INTERNATIONAL JOURNAL OF
Minerals, Metallurgy and Materials

Volume 32, Number 10, October 2025
Page 2469

Research Article https://doi.org/10.1007/s12613-025-3145-3

Factor analysis and machine learning for predicting endpoint carbon content in converter steelmaking

Lihua Zhao[1),✉], *Shuai Yang*[1)], *Yongzhao Xu*[1)], *Zhongliang Wang*[2)], *Xin Liu*[2)], *and Yanping Bao*[2)]

1) School of Metallurgical and Ecological Engineering, University of Science and Technology Beijing, Beijing 100083, China
2) State Key Laboratory of Advanced Metallurgy, University of Science and Technology Beijing, Beijing 100083, China

(Received: 6 January 2025; revised: 1 April 2025; accepted: 3 April 2025)

Abstract: The endpoint carbon content in the converter is critical for the quality of steel products, and accurately predicting this parameter is an effective way to reduce alloy consumption and improve smelting efficiency. However, most scholars currently focus on modifying methods to enhance model accuracy, while overlooking the extent to which input parameters influence accuracy. To address this issue, in this study, a prediction model for the endpoint carbon content in the converter was developed using factor analysis (FA) and support vector machine (SVM) optimized by improved particle swarm optimization (IPSO). Analysis of the factors influencing the endpoint carbon content during the converter smelting process led to the identification of 21 input parameters. Subsequently, FA was used to reduce the dimensionality of the data and applied to the prediction model. The results demonstrate that the performance of the FA–IPSO–SVM model surpasses several existing methods, such as twin support vector regression and support vector machine. The model achieves hit rates of 89.59%, 96.21%, and 98.74% within error ranges of ±0.01%, ±0.015%, and ±0.02%, respectively. Finally, based on the prediction results obtained by sequentially removing input parameters, the parameters were classified into high influence (5%–7%), medium influence (2%–5%), and low influence (0–2%) categories according to their varying degrees of impact on prediction accuracy. This classification provides a reference for selecting input parameters in future prediction models for endpoint carbon content.

Keywords: converter; endpoint carbon content; parameter classification; factor analysis; improved particle swarm optimization; support vector machine

1. Introduction

Precise control of carbon content is a critical aspect of the converter steelmaking process. Excessively high endpoint carbon content reduces the FeO content in the slag [1], whereas excessively low endpoint carbon content leads to over-oxidation of the molten steel and increases the deoxidation requirements in subsequent processes. In most Chinese enterprises, the management of this process typically employs a model of human–machine collaboration, but the proportion of human factors still remains relatively high. Techniques such as catch carbon practice, carbon pick-up practice, and high-carbon turndown practice are commonly employed to regulate the endpoint carbon content, which requires a high level of technical expertise from operators. However, manual judgment errors often lead to deviations in the endpoint carbon content, resulting in unnecessary resource and energy consumption. As a result, relying mainly on manual experience is insufficient to meet the demands of modern production [2–4].

With the continuous advancement of artificial intelligence technology, significant progress has been achieved in the intelligent transformation of the steel industry [5]. The development of converter endpoint prediction models has greatly enhanced the efficiency and quality of smelting processes. Currently, extensive research has been conducted on predicting endpoint carbon content in converters. For example, Zhang *et al.* [6] developed a converter carbon content prediction model by combining affinity propagation clustering with radial basis function networks to predict the endpoint carbon content of Q235B steel. This model offered a promising method for endpoint carbon content prediction. The model achieved a hit rate of 93.75% when the error range was within ±0.025%. Similarly, Liu *et al.* [7] established a prediction model for the endpoint carbon content of HRB4Nb-8 steel using the t-distributed stochastic neighbor embedding (t-SNE), particle swarm optimization (PSO), and back-propagation (BP) network. Their model achieved a hit rate of 98% when the error range was within ±0.02%. Compared with manual control, these models significantly improved prediction accuracy.

In practical production scheduling, steel grade transitions occur frequently, necessitating a broader application scope for endpoint prediction models. Li *et al.* [8] proposed a prediction model for endpoint carbon content in converters by integrating the BP network with a nonlinear least squares al-

---

✉ Corresponding author: Lihua Zhao E-mail: zhaolihua@metall.ustb.edu.cn
© University of Science and Technology Beijing 2025

gorithm. While the model achieved a hit rate of 80% within an error range of ±0.025%, its performance was suboptimal when processing large-scale data, and its hit rate remained relatively low. To enhance hit efficiency and model applicability for large-scale data, further studies were conducted. Liu *et al.* [9] applied principal component analysis (PCA) to reduce data dimensionality and developed a prediction model using a genetic algorithm-optimized backpropagation (GA–BP) network to forecast the endpoint carbon content across multiple steel grades. Wang *et al.* [10] built on a three-stage decarburization theory and considered factors such as oxygen lance height and top-blown oxygen flow rate. They employed the finite difference method to establish a continuous prediction model for carbon content in a 120-ton converter. The prediction results showed that the model achieved a hit rate of 85% within an error range of ±0.02%. Zhang *et al.* [11] compared the performance of the BP network, extreme learning machine (ELM), and support vector machine (SVM) in predicting the endpoint carbon content across multiple steel grades, concluding that SVM exhibited higher prediction accuracy. Similarly, Wang *et al.* [12] demonstrated that twin support vector regression (TSVR) outperformed the BP network in accurately predicting the endpoint carbon content for various steel grades simultaneously. Furthermore, Wang *et al.* [13] developed a model integrating grey relational analysis, partial correlation analysis, and TSVR to predict the endpoint carbon content for multiple steel grades. The model achieved a hit rate of 90% within an error range of ±0.005%. Gu *et al.* [14] developed a model combining case-based reasoning and long short-term memory to predict the endpoint carbon content of multiple steel grades in converters. The model achieved a hit rate of 91% within an error range of ±0.02%. In recent years, numerous novel methodologies have emerged for predicting the endpoint carbon content in converters. These include posterior probability and intra-cluster feature weight online dynamic feature selection algorithm [15], improved conditional generative adversarial network (ICGAN) [16], self-attention-based convolutional parallel network (SabCP) [17], and supervised weighting-based local structure preserving projection [18]. With continuous advancements in prediction methods, the hit rates of such models are steadily improving. In the metallurgical industry, we strive to address complex problems using relatively simple methods. Notably, many researchers have focused on optimizing SVM for endpoint carbon content prediction, as SVM performs well in addressing nonlinear, high-dimensional problems. However, it still requires improvement in processing large-scale data efficiently [19–23]. TSVR is capable of processing large-scale data, but its generalization capability can be impacted when the dataset becomes complex [24].

All the aforementioned models focus on the establishment of the input parameter system. The dimensionality and composition of the input parameters significantly influence the prediction accuracy of these models. Undeniably, as the number of steel grades being predicted increases, the number of influencing parameters also grows, resulting in higher data dimensionality. This increase, in turn, increases computation time and makes dimensionality reduction essential. Moreover, if the input data include numerous components unrelated to the endpoint carbon content, they can negatively affect prediction accuracy. However, previous studies often prematurely selected a single type of oxygen supply parameter—such as smelting parameters, raw material and auxiliary material parameters, oxygen consumption, or oxygen flow rate—as input parameters. The impact of input parameters on the prediction accuracy of endpoint carbon content was largely overlooked.

To address the inefficiency of SVM in processing large-scale data, the improved particle swarm optimization (IPSO) algorithm is used to optimize the parameter selection process and offers a viable solution. With the incorporation of inertia weight and dynamic acceleration factors, IPSO expands the search area and improves accuracy. Factor analysis (FA) helps clarify the intrinsic structural relationships within the data, simplifies data structures, and is well-suited for managing the high-dimensional, large-scale data typical of converter steelmaking processes. Moreover, IPSO–SVM has been applied in fields such as biology, medicine, geology, and the internet [25–33], but its application in metallurgy remains relatively limited. Therefore, the present study enhanced the PSO by introducing a nonlinear decreasing inertia weight and used the resulting IPSO to optimize the penalty factor and kernel parameters of the SVM. Combined with factor analysis, a prediction model for endpoint carbon content in converter steelmaking was developed. First, an input parameter system containing 21 parameters was constructed, and correlation analysis was performed on the preprocessed data. Additionally, FA was performed to reduce the dimensionality of the original variables. The performance of the FA–IPSO–SVM model was then evaluated using a test dataset. Finally, each parameter in the input parameter system was sequentially removed, and the FA–IPSO–SVM model was applied again to predict the endpoint carbon content. This process demonstrated that the varying significance levels of correlation between input parameters and endpoint carbon content influenced the hit rate of the model. According to the prediction results, the input parameters were categorized into high, medium, and low influence groups according to their impact on the hit rate of the endpoint carbon content model. This classification provided a reference for selecting input parameters in future endpoint carbon content prediction models. Accurately predicting the endpoint carbon content in converters is conducive to improving production efficiency, reducing costs, and decreasing energy consumption.

2. Analysis of factors influencing the endpoint carbon content in converter steelmaking and the development of a parameter system

The physicochemical reactions involved in converter steelmaking are highly complex. To develop an accurate model for predicting the endpoint carbon content, it is essential to identify the factors influencing endpoint carbon content and create a comprehensive parameter system. The steelmaking cycle and parameter distribution in a converter are illustrated in Fig. 1. Each stage before tapping includes factors that influence the endpoint carbon content. They can be classified into raw material and auxiliary material parameters, smelting parameters, and oxygen supply parameters based on their different characteristics. The process from lifting the ladle of hot metal to charging scrap and hot metal into the converter involves 11 raw material and auxiliary material parameters, including molten iron weight, molten iron temperature, carbon content, silicon content, manganese content, phosphorus content, sulfur content, the amount of scrap charged, the amount of lime, the amount of dolomite, and the amount of ore charged. Lime and dolomite not only act as slagging agents during the steelmaking process but also help adjust the temperature of the molten steel and remove phosphorus and sulfur [34–35]. Scrap and ore function as coolants to maintain the smelting environment's temperature. Therefore, the parameters must include the amounts of lime and lightly calcined dolomite added, as well as the amounts of scrap and ore charged. The carbon content of molten iron directly impacts the carbon content at the initial stage of smelting. Under the same smelting conditions, the higher carbon content in the molten iron increases the difficulty of decarburization. Therefore, the parameter system must include the carbon content of the molten iron.

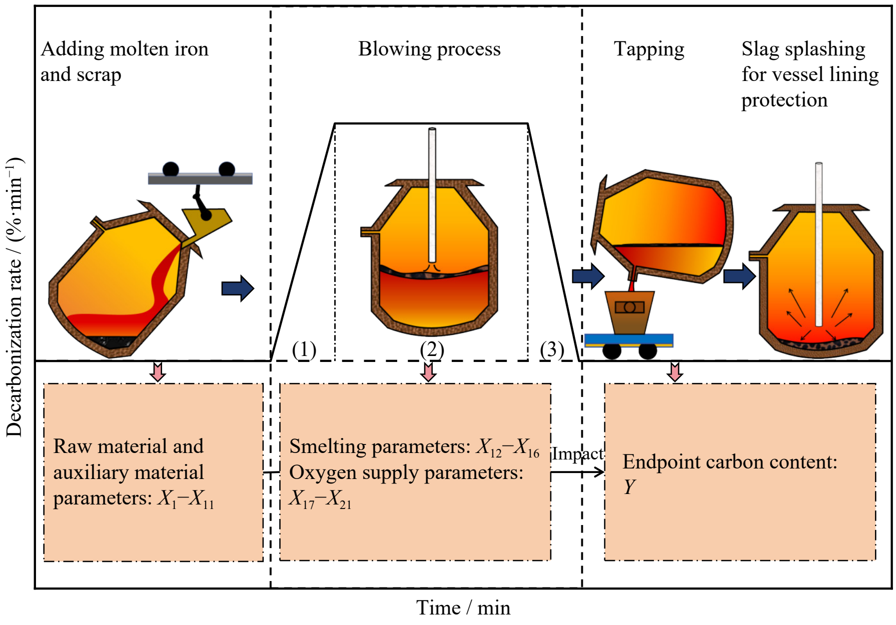

Fig. 1. Process and parameter distribution during converter smelting cycle. X_1—Molten iron weight; X_2—Molten iron temperature; X_3—C content of molten iron; X_4—Si content of molten iron; X_5—Mn content of molten iron; X_6—P content of molten iron; X_7—S content of molten iron; X_8—Amount of scrap added; X_9—Amount of lime added; X_{10}—Amount of dolomite added; X_{11}—Amount of ore added; X_{12}—TSC temperature; X_{13}—TSC carbon content; X_{14}—TSO temperature; X_{15}—Lance position; X_{16}—Bottom blowing time; X_{17}—Oxygen consumption; X_{18}—Oxygen pressure; X_{19}—Oxygen flow rate; X_{20}—Free oxygen content; X_{21}—Oxygen supply time. (1) Initial decarburization stage; (2) Intermediate decarburization stage; (3) Final decarburization stage.

The main task of converter steelmaking is decarburization. During the initial decarburization stage, the stronger affinity of [Si] and [Mn] with [O] compared with [C] with [O] causes [Si] and [Mn] to react with [O] first. This process consumes a significant amount of oxygen, leaves insufficient oxygen for the decarburization reaction, and results in a slower decarburization rate. As the contents of [Si] and [Mn] gradually decrease, more oxygen becomes available for [C], which accelerates the decarburization rate. In the final decarburization stage, the temperature of the molten bath increases, and the amount of (FeO) rises, leading to a slower decarburization rate. Additionally, the heat released during the decarburization process and the sensible heat of the molten iron itself serve as the two primary energy sources for converter steelmaking. Therefore, the parameter system must include the silicon and manganese contents of the molten iron, as well as its temperature. During the dephosphorization and desulfurization reactions at the steel–slag interface, CaO is consumed, and the presence of [P] and [S] in the molten iron can influence the progress of these reactions. Consequently, the phosphorus and sulfur contents of the molten iron should also be included in the parameter system.

The blowing process includes both smelting parameters and oxygen supply parameters. The smelting parameters consist of the temperature and carbon content measured by the Temperature-Sample-Carbon (TSC) probe, as well as the temperature measured by the Temperature-Sample-Oxygen (TSO) probe [36], lance position, and bottom blowing time during the blowing process of the converter. The oxygen supply parameters include oxygen consumption, oxygen pressure, oxygen flow rate, free oxygen content, and oxygen supply time. During the blowing process, the composition and temperature are measured through TSO and TSC probe. TSC probe measures the molten steel temperature when approximately 85% of the oxygen has been blown, while simultaneously determining the carbon content [37]. TSO probe measures the molten steel temperature and oxygen content at the end of the blowing process, which are used to determine the carbon content. If all indicators meet the endpoint requirements, tapping can proceed. Otherwise, additional blowing is necessary. Therefore, TSO probe measurements of temperature and elemental composition enable precise endpoint determination in converter smelting processes [38]. Temperature is a critical factor influencing the endpoint carbon content, as it directly affects the rate and equilibrium state of the decarburization reaction. During converter smelting, increasing the temperature can promote the oxidation of carbon and thus reduce the endpoint carbon content in the converter. Conversely, decreasing the temperature slows down the carbon oxidation reaction, resulting in a higher endpoint carbon content. To ensure a smooth and efficient blowing process without the need for multiple temperature measurements, only two temperatures can be collected using a sublance, which are TSC temperature and TSO temperature. The reason for considering the TSO temperature is that many steel plants cannot directly determine the carbon content when measuring the TSO temperature. Instead, they need to calculate the carbon content retrospectively based on the carbon-oxygen product. Therefore, these temperatures are included in the predictive indicator system. The bottom blowing time can influence the uniformity of the decarburization reaction as well as the distribution of temperature and composition within the molten bath, thereby affecting the endpoint carbon content. Additionally, TSC carbon content influences the subsequent blowing duration. Therefore, both bottom blowing time and TSC carbon content are incorporated into the indicator system. FeO in the slag plays a crucial role during converter smelting and acts not only as a reactant that affects the decarburization reaction but also influences the melting rate of lime. The FeO content in the slag is mainly adjusted by the lance position. In the initial stage of blowing, the lance position is raised to increase the FeO content in the slag to promote slag formation and facilitate the decarburization reaction. In the middle stage, the lance position is lowered to accelerate the decarburization rate. However, in the later stage, the decarburization rate slows down. Therefore, the lance position is included in the indicator system.

The relationship between oxygen and carbon content in the molten bath is closely interconnected [39–40], as reflected in the regulation of oxygen supply parameters during the blowing process. Proper oxygen control can enhance the kinetic conditions of the decarburization process, and effectively regulate the carbon content in the molten steel. At the endpoint of the blowing process, the oxygen supply parameters, available through the manufacturing execution system, include oxygen consumption, oxygen pressure, oxygen flow rate, free oxygen content, and oxygen supply duration. As such, these five parameters are included in the indicator system.

In summary, based on the analysis of converter smelting processes and previous research, 21 parameters have been selected to form the input parameter system. Among these, lance position, oxygen pressure, and oxygen flow rate were all time-series data. In this study, the lance position refers to the average lance position, the oxygen pressure refers to the average oxygen pressure, and the oxygen flow rate refers to the total oxygen flow during the blowing process.

3. Methods

3.1. Data collection

The raw production data used in this paper was sourced from a 120-t converter in a steel plant in China. The selected data did not distinguish between steel grades and was randomly collected, resulting in a dataset of 1056 heats. Some abnormal data were removed for the following reasons: (1) data was missing due to issues during data collection, transmission, or processing; (2) abnormal data was generated due to sensor malfunctions; (3) the final composition of the molten steel exceeded the required range. The statistical results of the processed data are presented in Table 1.

The selected parameters had different dimensions and units. To address the inconvenience caused by dimensional

Table 1. Statistical results of variables affecting endpoint carbon content

Symbol	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}	X_{19}	X_{20}	X_{21}	Y
Unit	t	°C	wt%	wt%	wt%	wt%	wt%	t	kg	kg	kg	°C	wt%	°C	m	min	m^3	MPa	$m^3 \cdot h^{-1}$	ppm	min	wt%
Minimum	90	1274	3.12	0.04	0.21	0.07	0.01	3.2	2342	0	0	1505	0.03	1556	1.25	12.85	4891	1.34	20260	129	12.55	0.01
Maximum	126	1410	4.94	0.78	0.64	0.13	0.08	30.2	9180	1206	5816	1674	0.89	1689	1.97	18.53	7142	1.63	26043	788	18.38	0.18
Mean	114	1340	4.41	0.31	0.33	0.11	0.02	22.9	4799	652	637	1579	0.63	1618	1.47	15.25	5946	1.58	23293	323	15.10	0.11
Range	36	136	1.82	0.74	0.43	0.06	0.07	27.0	6838	1206	5816	169	0.86	133	0.72	5.68	2251	0.29	5783	659	5.83	0.17
Standard deviation	4	24	0.24	0.11	0.06	0.01	0.02	4.1	1037	199	1256	93	0.54	20	16.17	50.09	282	0.01	567	114	47.65	0.03

differences and the varying ranges of certain indicators in the production data during model training, normalization was applied after removing abnormal values. In this paper, a linear transformation as shown in Eq. (1) was used to map the original values μ to the interval [0, 1], resulting in standardized values.

$$\mu^* = \frac{\mu - \mu_{\min}}{\mu_{\max} - \mu_{\min}} \tag{1}$$

where μ^* represents the standardized data, μ represents the data to be standardized, $\mu_{\max}$ represents the maximum value of the data after outliers are removed, and $\mu_{\min}$ represents the minimum value of the data after outliers are removed.

3.2. Model establishment

3.2.1. Correlation analysis

Pearson correlation analysis is commonly used to assess the correlation between two random variables [41–42]. It measures the degree of correlation by analyzing the elements of the variables that are related. The formula for calculating the correlation coefficient is shown in Eq. (2):

$$r = \frac{\sum_{i=1}^{n}\left(X_{p,i} - \overline{X}_p\right)\left(X_{q,i} - \overline{X}_q\right)}{\sqrt{\sum_{i=1}^{n}\left(X_{p,i} - \overline{X}_p\right)^2 \sum_{i=1}^{n}\left(X_{q,i} - \overline{X}_q\right)^2}} \tag{2}$$

where $X_{p,i}$ is the ith value of the pth variable, $\overline{X}_p$ is the mean of the pth variable, $X_{q,i}$ is the ith value of the qth variable, and $\overline{X}_q$ is the mean of the qth variable.

3.2.2. FA

Factor analysis proceeds from the dependency relationships within the correlation matrix of original variables, summarizing highly correlated variables into a small number of underlying, unobservable factors through analysis, thereby achieving the purpose of dimensionality reduction [43]. These factors are capable of explaining the correlations among variables and simplifying the data structure. Factor analysis is an extension of principal component analysis, with a greater emphasis on describing the correlational relationships among the original variables compared to principal component analysis. Assuming an observable random variable $\boldsymbol{z} = (z_1, z_2, \cdots, z_n)^{\mathrm{T}}$, each random variable is represented by m common factors, as shown in Eq. (3):

$$\begin{cases} z_1 = a_{11}F_1 + a_{12}F_2 + \cdots + a_{1m}F_m + \delta_1 \\ z_2 = a_{21}F_1 + a_{22}F_2 + \cdots + a_{2m}F_m + \delta_2 \\ \qquad \vdots \\ z_n = a_{n1}F_1 + a_{n2}F_2 + \cdots + a_{nm}F_m + \delta_m \end{cases} \tag{3}$$

where F_m is the common factor; δ_m is the unique factor, which representing the portion of the original variable that cannot be explained; a_{nm} is an element of the factor loading matrix, which indicates the contribution of the nth variable to the mth factor.

3.2.3. SVM

SVMs are commonly used to address non-linear regression prediction problems by incorporating an insensitive loss function, ε [44]. The sample space is mapped to a feature space of high or potentially infinite dimensions through the application of a nonlinear transformation. This transformation transforms the task of finding a complex hyperplane into solving a convex optimization problem within convex constraints [45–47].

In an SVM, a regression model is created by mapping the features into a high-dimensional space, as shown in Eq. (4):

$$f(x) = \boldsymbol{\omega}^{\mathrm{T}}\varphi(x) + b \tag{4}$$

where $\boldsymbol{\omega}$ is a vector of weight coefficients, $\boldsymbol{\omega}\in\mathbf{R}$; $\varphi(x)$ is a nonlinear mapping function; b is a constant, $b\in\mathbf{R}$.

The structural risk minimization principle is applied to find the optimal $\boldsymbol{\omega}$ and b, which transforms the estimation function into a minimization function, as shown in Eqs. (5)–(8):

$$\min_{\omega,b} \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{m}(\xi_i + \xi_i^*) \tag{5}$$

$$y_i - (\boldsymbol{\omega}\cdot\varphi(x_i) + b) \le \varepsilon + \xi_i \tag{6}$$

$$(\boldsymbol{\omega}\cdot\varphi(x_i) + b) - y_i \le \varepsilon + \xi_i^* \tag{7}$$

$$\xi_i, \xi_i^* \ge 0\,(i = 1, 2, \cdots, m) \tag{8}$$

where C is the penalty factor, ε is the insensitive loss function, ξ_i and ξ_i^* are the slack variables and denote the upper-bound exceedance error and the lower-bound exceedance error, respectively, x_i is the ith independent variable, and y_i is the ith dependent variable.

Through the application of the Lagrange multiplier and the introduction of the kernel function, the regression decision function $H(x)$ is given by Eq. (9):

$$H(x) = \sum_{i=1}^{n}(a_i - a_i^*)K(x_i, x_j) + b \tag{9}$$

where $(a_i - a_i^*)$ is the specified Lagrangian factor, $K(x_i, x_j)$ is the kernel function, and b is the deviation value.

In terms of kernel function selection, this paper chooses the Gaussian radial basis function as the kernel function in the SVM. The expression for this kernel function is given in Eqs. (10)–(11):

$$K(x_i, x_j) = \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\sigma^2}\right) \tag{10}$$

$$g = \frac{1}{2\sigma^2} \tag{11}$$

where σ is the width of the kernel function, and g is the kernel parameter.

3.2.4. Improved particle swarm optimization

The PSO algorithm is inspired by the food-searching behavior of certain animal populations [48]. In a d-dimensional search space, a swarm of particles is present, each characterized by three parameters: velocity, position, and fitness. First, a group of particles is randomly selected and initialized. Each particle moves through the d-dimensional target search space at a variable velocity, updating its position based on the information of other particles. The optimal individual position and the optimal global position are recorded during this process [49–50]. The following assumptions are made for the

PSO algorithm: In a d-dimensional search space with a certain number of particles, $\boldsymbol{V}$ represents the velocity of particle movement, and $\boldsymbol{L}$ denotes the position of the particle. At iteration k, the velocity and position information of the ith particle is shown in Eqs. (12)–(13):

$$\boldsymbol{L}_i^k = \left(\boldsymbol{l}_{i1}^k, \boldsymbol{l}_{i2}^k, \cdots, \boldsymbol{l}_{id}^k\right) \tag{12}$$

$$\boldsymbol{V}_i^k = \left(\boldsymbol{v}_{i1}^k, \boldsymbol{v}_{i2}^k, \cdots, \boldsymbol{v}_{id}^k\right) \tag{13}$$

where $\boldsymbol{l}_{id}^k$ is the position of the ith particle in the dth dimension at the kth iteration, and $\boldsymbol{v}_{id}^k$ is the velocity of the ith particle in the dth dimension at the kth iteration.

There are two key states of particles in the evolutionary process: the individual extremum and the global extremum. The individual extremum represents the best solution found by a single particle throughout its history of iteration. In contrast, the global extremum reflects the best solution found by the entire population during the optimization process. As the iteration progresses to the kth step, the historical best solution of each particle is denoted as $\boldsymbol{E}_i^k$, while the historical best solution of the entire population is denoted as $\boldsymbol{E}_g^k$, as shown in Eqs. (14)–(15):

$$\boldsymbol{E}_i^k = \left(\boldsymbol{e}_{i1}^k, \boldsymbol{e}_{i2}^k, \cdots, \boldsymbol{e}_{id}^k\right) \tag{14}$$

$$\boldsymbol{E}_g^k = \left(\boldsymbol{e}_{g1}^k, \boldsymbol{e}_{g2}^k, \cdots, \boldsymbol{e}_{gd}^k\right) \tag{15}$$

where $\boldsymbol{e}_{id}^k$ is the historical optimal solution of the ith particle in the dth dimension at the kth iteration, and $\boldsymbol{e}_{gd}^k$ is the historical optimal solution of the current population in the dth dimension at the kth iteration.

The position and velocity information of the ith particle are shown in Eqs. (16)–(17):

$$\boldsymbol{L}_i^{k+1} = \boldsymbol{L}_i^k + \boldsymbol{V}_i^{k+1} \tag{16}$$

$$\boldsymbol{V}_i^{k+1} = \boldsymbol{V}_i^k + c_1 r_1\left(\boldsymbol{E}_i^k - \boldsymbol{X}_i^k\right) + c_2 r_2\left(\boldsymbol{E}_g^k - \boldsymbol{X}_i^k\right) \tag{17}$$

where r_1 and r_2 are random numbers between [0, 1], and c_1 and c_2 are acceleration factors.

The classic PSO algorithm has two main limitations: it is prone to getting stuck in local optima, and it suffers from a lack of sufficient population diversity [51]. The inertia weight factor λ influences both the ability of the particles to find local and global optima. To address these two key drawbacks of the PSO algorithm, this study adopts a nonlinear decreasing inertia weight, as shown in Eq. (18). The velocity update for the ith particle, after introducing the inertia weight, is shown in Eq. (19):

$$\lambda = (\lambda_{\max} - \lambda_{\min}) \cdot \left(\frac{k}{k_{\max}}\right)^2 + (\lambda_{\min} - \lambda_{\max})\frac{2k}{k_{\max}} + \lambda_{\max} \tag{18}$$

$$\boldsymbol{V}_i^{k+1} = \lambda \boldsymbol{V}_i^k + c_1 r_1\left(\boldsymbol{E}_i^k - \boldsymbol{L}_i^k\right) + c_2 r_2\left(\boldsymbol{E}_g^k - \boldsymbol{L}_i^k\right) \tag{19}$$

where $\lambda_{\max}$ is the maximum value of λ, $\lambda_{\min}$ is the minimum value of λ, and $k_{\max}$ is the maximum iteration number.

3.2.5. Counting process

A prediction model for endpoint carbon content, based on FA integrated with IPSO–SVM, is developed to predict the endpoint carbon content during converter steelmaking. The process is illustrated in Fig. 2. First, the original data undergoes dimensionality reduction through FA processing. The reduced dataset, which consists of 1056 heats, is then divided into two categories, with a 7:3 split for the training and test sets. Specifically, 739 heats processed by FA are used to train the prediction model for endpoint carbon content, while the remaining 317 heats are used to test the model's performance.

3.3. Setting of model parameters

In the support vector machine algorithm, the penalty factor C is set to a search range of $0.1 \leq C \leq 100$, while the kernel parameter g, which controls the influence of a single training

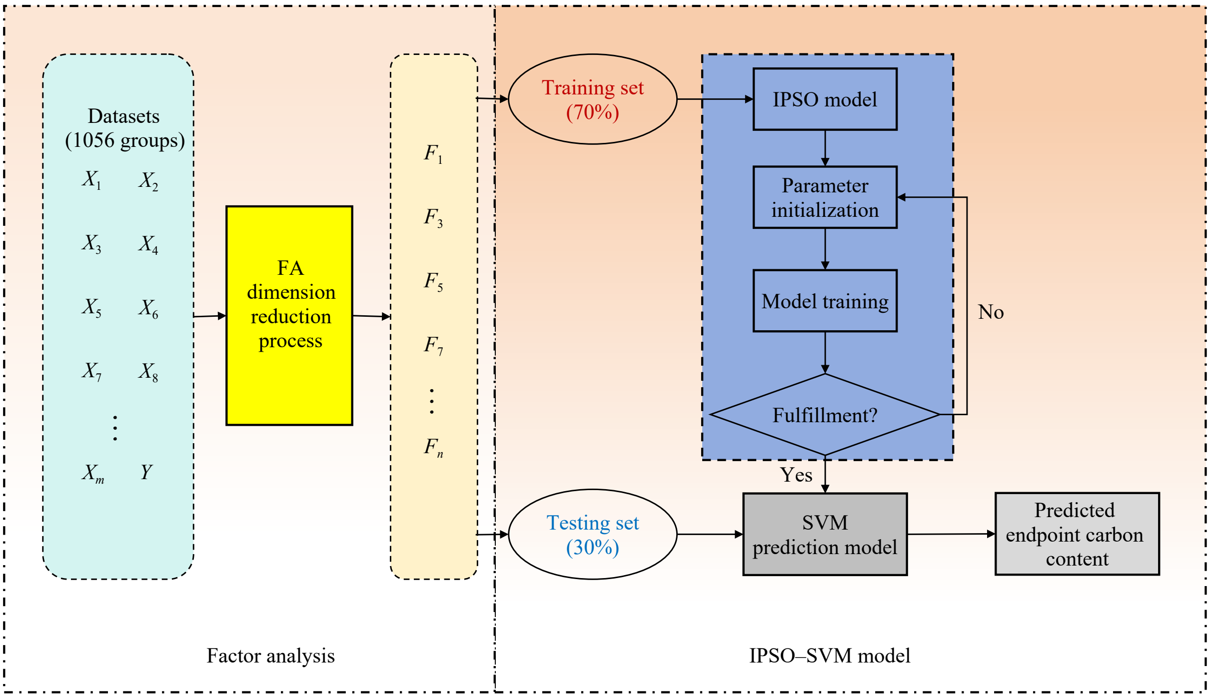

Fig. 2. Flowchart of model prediction process.

example, is set to $0.1 \le g \le 100$. For the PSO parameters, the initial value of the acceleration factor c_1 is set to 1.5, and the initial value of c_2 is set to 1.7. The maximum population size is set to 20 and the maximum number of iterations is set to 50. The maximum value of the inertia weight λ is set to 0.9 and the minimum value of λ is set to 0.4. The penalty factor C and kernel parameter g are critical parameters that directly influence the predictive accuracy of the model. The optimization algorithm is used to determine the optimal values for the penalty factor C and the kernel parameter g in the support vector machine. In this study, the root-mean-square error (RMSE) between the predicted and actual endpoint carbon content in the training set serves as the fitness function for optimizing the SVM model. The IPSO algorithm selects the parameters that minimize the RMSE on the training set as the optimal parameters. The iteration process is shown in Fig. 3, where it can be observed that after 18 iterations, the optimal fitness value stabilizes at 0.0092 and the optimal parameters are determined to be $C = 18.5154$ and $g = 0.3192$.

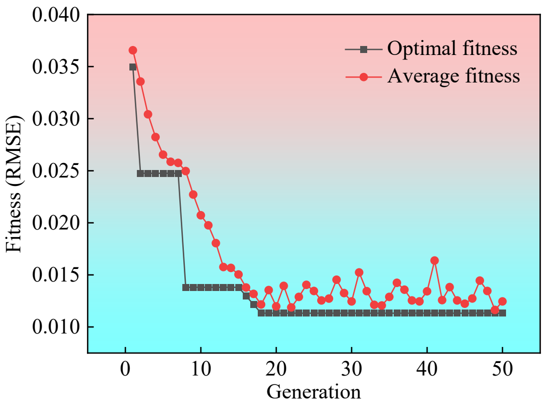

Fig. 3. Variation of fitness with the number of iterations.

3.4. Model evaluation

Five statistical evaluation metrics—namely, the coefficient of determination (R^2), mean absolute error (MAE), root-mean-square error (RMSE), mean absolute percentage error (MAPE), and hit rate—were used to assess the performance of the different models. The calculation formulas for these metrics are provided in Eqs. (20)–(24):

$$R^2 = \frac{\sum_{i=1}^{N}\left(Y_i - \overline{Y}\right)^2 - \sum_{i=1}^{N}\left(Y_i^* - Y_i\right)^2}{\sum_{i=1}^{N}\left(Y_i - \overline{Y}\right)^2} \tag{20}$$

$$\text{MAE} = \frac{\sum_{i=1}^{N}\left|Y_i^* - Y_i\right|}{N} \tag{21}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(Y_i^* - Y_i)^2}{N}} \tag{22}$$

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|Y_i - Y_i^*\right|}{Y_i} \tag{23}$$

$$\text{Hit rate} = \frac{N_{(|\text{Pre}-\text{Act}| \le \psi)}}{N_{\text{test}}} \times 100\% \tag{24}$$

where N is the number of data points after outlier removal; Y_i^* is the model's predicted value of the endpoint carbon content; Y_i is the true value of the endpoint carbon content; $\overline{Y}$ is the mean value of the endpoint carbon content; $N_{(|\text{Pre}-\text{Act}| \le \psi)}$ is the number of data points within the error range between the actual and predicted values, with $\psi = 0.01\%, 0.015\%, 0.02\%$ for endpoint carbon content prediction; N_{test} is the total number of test data points.

4. Results and discussion

4.1. Results of data analysis

4.1.1. Correlation analysis and applicability assessment

The correlation heatmap is shown in Fig. 4. The color gradient represents the degree of correlation. The interlaced dot matrix between two parameters turning red indicates a stronger negative correlation, while approaching white indicates a stronger positive correlation. An asterisk (*) denotes the significance of the correlation, with * representing $P <$ 0.05, which indicates a significant correlation between the parameters, and the absence of * representing $P > 0.05$. In the converter steelmaking process, various parameters reflected different aspects of the smelting process. For example, the amount of scrap, molten iron weight, and molten iron temperature reflected the raw material characteristics of converter steelmaking, while lance position and blowing time reflected the blowing characteristics. Correlation analysis was conducted on the selected 21 parameters; interactions and correlations of varying significance levels were consequently revealed. For example, a strong negative correlation was observed between the free oxygen content in the molten steel and the endpoint carbon content, which indicated that the endpoint carbon content decreased as the free oxygen content in the molten steel increased. Furthermore, a strong positive correlation was found between oxygen supply time and oxygen consumption, which suggested that longer oxygen supply time led to greater oxygen consumption.

The 21 parameters were ranked based on their absolute values of the correlation coefficients with the endpoint carbon content, resulting in the following order: X_{20}, X_{12}, X_4, X_{13}, X_3, X_{19}, X_1, X_{10}, X_9, X_{17}, X_{14}, X_5, X_8, X_7, X_{15}, X_{21}, X_{11}, X_{16}, X_6, X_2, X_{18}. Among these, X_{15}, X_{21}, X_{11}, X_{16}, X_6, X_2, and X_{18} did not show significant correlations with the endpoint carbon content.

The relevant parameters influencing the endpoint carbon content were input into SPSS 25 and yielded a Kaiser–Meyer–Olkin (KMO) value of 0.654, which was greater than 0.6 and indicated that factor analysis was appropriate. In Bartlett's test of sphericity, the P-value was 0, which was less than 0.05, suggesting that there were significant correlations among the variables. Therefore, it was appropriate to extract

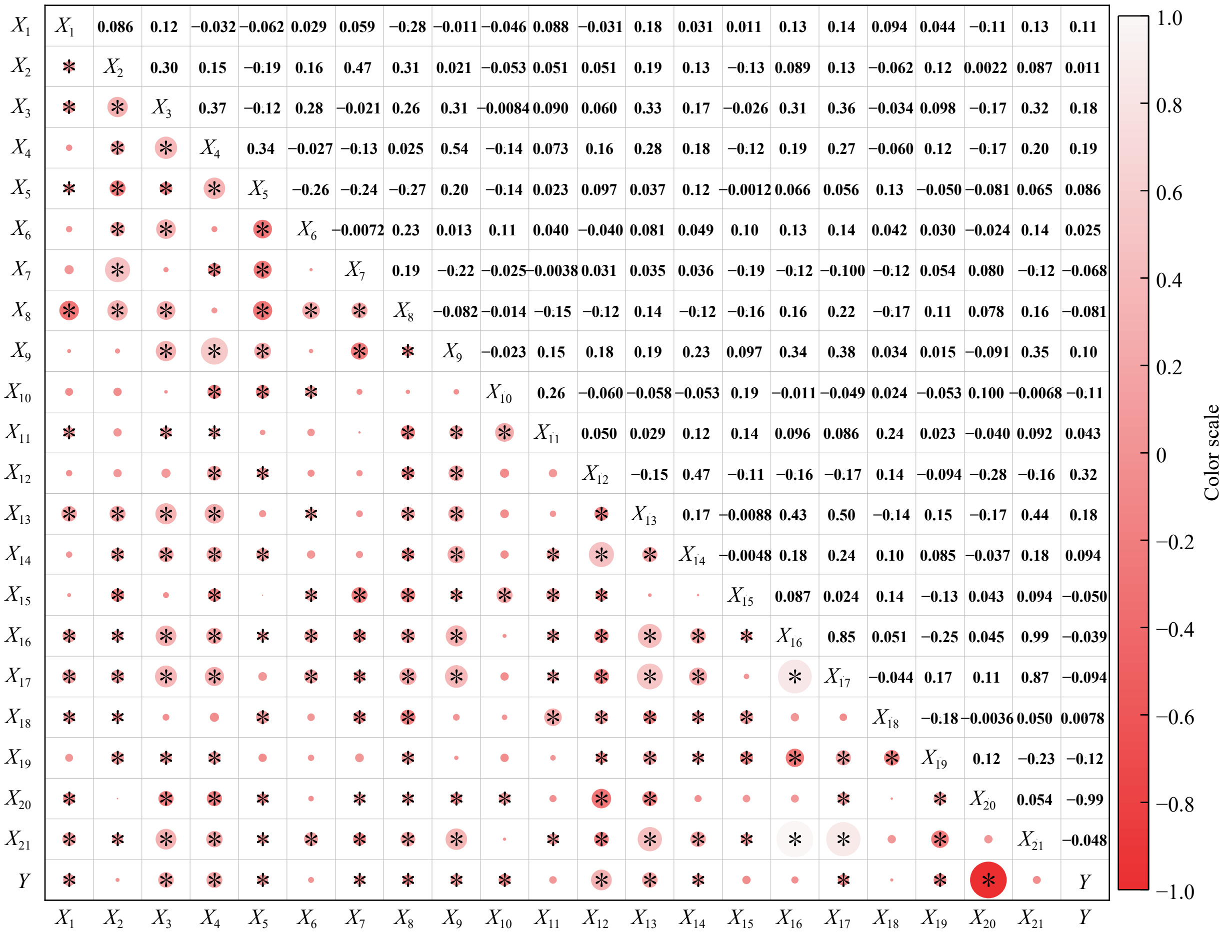

Fig. 4. Heat map of Pearson correlation coefficient.

common factors, which made factor analysis suitable for this study.

4.1.2. Extraction and representation of common factors

Principal component analysis was used to extract the variance of the common factors from the 21 variables, as shown in Table 2. Only X_6 had a low extraction value, while the variance of the common factors for the remaining 20 indicators was greater than 0.7, which indicated that the factor analysis was effective. The extraction of common factors and the total variance explained, as shown in Fig. 5, revealed that 13 common factors were extracted, with a cumulative variance contribution rate of 85.53%, which demonstrated the effectiveness of the analysis.

The expression of the common factor was derived from the score coefficient matrix, as shown in Table 3.

4.2. Comparison of FA–IPSO–SVM model with other established models

To evaluate the generalization performance of the FA–IPSO–SVM model, it was compared with other known models using 317 sets of untrained test data. Table 4 shows the performance of the various models, with the FA–IPSO–SVM model outperforming the IPSO–SVM, TSVR, and SVM models. The FA–IPSO–SVM model achieved the highest R^2 value of 0.960, and the lowest MAE, RMSE, and MAPE values of 0.005, 0.007, and 0.052, respectively. The

Table 2. Common factor variance of predictive evaluation indicators

Variable	Symbol	Starting	Extraction
Molten iron weight	X_1	1.000	0.935
Molten iron temperature	X_2	1.000	0.788
C content of molten iron	X_3	1.000	0.759
Si content of molten iron	X_4	1.000	0.821
Mn content of molten iron	X_5	1.000	0.715
P content of molten iron	X_6	1.000	0.654
S content of molten iron	X_7	1.000	0.762
Amount of scrap added	X_8	1.000	0.743
Amount of lime added	X_9	1.000	0.790
Amount of dolomite added	X_{10}	1.000	0.941
Amount of ore added	X_{11}	1.000	0.953
TSC temperature	X_{12}	1.000	0.942
TSC carbon content	X_{13}	1.000	0.867
TSO temperature	X_{14}	1.000	0.945
Lance position	X_{15}	1.000	0.988
Bottom blowing time	X_{16}	1.000	0.961
Oxygen consumption	X_{17}	1.000	0.920
Oxygen pressure	X_{18}	1.000	0.936
Oxygen flow rate	X_{19}	1.000	0.895
Free oxygen content	X_{20}	1.000	0.851
Oxygen supply time	X_{21}	1.000	0.979

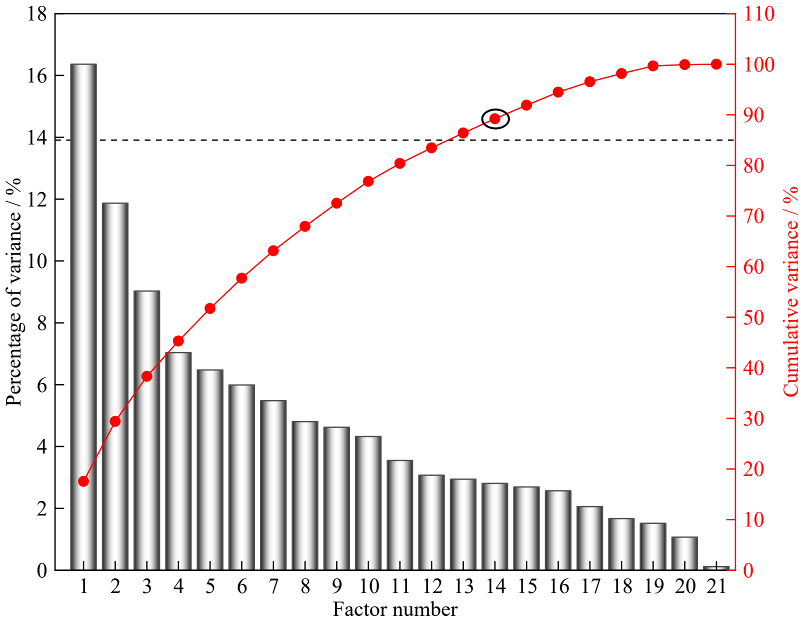

Fig. 5. Plot of factor variance and cumulative variance.

Table 3. Common factor expressions

Common factor	Common factor expressions
F_1	$-0.018X_2 - 0.017X_3 - 0.089X_4 + 0.013X_5 - 0.061X_6 + 0.019X_7 - 0.007X_9 - 0.042X_{10} + 0.033X_{11} - 0.045X_{12} + 0.035X_{13} - 0.013X_{14} - 0.031X_{15} + 0.353X_{16} + 0.269X_{17} - 0.035X_{18} - 0.212X_{19} + 0.089X_{20} + 0.347X_{21}$
F_2	$-0.065X_1 - 0.013X_2 + 0.350X_3 - 0.012X_4 - 0.273X_5 + 0.397X_6 - 0.125X_7 + 0.340X_8 - 0.005X_9 + 0.022X_{10} - 0.012X_{11} + 0.023X_{12} - 0.048X_{13} + 0.012X_{14} + 0.011X_{15} - 0.049X_{16} + 0.024X_{17} + 0.059X_{18} + 0.122X_{19} - 0.110X_{20} - 0.036X_{21}$
F_3	$0.030X_1 + 0.178X_2 + 0.110X_3 + 0.565X_4 + 0.149X_5 - 0.056X_6 - 0.095X_7 + 0.009X_8 + 0.520X_9 + 0.087X_{10} - 0.077X_{11} + 0.034X_{12} - 0.107X_{13} - 0.098X_{14} + 0.005X_{15} - 0.074X_{16} - 0.047X_{17} + 0.016X_{18} + 0.061X_{19} + 0.021X_{20} - 0.062X_{21}$
F_4	$0.011X_1 + 0.603X_2 - 0.043X_3 + 0.102X_4 + 0.018X_5 - 0.136X_6 + 0.573X_7 + 0.063X_8 - 0.010X_9 + 0.025X_{11} - 0.038X_{12} + 0.037X_{13} + 0.010X_{14} + 0.070X_{15} + 0.016X_{16} - 0.082X_{17} + 0.012X_{18} - 0.151X_{19} + 0.103X_{20} + 0.002X_{21}$
F_5	$0.040X_1 - 0.082X_3 - 0.031X_4 - 0.030X_5 - 0.095X_6 + 0.018X_7 + 0.035X_8 + 0.095X_9 + 0.02X_{10} - 0.062X_{11} + 0.074X_{12} - 0.071X_{13} - 0.072X_{14} + 0.023X_{15} - 0.057X_{16} + 0.227X_{17} + 0.095X_{18} + 0.499X_{19} + 0.749X_{20} - 0.014X_{21}$
F_6	$-0.036X_1 + 0.056X_2 - 0.009X_3 + 0.015X_4 + 0.328X_5 + 0.075X_6 - 0.029X_7 + 0.017X_8 - 0.285X_9 + 0.120X_{10} - 0.039X_{11} - 0.057X_{12} + 0.851X_{13} - 0.044X_{14} + 0.094X_{15} - 0.017X_{16} + 0.130X_{17} + 0.013X_{18} + 0.297X_{19} - 0.153X_{20} + 0.002X_{21}$
F_7	$0.917X_1 - 0.035X_2 + 0.091X_3 - 0.029X_4 - 0.154X_5 - 0.055X_6 + 0.015X_7 - 0.282X_8 + 0.099X_9 + 0.011X_{10} - 0.111X_{11} - 0.009X_{12} - 0.030X_{13} - 0.067X_{14} - 0.011X_{15} - 0.026X_{16} + 0.030X_{17} + 0.040X_{18} + 0.083X_{19} + 0.006X_{20} - 0.010X_{21}$
F_8	$0.053X_1 + 0.063X_2 - 0.017X_3 + 0.041X_4 + 0.210X_5 + 0.172X_6 - 0.003X_7 + 0.050X_8 - 0.039X_9 - 0.032X_{10} - 0.028X_{11} + 0.007X_{12} + 0.021X_{13} - 0.024X_{14} - 0.070X_{15} - 0.011X_{16} - 0.077X_{17} + 0.932X_{18} - 0.124X_{19} + 0.131X_{20} - 0.011X_{21}$
F_9	$0.027X_1 + 0.018X_2 + 0.011X_3 + 0.091X_4 + 0.007X_5 + 0.138X_6 + 0.016X_7 - 0.132X_8 + 0.057X_9 + 0.974X_{10} - 0.089X_{11} - 0.057X_{12} + 0.118X_{13} + 0.023X_{14} - 0.131X_{15} - 0.006X_{16} - 0.124X_{17} - 0.055X_{18} - 0.223X_{19} + 0.113X_{20} - 0.010X_{21}$
F_{10}	$-0.065X_1 - 0.044X_2 + 0.022X_3 - 0.095X_4 + 0.061X_5 + 0.217X_6 + 0.039X_7 - 0.232X_8 - 0.078X_9 + 0.019X_{10} - 0.072X_{11} - 0.026X_{12} - 0.038X_{13} + 0.965X_{14} - 0.039X_{15} - 0.018X_{16} + 0.020X_{17} - 0.040X_{18} + 0.065X_{19} - 0.082X_{20} - 0.020X_{21}$
F_{11}	$-0.001X_1 + 0.136X_2 - 0.023X_3 - 0.054X_4 + 0.065X_5 + 0.042X_6 - 0.002X_7 + 0.008X_8 + 0.047X_9 - 0.103X_{10} - 0.028X_{11} + 0.019X_{12} + 0.060X_{13} - 0.032X_{14} + 1.025X_{15} - 0.027X_{16} - 0.020X_{17} - 0.065X_{18} + 0.002X_{19} + 0.005X_{20} - 0.011X_{21}$
F_{12}	$-0.121X_1 - 0.020X_2 + 0.010X_3 - 0.102X_4 - 0.129X_5 - 0.077X_6 + 0.010X_7 - 0.040X_8 - 0.021X_9 - 0.080X_{10} + 0.990X_{11} + 0.018X_{12} - 0.059X_{13} - 0.065X_{14} - 0.011X_{15} + 0.011X_{16} + 0.115X_{17} - 0.011X_{18} + 0.215X_{19} - 0.164X_{20} + 0.004X_{21}$
F_{13}	$-0.008X_1 - 0.082X_2 - 0.024X_3 - 0.031X_4 - 0.261X_5 - 0.110X_6 + 0.013X_7 + 0.061X_8 + 0.172X_9 - 0.071X_{10} + 0.022X_{11} + 0.957X_{12} - 0.003X_{13} - 0.010X_{14} + 0.007X_{15} - 0.025X_{16} - 0.078X_{17} + 0.017X_{18} - 0.139X_{19} + 0.116X_{20} - 0.027X_{21}$

Table 4. Comparison of different models before factor analysis

Model	R^2	MAE	RMSE	MAPE
FA–IPSO–SVM	0.960	0.005	0.007	0.052
IPSO–SVM	0.903	0.008	0.010	0.074
TSVR [12]	0.870	0.009	0.012	0.088
SVM [11]	0.825	0.010	0.014	0.095

R^2 value of the FA–IPSO–SVM model was 0.057, 0.090, and 0.135 higher than those of the IPSO–SVM, TSVR, and SVM model, respectively. The MAE values were 0.003, 0.004, and 0.005 lower than those of the IPSO–SVM, TSVR, and SVM models, respectively. The RMSE values were 0.003, 0.005, and 0.007 lower than those of the IPSO–SVM, TSVR, and SVM models, respectively. The MAPE values were 0.022, 0.036, and 0.043 lower than those of the IPSO–SVM, TSVR, and SVM models, respectively.

Fig. 6 shows the hit rates of various models in predicting the endpoint carbon content in the converter within an error range of ±0.02%. As illustrated, the hit rate of the SVM mod-

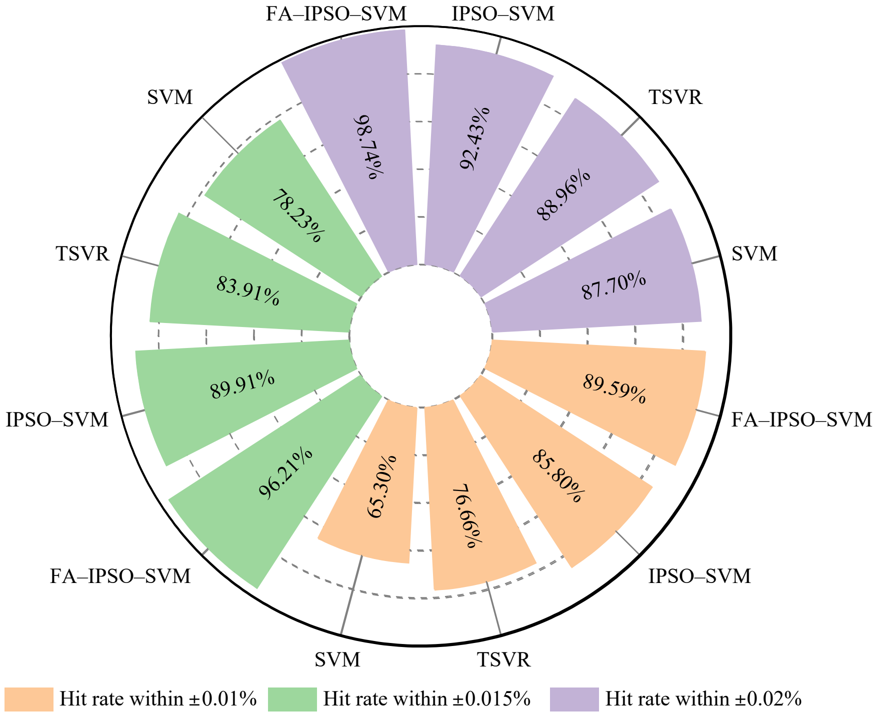

Fig. 6. **Prediction accuracies of different models.**

el was 87.70%, the lowest among the four models. The TSVR model achieved a hit rate of 88.96%, an improvement over the SVM model. The IPSO–SVM model demonstrated a hit rate of 92.43%, with predicted values closely matching the actual values, which indicated high prediction accuracy for the endpoint carbon content in the converter. The FA–IPSO–SVM model exhibited an extremely small difference between the predicted and actual values and achieved a hit rate of 98.74%, which was significantly higher than the other three models. Moreover, the model achieved hit rates of 89.59% and 96.21% within error margins of ±0.01% and ±0.015%, respectively, demonstrating robust predictive efficacy.

Fig. 7 shows the distribution of prediction results for various models within different ranges. As shown in Fig. 7(a), the prediction band width of the FA–IPSO–SVM model was 0.027, with a confidence band width of 0.003. Fig. 7(c)–(e) reveals that the prediction band widths for the IPSO–SVM, TSVR, and SVM model were 0.038, 0.044, and 0.049, respectively, with corresponding confidence band widths of 0.005, 0.006, and 0.007. Among these, the FA–IPSO–SVM model exhibited the narrowest confidence and prediction bands, indicating the smallest deviation of data points from the fitted curve. Fig. 7(b) demonstrates that, apart from four test data points with minor deviations, the remaining 313 test data points for the FA–IPSO–SVM model were distributed within ±0.02%, significantly outperforming the other three predictive models. Fig. 7(f) shows that the IPSO–SVM model had 293 data points falling within the ±0.02% range. Fig. 7(g) indicates that the TSVR model had 282 data points within the ±0.02% range. Fig. 7(h) reveals that the SVM model had 278 data points within the ±0.02% range, exhibiting the poorest predictive performance.

According to the above analysis, among the four prediction models for endpoint carbon content in the converter, the FA–IPSO–SVM model demonstrated superior prediction accuracy and stronger generalization ability. This advantage suggested that the FA–IPSO–SVM model was a feasible and reliable approach for predicting the endpoint carbon content and could meet the prediction requirements for various steel grades.

4.3. Analysis of parameters affecting prediction accuracy

The impact of input parameters on the prediction results of the endpoint carbon content for multiple steel grades was assessed through sequential removal of the 21 parameters and re-prediction using the remaining 20 parameters in the FA–IPSO–SVM model. The input parameters were then classified based on their influence on the prediction outcomes. The distribution of prediction accuracy errors for each parameter, with an error range of ±0.02%, is shown in Fig. 8. Fig. 8(a) shows that in zone I, the largest absolute error in the reduction of prediction accuracy occurs after removing X_{20}, followed by X_{12} with a value of 6.31%. In zone II, the absolute errors in the reduction of prediction accuracy ranged from a minimum of 3.16% to a maximum of 4.73%. In zone III, the maximum absolute error in the reduction of prediction accuracy was only 1.58%. These results aligned with the findings from the correlation analysis well, which suggested that the correlation and significance between input parameters and endpoint carbon content influenced prediction accuracy. The stronger the correlation and significance, the greater the impact; conversely, the weaker the correlation and significance, the smaller the impact.

From Fig. 8(b), the relative errors in the degree of prediction accuracy reduction ranged from 0.32%–6.63%. This

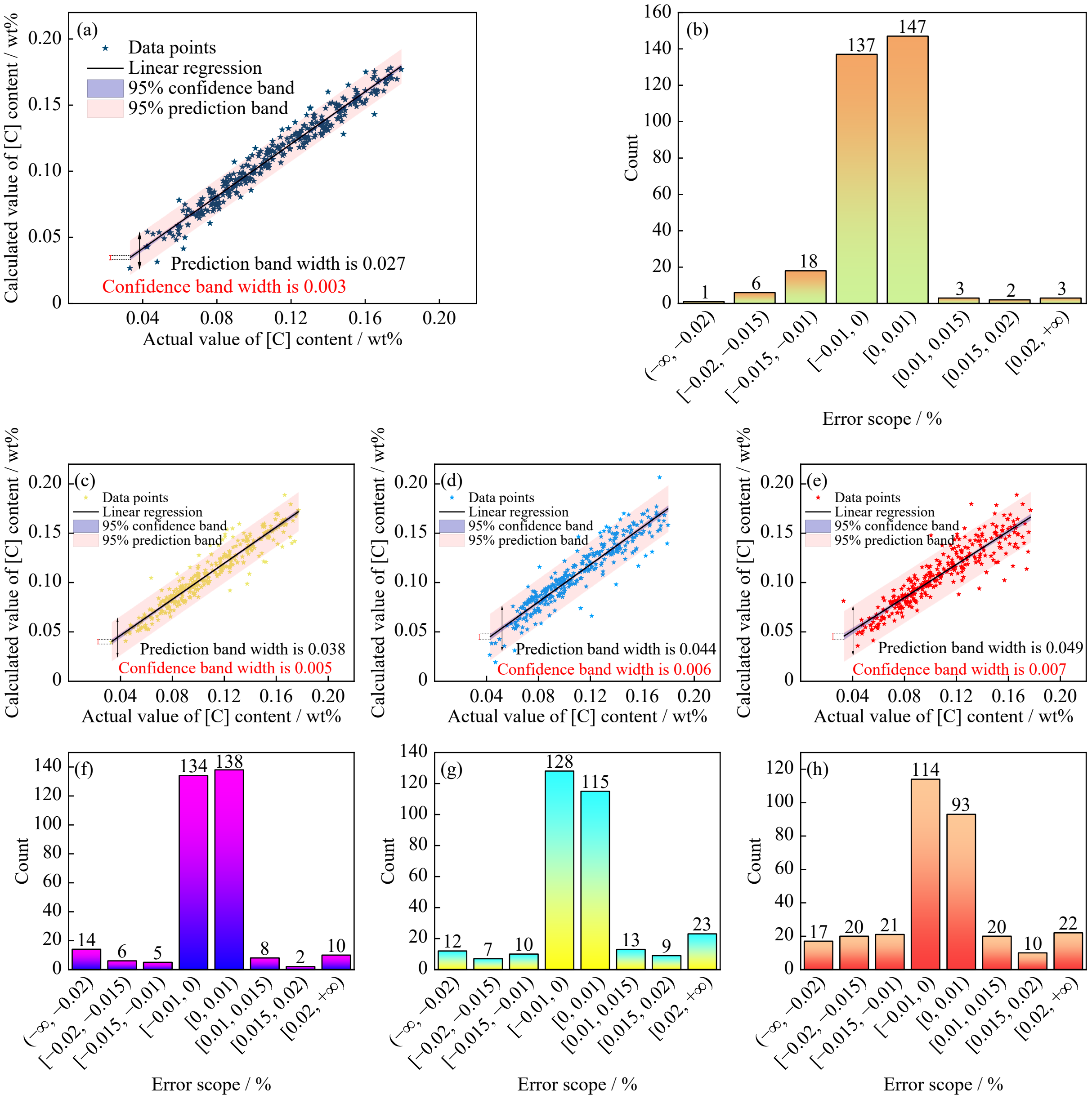

Fig. 7. Comparison between actual and predicted endpoint carbon contents obtained through different models: (a, b) FA–IPSO–SVM, (c, f) IPSO–SVM, (d, g) TSVR, and (e, h) SVM. (a, c–e) scatter plots and (b, f–h) bar charts.

study classified the area with relative errors between 0–2% as zone III (low influence zone), the area between 2%–5% as zone II (medium influence zone), and the area between 5%–7% as zone I (high influence zone). Fig. 8(c)–(e) shows that in zone I, raw material parameters, smelting parameters, and oxygen supply parameters each contribute one-third. In zone II, raw material parameters accounted for 75%, while smelting and oxygen supply parameters each accounted for 12.5%. In zone III, raw material parameters accounted for 42.9%, smelting parameters accounted for 28.6%, and oxygen supply parameters accounted for 28.6%.

Overall, Fig. 8 indicates that zone I included X_3, X_4, X_{12}, X_{13}, X_{19}, and X_{20}, zone II included X_1, X_5, X_7, X_8, X_9, X_{10}, X_{14}, and X_{17}, and zone III included X_2, X_6, X_{11}, X_{15}, X_{16}, X_{18}, and X_{21}. Removing parameters in zone I had a significant impact on the prediction results of endpoint carbon content while removing parameters in zone III had a minimal effect. When building a prediction model for the endpoint carbon content of multiple steel grades in a converter, parameters in zone I must be retained, parameters in zone II can be partially replaced, and parameters in zone III can be entirely substituted with other types of parameters to improve the hit rate of the prediction model.

5. Conclusions

According to the smelting data of various steel grades, a model was developed to predict the endpoint carbon content of multiple steel grades in converters by combining FA and IPSO–SVM. The conclusions are as follows.

(1) Through an analysis of the converter steelmaking process, a parameter system encompassing raw material parameters, auxiliary material parameters, smelting parameters, and oxygen supply parameters was constructed. Factor ana-

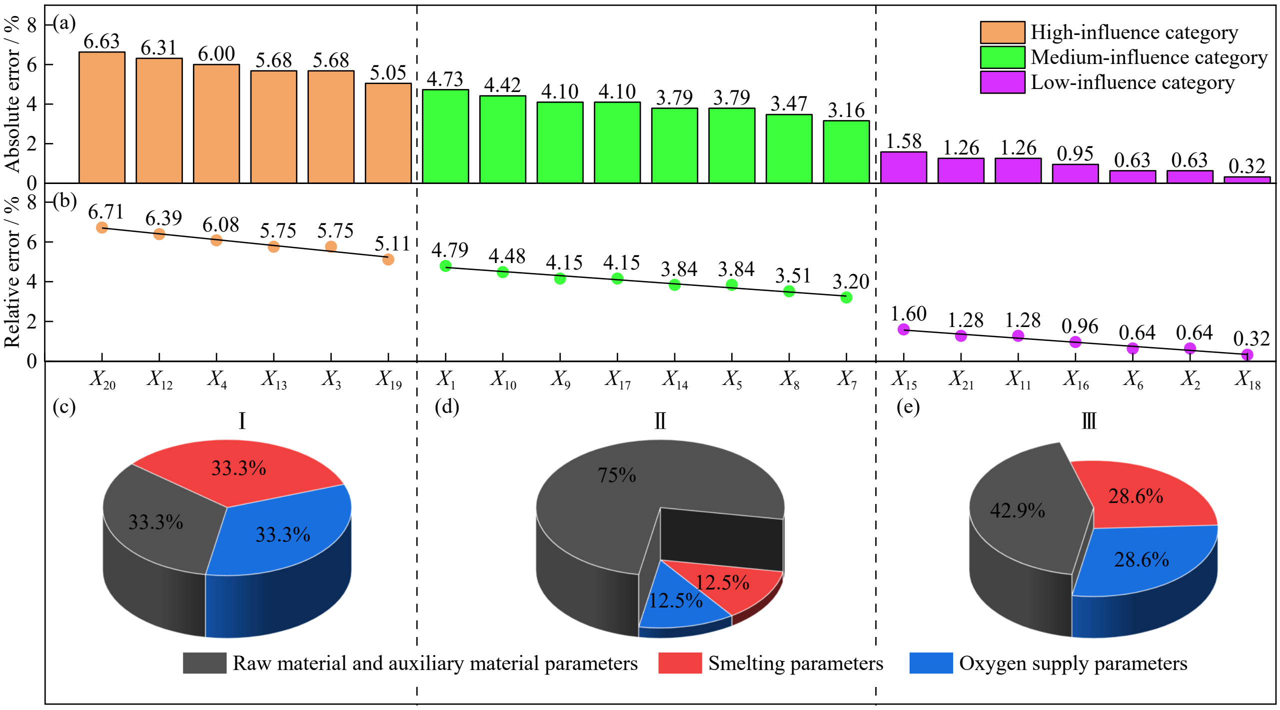

Fig. 8. Distribution of various metrics after parameter removal: (a) absolute error for predictive accuracy; (b) relative error for predictive accuracy; the proportion of various input parameters in the (c) high, (d) medium, and (e) low influence zones.

lysis was employed to reduce the dimensionality of the variables, resulting in the identification of 13 common factors with a cumulative variance contribution rate of 85.53%. This effectively mitigated collinearity among the data, laying a solid foundation for enhancing prediction accuracy.

(2) The production data were screened according to specification requirements, resulting in 1056 data samples. These samples were then shuffled, with 70% used for model training and 30% for model testing. The model performance was optimized by introducing a nonlinear decreasing inertia weight into the PSO component, and the penalty factor C and kernel parameter g in the SVM component were set to 18.5154 and 0.3192, respectively.

(3) The FA–IPSO–SVM model exhibited exceptional performance in predicting endpoint carbon content, achieving a hit rate of 98.74% within a ±0.02% error range, which surpassed methods reported in previous studies. Furthermore, using a parameter elimination approach coupled with the analysis of correlations between input parameters and endpoint carbon content, we categorized the influential parameters into high-, medium-, and low-impact groups. Specifically, high-impact parameters included X_3 (C content of molten iron), X_4 (Si content of molten iron), X_{12} (TSC temperature), X_{13} (TSC carbon content), X_{19} (oxygen flow rate), and X_{20} (free oxygen content); medium-impact parameters comprised X_1 (molten iron weight), X_5 (Mn content of molten iron), X_7 (S content of molten iron), X_8 (amount of scrap added), X_9 (amount of lime added), X_{10} (amount of dolomite added), X_{14} (TSO temperature), and X_{17} (oxygen consumption); and low-impact parameters encompassed X_2 (molten iron temperature), X_6 (P content of molten iron), X_{11} (amount of ore added), X_{15} (lance position), X_{16} (bottom blowing time), X_{18} (oxygen pressure), and X_{21} (oxygen supply time).

(4) Removing high-impact parameters resulted in a minimum reduction in the hit rate of 5.05%, by contrast, removing low-impact parameters led to a maximum reduction of just 1.58%. These findings demonstrate that high-impact parameters are pivotal to prediction accuracy and must be included, medium-impact parameters can be partially substituted, and low-impact parameters can be fully replaced with other parameters to enhance model precision. However, this conclusion may be influenced by factors such as noise and incomplete features in the smelting data, necessitating further mechanistic research to accurately analyze the factors affecting prediction accuracy.

Acknowledgements

This work was financially supported by the National Natural Science Foundation of China (No. 52174297).

Conflict of Interest

Yanping Bao is an editorial board member for this journal and was not involved in the editorial review or the decision to publish this article. The authors declare no conflict of interest.

References

[1] Y.J. Xia, H.B. Wang, and A.J. Xu, dmPINNs: An integrated data-driven and mechanism-based method for endpoint carbon prediction in BOF, *Metals*, 14(2024), No. 8, art. No. 926.

[2] M.Q. Gu, A.J. Xu, F. Yuan, X.M. He, and Z.F. Cui, An improved CBR model using time-series data for predicting the end-point of a converter, *ISIJ Int.*, 61(2021), No. 10, p. 2564.

[3] N. Madhavan, G.A. Brooks, M.A. Rhamdhani, B.K. Rout, and A. Overbosch, General heat balance for oxygen steelmaking, *J. Iron Steel Res. Int.*, 28(2021), No. 5, p. 538.

[4] X. Liu, Y.P. Bao, C.J. Hua, J.H. Chu, C. Gu, and M. Wang, Denaturation behavior of inclusions in industrial pure iron by calcium treatment, *J. Iron Steel Res. Int.*, 30(2023), No. 2, p. 249.

[5] Z.C. Xin, J.S. Zhang, Y. Jin, J. Zheng, and Q. Liu, Predicting the alloying element yield in a ladle furnace using principal component analysis and deep neural network, *Int. J. Miner. Metall. Mater.*, 30(2023), No. 2, p. 335.

[6] H.Y. Zhang, Q.L. Zhou, Z.X. Yuan, and Z.M. Liu, RBF neural network base on affinity propagation clustering and its application in BOF steelmaking, *J. Iron. Steel. Res.*, 26(2014), No. 1, p. 22.

[7] X. Liu, Y.P. Bao, L.H. Zhao, and C. Gu, Establishment and application of steel composition prediction model based on t-distributed stochastic neighbor embedding (t-SNE) dimensionality reduction algorithm, *J. Sustain. Metall.*, 10(2024), No. 2, p. 509.

[8] C.R. Li, H.W. Zhao, and Q. Yin, Prediction model of end-point carbon content for BOF based on LM BP neural network, *Adv. Mater. Res.*, 189-193(2011), p. 4446.

[9] Z. Liu, S.S. Cheng, and P.B. Liu, Prediction model of BOF end-point temperature and carbon content based on PCA-GA-BP neural network, *Metall. Res. Technol.*, 119(2022), No. 6, art. No. 605.

[10] D.Z. Wang, F. Gao, L.D. Xing, J.H. Chu, and Y.P. Bao, Continuous prediction model of carbon content in 120 t converter blowing process, *Metals*, 12(2022), No. 1, art. No. 151.

[11] C.C. Zhang, W.L. Dong, C.X. Ji, H.B. Li, B. Chen, and C.L. Zhao, Endpoint prediction of duplex process converter based on machine learning algorithm, *Steelmaking*, 39(2023), No. 4, p. 28.

[12] M. Wang, S.L. Li, C. Gao, and Y. Fan, End-point prediction TSVR model accuracy of 80 t BOF steelmaking, *Iron Steel*, 55(2020), No. 7, p. 53.

[13] M. Wang, C. Gao, X.G. Ai, B.P. Zhai, and S.L. Li, Whale optimization end-point control model for 260 tons BOF steelmaking, *ISIJ Int.*, 62(2022), No. 8, p. 1684.

[14] M.Q. Gu, A.J. Xu, H.B. Wang, and Z.T. Wang, Real-time dynamic carbon content prediction model for second blowing stage in BOF based on CBR and LSTM, *Processes*, 9(2021), No. 11, art. No. 1987.

[15] H.D. Wang, H. Liu, F.G. Chen, H. Li, and X.J. Xue, Endpoint carbon content and temperature prediction model in BOF steelmaking based on posterior probability and intra-cluster feature weight online dynamic feature selection, *High Temp. Mater. Process.*, 44(2025), No. 1, art. No. 20240067.

[16] Q.Q. Dong, M. Li, S.J. Hu, Y. Yu, and M.Q. Gu, A dynamic control model of basic oxygen furnace last blowing stage based on improved conditional generative adversarial network, *Metall. Mater. Trans. B*, 56(2025), No. 1, p. 638.

[17] T.Y. Xie, F. Zhang, Y.R. Li, Q. Zhang, Y.W. Wang, and H. Shang, Self-attention-based convolutional parallel network: An efficient multi-input deep learning model for endpoint prediction of high-carbon BOF steelmaking, *Metall. Mater. Trans. B*, 55(2024), No. 6, p. 4271.

[18] Y.K. Su, H. Liu, F.G. Chen, J.X. Liu, H. Li, and X.J. Xue, BOF steelmaking endpoint carbon content and temperature soft sensor model based on supervised weighted local structure preserving projection, *High Temp. Mater. Process.*, 43(2024), No. 1, art. No. 20240030.

[19] S.F. Ding, X.P. Hua, and J.Z. Yu, An overview on nonparallel hyperplane support vector machine algorithms, *Neural Comput. Appl.*, 25(2014), No. 5, p. 975.

[20] X.L. Li, Z.Y. Cao, and Y.L. Xu, Characteristics and trends of coal mine safety development, *Energy Sources Part A*, 47(2025), No. 1, p. 2316.

[21] X.L. Li, S.J. Chen, S. Wang, M. Zhao, and H. Liu, Study on in situ stress distribution law of the deep mine: Taking Linyi mining area as an example, *Adv. Mater. Sci. Eng.*, 2021(2021), No. 1, art. No. 5594181.

[22] W.L. Shen, G.C. Shi, Y.G. Wang, J.B. Bai, R.F. Zhang, and X.Y. Wang, Tomography of the dynamic stress coefficient for stress wave prediction in sedimentary rock layer under the mining additional stress, *Int. J. Min. Sci. Technol.*, 31(2021), No. 4, p. 653.

[23] X.G. Kong, D. He, X.F. Liu, *et al.*, Strain characteristics and energy dissipation laws of gas-bearing coal during impact fracture process, *Energy*, 242(2022), art. No. 123028.

[24] H.J. Huang, X.X. Wei, and Y.Q. Zhou, An overview on twin support vector regression, *Neurocomputing*, 490(2022), p. 80.

[25] Y. Wang, E.X. Du, S.Q. Yang, and L. Yu, Prediction and analysis of slope stability based on IPSO-SVM machine learning model, *Geofluids*, 2022(2022), art. No. 8529026.

[26] H. Zhang and Y.N. Li, Research on identification model of element logging shale formation based on IPSO-SVM, *Petroleum*, 8(2022), No. 2, p. 185.

[27] Q. Wang and X.M. Tian, Soft sensing based on EMD and improved PSO-SVM, *Appl. Mech. Mater.*, 291-294(2013), p. 2817.

[28] H. Mojahedi, A.B. Sangar, and M. Masdari, Towards tax evasion detection using improved particle swarm optimization algorithm, *Math. Probl. Eng.*, 2022(2022), art. No. 1027518.

[29] Y.B. Gao, J. Hua, G.W. Chen, L.P. Cai, N. Jia, and L.K. Zhu, Bi-directional prediction of wood fiber production using the combination of improved particle swarm optimization and support vector machine, *BioResources*, 14(2019), No. 3, p. 7229.

[30] Y.F. Ji, S.M. Ma, W. Peng, and Z.H. Cai, Research on alloy composition-process-wear properties of medium manganese steel based on machine learning, *Tribol. Int.*, 200(2024), art. No. 110164.

[31] Q.F. Zhang, S. Chen, and Z.P. Fan, Bearing fault diagnosis based on improved particle swarm optimized VMD and SVM models, *Adv. Mech. Eng.*, 13(2021), No. 6, art. No. 16878140211028451.

[32] D. Li and C. Chen, Research on exercise fatigue estimation method of Pilates rehabilitation based on ECG and sEMG feature fusion, *BMC Med. Inform. Decis. Mak.*, 22(2022), No. 1, art. No. 67.

[33] J.G. Zhou, F.T. Guang, and R.P. Tang, Scenario analysis of carbon emissions of China's power industry based on the improved particle swarm optimization-support vector machine model, *Pol. J. Environ. Stud.*, 27(2018), No. 1, p. 439.

[34] Y. Wang, S.F. Yang, J.S. Li, J. Feng, and F. Wang, Dephosphorization by double-slag process in converter steelmaking, *High Temp. Mater. Process.*, 37(2018), No. 7, p. 625.

[35] R.H. Zhang, J. Yang, H. Sun, and W.K Yang, Prediction of lime utilization ratio of dephosphorization in BOF steelmaking based on online sequential extreme learning machine with forgetting mechanism, *Int. J. Miner. Metall. Mater.*, 31(2024), No. 3, p. 508.

[36] L.Q. Zhang, Y. Min, C.J. Liu, J. Huang, C.X. Zheng, and W. Ma, Prediction of residual manganese content at the end point of converter smelting based on machine learning, *Steelmaking*, 40(2024), No. 5, p. 38.

[37] X.Z. Wang, M. Han, and J. Wang, Applying input variables selection technique on input weighted support vector machine modeling for BOF endpoint prediction, *Eng. Appl. Artif. Intell.*, 23(2010), No. 6, p. 1012.

[38] F. Gao, Y.P. Bao, M. Wang, Y. Liu, Y.S. Huang, and G.T. Sun, Prediction model of end-point phosphorus content of converter based on FA-ELM, *Iron Steel*, 55(2020), No. 12, p. 24.

[39] X. Shao, Q. Liu, Z.C. Xin, J.S. Zhang, T. Zhou, and S.S. Li, Hybrid model for BOF oxygen blowing time prediction based on

oxygen balance mechanism and deep neural network, *Int. J. Miner. Metall. Mater.*, 31(2024), No. 1, p. 106.

[40] Z.L. Wang and Y.P. Bao, Development and prospects of molten steel deoxidation in steelmaking process, *Int. J. Miner. Metall. Mater.*, 31(2024), No. 1, p. 18.

[41] C. Chen, N. Wang, and M. Chen, Prediction model of end-point phosphorus content in consteel electric furnace based on PCA-extra tree model, *ISIJ Int.*, 61(2021), No. 6, p. 1908.

[42] F. He and L.Y. Zhang, Prediction model of end-point phosphorus content in BOF steelmaking process based on PCA and BP neural network, *J. Process. Control*, 66(2018), p. 51.

[43] C. Zhang, Z.H. Xu, Y.Y. Lao, and Y. Li, Dynamic process fault detection and diagnosis method based on factor analysis: Application on the three-tank system process, *J. Chemom.*, 38(2024), No. 12, art. No. e3627.

[44] J.C. Li, N. Li, J.Z. Li, W. Wang, and H.L. Wang, Prediction of thermally modified wood color change after artificial weathering based on IPSO-SVM model, *Forests*, 14(2023), No. 5, art. No. 948.

[45] R.H. Zhang and J. Yang, State of the art in applications of machine learning in steelmaking process modeling, *Int. J. Miner. Metall. Mater.*, 30(2023), No. 11, p. 2055.

[46] Y.H. Liu, B. Yang, X.H. Han, *et al.*, Predicting laser penetration welding states of high-speed railway Al butt-lap joint based on EEMD-SVM, *J. Mater. Res. Technol.*, 21(2022), p. 1316.

[47] D. Yang, Y. Peng, T. Zhou, T. Wang, and G. Lu, Percussion and PSO-SVM-based damage detection for refractory materials, *Micromachines*, 14(2023), No. 1, art. No. 135.

[48] X.Q. Ren, S.L. Liu, X.D. Yu, and X. Dong, A method for state-of-charge estimation of lithium-ion batteries based on PSO-LSTM, *Energy*, 234(2021), art. No. 121236.

[49] T. Cuong-Le, T. Nghia-Nguyen, S. Khatir, P. Trong-Nguyen, S. Mirjalili, and K.D. Nguyen, An efficient approach for damage identification based on improved machine learning using PSO-SVM, *Eng. Comput.*, 38(2022), No. 4, p. 3069.

[50] H. Moayedi, M. Raftari, A. Sharifi, W.A.W. Jusoh, and A.S.A. Rashid, Optimization of ANFIS with GA and PSO estimating α ratio in driven piles, *Eng. Comput.*, 36(2020), No. 1, p. 227.

[51] J. Zhu, J. Wu, Z. Chen, L. Cao, M. Yang, and W. Xu, Research on system of ultra-flat carrying robot based on improved PSO algorithm, *Front. Neurorobot.*, 17(2023), art. No. 1294606.