

Development Status and Trend Forecast of China's Iron and Steel Industry Based on Multi-Algorithm Coupling Modeling

Tengyu Wang*, Yuting Ni

College of Metallurgy and Energy, North China University of Science and Technology, Tangshan 063210, Hebei, China

*Corresponding Author: 13363365179@163.com

ABSTRACT

Iron and steel metallurgy, as the most important weapon of the country, is ushering in a new round of good opportunities under the support of the national big infrastructure strategy. In this paper, we first use Python to crawl the data, and then use the data visualization software Emotional Sky visualization platform to get the big screen of steel industry data, and analyze the future development of the steel industry through the use of multi-algorithm coupled prediction method. The prediction of the future development of the iron and steel industry is realized through multi-algorithm coupling, and a new round of outlook on the iron and steel industry is proposed based on the prediction. Contribute to the early realization of efficient, intelligent and green development of China's iron and steel industry.

KEYWORDS

Low-carbon transition; "Dual-carbon" emission reduction; Iron and steel industry; Multi-algorithm coupled merit-based forecasting

1. INTRODUCTION

In the era of big data and internet, all kinds of big data in the steel industry are growing at an alarming rate, showing the trend of fragmentation and diversification [1]. How to effectively integrate these intricate data, and combine them with policies such as "dual-carbon emission reduction" and apply them to production and life, giving full play to the charm of big data, has become the focus of attention of all walks of life in this era.

At present, many enterprise platforms have launched different data visualization software (e.g., Huawei Cloud DLV software, EChart, Fine Report, etc.) to explore the value behind the data [2]. Through the collection, integration and analysis of data, we can make up for the steel industry's steel overcapacity, low product concentration, uneven distribution of metallurgical colleges and universities, the quality of practitioners needs to be improved, the level of equipment needs to be upgraded, the dependence on foreign mines is too high, and energy saving and emission reduction pressure is huge, etc [3]. At the same time, using the data visualization software Origin, the development of the iron and steel industry can be shown in the form of charts, which is helpful to understand the current situation of China's iron and steel industry intuitively, to explore the value behind the data to improve the scientific decision-making of the enterprise and the industry, and to serve the iron and steel industry [4].

Iron and steel metallurgy, as the most important weapon of the country, has ushered in a new round of opportunities under the support of the country's major infrastructure strategy [5]. Today's era is an

era of both opportunities and challenges, and it is especially important to clearly understand the current development of China's iron and steel industry in the midst of fierce competition [6].

2. PYTHON CRAWLER DATA COLLECTION

In order to ensure the accuracy and completeness of the data, the work uses Python software to crawl the data sources, including the National Bureau of Statistics, Metallurgy of China, China Metallurgical Society, and the World Iron and Steel Association and other websites. Part of the crawling code and results are shown in Figure 1 below.

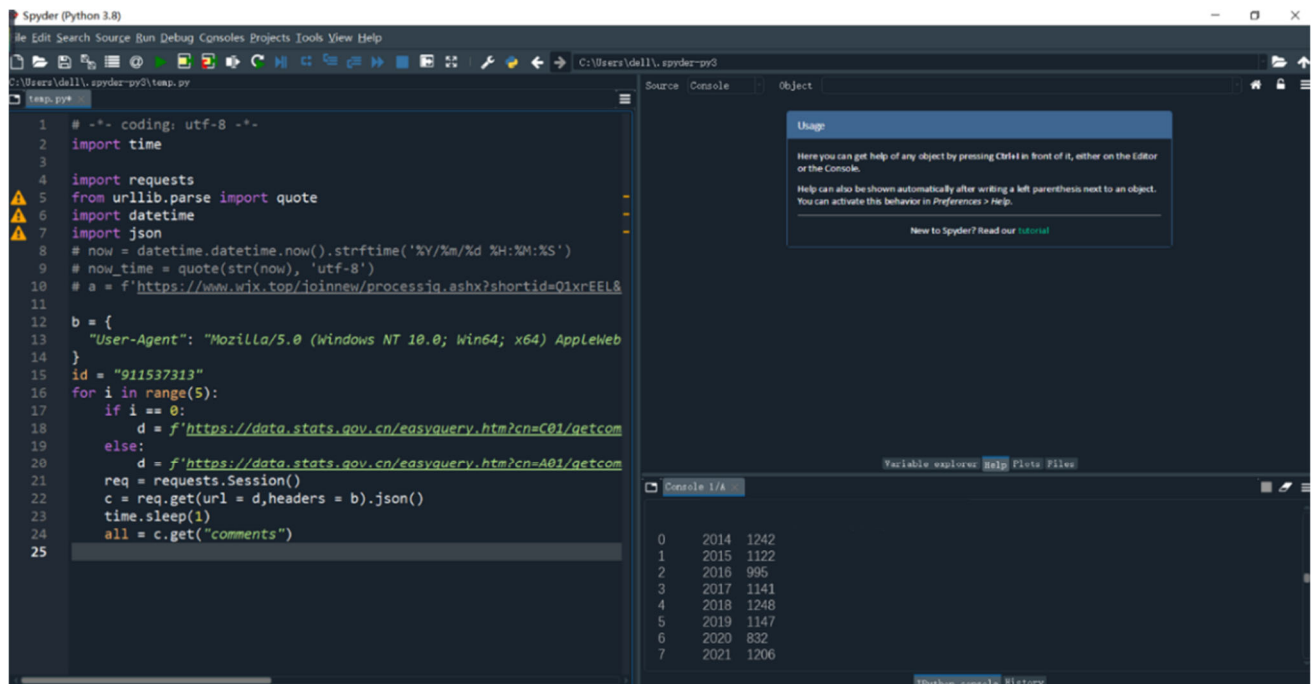


Figure 1. Python crawling data display chart

This part of the crawled data has the development status of China's iron and steel industry, the existing problems and predictions for the future, will be introduced in detail in the following article.

3. FORECASTING THE FUTURE OF THE STEEL INDUSTRY

3.1. Multi-algorithm coupling to predict steel data

We composed a coupling algorithm based on metabolism GM(1,1), ARIMA, BP neural network, and SVM regression algorithm based on Bayesian tuning parameter in order to forecast 7 steel data, i.e., combining the 2002-2022 Chinese steel imports, the number of Chinese iron and steel enterprises, the apparent consumption of steel, the crude steel production in China, the steel exports in China, the revenue of Chinese iron and steel industry enterprises, and the Chinese Steel Industry Enterprise Profit, utilizing the method of multi-algorithm coupling to forecast the 7 steel data from 2023-2027.

3.2. Multi-algorithm coupling design and prediction results

We take the first 80% of each item of steel data for the 20 years from 2002-2021 as the training set, and the second 20% as the test set, and use metabolic GM(1,1), ARIMA, BP neural network, and SVM regression algorithms based on Bayesian tuning parameter to make predictions in turn, observe

the RMSE prediction effect of the four algorithms in the test set, and select the best prediction effect algorithm as the final prediction algorithm that to predict each steel data in 2023-2026.

3.2.1. Metabolism GM(1,1) algorithm

Step1. Grade ratio test

In order to ensure the feasibility of the metabolic GM(1,1) algorithm, before the formal prediction, we need to carry out the level ratio test on the original sequence, we calculated all the level ratios according to the following formula, and found that they are all located in the tolerable coverage, so the level ratio test passed.

$$\lambda(k) = \frac{x^{(0)}(k-1)}{x^{(0)}(k)} (k = 2, 3, \dots, n) \quad (1)$$

Step2. Cumulative generation

We accumulate the original sequences to form the cumulative generation sequence.

Step3. Construct matrix Y0, B0, C0

We calculate the matrix Y0, B0, C0 by the following three formulas.

$$Y_0 = (x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n))^T \quad (2)$$

$$C_0 = \begin{pmatrix} a_0 \\ u_0 \end{pmatrix} \quad (3)$$

$$B_0 = -\frac{1}{2}[x^{(1)}(i) + x^{(1)}(i-1)], i = 2, 3, \dots, n \quad (4)$$

Step4. Calculate the development coefficient a_0 and gray role quantity u_0

Based on the principle of least squares, we calculate the development coefficient a_0 and the gray role quantity u_0 , and the calculation results are shown in the following Table 1:

Table 1 Metabolic GM(1,1) algorithm development coefficients a_0 and gray role quantities u_0

	a_0	u_0
1	0.043538	49.3368
2	0.048675	50.3968
3	0.049745	52.7466
...	...	...
...	...	...
14	0.047681	67.7913
15	0.04793	71.0051
16	0.047985	74.473

Step5. Predicting a new data

We predicted a new data based on the following formula, combined with the development coefficient a_0 and gray role quantity u_0 calculated in Step5.

$$x^{(1)}(k+1) = \left[x^{(0)}(1) - \frac{\hat{u}}{\hat{a}} \right] e^{-\hat{a}k} + \frac{\hat{u}}{\hat{a}}, k = 0, 1, 2, \dots, n \quad (5)$$

Step6. Delete the old data and add the predicted new data

Repeat Step1-Step6 until the end of the prediction work, the test set prediction effect is shown in Figure 2 below:

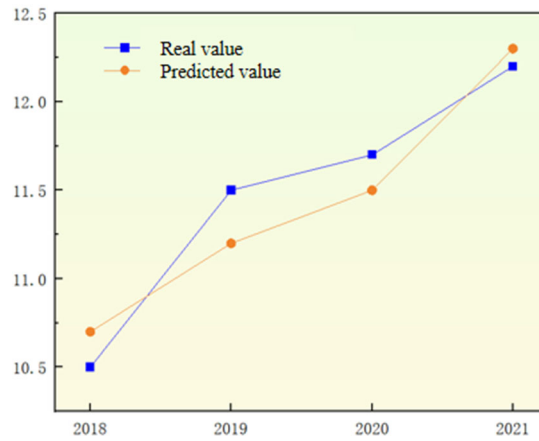


Figure 2. Metabolic GM(1,1) Algorithm Test Set Prediction Effectiveness

Step7. Model Test

We calculate the relative residual Q by the following formula, the result is 0.02340, the model test passed.

$$\varepsilon(k) = \frac{x^{(0)}(k) - \hat{x}^{(0)}(k)}{x^{(0)}(k)} \quad (6)$$

Step8. Calculate the RMSE value of prediction effect

We calculated the prediction effect of the metabolic GM(1,1) algorithm on the test set with an RMSE value of 0.2218 by using the following formula.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (7)$$

3.2.2. ARIMA algorithm

Step1. Smoothness test

The premise of ARIMA algorithm is that the time series is smooth, so before forecasting by ARIMA algorithm, we need to test the smoothness of the time series, we test the smoothness of the time series by ADF test and SPSS test. Our ADF test is 0 and SPSS test is 1, which does not pass the smoothness test and we need to further make the time series smooth by differencing [7].

We perform a first-order difference on the original series and perform the smoothness test again, and find that the ADF test is 1 and the SPSS test is 0, the test is passed.

Step2. Determine the order of ARIMA algorithm

We based on the AIC, BIC criterion for violent order, determine the order of ARIMA is (1,1,0).

Step3. Residual test

We carry out the residual test through the QQ diagram, the test results show that the residuals are in line with the normal distribution, and the residual test is passed.

Step4. ARIMA algorithm prediction

We carried out the prediction through Matlab's Forecast function, the prediction results of the test set are shown below in Figure 3, the RMSE value of the prediction effect is 0.3561.

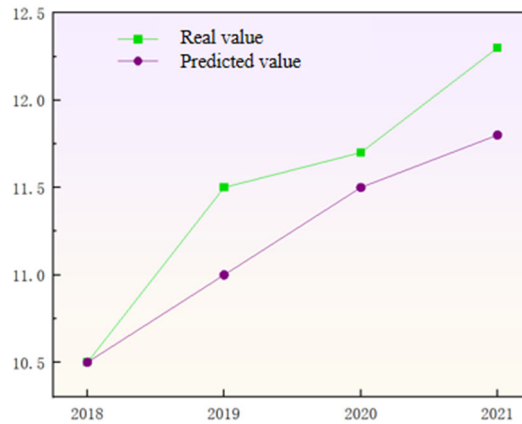


Figure 3. ARIMA algorithm test set prediction effect

3.2.3. BP neural network

Step1. Organize data

The BP neural network algorithm for time series prediction is essentially to transform the time series data into regression data type, followed by the establishment of regression model, to the regression model to input the past time series data, you can get to the required time series data. The data are organized as shown in the table below (yellow is the training set data, green is the test set data):

Table 2 BP neural network data organization

Serial number	x1	x2	x3	y	Serial number	x1	x2	x3	y
1	7.02	7.19	8.08	8.07	10	8.87	9.11	9.35	10.46
2	7.19	8.08	8.07	8.11	11	9.11	9.35	10.46	10.24
3	8.08	8.07	8.11	8.14	12	9.35	10.46	10.24	10.43
4	8.07	8.11	8.14	8.96	13	10.46	10.24	10.43	9.86
5	8.11	8.14	8.96	9.05	14	10.24	10.43	9.86	10.49
6	8.14	8.96	9.05	8.79	15	10.43	9.86	10.49	11.53
7	8.96	9.05	8.79	8.87	16	9.86	10.49	11.53	11.78
8	9.05	8.79	8.87	9.11	17	10.49	11.53	11.78	12.23
9	8.79	8.87	9.11	9.35					

Step2. Establishment of regression model

Based on the training set of Step1 collated data, after constructing and initializing the BP neural network, the training process first receives the data samples to be processed from the input layer and propagates them forward layer by layer through each hidden layer. In this process, each neuron sums the received weighted input with bias and transforms it into a nonlinear output through the activation function and passes it to the next layer. When the data flows through all the hidden layers to the output layer, the network generates a prediction of the input data [8].

This is immediately followed by the error computation phase, where we compare the actual predicted value with the true target value, and use a loss function (e.g., mean square error) to quantify the difference between the two and obtain the error value for the current batch or individual sample. Then we enter the backpropagation phase, where the error signal starts from the output layer and propagates to the input layer layer by layer in the reverse direction according to the principle of the gradient descent method. Each layer of neurons updates its corresponding weights and biases based on its contribution to the overall error (error gradient), allowing for a more accurate fit to the data in subsequent iterations.

This iterative training process continues until a predefined termination condition is reached, such as reaching a certain number of iterations, the model's performance on the validation set no longer improves significantly, or it converges to a certain threshold. Throughout the training cycle, we periodically evaluate the model's performance on the validation set to monitor overfitting and adjust hyperparameters such as learning rate accordingly. At the end of training, the generalization ability of the model is evaluated on an independent test set to ensure that the model not only performs well on the training data, but also achieves satisfactory results on new, unseen data.

Step3. Test set prediction

We predicted the test set prediction results through the regression model of Step2 and input the time series sequence data, as shown in Figure 3 below, the RMSE value of the test set prediction effect is 0.3158.

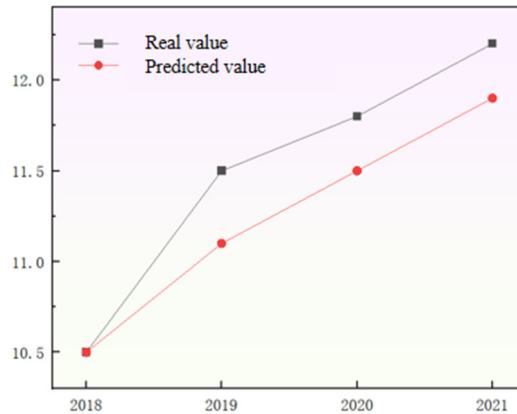


Figure 4. BP neural network algorithm test set prediction effect

3.2.4. SVM regression algorithm based on Bayesian modulation parameterization

Step1. Bayesian tuning to solve the optimal hyperparameters

We use Bayesian parameterization to search for the optimal hyperparameters in a set range of hyperparameters with box constraints 0.001-1000, Epsilon 0.024091-2401.9023, kernel function Gaussian, linear, quadratic, cubic, and normalized data true and false, and the optimal hyperparameters are obtained as follows: the kernel function is quadratic, the box constraints 0.3565, Epsilon 4.0525, and the normalized data false, Epsilon is 4.0525, and the normalized data is false.

Step2. Establish SVM by optimal hyperparameters

The SVM regression model is built by solving the optimal hyperparameters and combining the training set data.

Step3. Test set prediction

Through the regression model of Step2, input the time series sequence data, the prediction obtained the test set prediction results, as shown in Figure 5 below, the RMSE value of the test set prediction effect is 0.1636.

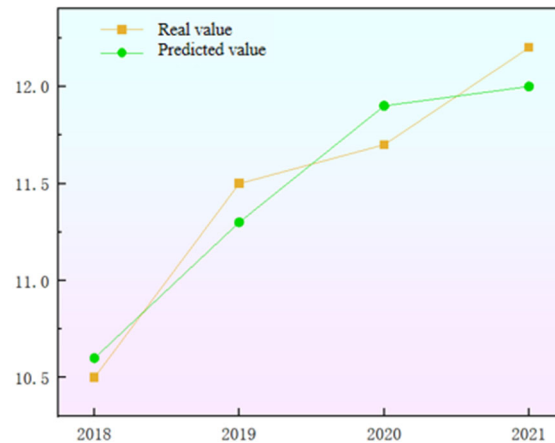


Figure 5. Test set prediction effect of SVM regression algorithm based on Bayesian tuning parameterization

3.2.5. Selecting the best prediction algorithm

The RMSE values of the prediction effect of the four prediction algorithms on the test set are 0.2218, 0.3561, 0.3158, 0.1636, so we choose the SVM regression algorithm based on Bayesian tuning parameterization as the final prediction algorithm for predicting the apparent consumption of iron and steel in the period of 2023-2027, and the results of the prediction are shown in Figure 6 below:

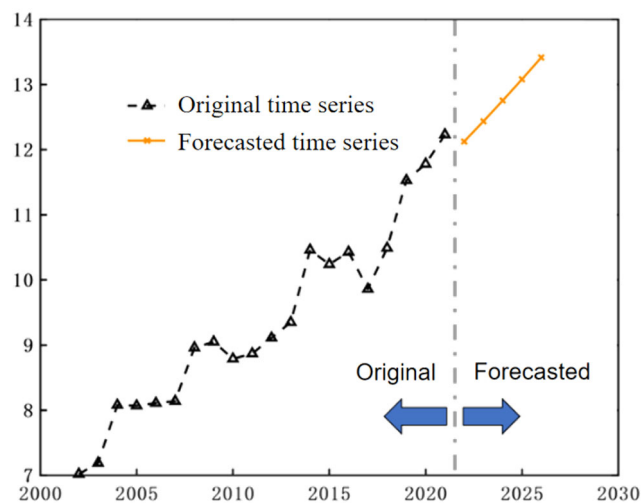


Figure 6. Forecast of apparent steel consumption

3.2.6. Forecasts for other indicators

The projections for the other 6 steel indicators are shown in the Table 3 below:

Table 3 Forecasts for the other 6 steel indicators

Indicators	China Steel Imports	Number of Steel Companies in China
Best prediction	Metabolism GM(1,1)	ARIMA algorithm
algorithm	algorithm	
2023	1123.37	7673.51
2024	1115.76	7625.63
2025	1113.43	7599.56
2026	1112.12	7564.54
2027	1108.62	7523.56
Indicators	China's Steel Exports	China's steel industry enterprise revenue
Best prediction	Metabolism GM(1,1)	ARIMA algorithm
algorithm	algorithm	
2023	8246.35	85213.27
2024	8438.13	88748.39
2025	8630.32	92657.37
2026	8756.56	96668.74
2027	8805.13	100759.29
Indicators	China's crude steel output	China's steel industry corporate profits
Best prediction	Metabolism GM(1,1)	ARIMA algorithm
algorithm	algorithm	
2023	10.17	3433.29
2024	11.39	3482.12
2025	11.65	3547.38
2026	12.24	3608.14
2027	12.67	3676.27

4. OUTLOOK FOR THE FUTURE DEVELOPMENT OF THE STEEL INDUSTRY

4.1. Green and low-carbon development is an inevitable trend

At the beginning of 2020, the China Environmental Protection Industry Association issued the "Technical Guidelines for Ultra-low Emission Retrofit of Iron and Steel Enterprises", which pointed out that all relevant industries should strengthen the prevention and control of pollution at the source, use low-sulfur coal, low-sulfur ores and other clean raw materials, and adopt advanced and new clean technologies in production and processing, so as to stifle air pollutants at the source.

4.2. The application of intelligent manufacturing in the iron and steel industry is becoming more and more extensive.

The robust development of intelligent manufacturing industrial structure in the iron and steel industry will strongly promote the iron and steel enterprises to shorten the R&D and production cycle of new products, reduce the normal operating costs of enterprises, improve the production and processing efficiency of products, optimize the process and performance of products, reduce the unnecessary

consumption of energy and resources, and enhance the working environment of employees in field operations, and so on. This year, with the integration of intelligent manufacturing into the iron and steel industry processing and manufacturing fields, more and more domestic steel enterprises began to devote themselves to the construction of intelligent manufacturing industry innovation, to successively achieve remarkable results. Famous domestic steel enterprises such as Baosteel and Nanshan Steel are actively exploring the application of intelligent manufacturing in the steel field.

5. SUMMARY OF THE CURRENT STATUS OF CHINA'S STEEL PANEL DATA INDUSTRY

In this paper, we first use Python to crawl the data, and then use the data visualization software Emotional Sky visualization platform to get a large screen of data in the steel industry, and analyze the future development of the steel industry through the use of multi-algorithm coupled prediction method.

1. Since entering the 21st century, China's iron and steel industry has undergone radical changes, and has become the world's largest producer and consumer of iron and steel materials. Among them, crude steel material has remained the first place in the world for many years, and the number and scale of iron and steel enterprises have been expanding continuously. China's iron and steel industry has shown strong vitality and competitiveness, and will continue to maintain a high level of development for some time to come.

2. Problems in the development process of the iron and steel industry are becoming more and more prominent, restricting the further development of China's iron and steel industry. Steel overcapacity, low industrial concentration, the number of metallurgical colleges and universities can not meet the development needs of regional capacity, the level of professional quality of employees to be improved, electric arc furnace steelmaking underutilization of iron ore resources, iron ore resources dependence is too high, energy saving and emission reduction pressure is huge and other issues. Relevant departments should strengthen macro-control, establish and improve the utilization of resources, environmental protection system, professional training mechanism, etc.; iron and steel enterprises should actively deal with the risks and enhance their competitiveness.

3. At present, the development trend of the iron and steel industry continues to develop in a favorable direction, including: "green and low-carbon policy, strong downstream industry driving the demand for steel, and extensive application of intelligent manufacturing". Through the multi-algorithm coupling to realize the prediction of the future development of the iron and steel industry, and according to the prediction to put forward a new round of outlook on the iron and steel industry. Contribute to the early realization of efficient, intelligent and green development of China's steel industry.

Through continuous adaptation and change, improve the status quo of China's iron and steel industry which is big but not strong, with insufficient high-end production capacity and excess low-end production capacity, and promote the transformation and upgrading of the industry to realize the development path of high efficiency, greening and intelligentization.

REFERENCES

- [1] MOTZ H , GEISELER J. Products of steel slags an opportunity to save natural resources [J].Waste Management,2001,21(3):285-293.
- [2] CHEN Wenying,YIN Xiang,MA Ding. A bottom-up analysis of China's iron and steel industrial energy consumption and CO2 emissions [J]. Applied Energy,2014,136:1174-1183.
- [3] ZHANG Qi,XU Jin,WANG Yujie,et al.Comprehensive assessment of energy conservation and CO2 emissions mitigation in China's iron and steel industry based on dynamic materialflows [J].Applied Energy,2018,209:251-265.

- [4] Wenqing Wu, Sitong An, Chia-Huei Wu, et al. An empirical study on green environmental system certification affects financing cost of high energy consumption enterprises-taking metallurgical enterprises as an example[J]. *Journal of Cleaner Production*, 2020, 244.
- [5] Xiyang Xiang, Qinghua Su, Zhongbo Hu. Non-equidistant grey prediction evolution algorithm: A mathematical model-based meta-heuristic technique.[J] *Swarm and Evolutionary Computation*, 2023, 78.
- [6] Chafak Tarmaninia, Nur Sarmab, Cenk Gezegin, et al. Short term load forecasting based on ARIMA and ANN approaches.[J] *ScienceDirect*, 2023, 9:550-557.
- [7] Huihui Yu, Anmin Yin, Zhi Xu, et al. Grain size characterization of TA1 with GA-BP neural network using laser ultrasonics.[J] *Optik*, 2023, 275.
- [8] Duong Tran Anh, Manish Pandey, Varun Narayan Mishra, et al. Assessment of groundwater potential modeling using support vector machine optimization based on Bayesian multi-objective hyperparameter algorithm.[J] *Applied Soft Computing*, 2023, 132.