



# When will China's industrial carbon emissions peak? Evidence from machine learning

Qiyang Ran<sup>1,2</sup> · Fanbo Bu<sup>3</sup> · Asif Razzaq<sup>4</sup> · Wenfeng Ge<sup>3</sup> · Jie Peng<sup>5</sup> · Xiaodong Yang<sup>3</sup> · Yang Xu<sup>3</sup>

Received: 11 December 2022 / Accepted: 3 March 2023 / Published online: 27 March 2023  
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

## Abstract

The manufacture of products in the industrial sector is the principal source of carbon emissions. To slow the progression of global warming and advance low-carbon economic development, it is essential to develop methods for accurately predicting carbon emissions from industrial sources and imposing reasonable controls on those emissions. We select a support vector machine to predict industrial carbon emissions from 2021 to 2040 by comparing the predictive power of the BP (backpropagation) neural network and the support vector machine. To reduce noise in the input variables for BP neural network and support vector machine models, we use a random forest technique to filter the factors affecting industrial carbon emissions. The statistical results suggest that BP's neural network is insufficiently adaptable to small sample sizes, has a relatively high error rate, and produces inconsistent predictions of industrial carbon emissions. The support vector machine produces excellent fitting results for tiny sample data, with projected values of industrial carbon dioxide emissions that are astonishingly close to the actual values. In 2030, carbon emissions from the industrial sector will have reached their maximum level.

**Keywords** Industrial carbon emissions · Carbon peaking · Neural network · Support vector machine

## Introduction

The third industrial revolution has severely impinged on the Earth's original ecology by bringing about massive consumption of fossil energy and triggering successive problems such as the greenhouse effect and global warming (Yang et al. 2022; Chai et al. 2022; Cheng et al. 2022). IPCC reports revealing that the global average temperatures will be 1.5 to 4.8 degrees warmer than pre-industrial revolution

levels by 2100 and that the impact of climate change on humankind's socioeconomic situation is expected to worsen if existing climate trends are not controlled (Mondal et al. 2022). The 2021 IPCC Sixth Assessment Report indicates that extreme weather is increasing globally and that global warming is irreversible for hundreds, if not thousands, of years (Pielke et al. 2022). Climate risk is a spatial and temporal externality that warrants joint efforts by the world's economies (Wang et al. 2020; Razzaq et al. 2022). The Kyoto Protocol, the first document in human history to limit greenhouse gas emissions through legislation, officially entered into force in 2005, kicking off a global focus on carbon reduction (Freedman and Jaggi 2011). The greenhouse effect issue is gaining attention from international parties as more than 170 worldwide signed the Paris Agreement in 2016 to jointly address global climate change after 2020 (Yan et al. 2021). As one of the world's highest energy-consuming and carbon-emitting economies, China has consistently adhered to resource conservation and environmental protection to drive its low-carbon economic development (Irfan et al. 2022; Hao et al. 2021a, b). In 2015, the Chinese government signed the document "Intensifying Actions to Address Climate Change" to the United Nations, stating that by 2030, the carbon emission intensity should be reduced by

Responsible Editor: Ilhan Ozturk

✉ Yang Xu  
yangxujedu@126.com

<sup>1</sup> Shanghai Business School, Shanghai 200235, China

<sup>2</sup> Center for Innovation Management Research of Xinjiang, Xinjiang University, 830047 Urumqi, China

<sup>3</sup> School of Economics and Management, Xinjiang University, Urumqi 830047, China

<sup>4</sup> Department of Business Administration, ILMA University, Karachi, Pakistan

<sup>5</sup> Jiangsu Institute of Industrial Development Research, Nanjing University of Finance and Economics, Nanjing 210000, China

60–65%. Moreover, President Xi Jinping gave an ambitious statement that “carbon emissions strive to peak by 2030, and work towards realizing carbon neutrality by 2060” (Tang et al. 2022; Sun et al. 2022). Moreover, to meet carbon reduction goals, a series of measures have been taken in China, including promoting electric energy and clean energy to replace fossil energy, establishing carbon trading pilots, and using market mechanisms and economic instruments. Although China’s carbon reduction governance has yielded specific positive achievements, carbon emissions are still increasing, and carbon reduction efforts are still under tremendous pressure (Razzaq et al. 2021; Li et al. 2021a, b; Shi and Xu 2022).

Relevant statistics demonstrate that China is the most diverse in the world regarding industrial categories (Yang et al. 2021; Zhao et al. 2022; Guo 2022). According to the IPCC report, the energy sector accounted for 35% of GHG (greenhouse gas) emissions in 2010, industry for 21%, transport for 14%, and construction for 6.4%, with industry accounting for 31% of GHG emissions and construction for 19% when indirect emissions are taken into account (Pielke et al. 2022). The report, therefore, states that industry and construction are crucial to achieving the target of a 45% reduction in carbon emissions by 2030 compared to 2010 (Wu et al. 2020; Shi et al. 2022a, b). As a large industrial country, the scale of the industry has expanded rapidly with economic development, with the value added of the sector increasing from 12 billion yuan in 1952 to 31.3 trillion yuan in 2020. However, because of the characteristics of late industrial development and irrational internal structure, enterprises represented by high energy consumption, high emission, and low mainly value-added constitute the subject of the current industrial system (Shi et al. 2022a, b). Since the industry is the largest energy-consuming and carbon-emitting sector among carbon emission sources, industrial carbon peaking and reduction have emerged as the cornerstone (Sun et al. 2021; Su et al. 2021). Therefore, it is essential to improve the measurement and prediction of industrial carbon emissions, which helps to promote ecological civilization roads and is a critical guide to further enhance carbon emission reduction policies and optimize the environmental protection policy system. So, what is the industry’s path to achieving carbon peaking and reduction? What are the factors that affect the industrial sector’s carbon peaking? The answers to the above questions have significant and positive implications for carbon peaking and carbon neutrality goals.

The main contributions of this paper are as follows. On the one hand, we use machine learning to filter the factors affecting industrial carbon emissions by random forest algorithm. The factors affecting industrial carbon emissions are complicated, and the nexus between each influencing factor and industrial carbon emissions is not linear. In previous research, the influencing factors of carbon emission are

determined mainly by simple qualitative division, and the prediction models applied are mostly simple linear models. For example, logistic, ARIMA (autoregressive integrated moving average model), and ridge regression models have been used to predict industrial carbon emissions. However, it is not easy to find the relationship between the influencing factors and industrial carbon emissions, and these models cannot accurately predict the future carbon emissions trend. Random forest algorithm introduces randomness, which is not prone to overfitting, and strong anti-noise capability capacity, and has high accuracy in screening out the influencing factors of industrial carbon emissions. On the other hand, the BP neural network and support vector machine are repeatedly trained and learned to reflect the pattern between the influencing factors and industrial carbon emissions, which can predict the carbon emission trend more accurately, aiming to provide direction for energy saving and emission reduction efforts.

## Literature review

As it stands, there are two main areas of interest in the study of industrial carbon emissions: the factors that influence industrial carbon emissions, forecasting methods development, and the examination of peak scenario outcomes.

The scholarly consensus is that carbon emissions have a substantial relationship with economic growth and energy use. According to the environmental Kuznets curve (EKC) theory, environmental pollution and economic development are shaped like an inverted U. Economic development is usually accompanied by environmental pollution problems in the early stages. Moreover, when economic development reaches a certain level, an environmentally sustainable economic development state will be formed due to the factors of low-carbon technology and environmental policies (Kaika and Zervas 2013). Meng et al. (2022) factor bias in technological progress is determined based on data from 34 industries in China from 2000 to 2015. They found that China first tends to consume energy to pursue capital between capital input and energy input, while it tends to save energy after the Eleventh Five-Year Plan. Moreover, environmental change is closely related to the level of economic development. At present, the environmental quality of the eastern provinces of China is higher than that of the central and western provinces, and the gap between regions is increasing (Guo 2022). At the same time, the improvement of digital technology also provides more powerful technical support for academic research, which can effectively integrate all kinds of information and use different data (Xiao and Liu 2022). As the pillar industry of a country (region), the carbon dioxide emission of enterprises has a crucial influence on regional carbon emission. Scholars’ studies

on carbon emissions' influencing factors show that factors such as industrial scale, industrial structure, technology, and environmental regulation significantly affect carbon emissions. Huang et al. (2019) argued that carbon emissions are significantly positively associated with the industrial scale of the ferrous metal industry.

In contrast, Appiah et al. (2019) argued that energy intensity is the main factor that affects industrial carbon emissions. Zhu et al. (2012), on the other hand, verified the mechanism of its influence on industrial carbon emissions from the perspective of industrial structure. Du et al. (2019) further argued the inhibitory effect of green technology on industrial carbon emissions, and carbon capture, utilization, and storage (CCUS) technology, as an important way of carbon sink, has also received more and more attention from industrial enterprises in recent years (Gür 2022). It has also become a consensus among scholars that the mitigation of industrial carbon emissions can be achieved by both command-based environmental regulations, mainly by administrative orders, and market-based environmental regulations, mainly by building carbon taxes and carbon markets (Wesseh et al. 2017; Zhang et al. 2020; Yang et al. 2020; Y. Wang et al. 2018).

When investigating carbon peak prediction, the academic community has mainly used parties with the LMDI model, Kaya constant equation model, IPAT model, STIRPAT model, and machine learning model. Ehrlich and Holdren (1971) first proposed the LMDI model to differentiate each variable separately while keeping other factor variables constant to determine the degree of influence of each factor change on the target quantity. Chen et al. (2020) and other scholars used this model to predict the peak CO<sub>2</sub> emissions of industrial and agricultural sectors in China. The IPAT model considers the growth in population size, the increase in material living standards, and the increase in resource exploitation as the root causes of the deteriorating resource and environmental problems. Wang et al. (2021a, 2021b) used this model to predict the peak path of carbon emissions in China's Bohai Sea Rim region. STIRPAT is an enhanced version of the IPAT model, which analyzes the connection between the human population, material wealth, technological advancement, and ecological conditions. Dalton et al. (2008) used the STIRPAT model and added scenario analysis to predict carbon emissions in the USA. The essence of the Kaya model is to explore the drivers of the amount of change, i.e., the percentage contribution of each driver of carbon emission change, which has the advantages of simple mathematical form, decomposition without residuals, and reliable explanatory power for the drivers of carbon emission change (Lu and Jiahua 2013). With the development of information technology applications in environmental governance, the role of the Internet in improving energy efficiency and reducing energy-saving potential (ESP) has

attracted more attention (Ren et al. 2022). Wu et al. (2022) measured the carbon emissions peak using the Kaya constant equation approach. A machine learning model is a file that can recognize specific patterns after training. The model can be trained with a set of data, providing it with an algorithm that the model uses to learn these data and make inferences. After introducing the model, it can be used to make inferences based on previously unseen data and to make predictions about those data (Wang et al. 2009). Some scholars believe that machine learning models have better prediction accuracy than multiple linear regression models and have applications in analyzing industrial carbon emissions and their prediction of it in different regions and industries (Leerbeck et al. 2020; Ağbulut and Consumption 2022). Chen et al. (2022) present two approaches to forecast parameters in the SABR model. The first approach is the vector autoregressive moving-average model (VARMA) for the time series of the in-sample calibrated parameters, and the second is based on machine learning techniques called epsilon-support vector regression. In addition, some other scholars have used CGE, logistic, and system dynamics theory models to predict the peak carbon emissions of different regions and industries, and their different scenarios have been analyzed (Meng and Niu 2011; Mirzaei and Bekri 2017; Beuséjour et al. 1995). Cheng et al. (2022) evaluate that the carbon trading pilot policy can promote the upgrading of industrial structure in the region and ultimately achieve the reduction of industrial carbon emissions.

The accuracy of scenario analysis methods relies heavily on human subjective judgment. A large number of studies have been conducted in the literature on industrial carbon emissions and carbon peaking, and relatively rich research results have been obtained, but there are still many shortcomings. On the one hand, the existing carbon emission measurement methods still need to be unified, and the mainstream prediction methods have high data requirements and are difficult to obtain. On the other hand, the existing carbon emission prediction models are either traditional regression-based models with low prediction accuracy or models with high requirements for a priori data. The influencing factors of industrial carbon emissions still need to be clarified, and the sample size is small. Therefore, this paper applies the energy consumption method to measure industrial carbon emissions from direct and indirect perspectives, screens out the factors that have a large impact on carbon emissions through the random forest method, and then predicts carbon emissions through a machine learning model. Due to the integrated algorithm, the random forest algorithm itself has better accuracy than most individual algorithms, so it is highly accurate, performs well on the test set, and does not easily fall into overfitting. For industrial research, the random forest has a certain anti-noise capability, which has some advantages over other algorithms. In general, there

is relatively little literature on the use of machine learning to study carbon peaking, and there is even less literature related to the study of industrial carbon peaking. Therefore, this paper can provide some reference for the achievement of China’s industrial carbon peaking goal by using machine learning methods and filling some research gaps.

## Study design

### Research strategies

Machine learning is an algorithm based on historical data and is trained iteratively to find patterns in the data. Machine learning is widely used for prediction and classification problems, where input variables are linked to output variables by specific functions. Machine learning aims to learn tasks by training the data and making them well-suited to predicting the data (Mason et al. 2018). In this study, the soft Matlab 2020 is used to implement BP neural network and support vector machine.

### BP neural network prediction

**Neural networks** A neural network is a parallel distributed processor built on individual neurons, each of which is a relatively independent learning unit. A neural network is similar to the human brain in that it can be trained repeatedly to gain a particular experience and to preserve the learned knowledge. Neural networks process data based on neurons and the principle of neural networks can be described as follows.

$$u_k = \sum_{j=1}^m w_{kj}x_j \tag{1}$$

$$y_k = \varphi(u_k + b_k) \tag{2}$$

where  $x_j$  denotes the input information,  $u_k$  denotes the neuron output information,  $w_{kj}$  denotes the weights of individual neurons, and  $b_k$  is the offset term.  $b_k$  can be adjusted to the neuron output data by taking the offset term into account.

Then, the resultant data can be expressed as follows:

$$v_k = u_k + b_k \tag{3}$$

The above three equations can be combined to obtain

$$v_k = \sum_{j=0}^m w_{kj}x_j \tag{4}$$

$$y_k = \varphi(v_k) \tag{5}$$

**BP algorithm** The error backpropagation algorithm is the basis of the BP neural network, which is a multi-layer feed-forward neural network based on this algorithm (Muruganandam et al. 2023). BP neural network is trained on the input data set, and the output results will be compared with the actual results, the difference between the output error (Kosarac et al. 2022). The backpropagation of the output error is performed, and the weights and thresholds of each neuron are corrected based on the output error to obtain the optimal result (Bai et al. 2023). The gradient descent method is the core of the error backpropagation algorithm. Following Kusumadewi et al. (2023) using the gradient descent method, the BP neural network can predict the mean square error to achieve the minimum. Three layers make up the BP neural network model: input, output, and implicit. An excitation function is used to reflect the mapping relationship between the input layer and the output. The model is trained with an empirical risk minimization strategy, and an error backpropagation algorithm continuously corrects the neuron weights and thresholds until the error value satisfies the desired error (Sonia et al. 2023). The training process of the BP neural network is as follows. BP neural network is an efficient and widely used intelligent algorithm in the field of classification and prediction, which is suitable for solving uncertainty problems with multiple influencing factors and constraints (Liang et al. 2023). The model features distributed storage, self-learning and adaptation, large-scale parallel processing, strong learning ability, and generalization ability. At present, the neural network has become a popular prediction method in many disciplines.

Let the input data set be  $N = (X_1, X_2 \dots X_m)$ . Mapping this sequence from the input layer to the implicit layer, the sequence becomes  $N = (X_1, X_2 \dots X_m)$ . After the computation, this sequence becomes  $\theta(w)$ . Then,  $\theta(w)$  is mapped from the implicit layer to the output layer. Finally, the computation is completed in the output layer and the output value is returned.

The output value is used to calculate the error, and the neuron weights and thresholds are continuously corrected by back propagation of the output error. The expressions of the weights are as follows.

Implicit layer:

$$\Delta w_{ij} = -\gamma \frac{\partial E}{\partial w_{ij}} = -\gamma \frac{\partial E}{\partial net_i} \frac{\partial net_i}{\partial w_{ij}} = -\gamma \frac{\partial E}{\partial y_i} \frac{\partial y_i}{\partial net_i} \frac{\partial net_i}{\partial w_{ij}} \tag{6}$$

Output layer:

$$\Delta w_{ki} = -\gamma \frac{\partial E}{\partial w_{ki}} = -\gamma \frac{\partial E}{\partial net_k} \frac{\partial net_k}{\partial w_{ki}} = -\gamma \frac{\partial E}{\partial y_k} \frac{\partial y_k}{\partial net_k} \frac{\partial net_k}{\partial w_{ki}} \tag{7}$$

Organized by:

$$\Delta w_{ki} = -\gamma \sum_{p=1}^p \sum_{n=1}^N (T_n^p - O_n^p) \cdot \varphi'(\text{net}_n) \cdot y_i \tag{8}$$

$$\Delta w_{ij} = -\gamma \sum_{p=1}^P \sum_{n=1}^N (T_n^p - O_n^p) \cdot \psi'(\text{net}_n) \cdot w_{ji} \cdot \varphi'(\text{net}_i) \cdot x_j \tag{9}$$

$k$  is the output layer node.  $j$  is the input layer node where  $i$  is the implicit layer node.  $c$  is the weight between the implicit layer node  $i$  and the input layer  $j$  ( $i = 1, 2, 3, \dots, p, j = 1, 2, 3, \dots, n$ ).  $w_{ki}$  is the weight between node implicit layer  $i$  to node output layer  $k$  ( $i = 1, 2, 3, \dots, p, k = 1, 2, 3, \dots, n$ ).  $x$  is the input data value of node  $i$  in the input layer.  $\varphi$  is the excitation function of the implicit layer.  $\psi$  is the excitation function of the output layer, and  $O^k$  is the output value of node  $k$  in the output layer.

Finally, the mean square error of the model output is calculated as follows.

$$E_k = \frac{1}{2} \sum_{j=1}^n (Y_j^n - y_j^n)^2 \tag{10}$$

The weights and thresholds are corrected according to the output error until the error meets the requirements.

### Support vector machine prediction

**Kernel function** The kernel function has to appear before the support vector machine, and then, scholars widely use its theory to support vector machines. The principle of the support vector machine is to map the data from low-dimensional to high-dimensional. If the number of dimensions of the data is large, the conversion from low to high dimensions may result in a “dimensional disaster.” This problem can be mitigated when using kernel functions. If there exists a function  $K(x, x_i)$  such that all the  $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}, x, y \in R$  s meet:

$$K(x, x_i) = \phi(x) \cdot \phi(x_i) \tag{11}$$

If the conditions of Mercer’s theorem are met, the function is a kernel function. There are four common types of kernel functions as follows.

Linear kernel function (*linear*)

$$K(x, x_i) = x \cdot x_i \tag{12}$$

Linear kernel functions have fewer parameters and are often used to solve linearly differentiable problems.

Polynomial kernel function (*poly*)

$$K(x, x_i) = [(x \cdot x_i) + 1]^q \quad (q = 1, 2, 3 \dots) \tag{13}$$

Polynomial kernel functions are widely used in nonlinear problems, which not only contain more parameters, but also the parameters have more influence on the model. And this type of kernel function is less suitable for large data sets.

Radial basis kernel functions

$$K(x, x_i) = \exp \left[ \frac{\|x - x_i\|^2}{2\sigma^2} \right] \tag{14}$$

Gaussian kernel functions, involving only one parameter, are often used to study nonlinear problems, can map data to infinite dimensions, and are computationally slow. Linear kernel functions are a case of radial basis kernel functions.

Sigmoid kernel function

$$K(x, x_i) = \tanh(\gamma(x \cdot x_i) + c) \tag{15}$$

Since the values of the parameters  $\gamma$  and  $c$  do not always meet Mercer’s theorem, such kernel functions are less frequently used.

**Support vector machine** There exists a hyperplane in the high-dimensional space such that the plane becomes the decision boundary when all sample points are closest to this plane. Suppose there are  $n$  samples and the original input data is  $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}, x, y \in R$ . The expression of linear regression in high-dimensional space is

$$f(x) = w \cdot \phi(x) + b \tag{16}$$

$W$  denotes the weight and  $b$  denotes the bias term.

Given that the error is  $e$ , the model prediction is  $f(\vec{x})$  and the actual value is  $y_i$ . When  $(\vec{x}) - y_i \leq e$ , no loss is calculated. When  $f(\vec{x}) > e$  and  $y_i > e$ .

$$\min_{w,b} \left( \frac{1}{2} \|\vec{w}\|^2 + C \sum_{i=1}^N L_\epsilon(f(\vec{x}) - y_i) \right) \tag{17}$$

where  $C > 0$  is the penalty parameter.  $L_\epsilon$  denotes the loss function:

$$L_\epsilon(Z) = \begin{cases} 0, & |Z| \leq \epsilon \\ |Z| - \delta, & \text{other} \end{cases} \tag{18}$$

For the optimization problem, introducing slack variables  $\zeta_i$  and  $\hat{\zeta}_i$ , the

$$\begin{cases} \min \frac{1}{2} \|w\|^2 + \sum_{i=1}^m \epsilon (f(x_i) - y_i) \\ y_i - (w \cdot \varphi(x_i) + b) \leq \epsilon + \xi \\ (w \cdot \varphi(x_i) + b) - y_i \leq \epsilon + \xi \\ \xi, \xi^n \geq 0 \end{cases} \tag{19}$$

The Lagrange function is introduced to solve for the above function.

$$L(w, \xi, \xi^n) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^M (\xi_i + \xi^n) - \sum_{i=1}^M \alpha_i (\xi_i + \epsilon + y_i - \langle w, x_i \rangle - b) \tag{20}$$

$$-\sum_{i=1}^M \alpha(\xi_i^n + \varepsilon + y_i - \langle w, x_i \rangle - b) - \sum_{i=1}^M (\eta_i \xi_i + \eta_i^n \xi_i^n) \tag{21}$$

where  $\langle w, x_i \rangle$  denotes the inner product.  $\alpha_i, \xi_i^n$ , and  $\eta_i$  are the Lagrangian multiplier. By the KKT condition, this study knows that:

$$\begin{cases} \alpha_i(f(\bar{x}_i) - y_i - \varepsilon - \xi_i) = 0 \\ \hat{\alpha}_i(f(\bar{x}_i) - y_i - \varepsilon - \xi_i) = 0 \\ \alpha_i \hat{\alpha}_i = 0, \xi_i, \hat{\xi}_i = 0 \\ (C - \alpha_i)\xi_i = 0, (C - \alpha_i)\hat{\xi}_i = 0 \end{cases} \tag{22}$$

Substituting the above equation into the kernel function:

$$f(\bar{x}) = \sum_{i=1}^n (\hat{\alpha}_i - \alpha_i)k(\bar{x}_i, \bar{x}) + b \tag{23}$$

where  $k(\bar{x}_1, \bar{x}) = \phi(\bar{x}_1)^T \phi(\bar{x})$ , the kernel function.

### Variable selection

#### Industrial carbon emission measurement

Direct emissions, or those that come directly from factories, are one kind of indirect emission. Carbon emissions from fossil energy sources such as crude oil, coke, crude oil, gasoline, kerosene, and diesel are the primary contributors to direct emissions. Carbon emissions from the use of fossil fuels to provide power and heat for factories are a major source of indirect emissions (Hao et al. 2021a, b). This is the model used to quantify carbon emissions from industry.

$$E = E_{dir} + E_{ind} = \frac{44}{12} \sum C_i \times \alpha_i + \frac{44}{12} \sum M_j \times \beta_j \tag{24}$$

In the above equation,  $E$  denotes total CO<sub>2</sub> emissions;  $E_{dir}$  denotes direct emissions;  $E_{ind}$  is indirect emissions;  $C$  and  $M$  denote direct and indirect energy consumptions, respectively; and  $\alpha$  and  $\beta$  are the carbon emission coefficients of that energy source. Considering the industrial consumption of energy and the availability of data, the industrial consumption of raw coal, coke, crude oil, gasoline, kerosene, diesel, fuel oil, liquefied petroleum gas, refinery dry gas, and natural gas is used to represent the industrial direct energy consumption, and the industrial heat and electricity consumption are used to measure the industrial indirect energy consumption. The specific carbon emission factors for each energy source are listed in the following (Table 1).

Table 1 indicates that the unit of average low level heat content is kJ/kg and the unit of average low-level heat content of natural gas is kJ/m<sup>3</sup>; the unit of discount standard coal factor is kgce/kg, the unit of natural gas discount standard coal factor is kgce/m<sup>3</sup>, the unit of heat discount standard coal factor is kgce/MJ, and the unit of electricity discount standard coal factor is kgce/kW-h. Unit calorific value of carbon content unit is tons of carbon/TJ. The unit of carbon emission factor is kg-CO<sub>2</sub>/kg, the unit of natural gas carbon emission factor is kg-CO<sub>2</sub>/m<sup>3</sup>, and the unit of thermal carbon emission factor is tCO<sub>2</sub>/GJ. As the form of power generation varies from region to region, the carbon emission factors vary, and the carbon emission factors for electricity in different regions are shown in the following.

Table 2 reveals that the carbon emission factors of electricity in each region are in units of kg-CO<sub>2</sub>/kW-h. The data in Table 1 and Table 2 are from the National Development and Reform Commission, General Rules for Calculating

**Table 1** Carbon emission factors for various types of energy

Energy	Mean low level heat output	Refractive index of standard coal	Carbon content per unit calorific value	Carbon oxidation rate	Carbon emissions coefficient
Raw coal	20,934	0.7143	26.37	0.94	1.90266
Coke	28,470	0.9714	29.5	0.93	2.86394
Crude oil	41,868	1.4286	20.1	0.98	3.02396
Gasoline	43,124	1.4714	18.9	0.98	2.92872
Kerosene	43,124	1.4714	19.5	0.98	3.02170
Diesel oil	42,705	1.4571	20.2	0.98	3.09976
Fuel oil	41,868	1.4286	21.1	0.98	3.17440
Liquefied petroleum gas	50,242	1.7143	17.2	0.98	3.10522
Refinery dry gas	46,055	1.5714	18.2	0.98	3.01194
Natural gas	38,979	1.3300	15.3	0.99	2.16485
Heat	–	0.03412	–	–	0.11
Electricity	–	0.1229	–	–	–

**Table 2** Carbon emission factors of different regions

Grid name	Covered area	Emission coefficients
North China Region	Beijing, Tianjin, Hebei, Shanxi, Shandong, Inner Mongolia	1.246
Northeast Region	Liaoning, Jilin, Heilongjiang	1.096
East China Region	Shanghai, Jiangsu, Zhejiang, Anhui, Fujian	0.928
Central China Region	Henan, Hubei, Hunan, Jiangxi, Sichuan, Chongqing	0.801
Northwest China	Shaanxi, Gansu, Qinghai, Ningxia, Xinjiang	0.977
Southern Region	Guangdong, Guangxi, Yunnan, Guizhou	0.714
Hainan	Hainan	0.917

**Table 3** Industrial carbon emissions by year

Year	Carbon emissions (million tons)	Year	Carbon emissions (million tons)
1991	158,270.5686	2006	459,799.396
1992	170,808.6905	2007	508,878.6702
1993	183,781.7623	2008	530,597.8257
1994	198,356.2619	2009	553,911.5128
1995	207,096.0643	2010	589,041.065
1996	211,176.4072	2011	660,455.7519
1997	205,830.3464	2012	677,655.9406
1998	211,557.4999	2013	694,568.9028
1999	215,146.3416	2014	702,507.9429
2000	217,781.0658	2015	701,737.8454
2001	235,046.1212	2016	684,302.1214
2002	254,098.8596	2017	715,309.2671
2003	302,704.2602	2018	750,801.9079
2004	356,092.4823	2019	765,870.7612
2005	416,798.8375	2020	783,876.9667

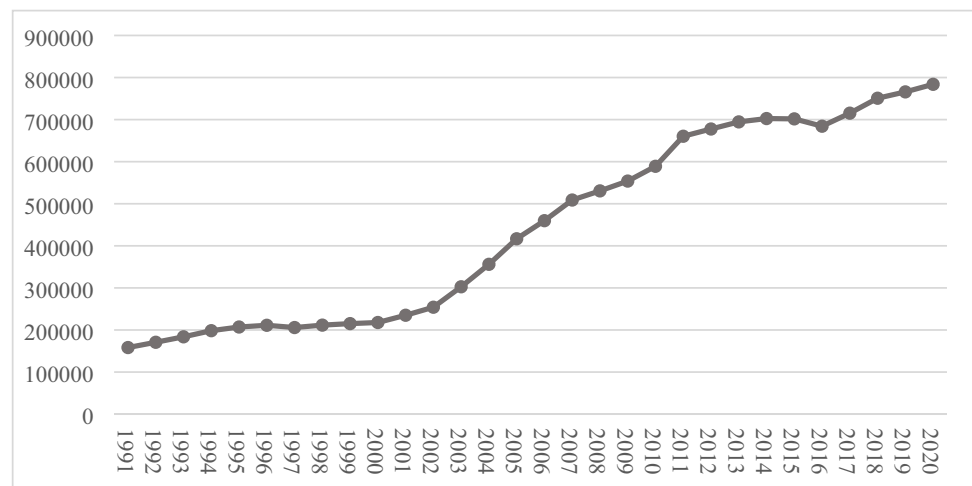
Comprehensive Energy Consumption (GB/T 2589-2020), and Guidelines for Preparing Provincial Greenhouse Gas

Inventories (NDRC Climate [2011] No. 1041), and the carbon emission coefficients in Table 1 are calculated as follows.

$$e_i = 10^{-9} \cdot g_i \times u_i \times o_i \times 10^3 \times \frac{44}{12} \quad (25)$$

In the above equation,  $e$  denotes carbon emission factor,  $g$  denotes average low-level heat generation,  $u$  denote carbon content per unit calorific value, and  $o$  denote carbon oxidation rate. According to the carbon emission calculation model in the above section and the carbon emission coefficients in Table 3, the carbon emission of China's industry for the period from 1991 to 2020 is measured, and the results are as follows:

Figure 1 illustrates that China's industrial carbon emission trends are divided into four main stages. From 1991 to 2000, industrial carbon emissions grew slowly and leveled off from 2001 to 2011; the growth of industrial carbon emissions accelerated, increasing from 254.99 million tons in 2001 to 660.456 million tons in 2011, an increase of about 260%. After China's accession to WTO in 2001, light industry and processing manufacturing industries developed rapidly, and China's share of the global manufacturing industry increased rapidly, but it also increased the consumption of coal, oil, and

**Fig. 1** Industrial carbon emission trends

other energy sources, which led to a rapid increase in total carbon emissions. From 2012 to 2016, China’s industrial carbon emissions rose slowly and trended downward. This is due to the fact that with economic development and frequent outbreaks of global climate extremes, addressing climate change has become a global consensus, China has increased its management of greenhouse gas emissions in the country, and total carbon emissions have shown a slowing trend. From 2017 to 2020, industrial carbon emissions rose rapidly. In recent years, affected by the Covid-19 epidemic, the world’s economies have come to a halt, and China, as a major manufacturing country, has seen its manufacturing capacity rise rather than fall, providing the world with a large supply of goods and raw materials and thus has seen a rapid increase in the growth rate of carbon emissions.

### Influencing factors of industrial carbon emissions

Different scholars use different terms to describe the variables that influence industrial carbon emissions. Therefore, filtering out the influencing factors of industrial carbon emissions is the key to accurate prediction of industrial carbon emissions. Learning from the studies of existing scholars, the factors influencing industrial carbon emissions are identified, the characteristics are screened according to the random forest model, the importance of the attributes of each variable is ordered, and ultimately, the industrial carbon emission influencing variables applicable to this study are selected. For this reason, this study’s focus on the scientific identification of the influencing elements of industrial carbon emissions and the appropriate screening is crucial, which determines the scientificity of the results. In this study, the initial selection of influencing variables is based on the IPAT constant model (Kim et al. 2020; Hussain et al. 2022; Zhang et al. 2022), and 12 indicators that have the potential to influence industrial carbon emissions are identified from three aspects: demographic, economic, and technological. The specific indicators are as follows.

Random forest (RFF) is a classifier containing multiple decision trees proposed by Breiman (2001). When using random forest for feature filtering, the training set of any decision tree is about two thirds of the entire training set, and the remaining one third of the data set is out-of-bag data. The out-of-bag data is computationally critical when filtering for feature importance (Altmann et al. 2010). All the nodes of the decision tree are randomly drawn  $\sqrt{n}$  features from the  $n$  vit levy set, and then, one of these features is selected based on the Gini gain maximization principle, and then, the parent  $n_p$  data is divided into the right child node  $n_r$ , and the left child  $n_l$ . The Gini gain maximization principle is as follows.

$$\Delta I_G = I_G(n_p) - p_l I_G(n_l) - p_r I_G(n_r) \tag{26}$$

The Gini gain is maximized when the above equation is maximized.  $I_G(n) = 1 - \sum_{c=1}^2 p_c^2$  is the Gini index of node  $n$ ,  $p_c$  is the proportion of  $c$  class samples on node  $n$ ,  $p_l$  is the proportion of data  $n_p$  distributed to  $n_l$ , and  $p_r$  is the proportion of data distributed to  $n_p$  by parent node  $n_r$ .

There are usually two measures for random forest-based feature filtering. One is the Gini index hair, and the other is measured by the correct classification rate of out-of-bag data. In this study, the Gini index method was used for feature importance screening, and the amount of Gini index variation was as follows.

$$IMP_{in}^{Gini} = I_G(n) - I_G(n_l) - I_G(n_r) \tag{27}$$

$IMP_{in}^{Gini}$  is the importance of feature  $x_i$  at a node, i.e., the amount of change in the Gini index before and after partitioning the data to the left and right children at this node. When feature  $x_i$  is partitioned in the  $m_{th}$  decision tree, the set of partitioned nodes is  $N$ . Then, the importance of  $x_i$  on decision tree  $m$  is.

$$IMP_{i\_m}^{Gini} = \sum_{n \in N} IMP_{in}^{Gini} \tag{28}$$

If there are a total of  $M$  decision trees in the entire random forest, the importance of feature  $x_i$  is

$$IMP_i^{Gini} = \frac{1}{M} \sum_{m=1}^M IMP_{i\_m}^{Gini} \tag{29}$$

Based on the Gini index method for feature importance screening, this study establishes a random forest algorithm for the 12 variables affecting industrial carbon emissions in Table 4 by R software. The importance of each variable is shown in Fig. 2 below.

Figure 2 explains the importance ranking of each influencing factor as follows: energy emission intensity > industrial employees > urbanization rate > total population > the number of industrial technical papers > gross domestic product > industrial GDP > economic development level > industrial secondary energy consumption > industrial primary energy consumption > industrial structure > industrial pull rate. To avoid redundancy of input variables, the top five influencing factors of importance are selected as input variables in this study to construct the industrial carbon emission prediction model.

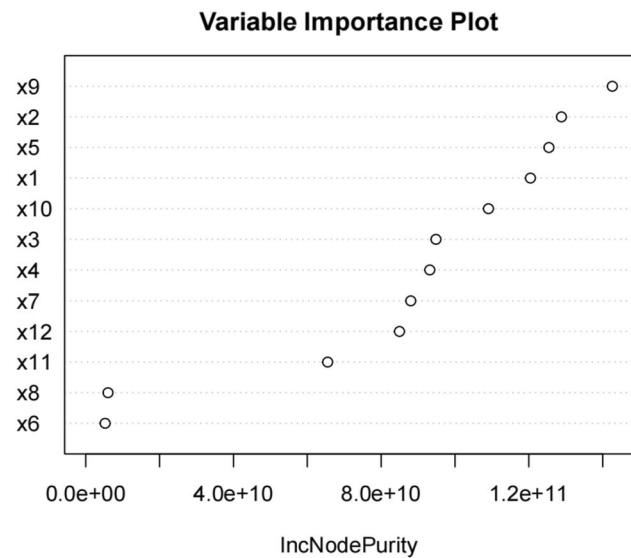
### Data sources

The National Bureau of Statistics, the China Statistical Yearbook, and the China Energy Statistical Yearbook provide information on industrial energy usage from 1991 to 2020.



**Table 4** Industrial carbon emissions influence factor selection

Type	Influence factors	Specific indicators	Units	Indicators code
Demographic	Total population	–	10,000 people	×1
	Number of employed employees in industrial enterprises	–	10,000 people	×2
Economic	GDP	–	100 million Yuan	×3
	Industrial GDP	–	100 million Yuan	×4
	Urbanization rate	–	%	×5
	Industrial pull rate	–	%	×6
	Economic development level	GDP/total population	10,000 Yuan/person	×7
	Industrial structure	Gross industrial product/GDP	%	×8
Technological	Energy emission intensity	Carbon emissions/GDP	Tons/10,000 Yuan	×9
	Number of industrial technology papers	–	Articles	×10
	Industrial primary energy consumption	–	Million tons of standard coal	×11
	Industrial secondary energy consumption	–	Million tons of standard coal	×12



**Fig. 2** The importance of each influencing factor

## Results and discussion

### Industrial carbon emission prediction result based on BP neural network

The industrial carbon emissions prediction model is constructed based on the influencing factors screened by random forest, i.e., five variables: energy emission intensity, industrial employees, urbanization rate, total population, and the number of industrial technical papers as input variables of the BP neural network model, using the data from 1991 to 2014 as the training set and the data from 2015 to 2020 as the test set to predict industrial carbon emissions.

### Selection of critical functions and parameters

The key to the BP neural network prediction model is selecting the activation function, training function, and other vital parameters. In this paper, the hyperbolic tangent S-type function is chosen as the transfer function for mapping the input layer to the hidden layer; the linear function is chosen as the activation function for mapping the hidden layer to the output layer and is chosen as the training function of the model.

The BP neural network’s learning rate affects the model’s training effect. When the learning rate is low, the model is trained slowly but with reasonable accuracy; however, when the learning rate is significant, the model is trained quickly but with poor convergence. The parameters used in this analysis were a learning rate of 0.1, a maximum number of training sessions of 1000, and an anticipated variation of 0.001 for the training objective.

Absolute error is followed as follows:

$$APE = |Y_i - \hat{Y}_i| \tag{30}$$

The relative error is followed as follows:

$$RPE = \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \times 100\% \tag{31}$$

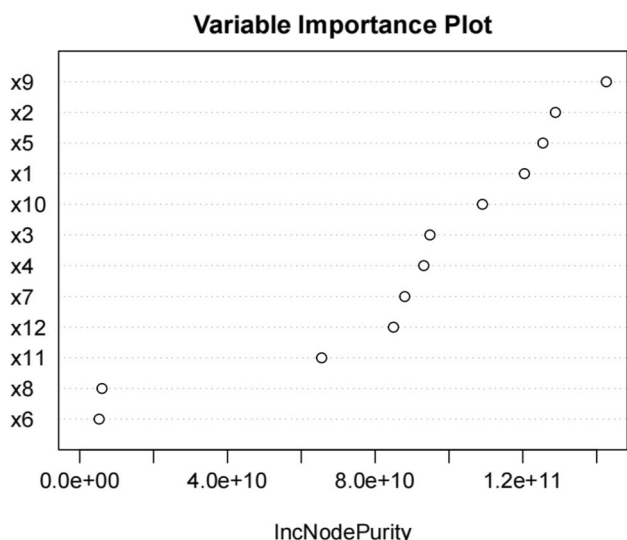
The mean square error is following as:

$$MSE = \frac{1}{N} \sum_{i=1}^N \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right|^2 \tag{32}$$

where  $Y_i$  is the actual value and  $\hat{Y}_i$  is the predicted value. The predicted results are fitted to the curves as follows:

**Table 5** BP neural network prediction errors

Year	Actual value	Predicted value	Absolute error	Relative error
2015	701,737.85	674,092.91	3.94%	27,644.93
2016	684,302.12	662,448.43	3.19%	21,853.70
2017	715,309.27	631,500.43	11.72%	83,808.84
2018	750,801.91	619,525.96	17.48%	131,275.95
2019	765,870.76	625,350.34	18.35%	140,520.42
2020	783,876.97	647,021.20	17.46%	136,855.77
mse	0.020111			
R <sup>2</sup>	0.98264			



**Fig. 3** Prediction fitting curve

The prediction results and error analysis are shown in Table 5. Figure 3 and Table 5 reveal that for the industrial carbon emissions data from 2015 to 2020, the prediction error is relatively small in the first 2 years, with an average relative error within 5%, when the model’s prediction results are relatively close to the relationship between various influencing factors and industrial carbon emissions. While the prediction results in the last 4 years are poor, with the error gradually becoming more extensive, with an average relative error of over 15%, and in 2019. The relative error of the predicted value reached 18.35%, which deviated significantly from the actual value. This indicates that although the BP neural network can reflect the pattern between individual influencing factors and industrial carbon emissions, the prediction results could be more satisfactory in actual forecasting. Overall, the average relative error of the BP neural network model for industrial carbon dioxide emissions is 12.02%, which indicates that the prediction values returned by the BP neural network deviate from the fundamental values, and the prediction accuracy is low. In addition, some of

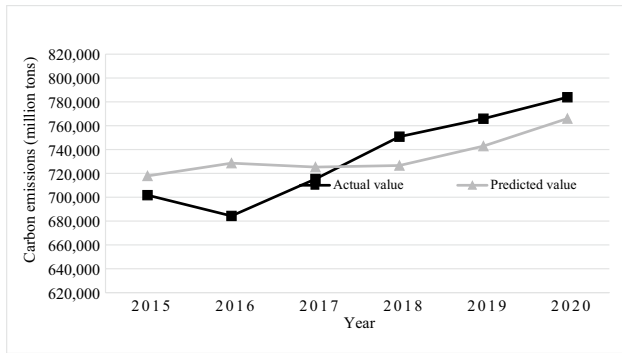
the initial parameters of the BP neural network are random, which causes the prediction results not to remain consistent each time.

### Industrial carbon prediction model based on network optimization method support vector

Firstly, the input variables should be selected. The five influencing factors that have a more significant impact on industrial carbon emissions as screened by random forest are used as the input variables of the model, i.e., five variables of energy emission intensity, industrial employees, urbanization rate, total population, and the number of industrial technical papers as the input variables of the support vector machine model. Secondly, different kernel functions significantly impact the learning effect of the support vector machine, so it is essential to choose the appropriate kernel function according to the actual problem. This study uses a Gaussian kernel function with few parameters and strong adaptability. The two main parameters for using a Gaussian kernel function are the penalty factor and the function setting *g* in the kernel function. This study uses the grid optimization method to select optimal and appropriate parameters automatically. The grid optimization method is an exhaustive search method of parameter value selection and is optimized by the cross-validation method. By arranging the possible parameter values and generating a “grid” of all combinations, cross-validation is used to evaluate each variety’s performance and automatically select the optimal combination of parameters. The regression equation is then used to predict the carbon dioxide emissions of the test set based on the regression equation.

When building an SVM, one typically uses data from 1991 to 2014 for training and then uses data from 2015 to 2020 for testing, with the input variables being energy emission intensity, industrial employees, urbanization rate, total population and a number of industrial technical papers, and the output variable being industrial carbon emissions. The fitted curves of the prediction results are as follows (Fig. 4).

Table 6 reveals that the absolute error of the prediction of industrial carbon emissions by the support vector machine model is below 7%, and the relative error of the prediction value for 2017 is only 1.40%, which is the smallest prediction error value. The support vector machine is able to faithfully portray the association between several variables and commercial carbon output and shows the reasonableness of using the five variables of energy emissions intensity, industrial employees, urbanization rate, total population, and the number of industrial technical papers as input variables for the model. The overall average error of the prediction results of the test set is 3.11%, indicating that the model prediction accuracy is very high and its prediction value is very close to the actual value.



**Fig. 4** Support vector machine industrial carbon emissions fitting curve

**Table 6** Support vector machine prediction errors

Year	Actual value	Predicted value	Absolute error	Relative error
2015	701,737.85	717,862.29	16,124.45	2.30%
2016	684,302.12	728,576.50	44,274.38	6.47%
2017	715,309.27	725,297.33	9988.06	1.40%
2018	750,801.91	726,607.01	24,194.90	3.22%
2019	765,870.76	742,882.97	22,987.79	3.00%
2020	783,876.97	765,989.92	17,887.05	2.28%
mse	0.0035042			
$R^2$	0.99486			

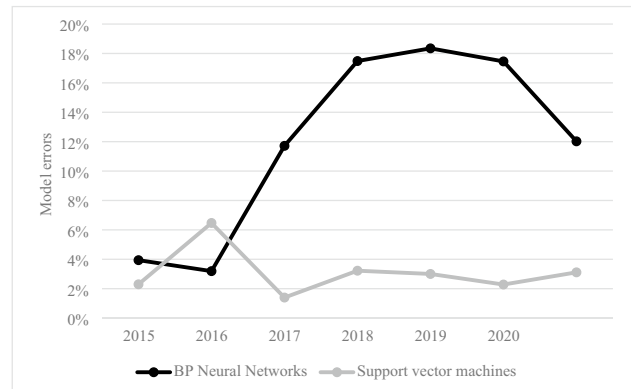
**Model prediction results in comparison**

In this study, the five influencing factors screened by the random forest model are used as input variables, and the measured industrial carbon emission data are used as output variables to construct a BP neural network prediction model and a support vector machine model to predict industrial carbon emissions from 2014 to 2020, respectively. To more comprehensively evaluate the prediction results of the two models, this study compares four aspects: the mean absolute error, mean relative error, mean square error, and  $R^2$  of the model predictions. The mean fundamental error can represent the actual prediction error of the model, and the mean relative error can characterize the accuracy and credibility of the model. The mean square error can be evaluated as the deviation between the measured and actual values, and  $R^2$  can measure the fitting effect of the model. The specific comparison results are as follows.

According to the findings shown in Table 7 and Fig. 5, the absolute and relative errors generated by the support vector machine model for the 2015–2020 industrial carbon emissions forecast results are much lower than those generated by the BP neural network. The mean square error and average relative error are likewise substantially fewer than those of the BP neural network, showing that the predicted

**Table 7** Comparison of model errors

Model type	BP neural networks	Support vector machines
Mean absolute error	90,326.60	22,576.11
Average relative error	12.02%	3.11%
mse	0.020111	0.0035042
$R^2$	0.98264	0.99486



**Fig. 5** Comparison of relative model errors

values of the support vector machine vary from the actual values far less often than those predicted by the BP neural network. The value of  $R^2$  for the support vector machine is closer to 1, indicating that the support vector machine model is a better fit.

When the accuracy of the two models’ predictions is compared, the support vector machine emerges as the clear winner over the BP neural network. The support vector machine accurately reflects the relationship between industrial carbon emissions and their influencing factors. For small-sample non-linear data sets, the support vector machine can return more accurate prediction results and also proves the shortcomings of the BP neural network in predicting small-sample problems. The support vector machine is more suitable for predicting industrial carbon emissions. Therefore, the support vector machine is chosen as the model for further industrial carbon emissions prediction.

**Industrial carbon emission prediction based on support vector machine decision model**

This study uses random forests to screen the impact of industrial carbon emissions. It uses five factors with more significant impact as input variables to construct BP neural network and support vector machine models to forecast industrial carbon dioxide emissions for the past six years. Firstly, the GM (1,1) approach is used to represent the elements that will

have an impact on industrial carbon dioxide emissions for the period 2021–2040. Because of this, this study predicts industrial carbon emissions from 2021 to 2040 utilizing a support vector machine model. The prediction results of energy emission intensity, industrial employees, urbanization rate, total population, and the number of industrial technical papers passed the test and satisfied the prediction requirements. The support vector machine prediction model was used to estimate industrial carbon dioxide emissions from 2021 to 2040, and the anticipated values of the five influential components discussed above were utilized as the input variables of the model, and the predicted carbon dioxide emissions data from 2021 to 2040 were obtained. Based on the predicted values output from the model, the effect of industrial carbon emission reduction in China was analyzed in terms of both industrial carbon dioxide emissions and emission intensity, and the specific trends are as follows.

Figure 6 reveals that actual industrialized carbon emissions are shown for 1991–2020, and predicted industrialized carbon emissions are offered for 2021–2040. Industrial carbon emissions from 2021 to 2040 show a trend of growth followed by a decline, peaking in 2030 and starting to decrease yearly after 2031. The prediction results report that China’s industrial carbon emissions will peak in 2030, which coincides with the milestone of achieving an overall “carbon peak” by 2030. To achieve the “carbon peak” target and actively respond to global warming, the industry, which is the critical area of carbon emissions in China, should actively adjust its development strategy, accelerate its energy structure adjustment, and contribute to the national emission reduction efforts.

### Conclusion and policy implications

The industry is a significant contributor to global warming and regional environmental degradation, and the measurement and prediction of industrial carbon emissions

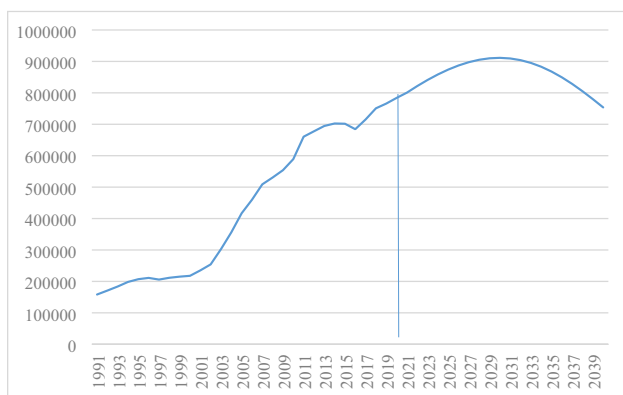


Fig. 6 Industrialized carbon emission prediction

are substantial for policy formulation and adjustment, as well as energy conservation and emission reduction efforts. This study defines the direct and indirect sources of industrial carbon emissions and measures industrial carbon emissions. The factors that affect industrial carbon emissions are selected by random forest feature filtering to obtain five factors that have an enormous impact on industrial carbon emissions. Using the above-influencing factors as input variables, a machine learning-based BP neural network and support vector machine prediction model is constructed to predict the industrial carbon emissions from 2015 to 2020. The GM (1,1) gray prediction method is applied to predict each input variable, and the industrial carbon emissions from 2021 to 2040 are predicted based on the support vector machine model. The main findings are as follows: random forest feature filtering, energy emission intensity, industrial employees, urbanization rate, total population, and industrial technology innovation are the five factors that strongly influence industrial carbon emissions. The average relative error of the BP neural network for industrial carbon emission prediction reaches 12.02%, much larger than the prediction error of the support vector machine. The predicted values of the support vector machine model are closer to the actual values, and the average relative error of the prediction is only 3.11%. Industrial carbon emissions will peak in 2030 and show a declining trend in the early years. As an essential component of carbon emissions, the industry is in line to achieve an overall “carbon peak” by 2030. Based on the influencing factors of industrial carbon emissions and the predicted results of the emission data, this study proposes the following implications:

1. Policymakers should seriously consider climate change and prioritize carbon emission reduction while developing the economy. Moreover, policymakers should increase their policy efforts and use strict environmental regulations to urge industrial enterprises to pay attention to energy conservation and emission reduction, take a new type of industrialization, promote the transformation of economic structure to low pollution, and realize the shift of the economy to green and low-carbon. Meanwhile, while formulating and enforcing policy, decision-makers must take into account not just economic but also regional demographic aspects. The government and industrial enterprises should also strengthen ties and cooperation to promote the people-oriented and low-carbon development concept and implement corresponding incentives and policy support for relevant energy conservation and emission reduction efforts.
2. Policymakers should also carry out publicity activities advocating the concept of green and low-carbon living to raise residents’ awareness and understanding of energy

conservation and environmental protection, thereby increasing the proportion of green energy in residents' consumption and promoting the transformation of the energy consumption structure to a cleaner one. Policymakers should use various information channels, such as official government websites and media, to improve the depth of public low-carbon concepts and use general supervision as the main force of supervision to achieve a people-oriented approach. Additionally, policymakers should improve the monitoring mechanism of low-carbon development, incorporate low-carbon indicators into the work assessment of relevant government departments, and make comprehensive considerations based on indicators such as energy and resource consumption and urban environmental improvement. Furthermore, policymakers should implement the development strategy of prioritizing new energy transportation and shared transportation and advocate low-carbon travel for the public to reduce motorized travel and carbon emissions.

3. Policymakers should reduce the use of coal, develop and utilize renewable and clean energy, and promote the transformation of energy structure to green and low-carbon. Next, policymakers should actively adjust the industrial layout and gradually outlaw high-energy and high-pollution industries with low-energy and high-efficiency industries. Simultaneously, policymakers should actively formulate policies conducive to industrial transformation and guide and constrain the development direction of industrial enterprises utilizing laws and taxes to transform and upgrade the industrial structure from low-end manufacturing to high-tech industries. Finally, in order to effectively reduce carbon emissions, policymakers should encourage the development of low-carbon energy-saving technologies through innovative investment, enhance the efficiency of energy utilization technologies, and construct a clean, low-carbon, safe, and efficient energy system.

**Author contribution** Fanbo Bu: conceptualization, project administration, writing — review and editing, writing — original draft, software, visualization, formal analysis. Wenfeng Ge: writing — original draft, writing — review and editing, formal analysis. Yang Xu: formal analysis, methodology, data curation, writing — review and editing, validation. Asif Razzaq: writing — review and editing, validation. Qiyang Ran: writing — review and editing, writing — original draft, conceptualization, methodology, funding acquisition, supervision. Xiaodong Yang: writing — review and editing. Jie Peng: writing — original draft.

**Funding** The authors received financial support from the National Natural Science Foundation of China (71463057) and the Key Project of the Key Research Base of Humanities and Social Sciences in Xinjiang ordinary universities (the “two mountains” theory and the research center of high-quality green development in South Xinjiang) — the research on the green development of the economy in South Xinjiang

under the background of “double carbon” (JDZD202205), the graduate research and innovation project of Xinjiang University (XJ2021G014, XJ2021G013, XJ2020G020, XJ2022G014), and a special project of the School of Business and Economics. The usual disclaimer applies.

**Data availability** Not applicable.

## Declarations

**Ethics approval and consent to participate** Not applicable.

**Consent for publication** Not applicable.

**Competing interests** The authors declare no competing interests.

## References

- Ağbulut Ü (2022) Forecasting of transportation-related energy demand and CO<sub>2</sub> emissions in Turkey with different machine learning algorithms. *Sustain Prod Consum* 29:141–157. <https://doi.org/10.1016/j.spc.2021.10.001>
- Altmann A, Toloşi L, Sander O, Lengauer T (2010) Permutation importance: a corrected feature importance measure. *Bioinformatics* 26(10):1340–1347. <https://doi.org/10.1093/bioinformatics/btq134>
- Appiah K, Du J, Yeboah M, Appiah R (2019) Causal relationship between industrialization, energy intensity, economic growth and carbon dioxide emissions: recent evidence from Uganda. *Int J Energy Econ Policy* 9(2):237. <https://doi.org/10.32479/ijeep.7420>
- Bai H, Cao Q, An S (2023) Mind evolutionary algorithm optimization in the prediction of satellite clock bias using the back propagation neural network. *Sci Rep* 13(1):2095. <https://doi.org/10.1038/s41598-023-28855-y>
- Beuséjour L, Lenjosek G, Smart M (1995) A CGE approach to modeling carbon dioxide emissions control in Canada and the United States. *World Econ* 18(3):457–488. <https://doi.org/10.1111/j.1467-9701.1995.tb00224.x>
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32
- Chai J, Wu H, Hao Y (2022) Planned economic growth and controlled energy demand: how do regional growth targets affect energy consumption in China? *Technol Forecast Soc Chang* 185:122068. <https://doi.org/10.1016/j.techfore.2022.122068>
- Chen X, Shuai C, Wu Y, Zhang Y (2020) Analysis on the carbon emission peaks of China's industrial, building, transport, and agricultural sectors. *Sci Total Environ* 709:135768. <https://doi.org/10.1016/j.scitotenv.2019.135768>
- Chen L, Zhu J, Yang C (2022) Forecasting parameters in the SABR model. *J Econ Anal* 1:66–78. <https://doi.org/10.58567/jea01010005>
- Cheng H, Liu X, Xu Z (2022) Impact of carbon emission trading market on regional urbanization: an empirical study based on a difference-in-differences model. *Econ Anal Lett* 1:15–21. <https://doi.org/10.58567/eal01010003>
- Dalton M, Neill B, Prskawetz A, Jiang L, Pitkin J (2008) Population aging and future carbon emissions in the United States. *Energy Econ* 30(2):642–675. <https://doi.org/10.1016/j.eneco.2006.07.002>
- Du K, Li P, Yan Z (2019) Do green technology innovations contribute to carbon dioxide emission reduction? Empirical evidence from patent data. *Technol Forecast Soc Chang* 146:297–303. <https://doi.org/10.1016/j.techfore.2019.06.010>
- Ehrlich PR, Holdren JP (1971) Impact of population growth: complacency concerning this component of man's predicament is unjustified and counterproductive. *Sci* 171(3977):1212–1217. <https://doi.org/10.1126/science.171.3977.1212>

- Freedman M, Jaggi B (2011) Global warming disclosures: impact of Kyoto protocol across countries. *J Int Financ Manag Acc* 22(1):46–90. <https://doi.org/10.1111/j.1467-646X.2010.01045.x>
- Guo K (2022) Spatial dynamic evolution of environmental infrastructure governance in China. *Economic Analysis Letters* 1(2):23–27. <https://doi.org/10.58567/ea101020004>
- Gür TM (2022) Carbon dioxide emissions, capture, storage and utilization: review of materials, processes and technologies. *Prog Energy Combust Sci* 89:100965. <https://doi.org/10.1016/j.pecs.2021.100965>
- Hao Y, Ba N, Ren S, Wu H (2021a) How does international technology spillover affect China's carbon emissions? A new perspective through intellectual property protection. *Sustain Product Consum* 25:577–590. <https://doi.org/10.1016/j.spc.2020.12.008>
- Hao Y, Zhang ZY, Yang C, Wu H (2021b) Does structural labor change affect CO<sub>2</sub> emissions? Theoretical and empirical evidence from China. *Technol Forecasting Soc Chang* 171:120936. <https://doi.org/10.1016/j.techfore.2021.120936>
- Huang J-B, Luo Y-M, Feng C (2019) An overview of carbon dioxide emissions from China's ferrous metal industry: 1991–2030. *Resour Policy* 62:541–549. <https://doi.org/10.1016/j.resourpol.2018.10.010>
- Hussain M, Mir GM, Usman M, Ye C, Mansoor S (2022) Analysing the role of environment-related technologies and carbon emissions in emerging economies: a step towards sustainable development. *Environ Technol* 43(3):367–375. <https://doi.org/10.1080/09593330.2020.1788171>
- Irfan M, Razzaq A, Sharif A, Yang X (2022) Influence mechanism between green finance and green innovation: exploring regional policy intervention effects in China. *Technol Forecasting Soc Chang* 182:121882. <https://doi.org/10.1016/j.techfore.2022.121882>
- Kaika D, Zervas E (2013) The environmental Kuznets curve (EKC) theory—part A: concept, causes and the CO<sub>2</sub> emissions case. *Energy Policy* 62:1392–1402. <https://doi.org/10.1016/j.enpol.2013.07.131>
- Kim J, Lim H, Jo H-H (2020) Do aging and low fertility reduce carbon emissions in Korea? Evidence from IPAT augmented EKC analysis. *Int J Environ Res Public Health* 17(8):2972. <https://doi.org/10.3390/ijerph17082972>
- Kosarac A, Mladjenovic C, Zeljkovic M, Tabakovic S, Knezev M (2022) Neural-network-based approaches for optimization of machining parameters using small dataset. *Materials* 15(3):700. <https://doi.org/10.3390/ma15030700>
- Kusumadewi S, Rosita L, Wahyuni EG (2023) Stability of classification performance on an adaptive neuro fuzzy inference system for disease complication prediction. *IAES Int J Artif Intell* 12(2):532. <https://doi.org/10.11591/ijai.v12.i2.pp532-542>
- Leerbeck K, Bacher P, Junker RG, Goranović G, Corradi O, Ebrahimi R, . . . Madsen H (2020) Short-term forecasting of CO<sub>2</sub> emission intensity in power grids by machine learning. *App Energy* 277:115527. <https://doi.org/10.1016/j.apenergy.2020.115527>
- Li R, Wang Q, Liu Y, Jiang R (2021a) Per-capita carbon emissions in 147 countries: the effect of economic, energy, social, and trade structural changes. *Sustain Prod Consum* 27:1149–1164. <https://doi.org/10.1016/j.spc.2021.02.031>
- Li Y, Yang X, Ran Q, Wu H, Irfan M, Ahmad M (2021b) Energy structure, digital economy, and carbon emissions: evidence from China. *Environ Sci Pollut Res* 28(45):64606–64629. <https://doi.org/10.1007/s11356-021-15304-4>
- Liang B, Liu J, You J, Jia J, Pan Y, Jeong H (2023) Hydrocarbon production dynamics forecasting using machine learning: a state-of-the-art review. *Fuel* 337:127067. <https://doi.org/10.1016/j.fuel.2022.127067>
- Lu Y, Jiahua P (2013) Disaggregation of carbon emission drivers in Kaya identity and its limitations with regard to policy implications. *Adv Clim Chang Res*. 9(3):210. <https://doi.org/10.3969/j.issn.1673-1719.2013.03.009>
- Mason K, Duggan J, Howley E (2018) Forecasting energy demand, wind generation and carbon dioxide emissions in Ireland using evolutionary neural networks. *Energy* 155:705–720
- Meng M, Niu D (2011) Modeling CO<sub>2</sub> emissions from fossil fuel combustion using the logistic equation. *Energy* 36(5):3355–3359. <https://doi.org/10.1016/j.energy.2011.03.032>
- Meng Y, Liu L, Xu Z, Gong W, Yan G (2022) Research on the heterogeneity of green biased technology progress in chinese industries—decomposition index analysis based on the slacks-based measure integrating (SBM). *J Econ Anal* 1:17–34. <https://doi.org/10.58567/jea01020002>
- Mirzaei M, Bekri M (2017) Energy consumption and CO<sub>2</sub> emissions in Iran, 2025. *Environ Res* 154:345–351. <https://doi.org/10.1016/j.envres.2017.01.023>
- Mondal SK, Huang J, Wang Y, Su B, Kundzewicz ZW, Jiang S, . . . Jiang T (2022) Changes in extreme precipitation across South Asia for each 0.5 C of warming from 1.5 C to 3.0 C above pre-industrial levels. *Atmos Res* 266:105961. <https://doi.org/10.1016/j.atmosres.2021.105961>
- Muruganandam S, Joshi R, Suresh P, Balakrishna N, Kishore KH, Manikanthan SV (2023) A deep learning based feed forward artificial neural network to predict the K-barriers for intrusion detection using a wireless sensor network. *Measurement: Sensors* 25:100613. <https://doi.org/10.1016/j.measen.2022.100613>
- Pielke Jr R, Burgess MG, Ritchie J (2022) Plausible 2005–2050 emissions scenarios project between 2 and 3 degrees C of warming by 2100. *Environ Res Lett*. <https://iopscience.iop.org/article/10.1088/1748-9326/ac4ebf/meta/2022.024027>
- Razzaq A, Ajaz T, Li JC, Irfan M, Suksatan W (2021) Investigating the asymmetric linkages between infrastructure development, green innovation, and consumption-based material footprint: novel empirical estimations from highly resource-consuming economies. *Resour Policy* 74:102302. <https://doi.org/10.1016/j.resourpol.2021.102302>
- Razzaq A, Sharif A, An H, Aloui C (2022) Testing the directional predictability between carbon trading and sectoral stocks in China: new insights using cross-quantilogram and rolling window causality approaches. *Technol Forecasting Soc Chang* 182:121846. <https://doi.org/10.1016/j.techfore.2022.121846>
- Ren S, Liu Z, Zhanbayev R, Du M (2022) Does the Internet development put pressure on energy-saving potential for environmental sustainability? Evidence from China. *J Econ Anal* 1:50–65. <https://doi.org/10.58567/jea01010004>
- Shi X, Xu Y (2022) Evaluation of China's pilot low-carbon city program: a perspective of industrial carbon emission efficiency. *Atmos Pollut Res* 13(6):101446. <https://doi.org/10.1016/j.apr.2022.101446>
- Shi R, Irfan M, Liu G, Yang X, Su X (2022a) Analysis of the impact of livestock structure on carbon emissions of animal husbandry: a sustainable way to improving public health and green environment. *Front Public Health* 145. <https://doi.org/10.3389/fpubh.2022.835210>
- Shi, X, Xu Y, Sun W (2022b) Evaluating China's pilot carbon Emission Trading Scheme: collaborative reduction of carbon and air pollutants. *Environ Sci and Pollut Res* 1–20.
- Sonia SE, Nedunchezian R, Ramakrishnan S, Kannammal KE (2023) An empirical evaluation of benchmark machine learning classifiers for risk prediction of cardiovascular disease in diabetic males. *Int J of Healthc Manag*, 1–16. <https://doi.org/10.1080/20479700.2023.2170006>

- Su X, Yang X, Zhang J, Yan J, Zhao J, Shen J, Ran Q (2021) Analysis of the impacts of economic growth targets and marketization on energy efficiency: evidence from China. *Sustainability* 13(8):4393. <https://doi.org/10.3390/su13084393>
- Sun Y, Duru OA, Razzaq A, Dinca MS (2021) The asymmetric effect eco-innovation and tourism towards carbon neutrality target in Turkey. *J Environ Manag* 299:113653. <https://doi.org/10.1016/j.jenvman.2021.113653>
- Sun Y, Razzaq A, Sun H, Irfan M (2022) The asymmetric influence of renewable energy and green innovation on carbon neutrality in China: analysis from non-linear ARDL model. *Renewable Energy* 193:334–343. <https://doi.org/10.1016/j.renene.2022.04.159>
- Tang C, Xue Y, Wu H, Irfan M, Hao Y (2022) How does telecommunications infrastructure affect eco-efficiency? Evidence from a quasi-natural experiment in China. *Technol Soc* 69:101963. <https://doi.org/10.1016/j.techsoc.2022.101963>
- Wang H, Ma C, Zhou L (2009) A brief review of machine learning and its application. Paper presented at the 2009 international conference on information engineering and computer science
- Wang Y, Yan W, Ma D, Zhang C (2018) Carbon emissions and optimal scale of China's manufacturing agglomeration under heterogeneous environmental regulation. *J Clean Prod* 176:140–150. <https://doi.org/10.1016/j.jclepro.2017.12.118>
- Wang LO, Wu H, Hao Y (2020) How does China's land finance affect its carbon emissions? *Struct Chang Econ Dyn* 54:267–281. <https://doi.org/10.1016/j.strueco.2020.05.006>
- Wang C, Zhao M, Gong W, Fan Z, Li W (2021a) Regional heterogeneity of carbon emissions and peaking path of carbon emissions in the Bohai Rim Region. *J Math* 1–13. <https://doi.org/10.1155/2021/3793522>
- Wang W-Z, Liu L-C, Liao H, Wei Y-M (2021b) Impacts of urbanization on carbon emissions: an empirical analysis from OECD countries. *Energy Policy* 151:112171. <https://doi.org/10.1016/j.enpol.2021.112171>
- Wesseh PK Jr, Lin B, Atsagli P (2017) Carbon taxes, industrial production, welfare and the environment. *Energy* 123:305–313. <https://doi.org/10.1016/j.energy.2017.01.139>
- Wu H, Xu L, Ren S, Hao Y, Yan G (2020) How do energy consumption and environmental regulation affect carbon emissions in China? New evidence from a dynamic threshold panel model. *Resour Policy* 67:101678. <https://doi.org/10.1016/j.resourpol.2020.101678>
- Wu H, Ba N, Ren S, Xu L, Chai J, Irfan M, ... Lu ZN (2022) The impact of internet development on the health of Chinese residents: transmission mechanisms and empirical tests. *Socio Econ Plan Sci* 81:101178. <https://doi.org/10.1016/j.seps.2021.101178>
- Xiao H, Liu J (2022) The impact of digital economy development on local fiscal revenue efficiency. *Econ Anal Lett* 1:1–7. <https://doi.org/10.58567/eal01020001>
- Yan Y, Wu C, Wen Y (2021) Determining the impacts of climate change and urban expansion on net primary productivity using the spatio-temporal fusion of remote sensing data. *Ecol Indic* 127:107737. <https://doi.org/10.1016/j.ecolind.2021.107737>
- Yang X, Su X, Ran Q, Ren S, Chen B, Wang W, Wang J (2022) Assessing the impact of energy internet and energy misallocation on carbon emissions: new insights from China. *Environ Sci Pollut Res* 29(16):23436–23460. <https://doi.org/10.1007/s11356-021-17217-8>
- Yang Y, Tang D, Zhang P (2020) Double effects of environmental regulation on carbon emissions in China: empirical research based on spatial econometric model. *Discret Dyn Nat Soc* 2020:1–12. <https://doi.org/10.1155/2020/1284946>
- Yang X, Wang W, Wu H, Wang J, Ran Q, & Ren S (2022) The impact of the new energy demonstration city policy on the green total factor productivity of resource-based cities: Empirical evidence from a quasi-natural experiment in China. *J Environ Plan Manag* 66(2):293–326. <https://doi.org/10.1080/09640568.2021.1988529>
- Zhang S, Chen W (2022) Assessing the energy transition in China towards carbon neutrality with a probabilistic framework. *Nat Commun* 13(1):1–15. <https://doi.org/10.1038/s41467-021-27671-0>
- Zhang Y-J, Liang T, Jin Y-L, Shen B (2020) The impact of carbon trading on economic output and carbon emissions reduction in China's industrial sectors. *Appl Energy* 260:114290. <https://doi.org/10.1016/j.apenergy.2019.114290>
- Zhao S, Tian W, Dagestani AA (2022) How do R&D factors affect total factor productivity: based on stochastic frontier analysis method. *Econ Anal Lett* 1:28–34. <https://doi.org/10.58567/eal01020005>
- Zhu Q, Peng X, & Wu K (2012) Calculation and decomposition of indirect carbon emissions from residential consumption in China based on the input–output model. *Energy Policy* 48:618–626. <https://doi.org/10.1016/j.enpol.2012.05.068>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.