fujian University of Technology
No.33 Xuefu South Road, Fuzhou, 350118, China
913193215@qq.com

Zhao-Lin Zhao

Department of Transportation
Fujian University of Technology
No.33 Xuefu South Road, Fuzhou, 350118, China
zhaozl@fjut.edu.cn

Jia-Cheng Yang

School of Mechanical and Automotive Engineering
Fujian University of Technology
No.33 Xuefu South Road, Fuzhou, 350118, China
yuukilight@outlook.com

Hong-Ji Ye

School of Mechanical and Automotive Engineering
Fujian University of Technology
No.33 Xuefu South Road, Fuzhou, 350118, China
949242185@qq.com

Xiang-Xu Ren

School of Mechanical and Automotive Engineering
Fujian University of Technology
No.33 Xuefu South Road, Fuzhou, 350118, China
ren_xiangxu@163.com

Ling-Hua Kong

School of Mechanical and Automotive Engineering
Fujian University of Technology
No.33 Xuefu South Road, Fuzhou, 350118, China
15392030898@163.com

*Corresponding author: Wei-Yu Hou

Received March 10, 2023, revised May 7, 2023, accepted July 31, 2023.

Abstract. *Metallurgical technology directly affects the development of multiple areas such as national defense, energy, and transportation. It is one of the foundations of modern industry and is of great significance for promoting economic development and enhancing national competitiveness. Traditional chemical analysis methods require multiple reagents and tedious operations, while Laser-Induced Breakdown Spectroscopy (LIBS) can quickly detect multiple elements by capturing the spectral signals of samples without any sample preparation. We collected spectral data for 9 selected steel samples, with 5 points selected for each sample and 500 hits for each point. We used maximum-minimum normalization and Principal Component Analysis (PCA) to denoise the spectral data and established Partial Least Squares Regression (PLSR), Support Vector Regression (SVR), and Extreme Learning Machine (ELM) as quantitative analysis models for Mn, P, S, and C in iron blocks. Results showed that the PLSR model predicted Mn, P, S, and C with R^2 values of 0.99917, 0.99532, 0.99471, and 0.99513, and R_{mse} values of 0.00064, 0.00028, 0.0023, and 0.00153, respectively. The PCA-SVR model predicted Mn, P, S, and C with R^2values of 0.99987, 0.99797, 0.99543, and 0.99983, and R_{mse} values of 0.00025, 0.00018, 0.0022, and 0.00028, respectively. The PCA-ELM model predicted Mn, P, S, and C with R^2 values of 0.87425, 0.79556, 0.80864, and 0.63416, and R_{mse} values of 0.01126, 0.00235, 0.00133, and 0.00966, respectively. Among all the models, the PCA-SVR model had the best performance. The research results indicate that the combination of PLSR and PCA-SVR models with LIBS can achieve high-precision quantitative analysis of steel elements and improve the efficiency of steel detection.*
Keywords: Laser-induced breakdown spectroscopy; Quantitative analysis; Partial least squares;Support vector regression; Extreme learning machine.

---

1. **Introduction.** Steel production capacity is an important manifestation of a country's level of modern industrial development and is the foundation of a modern state. Steel smelting is one of the most critical processes, and the content of various trace elements in steel, including C, Mn, P, S, and a small number of other elements, determines the steel's quality [1, 2]. The content of these elements directly affects the performance of high-end equipment, which requires precise control of trace elements. Therefore, quickly detecting and strictly controlling the content of various elements in steel has become a critical issue in steelmaking.

Although traditional offline testing methods can achieve high precision quantitative qualitative research of relevant samples, offline testing requires sampling pre-processing and a complex analysis process [3], which takes a long time. Moreover, the steel-making process needs to be stopped to take samples for testing, which increases the tediousness of smelting and the probability of impurity adulteration, easily leading to large gaps between the detected steel results and the actual results. The offline steel composition testing is far from being able to adapt to the requirements of modern gold product manufacturing. Therefore, new online testing methods are needed to reduce the testing process's time and complexity, control the trace element content from the source, and solve the painful problem of poor real-time steel testing in the industry.

Laser-induced breakdown spectroscopy (LIBS) is a fast spectroscopic technique that obtains the composition and content of a substance from the wavelength and intensity of atoms and ions in a laser-induced plasma spectrum [4]. It utilizes a high-energy-density pulsed laser to induce ionization on the surface of a material, resulting in the generation of laser-induced plasma. The plasma emission spectrum contains a line spectrum that carries a wealth of information about the elements of the sample and a continuous spectrum with background information [5].

After over 40 years of development, LIBS technology has become increasingly mature and has found wide applications in many fields, such as environmental detection [6, 7, 8],

the biomedical field [9, 10, 11], industrial metallurgy [12, 13, 14],, and food detection [15, 16] .

1.1. **Related work.** In recent years, machine learning algorithms have been combined with LIBS. Zhang used the SelectK-Best algorithm for feature selection and back propagation neural network (BPNN) for regression model training to determine the content of trace metal elements, achieving relative errors of prediction of 1.13%, 2.85%, and 7.20% for manganese, chromium, and nickel in steel, respectively, with an R^2 of over 0.996 [17]. He used a hybrid variable selection method of mutual information particle swarm optimization (MI-PSO) to achieve accurate screening of LIBS and Fourier transform infrared spectroscopy (FTIR) spectral feature variables, and constructed a mutual information particle swarm optimization kernel extremum learning machine (MI-PSO-KELM) model, resulting in R^2 values of above 0.96 for coal quality detection results [18]. Tian determined phosphorus in seafood using LIBS and developed univariate and multivariate regression models with PLS and SVM, respectively, resulting in an average RSD of 5.18% for SVM predictions and 9.40% for PLS; the results suggest that support vector machines may be more suitable for solving the non-linear behaviour in LIBS spectra caused by matrix effects [19]. Liang proposed the kernel limit learning machine (KELM) by extending ELM to replace the feature mapping in ELM with kernel matrices, and combined particle swarm optimization and KELM to achieve rapid identification of Salvia geographic regions, achieving a classification accuracy as high as 94.7% [20]. Zhang quantified carbon (C), hydrogen (H), and nitrogen (N) in coal using LIBS combined with machine learning, calibrated and predicted using linear regression, support vector regression, and random forest models, and the results showed that random forest outperformed the other models with R^2 values of 0.9844, 0.9625, and 0.9829, respectively [21]. These results demonstrate that machine learning algorithms combined with LIBS have wide-ranging applications in various fields, with impressive accuracy and precision.

1.2. **Innovation and contribution.** This experiment efficiently performed elemental analysis and prediction on steel samples containing multiple elements, with high accuracy and robustness. Preprocessing methods such as PCA dimension reduction were used to effectively extract features, reduce data dimensions, and improve model efficiency and accuracy. Finally, by comparing the prediction results of different machine learning algorithms, the optimal algorithm and parameter combination were identified to improve model performance and robustness. The study simplified the cumbersome and complex process of steel quantification testing, providing an effective method and approach for LIBS technology in metallurgical analysis and quality control.

2. System design.

2.1. **Installation design.** LIBS technology enables real-time, non-contact measurement and analysis of multiple elements in both light and heavy materials without requiring pre-production of samples. This allows for adjustment of elemental content during the smelting process to meet production standards, forming a closed-loop manufacturing control process. The experiment was conducted under standard atmospheric pressure in a confined, clean space. Figures 1 and 2 depict the physical and structural drawings of the LIBS equipment used, which consisted of a 4-channel fibre optic spectrometer (Spectrometer, AvaSpec-ULS2048-4-USB2, Avantes, Netherlands) and a Q-modulated Nd: YAG laser (Laser, Dawa-100, Beamtech Optronics, China) operating at a wavelength of 1064 nm with a single pulse energy of 50 mJ at a frequency of 5 Hz.

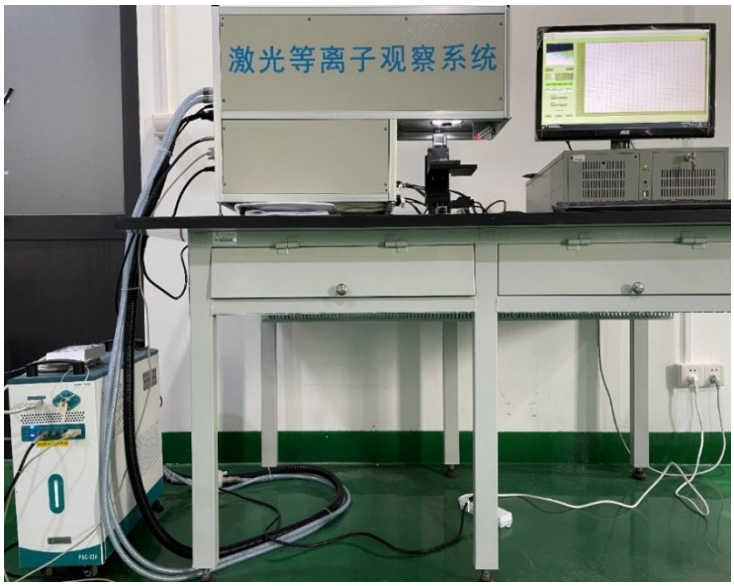

FIGURE 1. Physical view of LIBS equipment

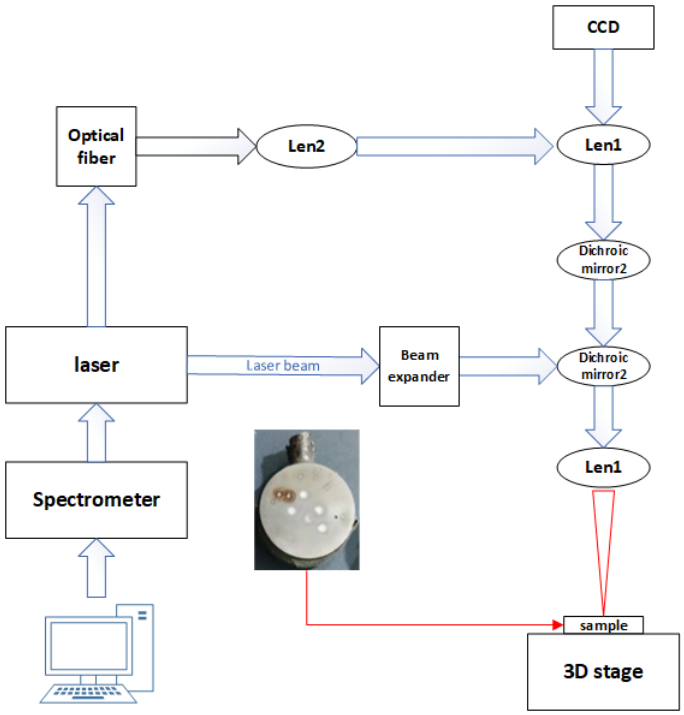

FIGURE 2. LIBS equipment construction diagram

The laser beam is expanded by a Beam expander and reflected by a Dichroic mirror 1, which is focused by a Lens 1 with a focal length of 100 mm onto the surface of the sample placed on a three-dimensional motorised translation table to produce a plasma.

The light emitted from the plasma passes through the converging lens 1 and the dichroic mirror 1, is reflected by the dichroic mirror 2, converges on the end face of the fibre via the converging lens 2 and is transmitted to the spectrometer for recording. The Dichroic mirror 2 has a reflection-to-transmission ratio of approximately 9:1, with a small amount of light passing through Lens 3 to be imaged on the camera to facilitate observation of the laser excitation position.

2.2. **Sample preparation and data acquisition.** To determine the steel content offline, iron samples were obtained from Minguang Iron and Steel Co. The samples were polished with sandpaper to remove surface impurities and oxidation, and placed on a three-dimensional electronically controlled moving platform. Five points were selected for each steel sample to be excited as a set of experiments, and the sample position was moved after each hit to ensure random sampling. The spectral acquisition was conducted with a wavelength range of 230-750 nm, an integration time of 1.05 ms, and a delay time of 1.28 μs.

FIGURE 3. Steel sample image

The experiments are grouped based on each iron block. This experiment is divided into 10 groups of experiments numbered S01, S02, ..., and S09. Each iron block has 5 randomly selected excitation points. With 9 groups of iron blocks, there are a total of 45 analysis points. For each analysis point, 500 spectra are collected, with 5 different excitation points for each iron block. The obtained spectra are averaged, resulting in 5 average spectra for each iron block, which correspond to each group of experiments. The experimental setup is fully controlled by software. The spectral information emitted by the plasma is divided by the spectrometer, converted into a digital signal by the detector, and the spectral data is stored and analyzed by the computer.

The exact composition of the 9 iron blocks was determined using an Optical Emission Spectrometer (OES), with the data obtained from the direct reading spectra serving as the actual results for the model. Table 1 shows the composition content of each element measured by the OES, with Fe accounting for the highest composition in the steel. To prevent the influence of Fe on the prediction results of the remaining four elements, Fe was excluded from the analysis. The remaining four elements were used as input data for subsequent analysis. When dividing the data into test and training sets, 20% of the samples were allocated as the test set, and a random number seed of 27 was used. Nine spectral data points were randomly selected from the 45 strikes for the test set, and the remaining 36 spectral data points were used for the training set.

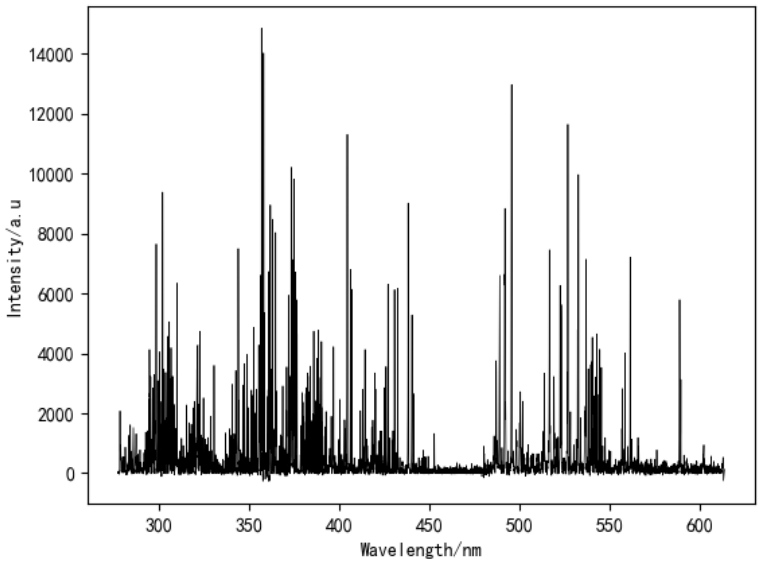

FIGURE 4. Average spectrum of sample S05

TABLE 1. Elemental content measured by direct reading spectrometers

Numb	Fe	Mn	P	S	C
S01	99	0.125	0.0193	0.0137	0.0812
S02	98.9	0.109	0.0154	0.0174	0.101
S03	98.8	0.147	0.0275	0.0102	0.0598
S04	98.9	0.191	0.0253	0.0095	0.0819
S05	99	0.117	0.0207	0.0186	0.0318
S06	99	0.129	0.0194	0.0141	0.0789
S07	99	0.132	0.0163	0.0117	0.0615
S08	98.9	0.142	0.0254	0.0101	0.0329
S09	98.9	0.142	0.0254	0.0101	0.0329

3. **Modeling.** Principal Component Analysis (PCA) is a widely used method for data dimensionality reduction in machine learning and data processing. The main idea of PCA is to map the original n-dimensional features onto k-dimensions, replacing the original n-dimensional features with new k-dimensional features [22].

By calculating the covariance matrix of the data, we obtain the eigenvalues and eigenvectors of the covariance matrix. We then select the matrix consisting of the eigenvectors corresponding to the K features with the largest eigenvalues, thereby achieving dimensionality reduction of the data features [23]. PCA regroups the original multiple indicators into a new set of indicators. This method not only reduces the interference of multivariate correlation but also captures the information of the original spectral data as much as possible. The results of the principal component analysis can be used to extract the important features of high-dimensional data and to reduce the computational overhead of the algorithm.

Spectral data pre-processing methods can optimize data and improve model prediction performance. Good spectral pre-processing techniques can reduce the effects of laser pulsation fluctuations, fluctuations, noise, and drift limitation on LIBS spectra [24]. To

improve the convergence speed of the model for subsequent data input, this study uses Min-Max Normalization (MMN) to scale the spectral intensities between the interval [0,1] [25]. For each one-dimensional feature x_i, where $i = 1, 2, ..., p$, the k - th sample eigenvalue x_{ik}, where $k = 1, 2, 3, ..., n$, is normalized.

$$\widehat{x_{lk}} = \frac{x_{ik} - \min(x_i)}{\max(x_i) - \min(x_i)} \tag{1}$$

The result is mapped between [0,1] in Equation (1), where $\min(x_i)$ and $\max(x_i)$ are respectively the minimum and maximum values of feature x_i across all samples.

To evaluate the accuracy of the model and to measure the experimental results indicators the coefficient of determination (R^2), and root mean square error (R_{MSE}) [26] are used as references to evaluate the accuracy of the regression prediction model, and the formulae for each indicator are shown below [27]. The formulae for each indicator are as follows:

$$R^2 = 1 - \frac{\sum \left(Y_i - \widehat{Y_l}\right)^2}{\sum \left(\widehat{Y_l} - \hat{Y}\right)^2} \tag{2}$$

In Equation (2) Y_i is the predicted mean value of the i-th sample. $\widehat{Y_l}$ is the Y_i the corresponding true value. $\widehat{Y_l}$ is the mean value of the sample.

$$R_{MSE} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{m}} \tag{3}$$

In Equation (3) y_i is the i-th predicted value, $\hat{y}_i$ is the y_i the corresponding true value m is the sample size.

In general, The smaller the value of R_{MSE}, the better the predictive power of the forecasting model. R^2 The closer the value is to 1, the better the model fits.

Partial Least Squares Regression (PLSR) is a multivariate statistical data analysis method that models multiple dependent variables on multiple independent variables. It is a generalization of the least squares method that can achieve quantitative analysis even when there are multiple correlations among independent variables. PLSR can more easily exclude noise in the independent variables and solve more difficult problems of multiple linear regression [28].

After data pre-processing and normalization, the spectral data is denoted as x_1, x_2, x_3, ..., x_n, and its composite indicator variable, the principal component of the raw data after dimensionality reduction, is denoted as y_1, y_2, y_3, ..., y_m (where $m \leq n$). The residual matrix is denoted as E_1, E_2, E_3, ..., E_m (where $m \leq n$). The relationship between the three can be written as:

$$\begin{cases} y_1 = l_{11}x_1 + l_{12}x_2 + l_{13}x_3 + \cdots + l_{1n}x_n + E_1 \\ y_2 = l_{21}x_1 + l_{22}x_2 + l_{23}x_3 + \cdots + l_{2n}x_n + E_2 \\ y_3 = l_{31}x_1 + l_{32}x_2 + l_{33}x_3 + \cdots + l_{3n}x_n + E_3 \\ \cdots \\ y_m = l_{m1}x_1 + l_{m2}x_2 + l_{m3}x_3 + \cdots + l_{mn}x_n + E_m \end{cases} \tag{4}$$

where the coefficient l_{i1} , l_{i2} , l_{i3} ,..., l_{in} ($i = 1, 2, 3, ..., m$) form the coefficient matrix.

Based on the principal component analysis to extract the most informative principal components reflecting the input data, a regression model between the element to be

measured Y and the spectral data principal component matrix X can be abbreviated as [29]:

$$Y = LX + E \tag{5}$$

In Equation (5) L is the coefficient matrix corresponding to the principal component matrix and E is the total residual matrix.

Support vector regression (SVR) is a linear regression method that uses limited sample information to ensure a non-linear transformation of input data into a decision function in a high-dimensional space, resulting in good learning and generalization capabilities and achieving optimal results [27]. The common kernel functions in SVR models include the polynomial kernel function (Poly), radial basis kernel function (RBF), linear kernel function (Linear), and Sigmoid kernel function, among others. For this study, the RBF kernel function was chosen, and the objective function of SVR is given by:

$$C_{svr} = \sum_{i' \in M_{SV}} \alpha i' \cdot k_{\text{libs}} \ (I''_{i'}, I) + b \tag{6}$$

In Equation (6), the M_{SV} is the set of support vectors; and $\alpha i'$ is the Lagrangian multiplier. i' is the i - th set of experimental data. $I''_{i'}$ is the suport vector; I is the input vector; b is a constant k_{libs} is the radial basis function, which can be expressed as [27] :

$$k_{libs}\big(I'_{i'}, I\big) = \exp\big(-\gamma \big| I'_{i'} - I \big|^2\big) \tag{7}$$

In Equation (7), the γ is the kernel function. The constraint on Equation (7) is $\sum_{i=1}^{n-m}(a_i - a_i^*) = 0$, where a_i^* is the Lagrangian multiplier, $0 \leq a_i^* \leq$, and C is the penalty function.

The SVR model has two important hyperparameters: C and γ. The penalty factor C reflects the degree of tolerance to errors, and its value affects the complexity and stability of the model. A larger value of C indicates a lower tolerance to model error, which may cause the model to overfit the training data. On the other hand, a smaller value of C indicates a greater tolerance to model error, which may cause the model to underfit the data. The parameter γ determines the width of the RBF kernel function and affects the smoothness of the decision boundary. A small value of γ creates a broad kernel, which can result in a smooth decision boundary, whereas a large value of γ creates a narrow kernel, which can result in a more complex and potentially overfitted decision boundary. Therefore, selecting appropriate values of C and γ is crucial for constructing an effective SVR model.

Extreme Learning Machine (ELM) is a machine learning method based on feed-forward neural networks [30]. Its basic principle is to randomly generate the weights of the connections between the input and hidden layers, and the biases of the nodes in the hidden layer, and set them without further adjustment, while ordinary feed-forward neural networks require continuous backpropagation iteration for parameter optimization. ELM significantly reduces computational effort compared to ordinary feed-forward neural networks and offers advantages such as fast learning speed and strong generalization ability [31]. The mathematical model of the extreme learning machine is shown below [32].

A single hidden layer feed-forward neural network with L neurons can be expressed as follows, for an arbitrary set of N distinct samples (x_i,t_i), where $x_i = [x_{i1}, x_{i2}, \ldots, x_{in}]^T \in R^n$ is the input quantity and $t_i = [t_{i1}, t_{i2}, \ldots, t_{im}]^T \in R^m$ is the output quantity [33]:

$$\sum_{i=1}^{L} \beta_i g\big(\mathbf{w}_i \cdot \mathbf{x}_j + \mathbf{b}_i\big) = \mathbf{o}_j, j = \mathbf{1}, \mathbf{2}, \ldots, \mathbf{N} \tag{8}$$

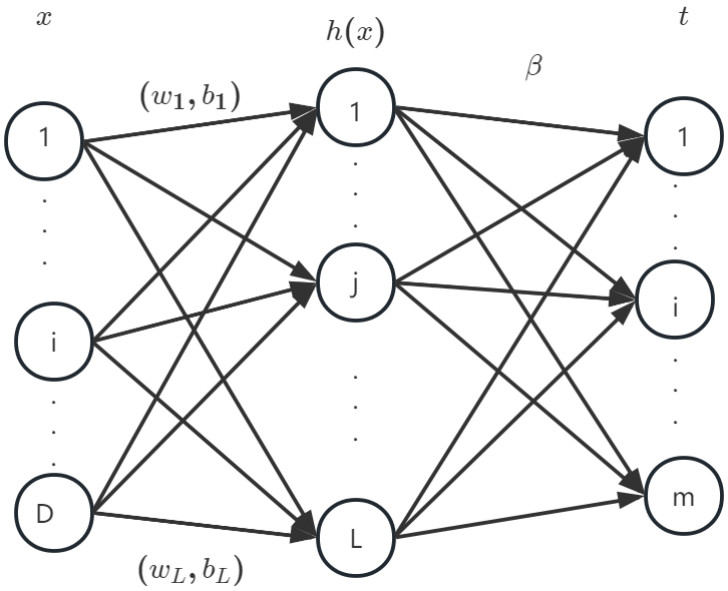

FIGURE 5. Structure of the ELM

In Equation (8), $g(x)$ is the activation function, $w_i = [w_{i1}, w_{i2}, \ldots, w_{in}]^T$ is the input weight, β_i is the output weight, b_i is the bias of the i - th hidden layer unit, and $w_i \cdot x_j$ is the inner product of w_i and x_j. The objective function of the extreme learning machine can be expressed as:

$$H\beta = \mathrm{T} \tag{9}$$

$$H(\mathbf{w}_1, \ldots, \mathbf{w}_L, \mathbf{b}_1, \ldots, \mathbf{b}_L, \mathbf{x}_1, \ldots, \mathbf{x}_L) = \begin{bmatrix} g(\mathbf{w}_1 \cdot \mathbf{x}_1 + \mathbf{b}_1) & \cdots & g(\mathbf{w}_L \cdot \mathbf{x}_1 + \mathbf{b}_L) \\ \vdots & \cdots & \vdots \\ g(\mathbf{w}_1 \cdot \mathbf{x}_N + \mathbf{b}_1) & \cdots & g(\mathbf{w}_L \cdot \mathbf{x}_N + \mathbf{b}_L) \end{bmatrix} \tag{10}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m} \qquad \mathrm{T} = \begin{bmatrix} \mathrm{T}_1^T \\ \vdots \\ \mathrm{T}_L^T \end{bmatrix}_{N \times m} \tag{11}$$

where H is the output of the hidden layer node, β is the output weight and T is the desired output. In the ELM algorithm, once the weights and the bias of the hidden layer are determined randomly, the output matrix of the hidden layer is uniquely determined and the learning process of ELM can be equated to finding the least squares solution, which can be expressed as:

$$\hat{\beta} = H^+ T \tag{12}$$

where H^+ is the Moore-Penrose generalised inverse matrix of the matrix H and the minimum value of the least squares solution is unique.

To improve the efficiency of the models and facilitate their input, the average spectral data obtained by striking the iron block was first subjected to maximum-minimum normalization, and the resulting normalized data was used as the input source for the three models.

PLSR has a data dimensionality reduction effect. During the training process of the PLSR model, the optimal number of data dimensions for reducing is determined. The accuracy of the results obtained in the training process varies greatly with different sizes

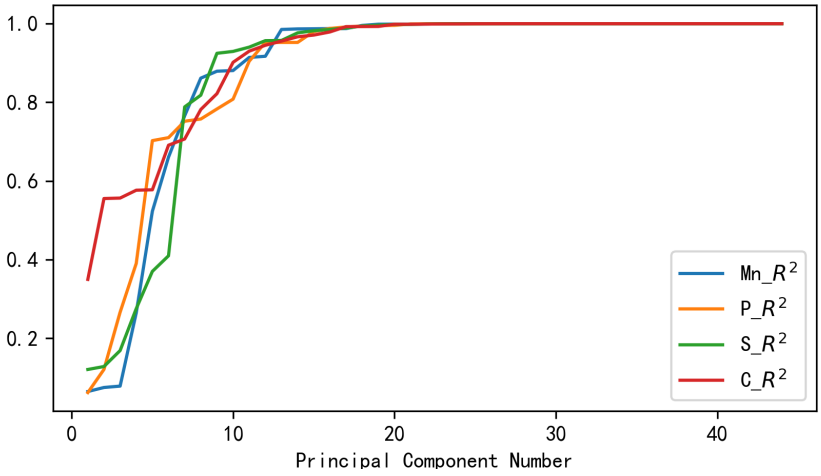

FIGURE 6. The four elements of the PLSR training

of dimensionality reduction. Through iterative calculations, it was found that the decision coefficient R^2 shows an increasing trend with the increase of data dimensionality, reaching its highest peak at a data dimensionality of 20 and then leveling off, as shown in Figure 6.

In the modeling process, PLSR obtains the weight coefficient of each feature value, which indicates its importance. To select the most important features, we sort the absolute value of the weight coefficients and select the top 20. The algorithm retains 20 components. The fit of the PLSR model to the predicted and actual values in the test set is illustrated in Figure 7.

The LIBS normalised dataset contains a vast amount of information that responds to the characteristics of the sample, and this data has high dimensionality. Principal Component Analysis (PCA) can effectively reduce the dimensionality of the high-dimensional data, transforming the high-dimensional data into variables with lower dimensionality and more focused responses to the original dataset. The data after the PCA dimensionality reduction process is then used as input for Support Vector Regression (SVR) and Extreme Learning Machine (ELM) to construct PCA-SVR and PCA-ELM models, respectively.

The number of principal components selected by PCA for data dimensionality reduction was 7. As shown in Figure 8, the cumulative contribution of these 7 principal components was over 90%, indicating that they can effectively represent the characteristic spectra of the iron blocks and all the information of the LIBS spectra. The eigenspace vectors of the samples were constructed based on these 7 principal components, and the resulting data matrix was used as the input of the SVR and ELM prediction models.

To determine the optimal combination of the model parameters C and γ in the SVR, a grid search method was used. The search interval for the parameters was $C \in [0, 50]$ and $\gamma \in [10^{-5}, 1]$. The optimal combination of parameters was determined by the grid search method to be $C = 8$ and $\gamma = 0.000002$. The prediction results of the model on the test set are shown in Figure 9.

The data processed by PCA were used as the input for the ELM model, and the model performed best when the number of hidden layer units was set to 3500 through training iterations. The activation function used in the model was Sigmoid. The prediction results from the model's test set are shown in Figure 10 below.

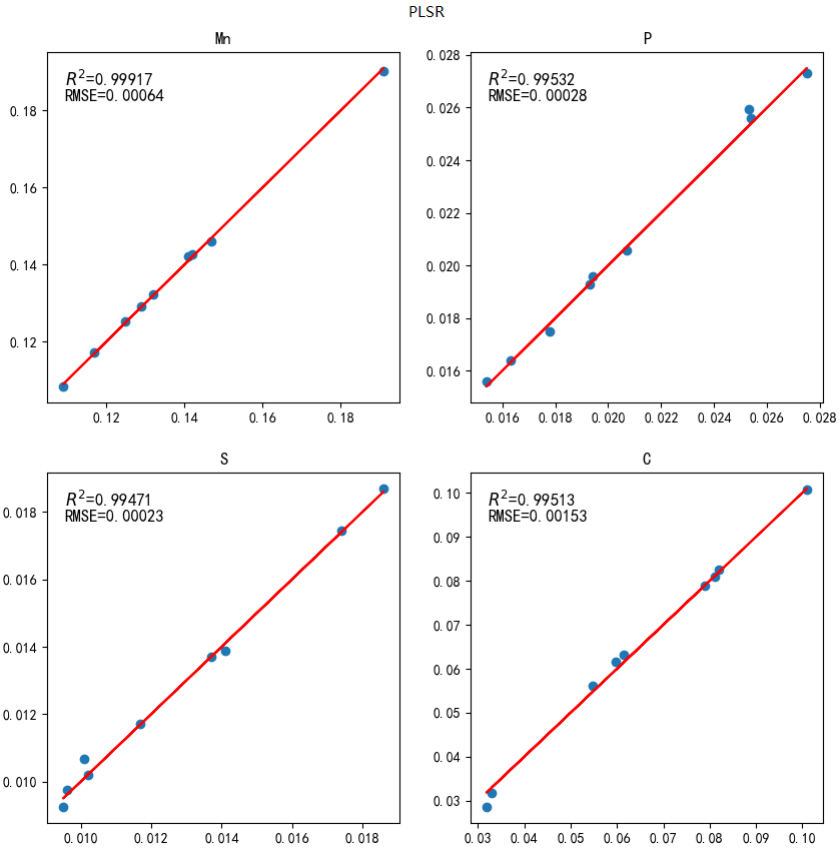

FIGURE 7. PLSR prediction results for the four elements

Figures 8 and 9 demonstrate the superior performance of PLSR and PCA-SVR models in predicting the elemental content, with high R^2 values close to 1 and low R_{mse} values. On the other hand, the PCA-ELM model exhibits better prediction accuracy than PLSR and PCA-SVR, indicating its potential as a reliable prediction model.

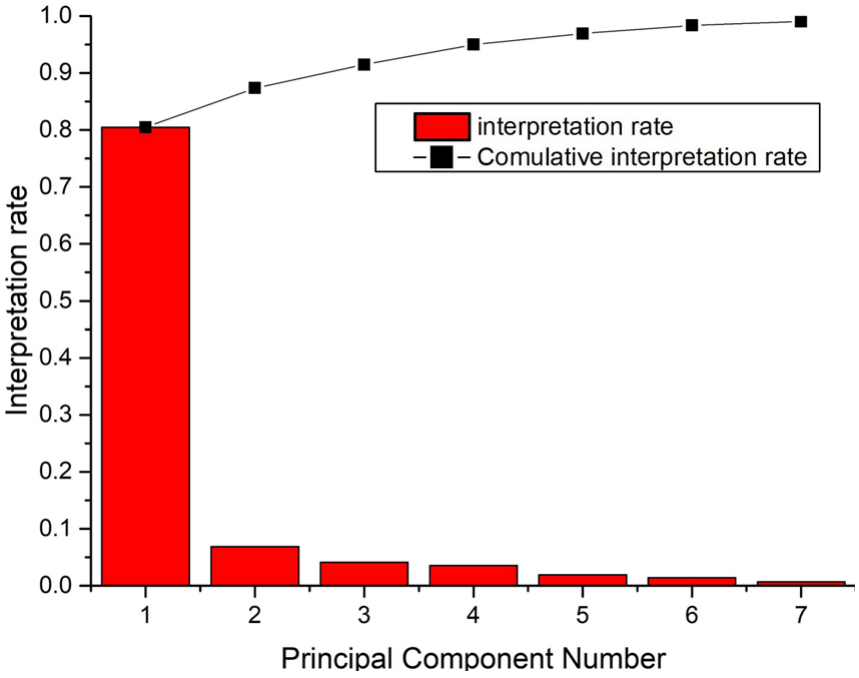

FIGURE 8. Percentage contribution of the seven selected principal components

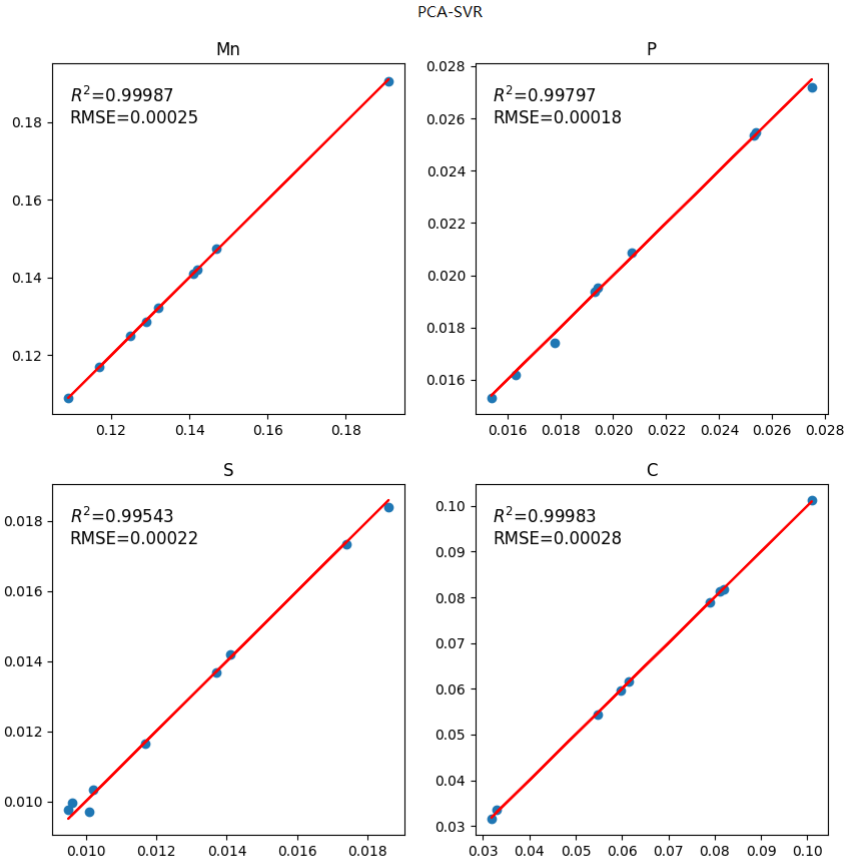

FIGURE 9. Prediction results of the PCA-SVR model for the four elements

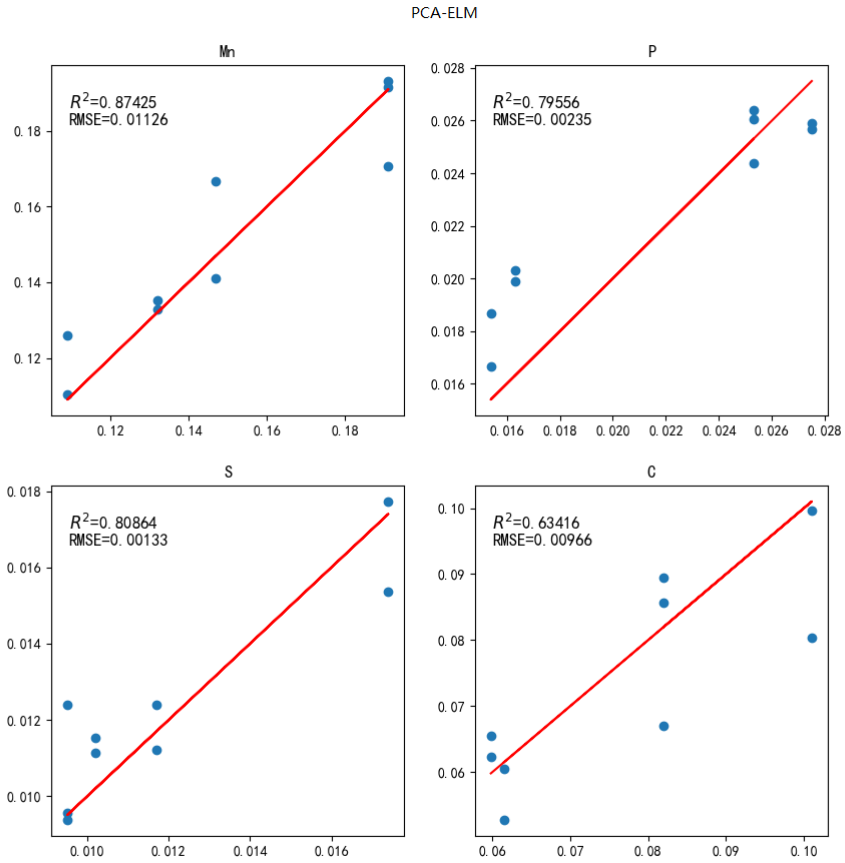

FIGURE 10. PCA-ELM prediction results for four elements

4. **Conclusions.** This article mainly introduces the quantitative analysis of the content of four elements (Mn, P, S, C) in iron blocks using laser-induced breakdown spectroscopy (LIBS) technology. Three prediction models were proposed: PLSR, PCA-SVR, and PCA-ELM. By preprocessing the data, such as maximum-minimum normalization and PCA dimensionality reduction, the input data for model training was obtained. The study found that the PLSR model performed best when 20 components were retained, and the R^2 values for predicting the four elements Mn, P, S, and C were 0.99917, 0.99532, 0.99471, and 0.99513, respectively, with R_{mse} values of 0.00064, 0.00028, 0.0023, and 0.00153, respectively. In the PCA-SVR, the grid search method was used to find the best parameter combination for the SVR model. The final model predicted R^2 values for the four elements Mn, P, S, and C were 0.99987, 0.99797, 0.99543, and 0.99983, respectively, with R_{mse} values of 0.00025, 0.00018, 0.0022, and 0.00028, respectively. In the ELM model, 3500 hidden layer units were selected, and the final predicted R^2 values for the four elements Mn, P, S, and C were 0.87425, 0.79556, 0.80864, and 0.63416, respectively, with R_{mse} values of 0.01126, 0.00235, 0.00133, and 0.00966, respectively. The results show that the PLSR and PCA-SVR models are better than the PCA-ELM model, with the PCA-SVR model having the highest prediction accuracy. Overall, this study provides an effective method for LIBS technology in element content prediction and demonstrates that combining machine learning algorithms with LIBS can achieve high-precision detection of steel element content. In the future, algorithms and methods can be further optimized to improve model performance and prediction accuracy, such as combining deep learning and other modern algorithms to improve model prediction accuracy and robustness. In addition, more sample data can be used to optimize and improve the model, and more efficient algorithms and tools can be developed to process large-scale data. At the same time, more in-depth and systematic research and exploration can be conducted for specific problems and fields, in line with practical application needs and research directions.

Acknowledgment. This work was supported by Science and Technology Innovation Key Project of Fujian Province,China (2021G02009); The author also thanks the reviewers for their valuable comments and suggestions, which have improved the manuscript. Ji-Shi Zheng, Ling-Hua Kong, and Zhao-Lin Zhao proposed the methodology and assisted in revising the manuscript. Wei-Yu Hou conducted the experiments and drafted the manuscript. Jia-Cheng Yang, Hong-Ji Ye, and Xiang-Xu Ren assisted in operating the equipment and provided constructive feedback.

REFERENCES

[1] A. N. Conejo, J.-P. Birat, and A. Dutta, "A review of the current environmental challenges of the steel industry and its value chain," *Journal of Environmental Management*, vol. 259, p. 109782, 2020.
[2] H. Townsend, "Effects of alloying elements on the corrosion of steel in industrial atmospheres," *Corrosion*, vol. 57, no. 06, 2001.
[3] X. Hao, W. Yin, M. Strangwood, A. Peyton, P. Morris, and C. Davis, "Off-line measurement of decarburization of steels using a multifrequency electromagnetic sensor," *Scripta Materialia*, vol. 58, no. 11, pp. 1033–1036, 2008.
[4] C. Pasquini, J. Cortez, L. Silva, and F. B. Gonzaga, "Laser induced breakdown spectroscopy," *Journal of the Brazilian Chemical Society*, vol. 18, pp. 463–512, 2007.
[5] 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," *TrAC Trends in Analytical Chemistry*, vol. 143, p. 116385, 2021.
[6] Y. Zhang, T. Zhang, and H. Li, "Application of laser-induced breakdown spectroscopy (libs) in environmental monitoring," *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 181, p. 106218, 2021.

[7] T. Chen, T. Zhang, and H. Li, "Applications of laser-induced breakdown spectroscopy (libs) combined with machine learning in geochemical and environmental resources exploration," *TrAC Trends in Analytical Chemistry*, vol. 133, p. 116113, 2020.

[8] K. Liu, D. Tian, C. Li, Y. Li, G. Yang, and Y. Ding, "A review of laser-induced breakdown spectroscopy for plastic analysis," *TrAC Trends in Analytical Chemistry*, vol. 110, pp. 327–334, 2019.

[9] B. Busser, S. Moncayo, J.-L. Coll, L. Sancey, and V. Motto-Ros, "Elemental imaging using laser-induced breakdown spectroscopy: A new and promising approach for biological and medical applications," *Coordination Chemistry Reviews*, vol. 358, pp. 70–79, 2018.

[10] S. Rehse, H. Salimnia, and A. Miziolek, "Laser-induced breakdown spectroscopy (libs): an overview of recent progress and future potential for biomedical applications," *Journal of Medical Engineering & Technology*, vol. 36, no. 2, pp. 77–89, 2012.

[11] R. Gaudiuso, N. Melikechi, Z. A. Abdel-Salam, M. A. Harith, V. Palleschi, V. Motto-Ros, and B. Busser, "Laser-induced breakdown spectroscopy for human and animal health: A review," *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 152, pp. 123–148, 2019.

[12] D. Bulajic, G. Cristoforetti, M. Corsi, M. Hidalgo, S. Legnaioli, V. Palleschi, A. Salvetti, E. Tognoni, S. Green, D. Bates *et al.*, "Diagnostics of high-temperature steel pipes in industrial environment by laser-induced breakdown spectroscopy technique: the libsgrain project," *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 57, no. 7, pp. 1181–1192, 2002.

[13] R. Noll, H. Bette, A. Brysch, M. Kraushaar, I. Mönch, L. Peter, and V. Sturm, "Laser-induced breakdown spectrometry—applications for production control and quality assurance in the steel industry," *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 56, no. 6, pp. 637–649, 2001.

[14] R. Noll, I. Mönch, O. Klein, and A. Lamott, "Concept and operating performance of inspection machines for industrial use based on laser-induced breakdown spectroscopy," *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 60, no. 7-8, pp. 1070–1075, 2005.

[15] M. Markiewicz-Keszycka, X. Cama-Moncunill, M. P. Casado-Gavalda, Y. Dixit, R. Cama-Moncunill, P. J. Cullen, and C. Sullivan, "Laser-induced breakdown spectroscopy (libs) for food analysis: A review," *Trends in Food Science & Technology*, vol. 65, pp. 80–93, 2017.

[16] R. A. Multari, D. A. Cremers, J. A. M. Dupre, and J. E. Gustafson, "Detection of biological contaminants on foods and food surfaces using laser-induced breakdown spectroscopy (libs)," *Journal of Agricultural and Food Chemistry*, vol. 61, no. 36, pp. 8687–8694, 2013.

[17] Y. Zhang, C. Sun, L. Gao, Z. Yue, S. Shabbir, W. Xu, M. Wu, and J. Yu, "Determination of minor metal elements in steel using laser-induced breakdown spectroscopy combined with machine learning algorithms," *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 166, p. 105802, 2020.

[18] 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)," *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 178, p. 106112, 2021.

[19] Y. Tian, Q. Chen, Y. Lin, Y. Lu, Y. Li, and H. Lin, "Quantitative determination of phosphorus in seafood using laser-induced breakdown spectroscopy combined with machine learning," *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 175, p. 106027, 2021.

[20] 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)," *Chemometrics and Intelligent Laboratory Systems*, vol. 197, p. 103930, 2020.

[21] L. Zhang and Y. Yang, "Simultaneous quantitative analysis of non-metallic elements in coal by laser-induced breakdown spectroscopy assisted with machine learning," *Optik*, vol. 222, p. 165449, 2020.

[22] A. Maćkiewicz and W. Ratajczak, "Principal components analysis (pca)," *Computers & Geosciences*, vol. 19, no. 3, pp. 303–342, 1993.

[23] F. Anowar, S. Sadaoui, and B. Selim, "Conceptual and empirical comparison of dimensionality reduction algorithms (pca, kpca, lda, mds, svd, lle, isomap, le, ica, t-sne)," *Computer Science Review*, vol. 40, p. 100378, 2021.

[24] T. Zhang, S. Wu, J. Dong, J. Wei, K. Wang, H. Tang, X. Yang, and H. Li, "Quantitative and classification analysis of slag samples by laser induced breakdown spectroscopy (libs) coupled with support vector machine (svm) and partial least square (pls) methods," *Journal of Analytical Atomic Spectrometry*, vol. 30, no. 2, pp. 368–374, 2015.

[25] J. Guezenoc, A. Gallet-Budynek, and B. Bousquet, "Critical review and advices on spectral-based normalization methods for libs quantitative analysis," *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 160, p. 105688, 2019.

[26] T.-Y. Wu, H. Li, and S.-C. Chu, "Cppe: An improved phasmatodea population evolution algorithm with chaotic maps," *Mathematics*, vol. 11, no. 9, p. 1977, 2023.

[27] Y. Yang, L. Ma, and C. Ma, "Quantitative analysis of liquid steel element in libs using svr improved by particle swarm optimization," *Laser Optoelectron. Progress*, vol. 57, p. 053002, 2020.

[28] X. Y. L. D. L. Dong and C. Tong, "Quantitative analysis of iron slurry based on laser induced breakdown spectroscopy combined with mutual information feature selection partial least squares method," *Metallurgical Analysis*, vol. 42, no. 1, pp. 18–24.

[29] C.-M. Chen, S. Lv, J. Ning, and J. M.-T. Wu, "A genetic algorithm for the waitable time-varying multi-depot green vehicle routing problem," *Symmetry*, vol. 15, no. 1, p. 124, 2023.

[30] J. S. Manoharan, "Study of variants of extreme learning machine (elm) brands and its performance measure on classification algorithm," *Journal of Soft Computing Paradigm (JSCP)*, vol. 3, no. 02, pp. 83–95, 2021.

[31] S. Ding, H. Zhao, Y. Zhang, X. Xu, and R. Nie, "Extreme learning machine: algorithm, theory and applications," *Artificial Intelligence Review*, vol. 44, pp. 103–115, 2015.

[32] Y. Ma, Y. Peng, and T.-Y. Wu, "Transfer learning model for false positive reduction in lymph node detection via sparse coding and deep learning," *Journal of Intelligent & Fuzzy Systems*, vol. 43, no. 2, pp. 2121–2133, 2022.

[33] A. L. H. P. Shaik, M. K. Manoharan, A. K. Pani, R. R. Avala, and C.-M. Chen, "Gaussian mutation–spider monkey optimization (gm-smo) model for remote sensing scene classification," *Remote Sensing*, vol. 14, no. 24, p. 6279, 2022.