https://doi.org/10.31449/inf.v49i34.8473 Informatica **49** (2025) 309–322 **309**

Carbon Emission Prediction in Industrial Zones Using IGWO-Optimized SVM and STIRPAT

Jianguan Xin[1], Jianzhong Yang[2], Huirong Chen[3]*
[1]Department of Information Engineering, Jiangxi Vocational Technical College of Industry & Trade, Nanchang 330103, China
[2]College of Electronic and Information Engineer, Beibu Gulf University, Qinzhou 535011, China
[3]College of Resource and Environment, Beibu Gulf University, Qinzhou 535011, China
E-mail: rongerdan@126.com
*Corresponding author

Keywords: improved grey wolf optimizer, support vector machine, stochastic impacts by regression on population, affluence, and technology, carbon emission projections

Received: March 4, 2025

With the continuous advancement of technology, carbon emission prediction results have become increasingly reliable. However, traditional carbon emission prediction methods face limitations in data and uncertainty, requiring substantial experimental resources and data. Therefore, the study optimizes the Support Vector Machine through the Improved Grey Wolf Optimizer. It combines the extended stochastic environmental impact assessment model to provide a framework for influencing factors and introduces randomness and regression analysis. This approach improves the accuracy and applicability of the fusion model in predicting carbon emissions in industrial zones. Experimental results show that, in the Hubei Province dataset, the proposed model achieves the smallest Mean Squared Error of 0.0075 among four models. The Root Mean Squared Error values of the individual Feedforward Neural Network and Multilayer Perceptron are 0.0101 and 0.0197 higher than that of the proposed model, respectively. Compared to existing single models, such as backpropagation neural networks, the Root Mean Squared Error values of the studied model is significantly reduced by 12%. These results indicate that the proposed prediction model demonstrates excellent timeliness in carbon emission forecasting. This capability provides policy makers with a variety of policy assessment tools to help develop more effective emissions reduction policies.

Povzetek: Hibridni model IGWO-SVM-STIRPAT omogoča bolj kvalitetno napoved emisij ogljika v industrijskih conah, ker združuje optimizacijo parametrov in regresijsko analizo.

1 Introduction

Since the end of the 20th century, the adverse effects of global warming have become increasingly significant. Carbon emissions have caused widespread effects on ecosystems and the economy. Climate change has also led to a rise in natural disasters, severely threatening agriculture and infrastructure [1]. To address this global challenge, governments and businesses worldwide have made reducing carbon emissions a common goal, with an increasing urgency for the implementation of emission reduction policies. Therefore, developing effective carbon emission prediction methods has become an important objective in current research. Although current carbon emission prediction methods are somewhat effective, their results are highly susceptible to various factors. In cases of data with low complexity, there is a risk of overfitting [2]. In recent years, Support Vector Machine (SVM) has been widely applied in carbon emission prediction research due to its strong generalization ability and suitability for nonlinear problems [3]. Compared to SVM, the Improved Grey Wolf Optimizer (IGWO) optimizes faster and adaptively adjust parameters based on model behavior [4]. Combining these two methods in carbon emission prediction allows for quicker solutions with fewer iterations and enhances the global search capability of individual algorithms by leveraging the strengths of different models [5]. Building on this, the study innovatively proposes integrating it with the scalable Stochastic Impacts by Regression on Population, Affluence, and Technology (STIRPAT) model. IGWO-optimized SVM enhances prediction accuracy through a non-linear kernel function and adaptive parameter adjustment. The STIRPAT model compensates for the deficiencies of traditional model to include qualitative factors, such as carbon policies, in prediction by introducing policy drivers. The proposed fusion model aims to overcome the limitations of traditional prediction methods, which are constrained by parameter dependencies and applicability, by improving overall prediction accuracy through the combination of multiple model strengths. This method is expected to offer valuable data insights to support low-carbon transformation and economic development, ensuring well-informed emission reduction policies and fostering a sustainable green economy.

2 Related works

With the continuous breakthroughs in machine learning and data mining technologies, SVM has demonstrated strong regression analysis capabilities in pattern recognition and classification problems. At the same time, the algorithm is efficient and robust when handling large datasets with many features. As a result, numerous scholars from both domestic and international research communities have conducted extensive studies leveraging this advantage. For example, Neethu's team proposed a novel classification method to address the shortcomings of traditional gesture classification methods. This approach applied a threshold segmentation method to finger segmentation based on SVM. The results showed that the sensitivity, specificity, and accuracy of the method all exceeded 90% [6]. Gangadevi and other researchers proposed a hybrid fruit fly optimization model based on SVM to accurately identify tomato diseases. The method simulated hybridization to address the issue of hyperparameters. Simulation experiments validated that the approach upon traditional algorithms while maintaining high precision [7]. Yu's team proposed a novel gain algorithm for model recognition, which optimized SVM combined with the conjugate direction method to enhance the functional gain of traditional algorithms. The results indicated that this algorithm outperformed other models in terms of timeliness [8]. Compared to SVM, STIRPAT is more flexible and scalable, as it does not require a large amount of experimental data for data evaluation. Shobande and Asongu proposed a combined algorithm based on STIRPAT to explore the relationship between information and communication technology and environmental sustainability. This method, when combined with time series data, mined the interactions between various indicators. The results showed that the approach effectively promoted the development of environmental sustainability [9]. Rasoulinezhad's team aimed to explore how carbon emissions and energy efficiency impact green finance, proposing the use of STIRPAT for experimentation. By analyzing the interactions between population, affluence, and technology, they concluded that there was no causal relationship among the variables in the short term [10].

In addition, in order to promote ecological civilization, low-carbon development has gradually become the main goal for future development. Regarding how to achieve accurate carbon emission predictions, scholars both domestically and internationally have proposed different perspectives. For example, Wang and other researchers proposed an extended planning model to address the impact of carbon emission trading price fluctuations on energy system planning. By analyzing device characteristics, they developed a planning model for carbon emission prediction. Simulation experiments verified the reliability of this prediction model [11]. Unsalan's team proposed a novel carbon emission prediction method to study the impact of carbon footprints on global warming. They employed various technological systems combined with SVM to analyze carbon emission data. The results showed that the network model exhibited excellent predictive performance and accuracy [12]. Tang and others proposed a combined algorithm based on a carbon emission planning model to address the issue of energy consumption optimization. By optimizing the configuration and predicting and optimizing carbon emissions, their method provided reliable data support for low-carbon transportation planning [13]. Amanatidou and colleagues proposed using a static floating chamber to collect carbon emissions and methane concentrations for accurate prediction and evaluation of greenhouse gas emissions. Their experiments validated the reliability of this method for assessing the environmental impact of reservoirs through the statistical correlation between gas and water characteristics [14]. Gao's team developed a new hybrid model to predict carbon emissions in the production industry and explored the solution space based on initial sequences and search strategies. Data results showed that the hybrid model had high accuracy and effectiveness [15]. A summary of the results of the available studies is shown in Table 1.

Table 1: Summary of available studies.

Research object	Author	Method	Advantage	Disadvantage
SVM	Neethu P S et al	Thresholding was used on the basis of the SVM	High sensitivity, specificity, and accuracy	Not suitable for complex backgrounds
	Gangadevi E et al	Hybrid fruit fly optimization model based on SVM	With a relatively high accuracy	High computational complexity
	Yu L et al	Optimized SVM combined with the conjugation orientation method	Have a better timeliness	Multi-classification problem is complicated
STIRPAT model	Shobande O A and Asongu S A	A binding algorithm based on the STIRPAT model	Flexibility and strong scalability	High data dependence
	Rasoulinezhad E et al	STIRPAT model	Extensive applicability	Limited handling of the nonlinear relationships
Carbon emission forecast	Wang L et al	Establish an extended planning model	Quantitative analysis ability is outstanding	The local optimal problem
	Unsalan K et al	Using different technical systems with the SVM	Strong classification ability	Parameter selection is sensitive

	Tang Y et al	A combined algorithm based on a carbon emission planning model	The model accuracy is high	The data dependency is relatively high
	Amanatidou E et al	Carbon emission and methane concentration were collected using a static floating chamber	Easy to operate	Spatial representation is limited
	Gao M et al	New hybrid prediction model for carbon emissions	High information utilization rate	Computational resources consume costly

Based on results of domestic and international scholars, current research methods for carbon emission prediction still face limitations in terms of a data quality and a lack of flexibility. The existing RF-SVM model is based on static data training and lacks dynamic adaptability. The neural network is prone to overfitting in low-dimensional temporal data. Some machine learning methods have weak policy variable modeling and a high dependence on data quality. Therefore, this research chooses to combine the optimized SVM with STIRPAT and integrate STIRPAT and machine learning features to build a "policy-technology-emissions" closed-loop feedback. The research aims to improve the timeliness of prediction and policy response ability through parameter optimization, mechanism innovation, and multi-source data collaborative governance, and then to promote the realization of low-carbon emission reduction.

3 Carbon emission prediction model based on the fusion of SVM and STIRPAT and its optimization

3.1 Construction of the industrial zone carbon emission prediction model based on IGWO-SVM

SVM is a model used for data classification in a supervised learning manner. It achieves better generalization ability and classification performance by maximizing the margin of the hyperplane. It is particularly effective for handling nonlinear classification problems and high-dimensional data [16]. However, when applied to large-scale datasets, the algorithm faces high computational complexity and requires significant machine memory during the training process. On the other hand, IGWO has the advantages of strong adaptability and fast convergence speed. It features a convergent factor and an information feedback mechanism that can be adaptively adjusted, making it easier to implement compared to SVM [17]. IGWO is an optimization algorithm that simulates the social behavior of wolf packs. It adjusts experimental parameters adaptively during the optimization process based on task adaptability, reducing parameters to lower the complexity of the model and obtaining the optimal solution in a shorter time. The working process is shown in Figure 1.

In Figure 1, IGWO first initializes parameters to define the number of gray wolves and the positions of the individual wolves. When the positions of the entire wolf pack meet the final condition, the optimal position of the gray wolf is obtained [18]. The dynamic weight adjustment strategy is used to output the objective function value of the optimal gray wolf. The Equation for calculating the distance between an individual and the prey is shown in Equation (1).

$$\vec{D} = \left|\vec{C}.\vec{X}_p(t) - \vec{X}(t)\right| \quad (1)$$

In Equation (1), t and $\vec{C}$ represent the current iteration number and the coefficient vector, respectively. $\vec{X}_p$ represents the prey's position vector, and $\vec{X}$ represents the gray wolf's position vector. When the initialized positions of the gray wolf pack do not meet the termination condition, the objective function value of the wolf pack needs to be calculated, and the positions of the wolves are updated. After obtaining the updated function value, the process returns to the initialization step of the wolf pack [19]. During IGWO operation, hyperparameter combinations for SVM are randomly generated through parameter initialization. The SVM model is then trained, and its fitness value is calculated. Then, the individual position is updated according to the hunting behavior of the gray Wolf, and finally the optimal hyperparameter combination is gradually approached by iterative updating the location of the gray Wolf, so as to realize the integration of IGWO optimized SVM. IGWO significantly improves the search efficiency by controlling the parameter random adjustment strategy, at the same time, by minimizing the MSE and RMSE as the optimization targets, the model parameters are dynamically adjusted to improve the prediction accuracy of the model. The study uses IGWO to optimize the SVM, resulting in IGWO-SVM. The implementation steps of this algorithm are shown in Figure 2.

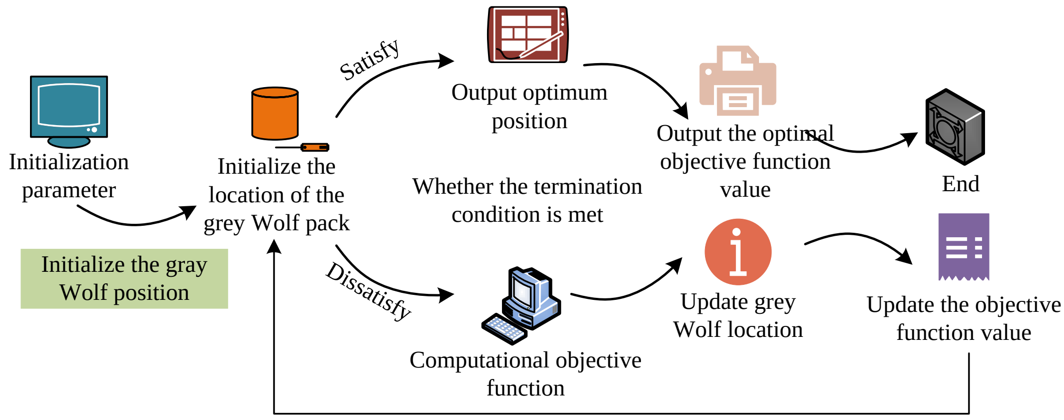


Figure 1: Working process of IGWO.

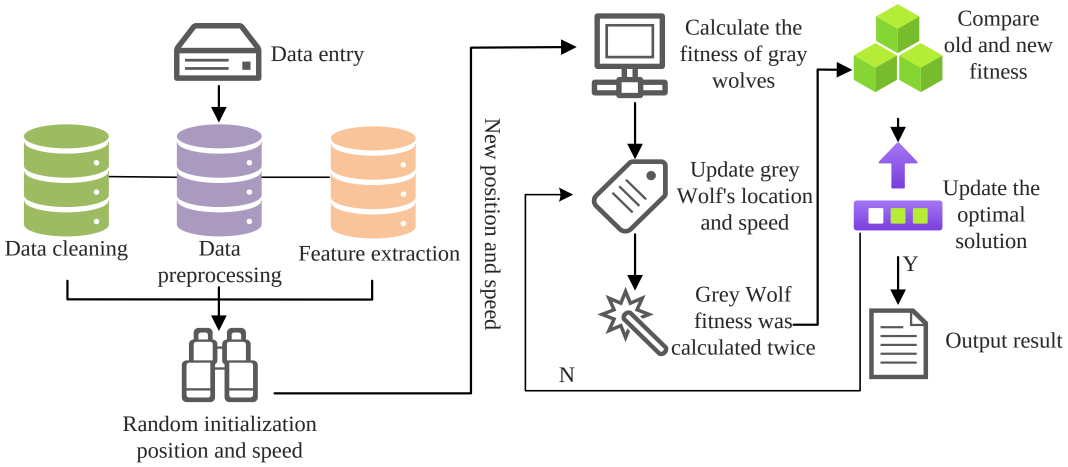


Figure 2: Implementation steps of IGWO-SVM.

As shown in Figure 2, the initial stage of IGWO-SVM involves preprocessing the input data and randomly initializing the data based on the position and velocity of the gray wolves. The fitness of the wolves is calculated using the new positions and velocities, and after a second calculation, the updated fitness is compared with the previous fitness to determine the optimal strategy that satisfies the termination condition. The Equation for calculating the individual gray wolf's tracking of the prey's position is shown in Equation (2).

$$\begin{cases} \vec{D}_\alpha = \left|\vec{C}_1 . \vec{X}_\alpha - \vec{X}\right| \\ \vec{D}_\beta = \left|\vec{C}_2 . \vec{X}_\beta - \vec{X}\right| \\ \vec{D}_\delta = \left|\vec{C}_3 . \vec{X}_\delta - \vec{X}\right| \end{cases} \tag{2}$$

In Equation (2), $\vec{D}_\alpha$ and $\vec{X}_\alpha$ represent the distance between α and other individuals, and α denotes the current position. $\vec{X}_\alpha$ and $\vec{X}_\beta$ represent the distance between β and other individuals, and β is the current position. $\vec{D}_\delta$ and $\vec{X}_\delta$ represent the distance between δ and other individuals, and δ is the current position. $\vec{C}_1$, $\vec{C}_2$, and $\vec{C}_3$ are random vectors. Individual tracking of prey location by grey wolves can shorten the optimization time of SVM parameters through a collaborative search mechanism and capture the non-linear relationship between carbon emissions and driving factors. The new position of the gray wolf individual is determined by the position-weighted average of α, β and δ, as shown in Equation (3).

$$\vec{X}(t+1) = \frac{\vec{X}_\alpha - A_1 \cdot \vec{D}_\alpha + \vec{X}_\beta - A_2 \cdot \vec{D}_\beta + \vec{X}_\delta - A_3 \cdot \vec{D}_\delta}{3} \tag{3}$$

In Equation (3), A_1, A_2 and A_3 coefficients representing the control step size. The convergence factors in IGWO-SVM are used to determine whether the algorithm reaches the optimal solution in the iteration process, and the information feedback mechanism feeds back information into the parameter update strategy by evaluating the quality of the current solution, and then guides the next search direction. IGWO-SVM dynamically adjusts the search step size and search range for gray wolf individuals by introducing adaptive weights and following convergence factors and information feedback mechanisms, and terminates the iteration early when the fitness value dose not significantly change in multiple successive iterations. Moreover, integrating IGWO can compensate for the lack of convergence performance and the need for a large amount of training data in SVM. The proposed IGWO-SVM carbon emission prediction model is shown in Figure 3.

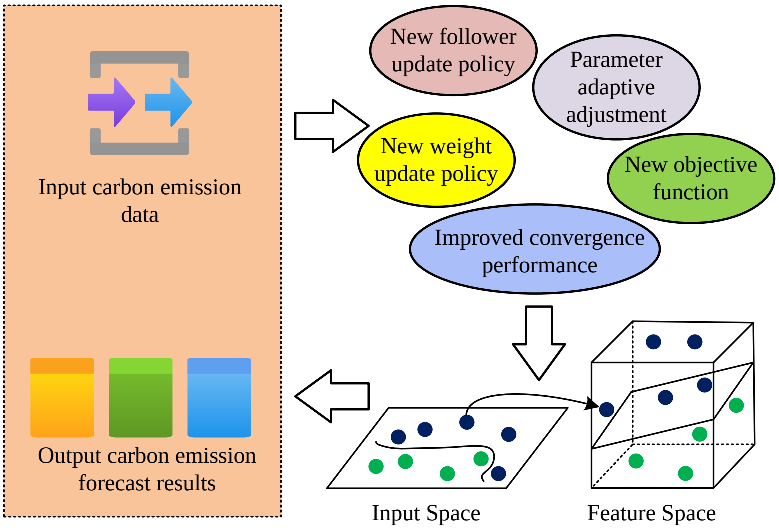


Figure 3: Working diagram of IGWO-SVM carbon emission prediction modeling.

As shown in Figure 3, the model for predicting carbon emissions first requires determining the target range and gathering relevant parameters. After initializing the parameters, the target function needed for the experiment is obtained. The new target function is then derived through adaptive parameter adjustments and weight updating strategies. This updated target function is used as the input vector, and SVM is employed to find a separating hyperplane in the feature space to classify the data. After data integration, the final carbon emission prediction results are obtained. The distance calculation Equation from any point to the hyperplane in the sample space is the core mathematical tool of SVM, which can allow the sample space to be divided into different regions through the hyperplane to capture the non-linear relationship between carbon emissions and driving factors. The distance from any point in the sample space to the hyperplane is calculated using the Equation in Equation (4).

$$r = \frac{\left|w^T + b\right|}{\|w\|} \tag{4}$$

In Equation (4), w^T and b represent the normal vector and the displacement term, respectively. w represents the normal vector of the hyperplane. The Equation for calculating the distance between the data points and the decision boundary is shown in Equation (5).

$$d_i = \frac{w^T x_i + b}{\|w\|} \tag{5}$$

In Equation (5), $\|w\|$ and x_i represent the Euclidean norm of the normal vector w and the data points, respectively. The Equation for calculating the individual gray wolf is shown in Equation (6).

$$\begin{cases} \vec{X}_1 = \vec{X}_\alpha - A_1 \cdot \vec{D}_\alpha \\ \vec{X}_2 = \vec{X}_\beta - A_2 \cdot \vec{D}_\beta \\ \vec{X}_3 = \vec{X}_\delta - A_3 \cdot \vec{D}_\delta \end{cases} \tag{6}$$

In Equation (6), $\vec{X}_1$, $\vec{X}_2$, and $\vec{X}_2$ represent different gray wolf individuals, determining the step size and direction in which the gray wolf individuals move towards α, β, and δ. By applying the IGWO-SVM model to carbon emission prediction, the model parameters are adaptively adjusted according to the different characteristics of carbon emissions, allowing the initial feature parameters to be more evenly distributed in the hyperplane solution space, thus improving the model's prediction accuracy and generalization ability. Next, the Gaussian kernel is preferentially selected as the kernel function, and the optimal value of the generalization and fitting ability of the equilibrium model is found by IGWO. The adjustment of the kernel function parameters was searched in a smaller range by IGWO to avoid the model being too sensitive to local data. This provides reliable data support for formulating and implementing carbon emission policies.

3.2 Optimization of the carbon emission prediction model combined with STIRPAT

Although applying IGWO to optimize SVM improves its sensitivity to parameters, the fusion model may get stuck in local optima in complex problems, which affects the model's prediction accuracy. Therefore, this study proposes further introducing STIRPAT to improve the prediction results based on the limitations of IGWO-SVM. By introducing randomness and regression methods, the research can flexibly handle the nonlinear relationship between carbon emissions and population, affluence and technology level, while the traditional IPAT model can only describe the linear relationship. And the STIRPAT model supports the addition of other variables, which can more comprehensively analyze the drivers of carbon emissions in industrial zones. Other models are usually limited to linear assumptions or single variables. Its components are shown in Figure 4.

As shown in Figure 4, the pressure in STIRPAT determines the environmental carrying capacity and the impact on the ecosystem, including factors such as climate change and environmental pollution. The four components—trend, pressure, impact, and response—together describe the comprehensive effect of human factors on the environment. By simulating and predicting the influencing factors, the model provides an important basis for ecosystem management [20]. The core calculation Equation of the STIRPAT model is shown in Equation (7).

$$I = aP^{b}A^{c}T^{d}e \tag{7}$$

In Equation (7), I represents environmental impacts, such as carbon emissions. a and P represent the model's elasticity coefficient and population size, while A and T represent affluence and technological level. b is the elasticity coefficient for population size, and c represents the elasticity coefficient for affluence. d and e represent the elasticity coefficient for technological level and the error term, respectively. By adjusting the exponents in the model and adding new variables, the model can be made suitable for different research needs, offering excellent flexibility and scalability. The mechanism of incorporating STIRPAT to improve the carbon emission prediction model is shown in Figure 5.

As shown in Figure 5, the IGWO-SVM carbon emission prediction model defines and normalizes the feature space vectors output by SVM, and after variable transformation, the STIRPAT model is used to improve the fusion model through model selection and fitting to determine the independent and dependent variables. The study constructs a regression model using linear regression methods, and finally evaluates the model's goodness of fit and predictive ability through residual analysis tools, thus obtaining the prediction results of the model. The error Equation is calculated as shown in Equation (8).

$$\begin{cases} Error(s \mid X, y) = (Xs - y)^{T(Xs-y)} \\ X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{m1} & X_{m2} & \cdots & X_{mn} \end{bmatrix} \\ s = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix} \end{cases} \tag{8}$$

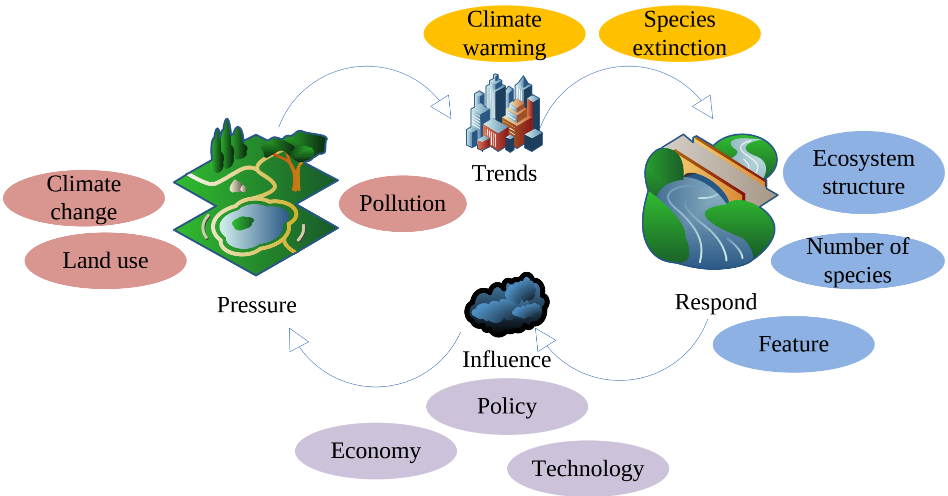


Figure 4: Components of STIRPAT.

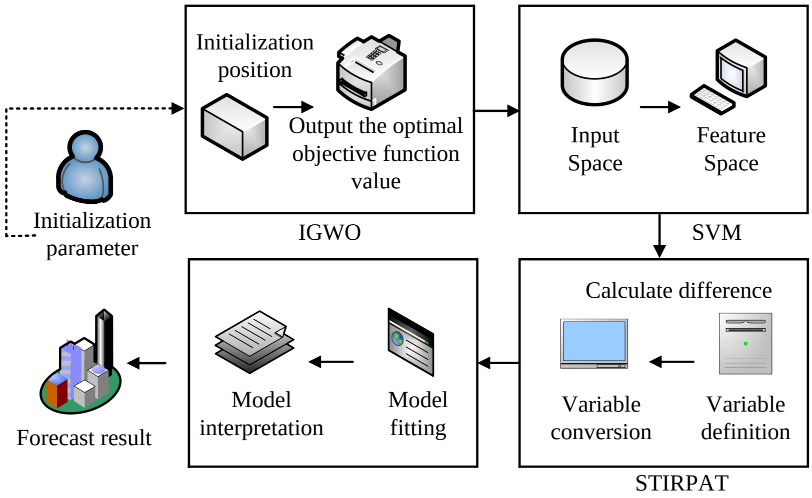


Figure 5: Workflow of the carbon emissions forecasting model introduced into STIRPAT.

In Equation (8), X and y represent the sample input matrix of $m \times n$ and the column vector of $m \times 1$, respectively. s represents the column vector of $n \times 1$. m and n are the number of samples and the number of features for each sample, respectively. The calculation Equation for the objective function is shown in Equation (9).

$$R = \min \frac{1}{2} w^T w + C \sum_{i=1}^{n} \gamma_i \tag{9}$$

In Equation (9), C and γ represent the penalty factor and slack variables, respectively. w weight the impact of different characteristics on carbon emissions in the carbon emission prediction model. By optimizing the normal vector, the relationship between carbon emissions and drivers can be fitted more accurately, and adjusting the penalty factor can avoid model overfitting and ensure the reliability of the prediction results. Minimizing the normal vector and penalty factor optimize the feature weights and reduce the computational complexity of the model. This model not only reduces the difficulty of data acquisition but also enhances the model's flexibility through regression analysis, effectively compensating for the limitations of traditional prediction models influenced by data quality. The study selected the filtering method as a feature selection technique, evaluating feature importance based on statistical indicators such as variance and screening it independently of the model. The carbon emission prediction model combining IGWO-SVM and STIRPAT is shown in Figure 6.

From Figure 6, it can be observed that after inputting carbon emission and influencing factor data into the model, the first step is to initialize the data through IGWO-SVM and output the optimal objective function value. This function value forms the optimal hyperplane in the feature space after maximizing the classification margin. The second part, the STIRPAT module, mainly standardizes the data, determines the distribution of parameters to be optimized, and optimizes the parameters using a random search method. Finally, after result verification, the predicted value of carbon emissions is output. The calculation Equation for the inner product kernel function is shown in Equation (10).

$$K(c_i, c_j) = \exp(-y \| c_i - c_j \|^2) \tag{10}$$

In Equation (10), c_i and c_j represent the input factors and the expected values, respectively. The calculation Equation for the nonlinear function is shown in Equation (11).

$$f(x) = \sum_{sv} (\alpha_i - \alpha_i^*) K(c_i g c_j) + v \tag{11}$$

In Equation (11), α_i and α_i^* represent the support vector parameters, while v and $K(c_i g c_j)$ represent the constant term and the inner product kernel function, respectively. The calculation Equation for the function margin is shown in Equation (12).

$$\hat{\gamma} = a(w^T x + v) \tag{12}$$

In Equation (12), $w^T x + b$ represents the distance from point T to the hyperplane. The integration of IGWO-SVM and STIRPAT mainly builds the STIRPAT model using historical data, outputs the quantitative relationship between carbon emission and each factor as the input feature of SVM, and then trains the SVM model optimized by IGWO. The optimization of the fusion model by STIRPAT not only enhances the model's flexibility and comprehensiveness, but also compensates for the lack of timeliness in single prediction algorithms. This makes the prediction results more accurate, greatly improving the timeliness and transparency of carbon emission data.

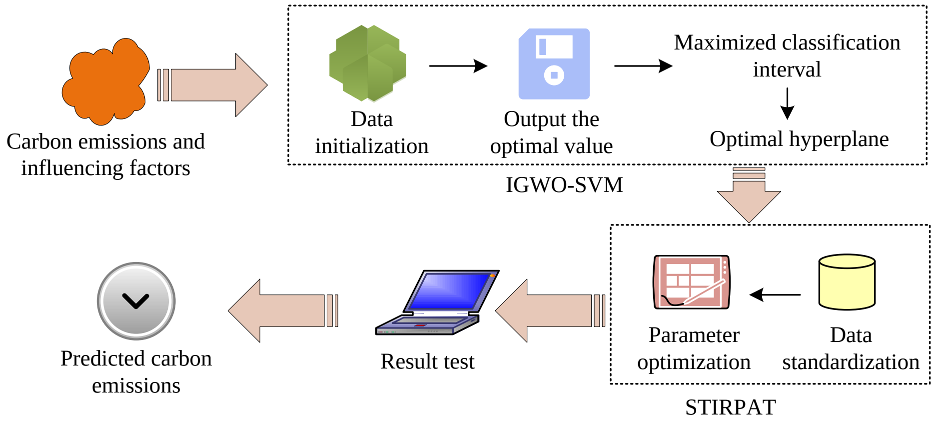


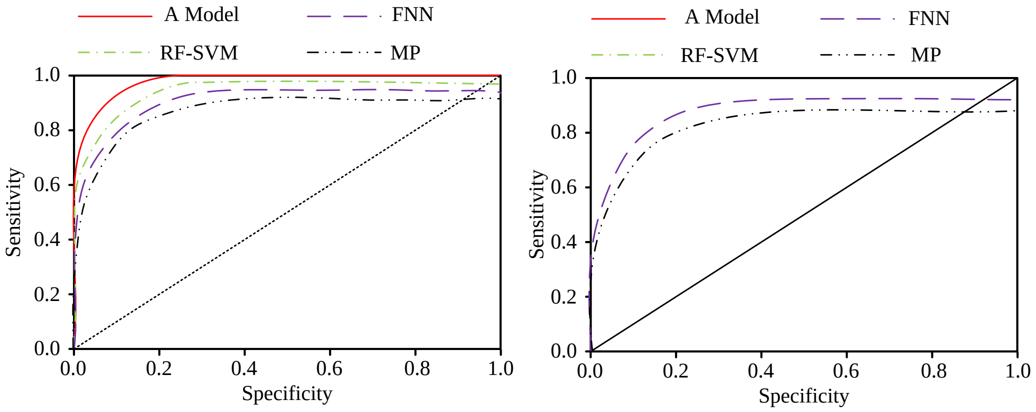
Figure 6: Final carbon emissions forecasting model workflow diagram.

Design the rolling verification mechanism for time-dependent nature of carbon emission data, and conduct the cross-validation of the model evaluation by simulating the actual prediction scenario. The penalty factor of SVM is dynamically adjusted within the range of 0.1~100 through IGWO to balance model complexity and training error. The optimization interval of Gaussian kernel parameters and the value of ε in the regression task are set to 0.001-1001000 and 0.01-0.1 respectively, to avoid overfitting of the model to industrial data noise. The population elasticity coefficient and the technical elasticity coefficient of the STIRPAT model are constrained between 0.8 and 1.5 and 0.3~0.9, respectively. To enhance the adaptability of the model to different policy strengths, the study treats the policy drivers by normalization and maps to an interval of 0 to 1.Set the IGWO algorithm to stop the calculation when the fitness value change rate is below 110-4 after 20 consecutive iterations.

4 Performance verification of the industrial zone carbon emission prediction based on the fusion model

4.1 Comparative performance analysis of the fusion model

In order to validate the performance of the STIRPAT combined with IGWO-SVM hybrid model, the study named this fusion model as Model A, and selected Feedforward Neural Network (FNN), Random Forest and Support Vector Machine (RF-SVM), and Multilayer Perceptron (MP) as comparison prediction models. FNN, RF-SVM and MP all have strong non-linear modeling capabilities, and can compare the performance with the integrated model of IGWO-SVM and STIRPAT in complex data scenarios, and choosing different types of models for comparison can verify the superiority of the integrated model from multiple perspectives. The study selected the programming language and TensorFlow 2.5 to implement data preprocessing and model implementation, using AMD Ryzen 7 and NVIDIA RTX 3060 for the hardware configuration, with a memory size of 32GB. The four models had 400 iterations, with a population size of 35. The 400 iterations ensure that the algorithm reaches the convergence state when optimizing the model parameters and effectively controls the computing resource consumption. Setting the overall size set to 35 ensures that the algorithm maintains sufficient population diversity within the search space, reducing the likelihood of a local optimal solution. The carbon emission prediction results for different provinces and years were compared and analyzed. The experimental datasets included the China Emission Accounts and Datasets (CEADs) dataset and the Multi-resolution Emission Inventory for China (MEIC) dataset. The CEADs dataset served as the training set, containing carbon accounting inventories at multiple scales, and was used to verify the performance of the fusion model. The MEIC dataset, as the testing set, covered carbon emission data from various industry sectors and was used to test the accuracy of the model in real-world applications. The CEAD and MEIC datasets adopt the proportional partition of 80% training set and 20% test set, combined with cross-validation to ensure that the model has good generalization ability while fully learning the data features. To ensure fairness in the experiment, the same parameter preprocessing was applied to all four models. CEADs data set in the process of pretreatment, the first 3 σ criteria to identify and eliminate the outlier indicators such as carbon intensity, through time series interpolation processing local missing, then the carbon emissions data and industrial economic indicators features cross and build multidimensional driver, normalized after the carbon emissions sequence decomposition into high frequency and low frequency mode. In the preprocessing of MEIC data set, spatial weighting method was used to aggregate the grid data to regional scales, divide carbon emission sources according to industrial sub-industries, and extract industry characteristic parameters as input variables for the model. Finally, the carbon emission trend differences of the data set was compared by cross-verification, and the systematic deviation was corrected by Bayesian network. And the ROC curves for the four models in the CEADs and MEIC datasets are shown in Figure 7.



(a) ROC curves of the four models in the CEADs dataset

(b) ROC curves of the four models in the MEIC dataset

Figure 7: ROC curves of the four models in CEADs dataset and MEIC dataset.

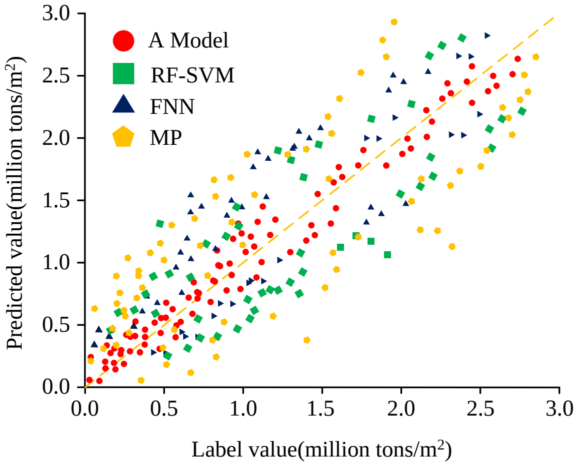
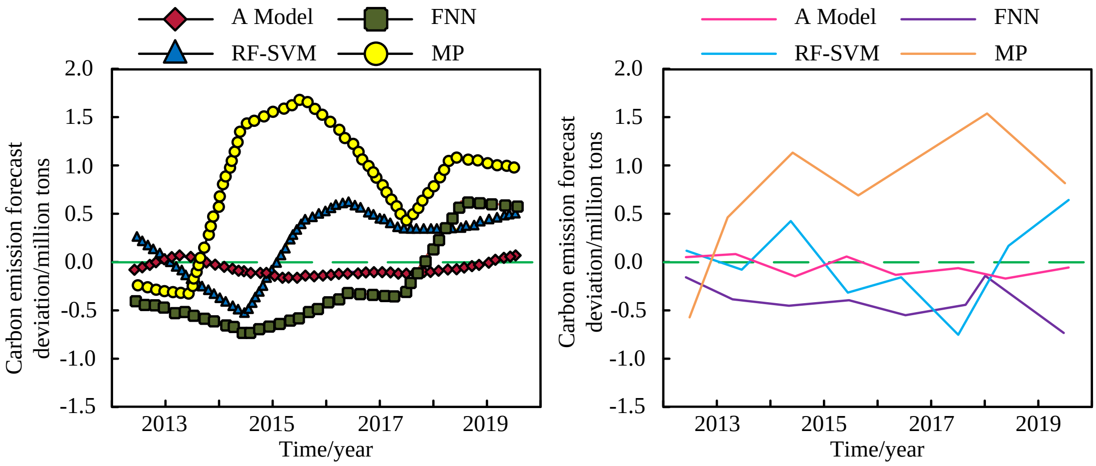


Figure 8: Prediction results of the four models in the CEADs dataset.

As shown in Figure 7(a), in the CEADs dataset, the sensitivity curves of all four models increased as specificity increased. The ROC curve of Model A was closest to the upper left corner, and the Area Under Curve (AUC) value was the largest, reaching 0.95. The AUC value of the RF-SVM model was relatively lower than that of Model A, at 0.91, while the AUC values of the FNN and MP models were noticeably different from that of Model A, at 0.86 and 0.82, respectively. As shown in Figure 7(b), Model A performed well in the MEIC dataset, with an AUC value of 0.94, while the AUC values of the other three comparison models were all lower than that of Model A. The AUC values of the RF-SVM and FNN models were 0.89 and 0.84, respectively, while the MP model had the lowest AUC value, at only 0.80. In summary, from Figure 7, it can be seen that in different datasets, the ROC curve of the proposed Model A was the closest to the left side, and the area under the curve was the largest, demonstrating excellent classification performance compared to traditional algorithms. The CEADs dataset contained rich data samples, and by transforming the samples, the number of training samples was increased, reducing the risk of model overfitting. Therefore, the experiment used this as input data, and the carbon emission prediction values output by the four models were analyzed against the dataset labels, as shown in Figure 8.

As shown in Figure 8, Model A exhibited relatively good prediction results in the CEADs dataset, when the label value is 0~1.0million tons / m2, the predicted value of carbon emission is close to the label value is relatively dense. RF-SVM prediction results were more scattered compared to Model A, and when the label value was 1.5 million tons/m2, its carbon emission prediction reached 1.92 million tons/m², deviating significantly from the label value. At the same time, both FNN and MP performed worse than Model A. When the carbon emission label value was 1.5 million tons/m2, FNN's carbon emission prediction deviated to 2.06 million tons/m², which was 0.55 million tons/m² higher than Model A's prediction. Based on the data results, it can be observed that Model A exhibited significant prediction accuracy among the four algorithms and maintained good performance even when the label values varied. For a comprehensive evaluation of the overall performance of the fusion model, the study also analyzed the carbon emission prediction deviations of the four models in the training and testing datasets, as shown in Figure 9.



(a) Carbon emission prediction error values of the four models in the CEADs dataset

(b) Carbon emission prediction error values of the four models in the MEIC dataset

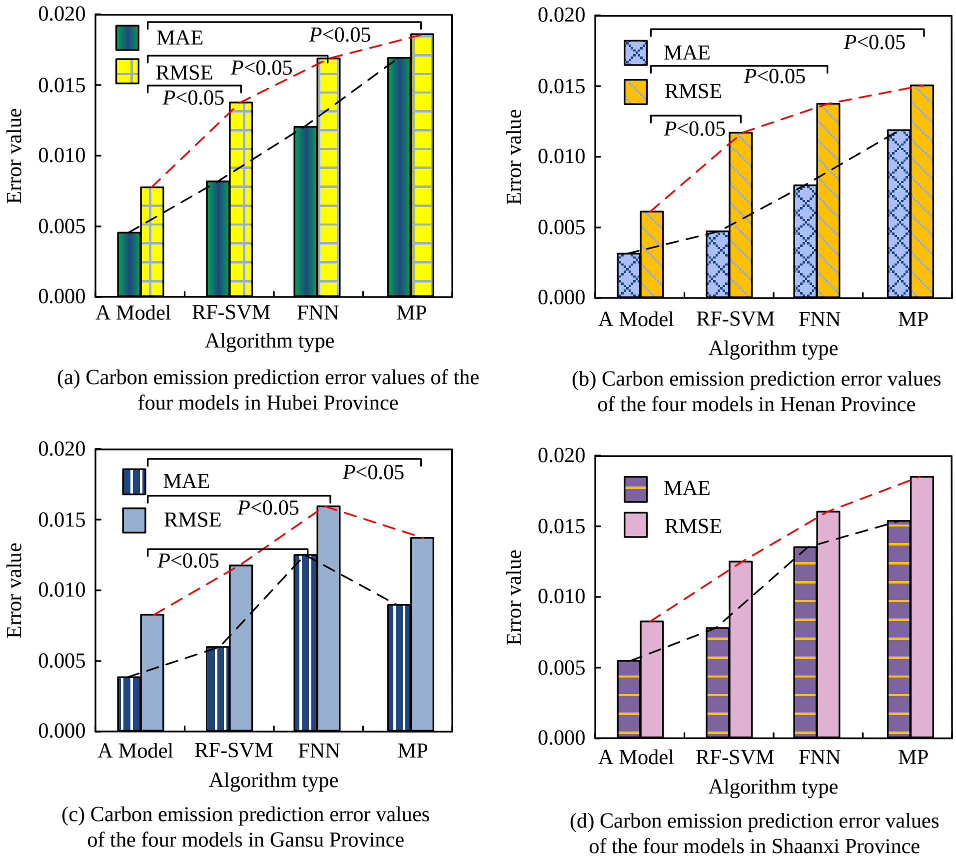
Figure 9: Carbon emission prediction deviation of four models in training set and test set.

As shown in Figure 9(a), in the CEADs dataset, Model A's carbon emission predictions were the closest to the actual values, resulting in the smallest data deviation. When the year was 2014, Model A's prediction deviation was only -0.003, while MP's prediction deviation reached 1.46 in 2016. The deviation curves of the comparison algorithms also fluctuated more significantly, and their overall performance was inferior to that of Model A. According to Figure 9(b), Model A achieved the minimum carbon emission prediction deviation of -0.009 in 2016. The corresponding prediction deviations for RF-SVM and FNN were -0.27 and -0.48, while MP's prediction deviation reached 0.78. Based on the data results, it can be concluded that the proposed Model A demonstrated excellent prediction performance and accuracy when compared to the other algorithms.

4.2 Verification of the prediction effect of the fusion model in practical application scenarios

To further validate the performance of the proposed fusion model in real-world carbon emission prediction scenarios, the study selected carbon emission data from four provinces—Hubei, Henan, Gansu, and Shaanxi—for the experiment, and assigned different sample numbers to them. The sample number range of Hubei Province is 001-100, and the data covers major carbon emission industries such as energy production and manufacturing. The sample number range of Henan Province is 101-200, which includes the carbon emission data of some emerging industries. The sample number of Gansu Province is 201-300, and the sample data is mainly energy production and mineral resources development. The sample number of Shaanxi Province is 301-400. The data includes carbon emission data mainly from high-tech industries and traditional energy industries. The difference in the accuracy of carbon emission prediction between different provinces and cities mainly stems from the combined impact of data quality, regional policy and model adaptability. The carbon emission prediction accuracy of the four models in various provinces is shown in Figure 10.

As shown in Figure 10, the four models demonstrated varying levels of accuracy in carbon emission predictions across different provinces and cities, the root mean square error (Root Mean Squared Error, RMSE) is statistically significant. In Hubei, Model A and RF-SVM had Mean Absolute Errors (MAE) of 0.0048 and 0.0076, respectively, while FNN and MP had MAEs of 0.0121 and 0.0173, respectively. Among the four models, MP had the highest Root Mean Square Error (RMSE) at 0.182, while Model A's RMSE was only 0.0047. In Henan and Gansu, Model A's MAEs were 0.0032 and 0.0041, respectively, with the MAE and RMSE of the other three comparison models being higher than those of Model A. In Shaanxi, Model A had the smallest MAE and RMSE at 0.0052 and 0.0073, respectively, showing the largest difference compared to MP's errors. These results indicated that Model A exhibited the smallest prediction errors among the four algorithms, making it more accurate in predicting carbon emissions than traditional methods. The adaptability of the four models in Hubei and Shaanxi is shown in Figure 11.



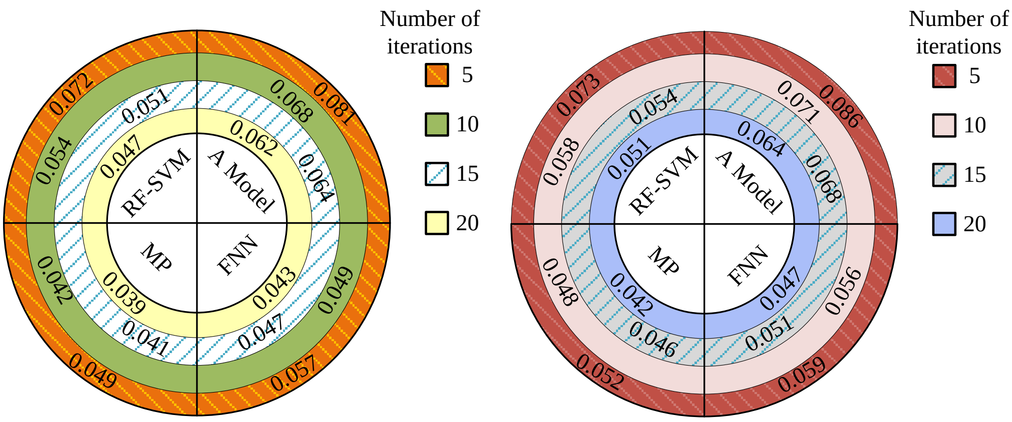
(a) Carbon emission prediction error values of the four models in Hubei Province

(b) Carbon emission prediction error values of the four models in Henan Province

(c) Carbon emission prediction error values of the four models in Gansu Province

(d) Carbon emission prediction error values of the four models in Shaanxi Province

Figure 10: Comparison of carbon emission forecasting accuracy in different provinces and municipalities.



(a) The fitness of the four models in Hubei Province (b) The fitness of the four models in Shaanxi Province

Figure 11: The fitness of four models in different provinces and cities.

As shown in Figure 11(a), in Hubei, when the number of iterations was 5, the fitness values of Model A and RF-SNN were 0.081 and 0.072, respectively. Compared to Model A, the fitness of FNN decreased by 0.024. MP had the lowest fitness, with a value of only 0.049. When the number of iterations was 20, Model A's fitness was 0.023 higher than that of MP, and the fitness values of the other two comparison algorithms were lower than that of Model A. In Figure 11(b), in Shaanxi, the fitness of Model A was higher than that of the comparison models. When the number of iterations was 20, MP's fitness was only 0.042, while Model A's fitness reached 0.064. Overall, Model A demonstrated the best fitness across different data samples, showing an improvement in performance compared to traditional algorithms. The prediction accuracy of Model A and comparison models on the CEADs dataset, MEIC dataset, and sample data from 2018 is shown in Table 2.

Table 2: Prediction accuracy of four models in CEADs dataset and MEIC dataset.

Data set	Model	MSE	RMSE	MAE	MAPE	Training time /s
Sample data of Hubei Province	A Model	0.0075	0.0866	0.0086	0.0186	64.39
	RF-SVM	0.0126*	0.1122*	0.0127*	0.0249*	116.32
	FNN	0.0153*	0.1237*	0.0164*	0.0275*	122.08
	MP	0.0169*	0.1300*	0.0178*	0.0304*	129.61
Sample data of Henan Province	A Model	0.0062	0.0787	0.0092	0.0191	51.37
	RF-SVM	0.0134*	0.1158*	0.0143*	0.0284*	114.58
	FNN	0.0159*	0.1261*	0.0172*	0.0291*	119.34
	MP	0.0156*	0.1249*	0.0186*	0.0327*	125.79
Gansu province sample data	A Model	0.0079	0.0889	0.0081	0.0142	59.05
	RF-SVM	0.0143*	0.1196*	0.0119*	0.0266*	114.87
	FNN	0.0152*	0.1233*	0.0167*	0.0293*	126.43
	MP	0.0161*	0.1269*	0.0206*	0.0375*	131.82
Shaanxi province sample data	A Model	0.0074	0.0860	0.0079	0.0148	62.39
	RF-SVM	0.0141*	0.1187*	0.0117*	0.0264*	118.74
	FNN	0.0157*	0.1253*	0.0165*	0.0296*	123.49
	MP	0.0169*	0.1300*	0.0078*	0.0341*	126.13

Note: * represents P <0.05, meaning that the results are statistically significant.

The data results in Table 2 showed that Model A demonstrated excellent prediction performance across all datasets. In the Hubei dataset, the RMSE of Model A was the smallest among the four models, with a value of 0.0075, while the RMSE of FNN and MP were higher than Model A by 0.0101 and 0.0197, respectively. In the Henan dataset, the MAE of Model A was 0.0092, and the errors of the other three comparison algorithms were all higher than that of Model A, with the MP's Mean Absolute Percentage Error (MAPE) reaching 0.0327. In the Gansu dataset, Model A also exhibited good accuracy, with an MAE of only 0.0081, while the MAEs of RF-SVM and FNN were 0.0119 and 0.0167, respectively, both higher than that of Model A. From the perspective of training time in different datasets, the training time of the proposed model was less than 100s, while the training time of the comparison model reached 129.61s, and according to the experimental results, the obtained error data were statistically significant. According to the experimental results, the obtained study data are statistically significant. Overall, the proposed fusion model performed excellently in various aspects, providing precise prediction performance on different sample data. It improved the applicability of traditional algorithms in large-scale big data applications. Moreover, applying this hybrid model to carbon emission prediction has provided a new approach for scientifically and rationally formulating emission reduction policies.

5 Discussion

As a key part of promoting green low-carbon transformation and realizing the goal of "dual carbon", carbon emission prediction can provide scientific basis for relevant institutions to formulate emission reduction policies by predicting the future carbon emission trend and peak time. According to the ROC curves of the proposed prediction model in different datasets, it can be found that the proposed model has excellent classification performance compared with the traditional comparison algorithm, and can still maintain good prediction performance under the condition of label value changes. Moreover, the predicted carbon emission value of the proposed model is the closest to the actual value, and the data deviation generated is the smallest. In the actual application scenario, the fitness of the proposed model in different data samples is the best. The fusion model dynamically screens the subset of features by combining the IGWO algorithm to ensure that the model can still run efficiently on large datasets, and the normalization in data preprocessing enables the model to adapt to a larger scale of data input.

The RF-SVM model has high parameter sensitivity when processing high-dimensional data, which is prone to overfitting risk. It is difficult for MP model to deal with the complex interaction effect between carbon emission and multiple factors, and it lacks adaptability to external variables, resulting in a large prediction error. The dynamic adjustment of parameters by IGWO in the proposed model effectively balances the bias and variance, making the model maintain a low error on different datasets. Moreover, STIRPAT has enhanced the model's ability to explain complex systems by introducing macro-driving factors, thus reducing the prediction error caused by sample bias. In terms of bias-variance tradeoff, the dynamic feature selection mechanism of the proposed model can automatically identify key drivers and reduce the prediction bias caused by feature redundancy or missing. The implementation of the research model in the real-world industrial environment or policy making background can be implemented through model verification optimization and policy coordination, so as to provide a quantitative decision tool for "two-carbon targets".

Although the IGWO-SVM-STIRPAT model shows excellent performance in the prediction of carbon emission in industrial zones, the noise data may interfere with the feature selection process of the IGWO algorithm and lead to the inaccurate selected feature subset, which affects the explanatory power and prediction effect of the fusion model. The model prediction should consider the short-term emission reduction targets and emphasize the balance between economic and social development in the industrial areas, and eliminate the variables that may cause bias to ensure the fairness of the prediction results. The fusion model will meet future challenges posed by the challenges of future carbon policy changes and new industrial data through carbon policy scenario simulation and adaptive optimization of the model.

6 Conclusion

To address the issues of low accuracy, limitations, and uncertainty in traditional carbon emission prediction models, this study proposed the IGWO-SVM, a novel predictive model combined with STIRPAT. This model achieves accurate carbon emission prediction and enhances the overall applicability and robustness of the model compared to traditional single algorithms. Experimental data results showed that in the CEADs dataset, as specificity increased, the sensitivity curve of the four models also rose. The ROC curve of Model A was closest to the top-left corner, with the largest AUC under the curve, reaching 0.95. The AUC value of RF-SVM was slightly lower than that of Model A, at 0.91, while the AUC values of FNN and MP showed significant differences compared to Model A, at 0.86 and 0.82, respectively. The prediction results of RF-SVM were more dispersed than those of Model A. When the label value was 1.5 million tons/m2, the carbon emission prediction of RF-SVM reached 1.92 million tons/m², deviating significantly from the label value. Similarly, both the FNN and MP had inferior prediction performance compared to Model A. When the carbon emission label value was 1.5 million tons/m2, the prediction of FNN deviated to 2.06 million tons/m², which was 0.55 million tons/m² higher than the prediction of Model A. Overall, the proposed fusion model demonstrated excellent accuracy and prediction performance in carbon emission prediction, with the best adaptability across different data samples. Compared to traditional algorithms, its performance showed notable improvements. The proposed model provides policy makers with scientific and interpretable carbon emission prediction tools through the elastic coefficient analysis of STIRPAT, feature selection optimized by IGWO and high-precision prediction of SVM. However, its classification performance still has room for improvement compared to the other three models, so further optimization and refinement of this fusion algorithm are necessary in future work.

References

[1] Yuting Luo. Artificial algae-optimized support vector machine for sentiment analysis of network public opinion. Informatica, 48(21):167-175, 2024. https://doi.org/10.31449/inf.v48i21.6620

[2] A. J. Gnanamalar, R. Bhavani, A. S. Arulini, and M. S. Veerraju. CNN-SVM based fault detection, classification and location of multi-terminal VSC-HVDC system. Journal of Electrical Engineering & Technology, 18(4):3335-3347, 2023. https://doi.org/10.1007/s42835-023-01391-5

[3] Hassane Ezziane, Hamza Houassine, Samir Moulahoum, and Moustafa Sahnoune Chaouche. A novel method to identification type, location, and extent of transformer winding faults based on FRA and SMOTE-SVM. Russian Journal of Nondestructive Testing, 58(5):391-404, 2022. https://doi.org/10.1134/S1061830922050047

[4] Ahmad Al Badawi, Ling Chen, and Saru Vig. Fast homomorphic SVM inference on encrypted data. Neural Computing and Applications, 34(18):15555-15573, 2022. https://doi.org/10.1007/s00521-022-07202-8

[5] Fenfen Cheng, and Junfeng Yao. The effect of the digital economy on carbon emissions in China's construction industry: Evidence from spatial econometric analysis. Environmental Science and Pollution Research, 31(54):63085-63097. https://doi.org/10.1007/s11356-024-35426-9

[6] P. S. Neethu, R. Suguna, and S. Palanivel Rajan. Performance evaluation of SVM-based hand gesture detection and recognition system using distance transform on different data sets for autonomous vehicle moving applications. Circuit World, 48(2):204-214, 2022. https://doi.org/10.1108/CW-06-2020-0106

[7] E. Gangadevi, R. Shoba Rani, Rajesh Kumar Dhanaraj, and Anand Nayyar. Spot-out fruit fly algorithm with simulated annealing optimized SVM for detecting tomato plant diseases. Neural Computing & Applications, 36(80):4349-4375, 2024. https://doi.org/10.1007/s00521-023-09295-1

[8] Lang Yu, Xin Ma, and Shengjie Li. A fast conjugate functional gain sequential minimal optimization training algorithm for LS-SVM model. Neural Computing & Applications, 35(8):6095-6113, 2023. https://doi.org/10.1007/s00521-022-07875-1

[9] Olatunji A. Shobande, and Simplice A. Asongu. Searching for sustainable footprints: Does ICT increase CO2 emissions. Environmental Modeling & Assessment, 28(1):133-143, 2023. https://doi.org/10.1016/j.eneco.2019.04.014

[10] Ehsan Rasoulinezhad, and Farhad Taghizadeh-Hesary. Role of green finance in improving energy efficiency and renewable energy development. Energy Efficiency, 15(2):13-24, 2022. https://doi.org/10.1007/s12053-022-10021-4

[11] Lei Wang, Wei Dai, Liuzhu Zhu, Xuli Wang, Chenxu Yin, Hao Cong, Tiancheng Shi, Xianjun Qi, and Rui Bi, Multi-objective expansion planning of park-level integrated energy system considering the volatility trend of CETP. IET Generation, Transmission & Distribution, 16(6):1225-1243, 2022. https://doi.org/10.1049/gtd2.12363

[12] Omer Faruk Cansiz, Kevser Unsalan, and Fatih Unes. Prediction of CO2 emission in transportation sector by computational intelligence techniques. International Journal of Global Warming, 27(3):271-283, 2022. https://doi.org/10.1504/ijgw.2022.10048914

[13] Yang Tang, Weiwei Liu, Saurabh Singh, Osama Alfarraj Saurabh Singh, and Amr Tolba. Passenger

flow forecast for low carbon urban transport based on bi-level programming model. Journal of Internet Technology, 24(5):1067-1077, 2023. https://doi.org/10.53106/160792642023092405005

[14] Elisavet Amanatidou, Georgios Samiotis, Eleni Trikoilidou, L. Tsikritzis, and Nikolaos Taousanidis. Centennial assessment of greenhouse gases emissions of young and old hydroelectric reservoir in mediterranean mainland. Journal of Environmental Informatics, 41(1):27-36, 2023. https://doi.org/10.3808/jei.202300485

[15] Mengxing Gao, ChenGuang Liu, Zigui Liu, and Xi Chen. A tabu search algorithm for the unrelated parallel machine scheduling problem with varied carbon emission constraints in different time intervals. The Journal of the Operational Research Society, 75(6):1111-1125, 2024. https://doi.org/10.1080/01605682.2023.2233555

[16] Xuefei Xie, and Lin Chen. Dynamic programming-based energy optimization for pure electric commercial vehicles with predictive cruise control. Informatica, 48(22):129-145, 2024. https://doi.org/10.31449/inf.v48i22.6922

[17] Jieqiong Zhang. Optimizing the analysis of energy plants and high-power applications utilizing the energy guard ensemble selector (EGES). Informatica, 49(10):113-126, 2025. https://doi.org/10.31449/inf.v49i10.7264

[18] Jinghan Zhou, Yongxiu He, Yuan Lyu, Kehui Wang, Yiran Che, and Xiaoqing Wang. Long-term electricity forecasting for the industrial sector in western China under the carbon peaking and carbon neutral targets. Energy for Sustainable Development, 73(3):174-187, 2023. https://doi.org/10.1016/j.esd.2023.02.003

[19] Rajkumar Palaniappan. Comparative analysis of support vector machine, random forest and k-nearest neighbor classifiers for predicting remaining usage life of roller bearings. Informatica, 48(7):39-52, 2024. https://doi.org/10.31449/inf.v48i7.5726

[20] Qiankun Jiang, and Haiyan Wang. Risk assessment method for new energy vehicle supply chain based on hierarchical holographic model and matter element extension model. Informatica, 49(7):49-62, 2025. https://doi.org/10.31449/inf.v49i7.6953