sustainability

MDPI

Review

A Review of Macroscopic Carbon Emission Prediction Model Based on Machine Learning

Yuhong Zhao [1,2], Ruirui Liu [1,2], Zhansheng Liu [1,2,*], Liang Liu [1,2], Jingjing Wang [1,2] and Wenxiang Liu [1,2]

[1] Faculty of Architecture, Civil and Transportation Engineering, Beijing University of Technology, Beijing 100124, China
[2] Key Laboratory of Urban Security and Disaster Engineering of Ministry of Education, Beijing University of Technology, Beijing 100124, China
* Correspondence: liuzhansheng@bjut.edu.cn

Abstract: Under the background of global warming and the energy crisis, the Chinese government has set the goal of carbon peaking and carbon neutralization. With the rapid development of machine learning, some advanced machine learning algorithms have also been applied to the control and prediction of carbon emissions due to their high efficiency and accuracy. In this paper, the current situation of machine learning applied to carbon emission prediction is studied in detail by means of paper retrieval. It was found that machine learning has become a hot topic in the field of carbon emission prediction models, and the main carbon emission prediction models are mainly based on back propagation neural networks, support vector machines, long short-term memory neural networks, random forests and extreme learning machines. By describing the characteristics of these five types of carbon emission prediction models and conducting a comparative analysis, we determined the applicable characteristics of each model, and based on this, future research ideas for carbon emission prediction models based on machine learning are proposed.

Keywords: macroscopic carbon emission; prediction model; machine learning

Citation: Zhao, Y.; Liu, R.; Liu, Z.; Liu, L.; Wang, J.; Liu, W. A Review of Macroscopic Carbon Emission Prediction Model Based on Machine Learning. *Sustainability* **2023**, *15*, 6876. https://doi.org/10.3390/su15086876

Academic Editor: Lifeng Wu

Received: 18 March 2023
Revised: 7 April 2023
Accepted: 13 April 2023
Published: 19 April 2023

1. Introduction

As the Paris Agreement proposes to curb global warming by controlling greenhouse gas emissions, more and more people are beginning to pay attention to carbon emissions. Robust regional changes in extreme temperatures and precipitation with cumulative CO_2 emissions [1] and extreme climate can lead to a decrease in regional ecosystem carbon stocks, which leads to an imbalance in the ecosystem [2]. Indeed, because of globalization, major climate disruptions in some countries can strongly affect others owing to political unrest, migration, impacts on global food production, supply chains and trade for instance [3–5]. Zhang et al. [6] analyzed the driving factors affecting China's carbon dioxide emissions by using the LMDI method and concluded that energy intensity was the primary indicator that reduced CO_2 emissions. The carbon emission prediction model relies on converting energy data, whether directly or indirectly, to estimate carbon emissions. This approach not only helps us understand the impact of energy consumption on the environment, but it also has the potential to drive innovation in the energy sector. By using carbon emissions as a unified measure of various energy variables, we can gain a more comprehensive and intuitive understanding of energy use, beyond what traditional monitoring methods that focus on individual energy sources can provide. Studies have shown that effective policy regulation can successfully control carbon emissions [7], and a precise and efficient carbon emission prediction model can be a valuable tool in shaping government strategies for future regulation.

We used the "carbon emission prediction model" as the primary search keyword on the Web of Science and manually analyzed relevant papers from 1998 to 2021. Figure 1

shows that the number of papers related to carbon emission prediction models has been rapidly increasing each year, particularly since 2005 and especially in recent years.

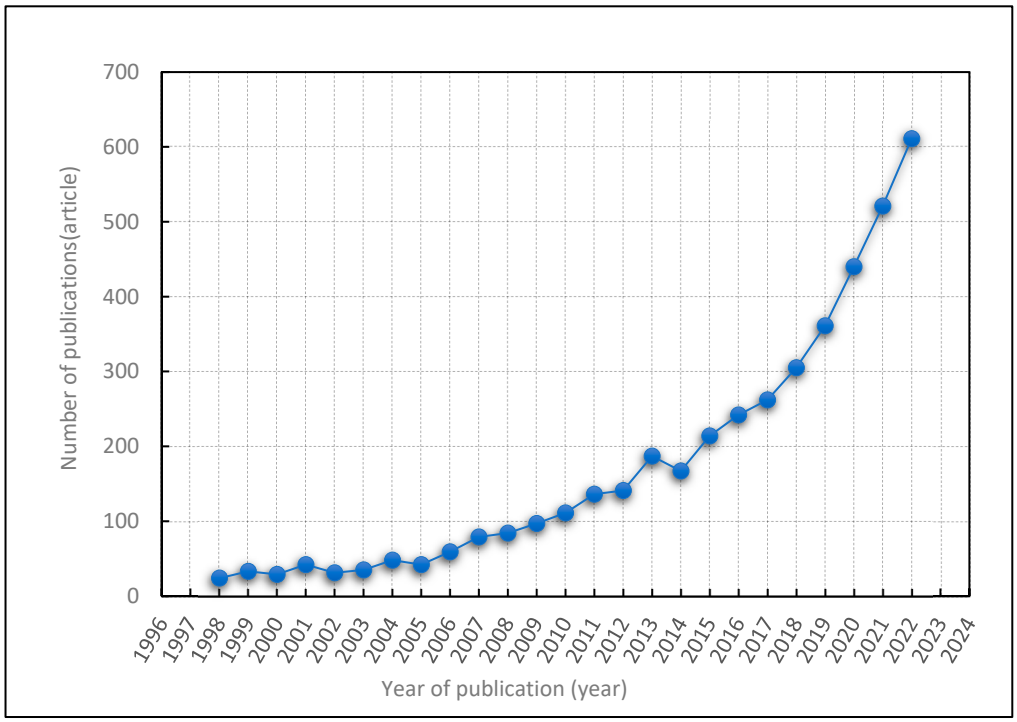

Figure 1. General trend of relevant papers on carbon emission prediction model (source: Web of Science).

Currently, many experts and scholars are devoted to conducting carbon emission calculations and creating prediction models. Several models have been proposed, including the Kaya (Japanese scholar Yoichi Kaya) model [8], Computable General Equilibrium (CGE) model [9], production function theory [10], Logarithmic Mean Divisia Index (LMDI) method [11] and other carbon emission calculation models, as well as the GM (1,1) model (GM (1,1), which refers to the first-order differential equation to establish a model for a variable), multiple linear regression model, differential integrated moving average autoregressive model (ARIMA), scalable random environmental impact assessment model (STIRPAT), system dynamics model (SD) and other carbon emission prediction models [12]. However, these models use traditional mathematical model methods to solve the problem of carbon emissions, which are relatively poor in accuracy and efficiency and cannot be efficiently used to achieve the established goals of carbon emission calculations and predictions. A keyword search on the Web of Science core collection using the term "carbon emission prediction model" and a keyword burst analysis performed through CiteSpace [13] revealed that recent research on carbon emission prediction models has focused mainly on four key areas: machine learning, renewable energy, carbon market and deep learning (Figure 2). These are all currently hot areas of carbon emission research.

During the seventy-fifth session of the United Nations General Assembly, China made a commitment to reach carbon reduction targets for the first time; namely, they aimed to peak carbon emissions by 2030 and achieve carbon neutrality by 2060 [14]. As the world's largest emitter, accounting for approximately 30% of global carbon emissions, China is under immense pressure to cut its carbon footprint. Fortunately, the development of machine learning has led to the maturation of corresponding methods, which have extended beyond the realm of computer science and found success in other fields. In particular, these methods have shown promise in the field of carbon emission prediction, which provides a valuable research direction for the global effort to reduce carbon emissions.

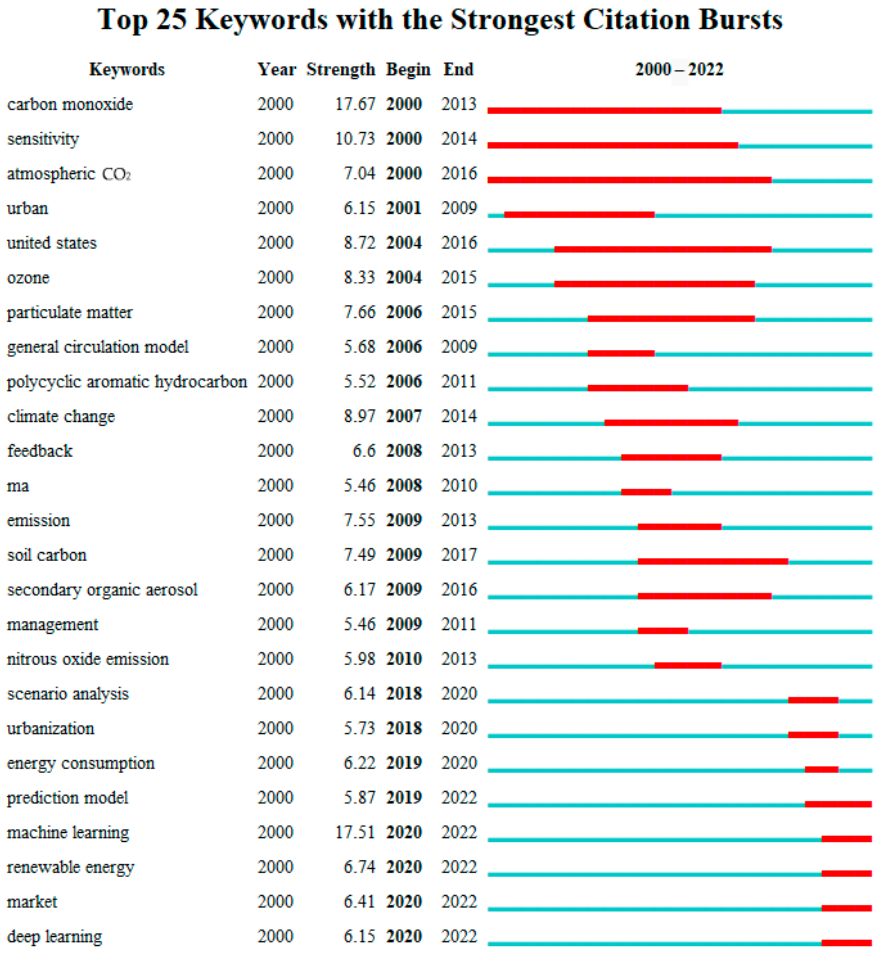

Figure 2. Keyword mutation (source: Web of Science, construction by CiteSpace).

1.1. Development Status

Traditional prediction models for carbon emissions can be broadly divided into several categories, including the grey prediction model, time series model, multiple regression model, logistic regression model, economic analysis model and others. Among them, the grey prediction model is particularly effective in analyzing uncertain factors. Typically, the GM (1,1) grey correlation model is used for carbon emission prediction, and the resulting prediction curve is generally smooth. However, this model is unable to predict nonsmooth, discrete curves. The time series method is another commonly used approach, which reflects the time variation law of carbon emissions. This method is suitable for data that are relatively stable and have roughly linear changes, but it cannot process nonlinear data. The multiple regression model is one of the earliest models applied to the prediction of carbon emissions. The core is to establish the relationship function between multiple carbon-emission-influencing factors and carbon emissions, as shown in Equation (1) [12]. The logistic model is an advanced version of the multiple regression model, which can overcome the defects of non-normal random error terms, heteroscedasticity and the regression equation limitations of the multiple regression model, but the function interpretation is poor. The economic analysis model mainly explores the relationship between the degree of economic development and carbon emissions from an economic standpoint in order to facilitate carbon emission prediction. Of all the current models, only the grey prediction model is still undergoing continuous optimization and application, while the other traditional models are gradually being abandoned due to their high complexity and low efficiency:

$$y(t) = a_0 + a_1x_1(t) + a_2x_2(t) + \cdots + a_nx_n(t) + a(t) \quad (1)$$

where $y(t)$ is the predicted value of the carbon emissions and $x_1(t), x_2(t) \cdots x_n(t)$ are the factors that affect carbon emissions. $a_1, a_2 \cdots a_n$ are the regression coefficients of the influencing factors. $a(t)$ is a random variable, the mean value must be equal to 0 and the variance must be constant.

As a new topic in modern times, machine learning plays an important role in information technology, especially in the field of artificial intelligence. Its essence is to guide the machine with many data and rules so that it can judge and predict new data. It is a model that imitates human learning behavior. The ability to analyze data determines the accuracy of the prediction. Tracing the history of machine learning, it originated in the 1940s. In 1943, neuroscientist McCulloch and mathematician Pitts published a paper title "Logical Calculus of Inner Thoughts in Neural Activities" and first proposed the MCP model. The principle was to transform various types of data into the information that we need after weighing. The embryonic form of neural networks began to emerge, which was also the earliest model of machine learning. After decades of development, machine learning has evolved to a comprehensive and systematic system that includes decision trees, K-nearest neighbor, logistic regression, BP neural network, perceptual vector machines, long short-term memory neural networks, deep learning, transfer learning and other methods, and new methods are still continuing to be derived. A back propagation neural network (BPNN) is a concept proposed by Rumelhart and McClelland in 1986. It is a multilayer feedforward neural network trained by an error back propagation algorithm. Because of its simple, easy-to-use and efficient attributes, BP neural networks have been widely promoted and applied. It can solve most of the problems in reality. However, it also has the disadvantages of gradient disappearance and gradient explosion. The support vector machine (SVM) proposed in the mid 1990s makes up for this defect well. An SVM is a kind of generalized linear classifier that classifies data through supervised learning. Its decision boundary is the maximum margin hyperplane of learning samples. SVMs use the hinge loss function to calculate the empirical risk and add a regularization term to the solution system to optimize the structural risk. It is a sparse and robust classifier. An SVM uses one of the common kernel learning methods, and it can be used to classify nonlinearly via the kernel method. Similar to an SVM, an extreme learning machine (ELM) has been proposed to improve the learning efficiency of BPNNs. Different from SVMs, the learning process of ELMs includes random weight and calculation weight. Although this structure has poor interpretability, its prediction accuracy is relatively high. There are also recurrent neural networks that belong to neural networks. The recurrent network has the problem of gradient disappearance. Long short-term memory (LSTM) neural networks, as an advanced stage of the recurrent network, solve the problem of the gradient disappearance of the recurrent neural network. Because it is usually used to process serial data, and most of the data related to carbon emissions are also time series data, many scholars try to use LSTM to predict carbon emissions, which also has good results. Random forest (RF) is a common application algorithm in the field of machine learning. It combines a large number of decision trees to form a new model. By splitting and optimizing complex data, each decision tree is independently calculated and analyzed. Finally, the analysis results are synthesized. This method greatly improves the speed of the operation. However, most of its current applications in the field of carbon emission prediction focus on data classification, and there are relatively few applications at the prediction level.

The traditional model is less integrated with today's emerging technologies. Most of these models rely on artificial empirical formulas and mathematical model-based algorithms for prediction, which are highly dependent on people and often have slow update rates. However, the emergence of machine learning has given researchers a new direction. Machine learning can perform self-learning based on actual case data, and its update rate is much faster than that of mathematical models, which reduces the dependence on people. A large number of new algorithms have been applied to the field of carbon emission predictions and have shown great promise. Machine learning models not only outperform traditional prediction models in terms of accuracy but also demonstrate sig-

nificant advantages in terms of efficiency. Table 1 compares the traditional model with the intelligent model and indicates that traditional models require significant experience in carbon emission modeling, which makes it more demanding for people. In contrast, intelligent models rely on data and algorithm improvements, which leads to an improved fitting efficiency and accuracy.

Table 1. Comparison between traditional prediction model and intelligent prediction model. Own elaboration.

	Traditional Model	Intelligent Model
Forecast function generation	Analyze, verify and modify all kinds of data by using manual experience and mathematical analysis methods to obtain corresponding prediction functions	Import the corresponding carbon emission data into the program for self-learning and finally generate the corresponding prediction function
Application difficulties	Highly demanding on the experience of researchers, too computationally intensive or even unpredictable for some complex problems	High accuracy requirements for data samples, requiring personnel with the appropriate level of programming

1.2. Mining Common Machine Learning Algorithms

Using 'carbon emission' and 'prediction' as keywords to search the core collection of Web of Science, 144 papers related to carbon emission prediction published from 1 January 2015 to 1 September 2022 were manually selected. The results are presented in Figure 3 (only the main keywords are shown). The most prominent keywords were artificial neural network followed by extreme learning machines, support vector machines and random forest.

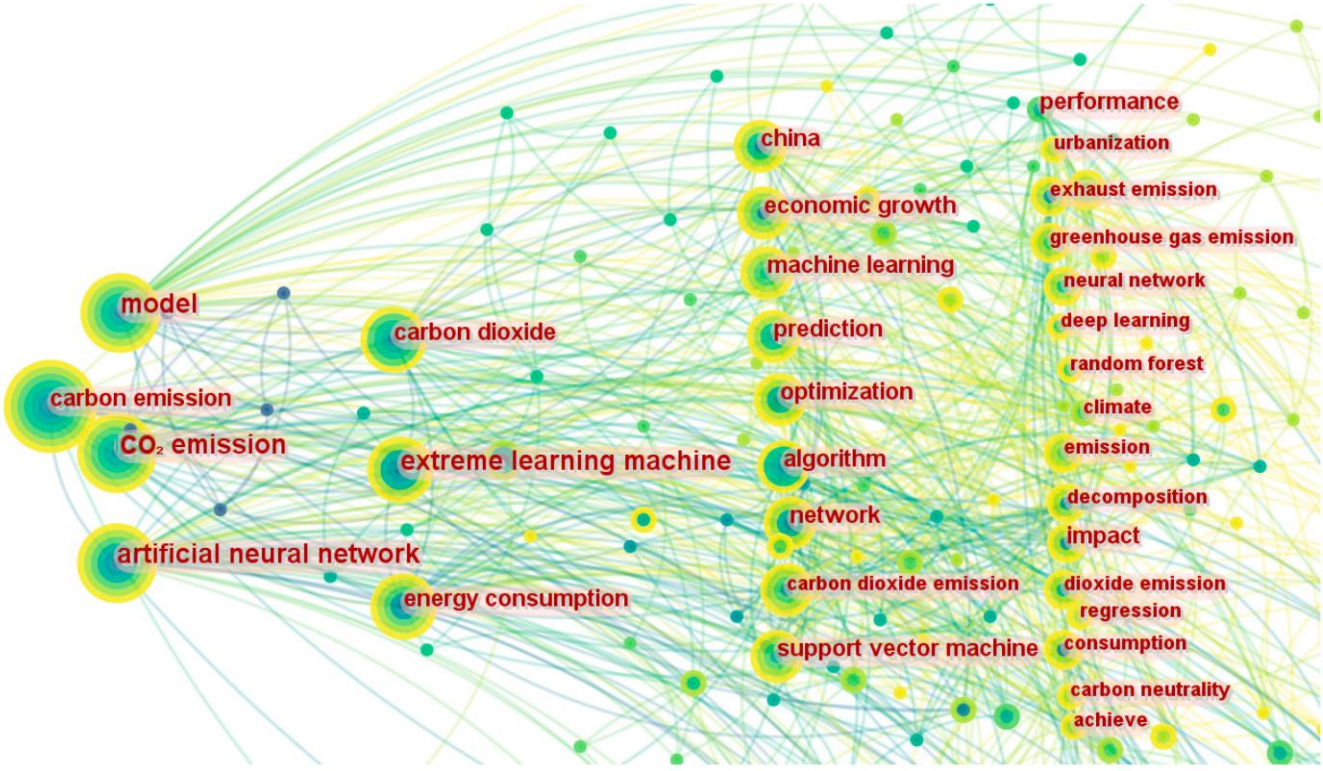

Figure 3. Keyword co-occurrence (source: Web of Science, construction by CiteSpace).

However, given the broad scope of artificial neural networks, we conducted a further manual search using "carbon emission" and "prediction" as keywords in the Web of Science and manually selected 112 articles related to carbon emission prediction published after 1 January 2020 to statistically analyze their topics (Figure 4). Our findings indicated that LSTM has gained acceptance among many researchers, followed by BPNN, ELM, SVM, SVR and RF. Please note that the "ANN" in Figure 4 refers to the general term of other neural networks in addition to those already presented.

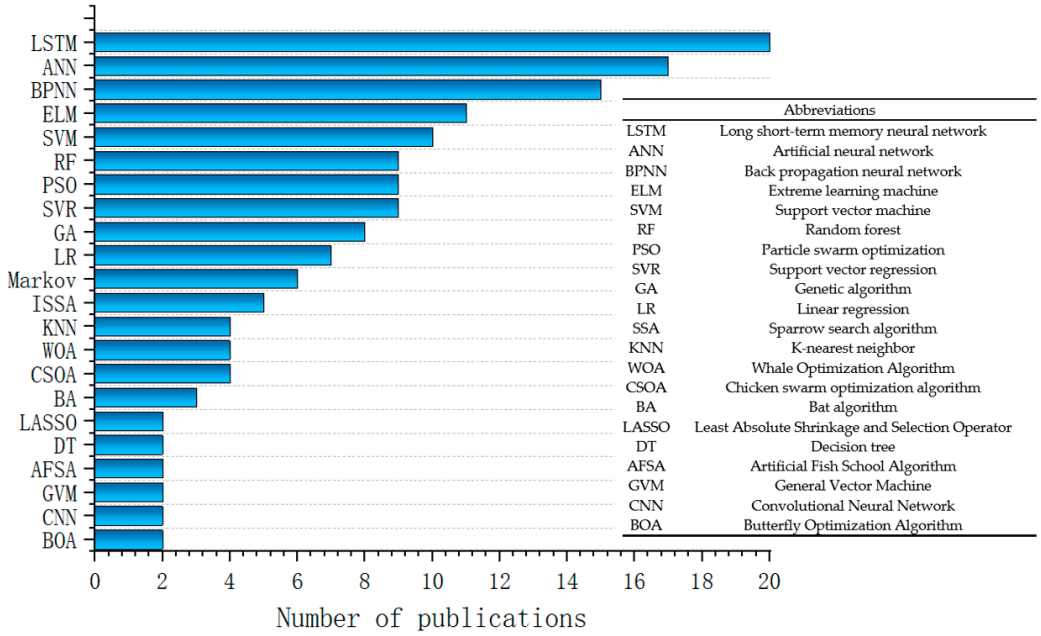

Figure 4. Topic distribution of papers related to carbon emission prediction model (source: Web of Science).

After comparing the outcomes of using the paper retrieval software with the outcomes when using manual retrieval, we observed that the results were largely similar. From extensive literature reviews, we discovered that BP neural networks, support vector machines, long short-term memory neural networks, random forests and extreme learning machines demonstrate superior prediction capabilities. Therefore, this paper principally focuses on exploring the carbon emission prediction models related to these five machine learning algorithms.

Our goal is to propose a comprehensive review that summarizes and evaluates the current research on carbon emission prediction models based on machine learning. The specific objectives of this review are as follows:

1. To compare the advantages and disadvantages of traditional carbon emission prediction models with intelligent carbon emission prediction models based on machine learning, and to indicate the significant advantage of machine learning in the field of carbon emission prediction.
2. To identify and describe the characteristics of the five mainstream carbon emission prediction models based on machine learning through extensive literature review, and to clarify their applicable characteristics through a comparative analysis.
3. To provide insights into the current research status and application characteristics of carbon emission prediction models based on machine learning, and to consider the future development path of this field.

2. Driving Factors

When predicting carbon emissions, higher is required of the original data, so it is crucial to carefully select and calculate the data used. An accurate and comprehensive

variable data set will bring a qualitative leap to the accuracy of the final prediction results, so knowing how to better determine the factors relative to carbon emissions is very important. Nowadays, many carbon emission prediction models based on machine learning are developed based on the basic theory and data of traditional models. The data generated by traditional carbon emission models are used for autonomous learning so that the prediction results are more intelligent, which greatly improves the prediction efficiency and reduces the complexity. Therefore, both the traditional prediction models and the intelligent prediction models need relatively complete original data to support.

The typical driving factors of most carbon emission prediction models are mainly divided into these categories: economic development level, industrial structure, urbanization level, population size, energy consumption and technology development. There are also some scholars that have expanded the driving factors, such as traffic load [15,16], import and export scale [17], education level [18] and so on. In addition, there are some scholars that have refined the typical driving factors, such as the level of economic development, which is subdivided into GDP, per capita GDP, per capita disposable income, per capita consumption expenditure and so on. These refinements improve the accuracy and reliability of the basic data used in carbon emission prediction models.

When data are related on the national level, the drivers can be classified and obtained. For example, data on GDP published by national and international organizations can be directly used to determine the level of economic development, while data related to the national census can be used to determine population size. National data are generally more comprehensive and available. At the regional level, relevant data can typically be obtained from state-published sources, but they may not always be comprehensive. In such cases, it may be necessary to use data provided by unofficial organizations or institutions, which can be obtained through screening and identification. At the family level, the data are typically more refined, which makes it difficult to obtain them from official channels. Generally, sample questionnaire surveys are used to obtain the required data, such as income level and population composition. These three levels of data are usually not fused. On the one hand, the sample size supported by the prediction model is limited, and excessive calls may lead to some unpredictable problems. On the other hand, the three levels of data belong to different dimensions, and there is a great correlation between the data. Mixed calls may cause repeated calls, which makes the prediction results abnormal for the characteristics of a certain type of data. The data corresponding to these factors were normalized and imported into the intelligent prediction model as training samples and test samples, which could speed up the generation of the prediction model. Equation (2) is the normalized basic formula:

$$y = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{2}$$

where y is the normalized value, x is the prenormalized value, $x_{\min}$ is the minimum value among values of the same type and $x_{\max}$ is the maximum value among values of the same type.

Numerous scholars rely on the official statistical yearbook to obtain credible data on carbon emissions and their driving factors, whether through direct or indirect means. The primary source of carbon emission data is the official energy statistics, which are then converted into carbon emissions using the calculation method outlined in the IPCC Guidelines. This process ensures the reliability and accuracy of the data.

Some scholars have made unique innovations by bypassing numerous research ideas. For example, Wen and Cao [19] focused on predicting the carbon emissions of single buildings by collecting building data with sensors and converting them into corresponding carbon emission values. They predicted and analyzed the carbon emissions of single buildings to explore carbon reduction paths. Kong et al. [20], on the other hand, boldly set aside the traditional independent variable–dependent variable data set by using an ensemble empirical mode and variational mode decomposition to decompose the carbon emission data into 11 IMF components. They then mined the law of components and total

carbon emission, which had a good effect. These data-level innovations were based on the high fitting of the final prediction model, which will provide future researchers with inspiration and valuable insights.

3. Carbon Emission Prediction Model Based on Machine Learning

3.1. Prediction Model of Carbon Emission—BP Neural Network

3.1.1. Introduction

The backpropagation (BP) algorithm utilizes the gradient descent method to train a neural network and employs empirical risk minimization. The error backpropagation algorithm is a critical component of the BP neural network, which trains a multilayer feed-forward neural network with an input layer, hidden layer and output layer. The BP neural network minimizes the mean square error of the model via the gradient descent method to meet the preset accuracy requirements. The BP algorithm involves two processes: First, it uses forward propagation to calculate the difference between the actual and expected output values, known as the output error. Then, the error reverse propagation process assigns the output error to each layer of neurons, which corrects the weight and threshold of each neuron. The BP algorithm is widely used to deal with nonlinear problems due to its excellent nonlinear mapping ability. However, it has some limitations when applied to carbon emission prediction models, such as a tendency to fall into the local minimum, which does not guarantee that the final solution is the global optimal solution. Additionally, it can have a slow convergence speed and a long training time due to a small learning rate.

The BP neural network is composed of three layers: the input layer, hidden layer, and output layer. The input layer receives the necessary information and transfers it to the active function in the hidden layer for processing. The processed results are then fed back through the output layer. Figure 5 depicts the topological structure of the BP neural network. The core component of this network is the hidden layer, which acts as a vital bridge between the input and output layers and serves as a hub for processing various types of information:

$$y = \omega_1 x_1 + \omega_2 x_2 + b \quad (3)$$

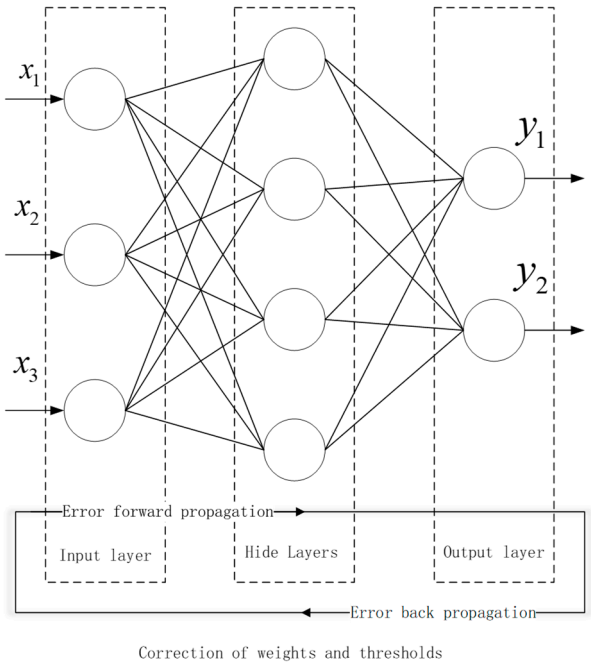

Figure 5. Structure diagram of BP neural network. Own elaboration.

Equation (3) is a simple neural network control function. y is the output value; x_1 and x_2 are the input values; ω_1 and ω_2 are the weights of the corresponding input values,

which are mainly used to process the input values to reach the predicted values; and b is the offset, which is mainly used to fine tune the input values after weight conversion to make them closer to the actual values.

The number of neurons in the hidden layer of a BP neural network influences the ability achieve an optimal performance. Using too few neurons in the hidden layer can result in underfitting, where the network fails to capture the complexity of the data. Conversely, using too many neurons can lead to overfitting, where the network becomes too complex and starts to memorize noise in the training data, which leads to poor generalization on the new data. Therefore, finding the optimal number of neurons in the hidden layer is a critical step in designing a BP neural network. By carefully selecting the number of neurons in the hidden layer, we can strike a balance between the model complexity and generalization performance, which will lead to a more effective and efficient network.

3.1.2. Application

The BP neural network is a multilayer feedforward network trained using the error backpropagation algorithm. It is highly flexible and widely used in applications such as damage detection, fault diagnosis and performance prediction. Recently, it has also been increasingly applied to carbon emission prediction with great success.

From the analysis of the existing research (Table 2), the factors influencing carbon emissions can be divided into four categories: population, economic, energy and resource. To select appropriate indicator parameters, the correlation analysis method, the gray correlation method or ridge regression method can be used. It is important to select the right number of indicators to balance accuracy and computational efficiency. Typically, five to eight factors with the greatest influence are selected, while the remaining factors with less influence are taken into account through the error term in the prediction process:

$$m = \sqrt{n+l} + \alpha \tag{4}$$

where m represents the number of neurons in the hidden layer; n and l are the number of neurons in the input layer and the number of neurons in the output layer, respectively; and α is a range constant between 1 and 10.

In the analysis of the BP neural network, the hidden layer played a particularly important role as it directly affected the accuracy of the predictions. Currently, when establishing the structure of a BP neural network to solve specific problems, the number of neurons in the hidden layer is mostly determined through experience. Based on existing research, the most efficient method for estimating the number of hidden layer neurons is through Equation (4), which provides a range of possible values. The optimal number of hidden layers can then be determined step by step within this range through testing.

Apart from improving the structure of the BP neural network to enhance the prediction accuracy, optimization algorithms can also be used to optimize its weights and thresholds. The commonly used optimization algorithms include particle swarm optimization (PSO), genetic algorithms (GA) and others. With the use of these algorithms, the coefficient of determination of the optimized prediction result can reach as high as 0.95, which indicates a relatively high degree of fitting. However, in cases where some influencing factors are relatively simple and not fully considered, the prediction accuracy may be low. Therefore, the final results of the BP model are greatly affected by the selection of the influencing factors and have higher requirements for the original data.

The carbon emission prediction model based on the BP neural network has been significantly enhanced and improved with the contributions of many researchers, particularly in terms of accuracy and operational speed. Given the high demand for data accuracy in the BP neural network prediction model, some scholars have implemented various approaches to enhance the accuracy of the original carbon emission data. Specifically, the selection method of the influencing factors is optimized, and the data selection of the identified influencing factors is considered to ensure the authenticity and effectiveness of the model.

With the optimization of the BP neural network structure and prediction algorithm, the prediction accuracy was significantly improved.

Table 2. Research on carbon emission prediction model based on BP neural network (source: Web of Science).

Reference	Data Sources	Prediction Method Selection	Prediction Accuracy
[21]	Taking permanent resident population, age of head of household, family income, housing area and current housing value as the indicators of household consumption carbon emissions, a total of 307 samples of 14 household structures were selected as the research objects.	The BP model. The number of hidden layer nodes is 8.	R^2 = 0.9456, MSE = 0.0223
[16]	The grey correlation method was used to identify the main factors that influenced each stage of the building's life cycle, which were then used to construct the STIRPAT model. These influencing factors included the size of the resident population, the degree of urbanization, per capita GDP, the added value of the tertiary industry, the average distance of steel production and road transportation and the labor productivity of the construction enterprises. Carbon emissions were determined primarily by the sum of energy-related carbon emissions, electricity-related carbon emissions and thermal-related carbon emissions.	The GA-BP model. The number of hidden layer nodes is 15.	R^2 = 0.944, MSE = 0.034694
[22]	Based on the industry's attributes and existing research results, the influencing factors of carbon emissions in the heavy chemical industry with a high carbon emission factor were selected.	The PSO-BP model. The number of hidden layer nodes is 9.	R^2 = 0.9985, MSE = 0.4139
[23]	The EEMD method was used to decompose the daily calculated carbon emission data into six modal functions and a residual sequence.	The PSO-BP model. The number of hidden layer nodes is 6.	R^2 = 0.9507, MSE = 0.1177, MAPE = 0.093
[24]	After referring to the relevant literature, five influencing factors related to carbon emissions were established: coal consumption, crude oil consumption, natural gas consumption, population and GDP. Carbon emissions were calculated using data from the official statistical yearbook.	The IPSO-BP model. The number of hidden layer nodes is 5.	$R^2 > 0.9$
[25]	The Building Energy Consumption (BEC) and Building Environment Factor (BEF) data for Central and Eastern Changxing City, Northeast Zhejiang Province, China, from 2018 were used. The BEF data included buildings, blocks, population, industry and information points provided by various government departments in Changxing.	The BP model. The number of hidden layer nodes is 10.	10% < MAPE < 20%

3.2. Prediction Model of Carbon Emission—SVM

3.2.1. Introduction

The support vector machine (SVM) is a powerful supervised learning method that finds broad application in statistical classification and regression analysis. As a generalized linear classifier, an SVM possesses the unique ability to minimize empirical error while maximizing the geometric edge region. As a result, SVMs are often referred to as maximum edge region classifiers.

SVMs map input vectors into a higher-dimensional space where a maximum margin hyperplane is established. This hyperplane separates the data into two classes and is defined by two parallel hyperplanes on either side. The SVM finds the hyperplane that maximizes the distance between the parallel hyperplanes, which is called the margin. It is assumed that a larger margin results in a better generalization performance. Figure 6 illustrates the structure diagram of a support vector machine.

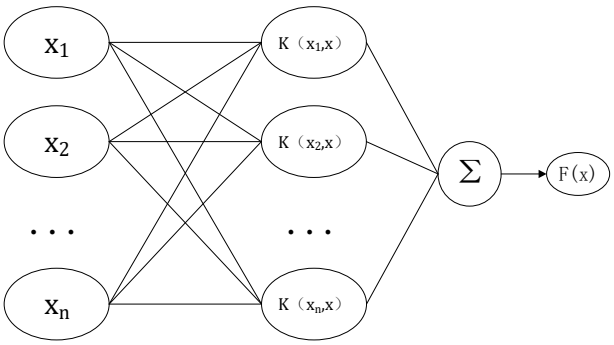

Figure 6. Structure diagram of support vector machine. Own elaboration.

The main idea of an SVM can be summarized as two points:

1. It is analyzed for linearly separable cases. For linearly nonseparable cases, the linearly nonseparable samples in the low-dimensional input space are transformed in the high-dimensional feature space by using the nonlinear mapping algorithm so that they are linearly separable, which makes it possible for the high-dimensional feature space to use the linear algorithm to analyze the nonlinear characteristics of the samples.
2. Based on structural risk minimization theory, it is used to construct the optimal partition hyperplane in the feature space so that the learner can be globally optimized, and the expected risk in the whole sample space satisfies a certain upper bound with a certain probability.

3.2.2. Application

An SVM is a powerful machine learning method rooted in statistical learning theory. It is particularly effective in function approximation and regression estimation, and it has found numerous applications in pattern recognition, including but not limited to portrait recognition, text classification, handwritten character recognition and bioinformatics. SVMs have low requirements for samples and they are suitable for limited samples. Theoretically, they can obtain the global optimal point, and the computational complexity is independent of the sample dimension.

The carbon emission prediction model based on SVMs is similar to the carbon emission prediction model based on BP neural networks in terms of data selection, but it has great adaptability when it comes to the selection of the number of influencing factors, which ranges between two and nine (Table 3). In other words, it can handle a wider range of data that a BP neural network cannot handle and still maintains a high level of prediction accuracy. The prediction accuracy mainly depends on the selection of penalty factor C and kernel function parameter γ. The value of penalty factor C is mainly determined by the type of predicted data. For each misclassified point, C is used as the dimension for punishment. When the C is larger, the number of misclassified points is less, and the fitting effect is

improved. However, infinitely expanding C can result in overfitting. The Radial Basis Function (RBF) is usually selected as the kernel function of the carbon emission prediction model based on a support vector machine, and its overall prediction effect is better than the prediction model using other kernel functions. Equation (5) is the basic expression form of the RBF kernel function. The RBF function comes with the super parameter γ, which is generally small, and similar to the role of penalty factor C, it is used to adjust the degree of fitting. So, C and γ are usually adjusted simultaneously to find the optimal parameter value. The initial range is $0.0001 < \gamma < 10$ and $0.1 < C < 100$. The optimal parameters can be achieved through manual debugging or using corresponding tuning algorithms. At present, most tuning parameters depend on relevant optimization algorithms, such as the improved chicken swarm optimization algorithm (ICSO), sparrow search algorithm (SSA), firefly optimization algorithm (FFA), etc. Through a literature search, it was found that the prediction accuracy of the carbon emission prediction model based on a SVM was generally controlled between 0.90 and 0.97:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (5)$$

where x_i and x_j are the input vectors calculated from the data set; γ is the kernel function parameter of RFB, which can be generally expressed as $1/2 * \sigma^2$; and σ is the width parameter of the function.

Table 3. Research on carbon emission prediction model based on support vector machine (source: Web of Science).

Reference	Data Sources	Prediction Method Selection	Prediction Accuracy
[26]	The data set was derived from the database provided by the World Bank for EU countries. The input variables included rural population growth, rural population, urban population growth, urban population and total population growth. The output variables included carbon emissions from three types of sources: gases, solids and liquids.	The FFA-SVM model depends on three parameters of C, ε and γ that have a great influence on the prediction accuracy, and this model is optimized by firefly optimization algorithm.	The coefficients of determination of the predicted carbon emissions of gas, liquid and solids were 0.9015, 0.9348 and 0.9261, respectively
[18]	The data set was derived from the data related to Shanghai from 2000 to 2016 in the official statistical yearbook. Firstly, the grey correlation method was used to analyze the correlation between the influencing factors and carbon emissions, which resulted in the selection of 18 preliminary indicators. Then, four principal components (economic level, living standard, social condition and education level) were extracted via a principal component analysis as input indicators. Carbon emissions were calculated from the corresponding energy data and were used as output indicators.	The ICSO-SVM model uses the improved chicken swarm optimization algorithm to optimize the weight and threshold of the SVM. The minimum fitness was 0.0441, the corresponding penalty factor C was 77.9690 and the kernel function parameter γ was 0.1013.	MAPE = 1.21%, RMSE = 0.4346

Table 3. *Cont.*

Reference	Data Sources	Prediction Method Selection	Prediction Accuracy
[27]	The data were derived from statistics released by Henan Province covering the years 1990 to 2015. The influencing factors considered included the total population of Henan Province, the proportion of the secondary industry, the ratio of coal consumption, the urbanization rate and GDP. Carbon emissions were calculated based on the energy consumption data.	The SVM model. The best optimized parameter penalty factor C was 9.18959 and kernel function parameter γ was 0.0358968.	The coefficients of determination for the training and test sets were 0.9952 and 0.9543, respectively.
[28]	The data used in this study were obtained from the 'China Statistical Yearbook,' 'China Construction Industry Statistical Yearbook' and 'China Energy Statistical Yearbook' from 1995 to 2016. Based on a grey correlation analysis, the completion area of the construction projects, GDP, total output value of the construction industry, labor productivity of the construction industry, number of employees in construction enterprises and primary energy consumption in the construction industry were selected as the main influencing factors of carbon emissions in the construction industry.	The FCS-SVM model uses the fuzzy cuckoo search algorithm to determine its optimal parameters.	MAPE = 0.003 R^2 = 0.9581
[29]	Through a Granger causality analysis, the influencing factors of carbon dioxide emissions related to primary industry, secondary industry, tertiary industry and residential consumption were tested. Finally, four categories of carbon dioxide emissions and the corresponding influencing factors were selected for a predictive analysis.	The LSSVM model. The final optimized regularization parameter C and kernel parameter σ^2 were 152.469 and 0.027, respectively.	MAPE = 0.16% MaxAPE = 0.204% MDAPE = 0.189% RMSE = 0.001

SVMs have a rigorous theoretical and mathematical foundation. They boast a stronger generalization ability and are capable of achieving global optimal solutions, and they also perform well in small-sample prediction. In the applications discussed above, researchers have mainly focused on optimizing the SVM algorithm to improve the accuracy and efficiency of the prediction model, while a minority of scholars have used combined models to meet the accuracy improvement requirements.

3.3. Prediction Model of Carbon Emission—LSTM

3.3.1. Introduction

LSTM is a prediction tool based on time series. It is an advanced algorithm model of RNN, which objectively makes up for the shortcomings of RNN in some aspects. With the ability to predict relevant data for the next time period through a period of existing data, LSTM has achieved good application results. Its selective memory function enables it to handle data with a large sample size without the problem of RNN gradient disappearance or gradient explosion.

In the structure diagram of long short-term memory neural networks (Figure 7), the network is divided into the input gate, forget gate and output gate. The forget gate automatically filters the information from the previous time period and selectively forgets irrelevant information, retaining only the useful information. This operation significantly

reduces the amount of information and thereby reduces the amount of computation needed. The input gate combines the newly obtained information with the original information, and the output gate processes the existing information and outputs the corresponding results. The key to this structure is the forget gate, which distinguishes LSTM from RNN. The forget gate allows the LSTM network to adjust the weights to stabilize the value obtained after backpropagation in the normal range. This feature is crucial for ensuring the stability of the network and preventing issues such as gradient vanishing or exploding.

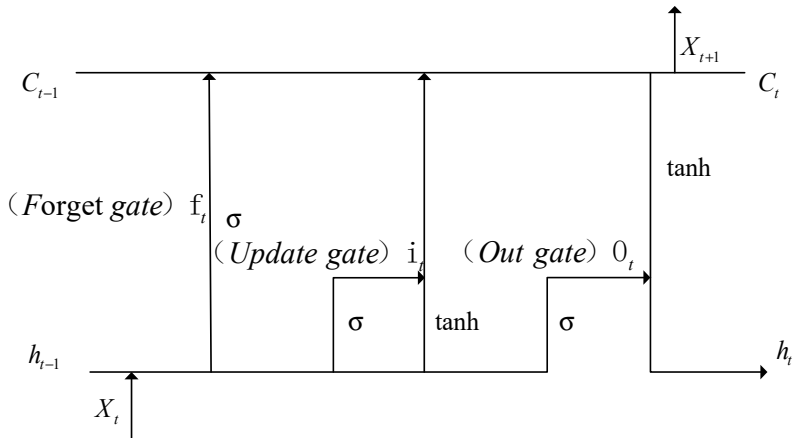

Figure 7. Structure diagram of long short-term memory. C_{t-1}, C_t is the cell state vector; h_{t-1}, h_t is the hidden state vector; X_t, X_{t+1} is the input vector; σ is the sigmoid function. Own elaboration.

3.3.2. Application

The good performance of LSTM when dealing with gradient disappearance and a gradient explosion will cause it to gradually replace the original recurrent neural network when predicting time series data. The most classic application is the prediction of stock trends, and of course, it has shown promising results in many other time-series-related aspects. In recent years, researchers have used LSTM to predict carbon emissions with good success.

As outlined in Table 4, the optimization of the carbon emission prediction model based on LSTM mainly focuses on data mining. As it predicts time series data, it can rely on historical carbon emission data to make predictions. Generally, carbon emission time series data are predicted by analyzing the dates of several time series before the target time series, but such prediction results are often less convincing. Therefore, scholars generally combine relevant influencing factors or other optimization algorithms to improve the prediction effect of the LSTM model. Since LSTM is a machine learning algorithm that has emerged in recent years, the method of optimizing itself is still relatively rare. Hence, the prediction accuracy of LSTM mainly depends on the integrity and accuracy of the data, and the data should conform to the timing rules. The commonly used data analysis method of the carbon emission prediction model based on LSTM is mainly used to analyze the influence degree of the carbon emission influencing factors. Similar to the data processing method of most prediction methods, the original data predicted by LSTM needs to be simplified to reduce data redundancy. The verification of the prediction accuracy by LSTM is generally conducted by other prediction algorithms as comparison objects. However, the prediction methods used for comparison are more commonly used, which are not targeted and cannot explain the actual utility of the prediction.

The original data on the carbon emissions were obtained by converting the relevant statistical data, and the different calculation standards of different researchers also led to different data being obtained, which also caused corresponding differences in their prediction results. Therefore, the previous improvement in carbon emission prediction based on LSTM mainly focuses on data processing and integration.

Table 4. Research on carbon emission prediction model based on long short-term memory (source: Web of Science).

Reference	Data Sources	Prediction Method Selection	Prediction Accuracy
[30]	By referring to the existing literature and STIRPAT model, six influencing factors were determined as input variables: urbanization, per capita GDP, industrial structure, energy consumption, energy intensity and population density. The research objects were 30 provinces in China. The relevant data were derived from the 'China Statistical Yearbook' and the 'China Energy Statistical Yearbook'. The carbon emission data were derived from the energy consumption data combined with the 'IPCC National Greenhouse Gas Inventories Guidelines (IPCC, 2006)'.	The LSTM-STIRPAT model.	MAPE of carbon emission projection for provinces is between 2.1% and 6.6%
[31]	The data used in this study were official statistical data. The researchers employed a quadratic assignment process regression analysis (QAP) to analyze the influencing factors of carbon emissions in the Yellow River basin from the perspective of regional differences. The analysis led to the identification of five factors that affect carbon emissions: population, GDP, industrial structure, urbanization rate and energy intensity. The estimation method used for the carbon emissions is in line with the recommendations of the Intergovernmental Panel on Climate Change (IPCC).	The SSA-LSTM model.	MAE = 30.90, RMSE = 36.67, MAPE = 0.0099
[32]	The data were sourced from the 'China Statistical Yearbook', while the carbon emission data pertained to the carbon emission values of China from 1965 to 2014, as released by the World Bank.	The KLS model.	MSE = 0.0039, MAE = 0.06, MAXE = −0.089
[33]	The object of prediction was the carbon emissions of the port of Los Angeles. By referring to the STIRPAT model, researchers have selected relevant indicators of carbon emissions, including port throughput, carbon emission intensity and historical carbon emission data.	The STIRPAT-ARIMAX-LSTM integrated model.	RMSE = 0.0145, MAPE = 7.9306, MDA = 0.685
[34]	The AIS data of LNG carriers provided by exactEarth were optimized by using the cubic spline interpolation method. The optimized AIS data were combined with the program recommended by the International Towed Tank Conference (ITTC) to calculate the carbon dioxide emission of the ship.	The LSTM model.	The difference between the actual and predicted values of total carbon dioxide emissions is 0.022.

3.4. Prediction Model of Carbon Emission—RF

3.4.1. Introduction

Random forest is an advanced extension of decision trees that utilizes multiple decision trees to make more accurate predictions. Each decision tree in the random forest uses its own decision criteria to make a prediction. The final decision is made by summarizing

the predictions of all the decision trees in the forest. Each individual decision tree in a random forest is generally divided into two types. On the one hand, there is data selection. Generally, each decision tree will randomly select different sample sets from the total samples for operation, so each decision tree has different characteristics. On the other hand, there is feature selection. The efficiency and accuracy of the operation can be improved through random sampling of the features in the feature set.

As shown in Figure 8, each decision tree in a random forest runs independently. Because the data set or feature set of each decision tree in the random forest is smaller, its operational efficiency is faster than that of a decision tree with a complete sample set and complete feature set. Additionally, each decision tree has its own unique characteristics. During the integration stage, the results of each decision tree are combined to determine the final outcome. The structure diagram of a random forest reveals that it can simplify complex problems by analyzing the characteristics of multidimensional data through each decision tree. Finally, the final results are determined through comprehensive comparison and screening, which results in a significantly improved accuracy and operational speed with a high degree of interpretability. However, the results can be too general to identify subtle changes or accurately detect abnormal points in the data.

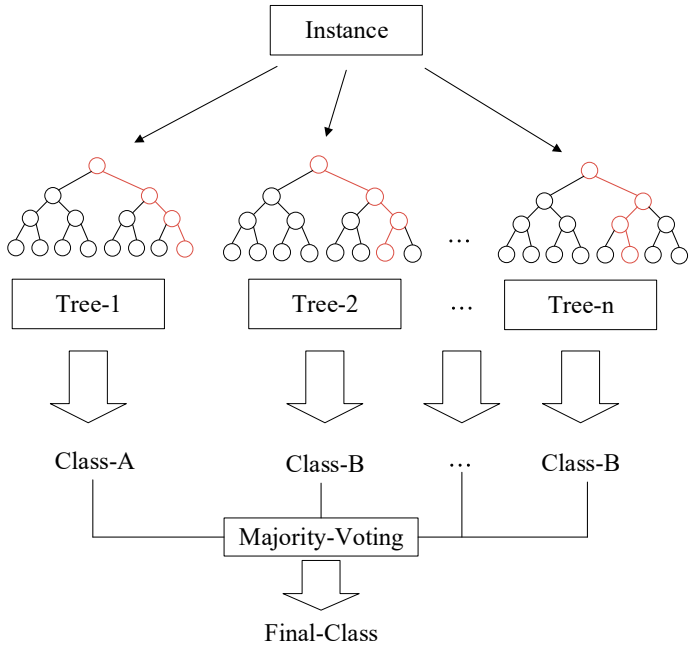

Figure 8. Structure diagram of random forest. Own elaboration.

3.4.2. Application

From the existing research (Table 5), the carbon emission prediction model based on random forest is widely used at the level of single buildings. Such refined predictions involve many influencing factors, and the random forest algorithm is particularly suitable for processing multidimensional data with high efficiency. However, when compared with other prediction methods, its accuracy is slightly lower. Although its accuracy is not as high as some existing high-precision models, its fast operation speed is particularly important in the era of big data, which makes it highly promising for broad applications.

Furthermore, the random forest algorithm performs well when it comes to screening influencing factors. Its multidecision tree structure enables the determination of the overall feature contributions by integrating the contribution of each decision tree classification feature. This makes it effective at identifying several types of influencing factors that have a greater impact on carbon emissions and can serve the prediction model well.

Table 5. Carbon emission prediction model based on random forest (source: Web of Science).

Reference	Data Sources	Prediction Method Selection	Prediction Accuracy
[35]	The data used were obtained from the Open Data Inventory of Anthropogenic Carbon Dioxide (ODIAC) of the China Environmental Research Institute, which provided a global dataset of carbon dioxide emissions resulting from fossil fuel combustion. The three-dimensional building information was obtained from the Gaode map. A Pearson's correlation test was used to examine the relationship between spatial factors and carbon emissions, with relevant factors being used as input variables and the corresponding carbon dioxide emissions being used as the output variable.	The RF model. The baseline model, which considers the architectural structure factors of previous studies, was compared with the improved model, which took into account both the previous research and potential architectural structure factors. The results showed that the improved model had a higher prediction efficiency.	R^2 = 0.9392, RSE = 32.53%, RAE = 28.50%
[36]	The data were derived from 38 buildings located in the Pearl River Delta region of China. After analyzing and screening the influencing factors of carbon emissions during the construction stage in the previous literature, the final factors were determined to be the foundation area, ground area, underground area, building height, number of floors on the ground and basement depth. The carbon emissions were calculated by using the quota method.	The RF model. The minimum number of nodes (nodesize) in each tree was set to 5. The optimal number of decision trees (ntree) was determined to be 124, and the number of randomly selected variables (mtry) for growing each tree was set to 5.	R^2 = 0.6403, MSE = 0.7649

3.5. Prediction Model of Carbon Emission—ELM

3.5.1. Introduction

The extreme learning machine (ELM) was introduced in 2004 by Guang-Bin Huang, Qin-Yu Zhu and Chee-Kheong Siew from Nanyang Technological University. Their paper was published in the IEEE International Joint Conference [37]. Their aim was to enhance the Backward Propagation algorithm to improve the learning efficiency and enact a simpler learning parameter setup. The structure diagram of the extreme learning machine (Figure 9) is distinct in that the weights of the hidden layer nodes are randomly or artificially assigned and remain fixed without any updates. During the learning process, only the output weights were computed and optimized. The optimization objective is expressed in Equation (6), and the optimal solution is attained through the least-squares method [38]:

$$\text{argmin}\|H\beta - T\|^2 \tag{6}$$

where H is the output matrix, T is the training target, $||$ is the Frobenius norm of the matrix elements and β is the output weight.

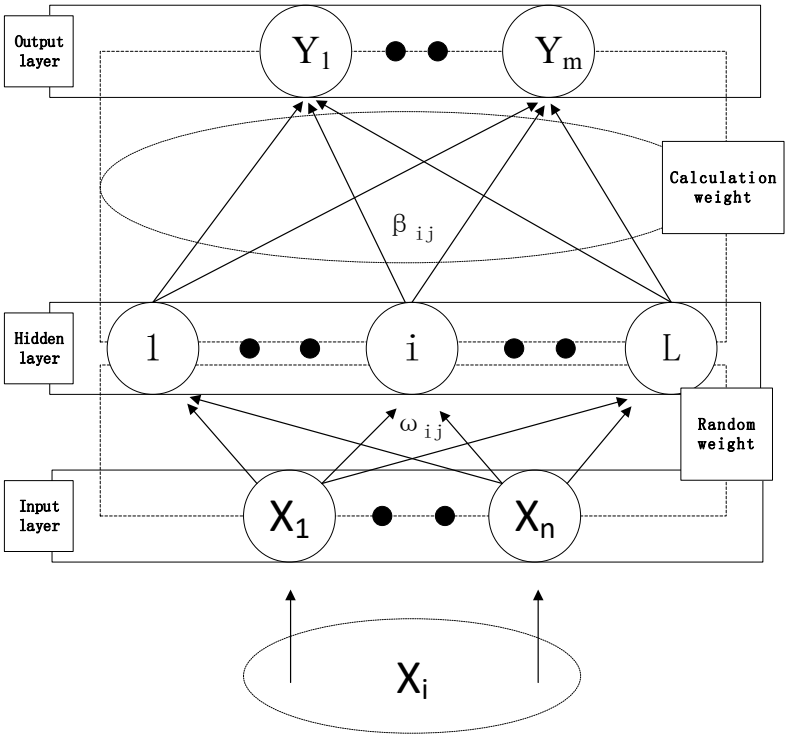

Figure 9. Structure diagram of extreme learning machine. $X_1, \cdots, X_n$ is the date of input layer; $Y_1, \cdots, Y_m$ is the date of output layer; $1, \cdots, i, \cdots, L$ is the date of hidden layer; $\omega_{ij}(i = 1, \cdots, n, j = 1, \cdots, L)$ is the connection weight between the input layer and the hidden layer; $\beta_{ij}(i = 1, \cdots, L, j = 1, \cdots, m)$ denotes the connection weight between the hidden layer and the output layer. Own elaboration.

3.5.2. Application

ELMs have numerous applications in computer vision and bioinformatics and are also widely used to solve regression problems in Earth science and environmental science [39]. In the field of image processing, ELMs have proven to be effective at transforming low-resolution images into high-resolution ones and identifying surface types in remote sensing images. In the biological sciences, ELMs are utilized to predict protein interactions. Due to their ability to generalize, ELMs have been employed to address prediction problems involving nonlinear processes and insufficient observational data in Earth sciences. Successful examples include predicting daily river runoff, wind speed, the drought index and carbon emissions.

As shown in Table 6, data processing for the extreme learning machine typically involves factor analysis. Due to the large size of the data processed by the extreme learning machine, manually selected factors may exhibit some correlation, which could introduce some deviations into the prediction results. Therefore, the data are generally classified using factor analysis methods to reduce the correlation between the factors. This approach can also simplify the input structure and improve the operational efficiency. In the algorithm layer, the extreme learning machine is generally determined as the optimal algorithm by comparing it with other prediction algorithms. In addition, in the process of comparison, the prediction model is also optimized by optimization algorithms such as the genetic algorithm and particle swarm optimization algorithm to improve the overall prediction accuracy.

In terms of the predictive accuracy, the current carbon emission prediction model based on ELM achieved an average absolute percentage error and root mean square error of approximately 1%, which demonstrated a reliable performance that ranked in the upper-middle level among the other prediction models.

Table 6. Research on carbon emission prediction model based on extreme learning machine (source: Web of Science).

Reference	Data Sources	Prediction Method Selection	Prediction Accuracy
[40]	The information was obtained from relevant data in the China Energy Statistics Yearbook and China Statistical Yearbook covering the period from 2000 to 2017. Thirteen factors deemed to affect traffic carbon dioxide emissions were subjectively selected. Through the factor analysis, the four main factors of economic development, traffic structure, energy consumption structure and population size were identified as the input variables. Nine types of energy consumption, including coal, gasoline, kerosene, fuel oil, crude oil, liquefied petroleum gas, natural gas and electricity, were selected to calculate the corresponding traffic carbon dioxide emissions.	The GA-ELM model. BPNN model, ELM model, GA-BP model and GA-ELM model were used for prediction, and the predicted values of each algorithm were compared with the actual values. The results showed that the fitting degree between the predicted values and the real values of the GA-ELM model was the highest.	MAPE = 1.4594%, RMSE = 35.356
[41]	All the data on the carbon-emission-influencing factors from 2000 to 2017 were derived from the 'China Statistical Yearbook (2018)', and the corresponding carbon emission data was obtained from the IEA (2019) report. The average influence value method was used to analyze the degree of influence of 13 preliminary indicators. After adjusting the size of the indicators in the same proportion, the MRFO-ELM prediction model was input.	The MRFO-ELM model. The optimal fitness of MRFO-ELM decreased from 0.0583 to 0.0186 in 20 iterations, which had a faster optimization speed and higher accuracy.	MAPE = 1.25%, RMSE = 11.34%
[20]	This paper selected China's daily carbon emission data from 1 January 2019 to 18 June 2021 as a sample set for empirical analysis. The data were derived from "https://carbonmonitor.org.cn/downloads/ (accessed on 28 June 2021)". The daily carbon emission data were decomposed via ensemble empirical mode decomposition, and then the possible input variables were screened by using a partial autocorrelation coefficient method. Finally, the ReliefF algorithm was used to determine the five highest weight sequences as the final input variables.	The ISSA-ELM model. Through a comparison of the results obtained by using various models including BP, PSO-BP, ELM, PSO-ELM, SSA-ELM, ISSA-ELM, EMD-ISSA-ELM, EEMD-ISSA-ELM, ICEEMDAN-ISSA-ELM and ISSA-ELM, it was demonstrated that the ISSA-ELM model was the most appropriate for predicting daily carbon emissions.	R^2 = 0.9968, MAPE = 0.0040, RMSE = 174.759

Table 6. *Cont.*

Reference	Data Sources	Prediction Method Selection	Prediction Accuracy
[15]	Based on the data from the China Statistical Yearbook and Hebei Economic Yearbook, this paper analyzed the carbon dioxide emissions and its influencing factors in Hebei Province from 1995 to 2014. In view of the factors affecting CO_2 emissions in Hebei Province, considering the economy, energy and other aspects, 22 indicators were preselected, and 6 categories (actually 8 categories) were obtained by factor analysis as the input of the prediction model, which were output value factor F11, export factor F12, energy consumption factor F2, general investment factor F31, pollution investment factor F32, consumption factor F4, city factor F5 and traffic factor F6. According to the coefficients published in the 'National Greenhouse Gas Inventories Guide (IPCC, 2006)', the CO_2 emissions were calculated per unit of different energy consumption types converted to standard coal.	The PSO-ELM model. The PSO-ELM model has 20 hidden layer nodes. Compared with the forecasting results of the BPNN and ELM, the proposed PSO-ELM model had the best fitting curve.	RMSE = 91.63 (million tons), MAPE = 0.31%

4. Applicable Characteristics of Common Intelligent Prediction Models

In the field of machine learning, BP neural networks were early algorithms. After continuous development, they have been applied in many areas, including function approximation, pattern recognition, classification and data compression. In the context of carbon emission prediction, the BP neural network is primarily used for function approximation. However, it still has some drawbacks, such as a slow learning speed, a tendency to fall into a local minimum and the absence of a theoretical basis for the number of neurons. To some extent, the support vector machine can make up for the shortcomings of the BP neural network. It has the advantages of having a faster operation speed and wider data types while retaining the strong nonlinear processing ability of the BP neural network. Moreover, since the prediction process requires the feature vector to be filtered, the data-size requirement was not particularly significant, and an SVM can provide better prediction results regardless of whether the data are small or large. LSTM is a recurrent neural network, which solves the problem of gradient disappearance by increasing weight, and the existence of a forgetting gate can make it more efficient at processing data. It performs well when the data have time characteristics, which makes it particularly useful with carbon emission data, which typically exhibit temporal characteristics. Random forest is a combination of multiple decision trees that randomly select data samples or features, which makes it useful to identify the feature weight. This approach can be used to analyze the factors affecting carbon emission or as a prediction algorithm for carbon emissions. ELMs use a random mapping method to convert the input layer to the hidden layer, which makes it distinct from the other four algorithms. The weight from the hidden layer to the output layer is then calculated, and the prediction effect is mainly due to its unique structure. Although the results may be less interpretable, various studies have shown that the prediction effect is not inferior to or even better than other prediction models.

The data structure of the prediction model consists of independent and dependent variables, with the dependent variable relying on data from multiple dimensions. When the research focuses on the optimization of the algorithm, the independent variables are typically based on the traditional carbon emission prediction model. The causal index of the data layer is established by the influencing factor information of the traditional model, and then the optimization of the algorithm is used to improve the prediction accuracy. When focusing on the data layer, each traditional model can serve as a reference point. By analyzing the factors used in various traditional models, factor selection and data optimization can be combined to improve the data layer. Additionally, some studies determine the appropriate method by selecting the best model for different types of data to achieve the desired prediction effect. Table 7 shows the data types that have a good application effect for the five commonly used prediction models.

Table 7. Comparison of intelligent prediction models. Own elaboration.

	BPNN	**SVM**	**LSTM**	**RF**	**ELM**
Data Type	Medium-size samples and high-precision data	Large and small sample data	Large sample data based on time series	Complex multidimensional data	Large sample data with clear causal relationship
Data Sources	At present, the most common data sources are official statistical data, and some data sources are research and investigations, data integration, internal supply of the industry, data monitoring websites, etc.				
Prediction Accuracy	R^2 is generally around 0.95, and MSE is generally less than 0.8.	R^2 is between 0.90 and 0.95, RMSE < 0.5.	MAPE is between 2% and 3%.	The prediction accuracy varies greatly, and the overall prediction performance is poor.	MAPE is between 0.3% and 1.5%.

All prediction models share many commonalities at the data-processing level, such as conducting a correlation analysis between factors and an analysis of the degree of influence between factors and carbon emissions, to determine the final input parameters. Regarding the input parameters, good prediction results can only be obtained when the parameters fall within a certain influence range and are suitable for the model, and the choice of data analysis methods will differ. Typically, available combinations of influencing factors are screened, with higher-influence factors being selected. When factors have a large correlation, a factor analysis is used to reduce the impact of the correlation on the prediction results. Table 8 summarizes the main optimization methods used at the data layer.

At the algorithm optimization level, the focus is mainly on optimizing each variable that determines the final prediction result, such as the weight and threshold of the algorithm. Generally, SVMs have a high degree of adaptability to various types of data. From its structure, it can be observed that it predicts by screening feature vectors, which indicates that there are many choices in the process of feature vector screening, and the optimization of the weight and threshold after feature vector selection is also a crucial aspect. Therefore, research in the field of carbon emission prediction models based on SVMs is extensive, while research on carbon emission prediction models based on BPNNs, ELMs and LSTM is mainly focused on data mining and processing. Currently, there is limited reference for carbon emission models based on random forest in both the data and algorithm layers. However, the existing research suggests that a random forest can achieve a high prediction accuracy and has a significant advantage in processing speed, which makes it a worthwhile topic of discussion. Table 9 summarizes the primary optimization methods used in related algorithms.

Table 8. Optimization method of data layer (source: Web of Science).

BPNN	Random forest [42], Rough set theory [43], empirical mode decomposition [23,44], principal component analysis [45], Mean impact value method [46], Grey relational analysis [16,47,48], Bivariate correlation analysis [49], Generalized fisher index decomposition [50], Influence coefficient method [22]
SVM	Principal component analysis [19,29], Grey relational analysis [18,51,52], factor analysis [53], Detection of steady state [54], Cointegration test [54], Granger causality test [54–56], random forest [57], Impulse response function [58], Variance decomposition [58], Bivariate correlation analysis [19,29], Copula function [59], ridge regression analysis [32]
LSTM	Principal component analysis [33,60], Grey relational analysis [60], ensemble empirical mode decomposition [61], variational mode decomposition [61], multiple linear regression [33], regression analysis of quadratic assignment process [31], ARDL boundary test [17]
RF	Generalized additive models [62], Cluster analysis [63], random forest [63], Pearson test [35]
ELM	Bivariate correlation analysis [8,64], factor analysis [8,64,65], Linear analysis [64,65], random forest [66], Mean impact value [41,67], ensemble empirical mode decomposition method [20], partial autocorrelation function [20,68], ReliefF algorithm [20], Logarithmic average division [67], principal component analysis [68], Pearson test [68]

Table 9. Optimization method of algorithm layer (source: Web of Science).

BPNN	Particle swarm optimization algorithm [22,42,48], genetic algorithm [16,69], Improved particle swarm optimization algorithm based on noninertial weight coefficient [49,50,70,71]
SVM	Particle swarm optimization algorithm [29,72–74], Firefly Algorithm [26], FCS Algorithm [28], chicken swarm optimization algorithm [19], Fruit Fly Algorithm [53], Lion Optimizer [75], genetic algorithm [55,75], Grey Wolf Optimizer [76], Shuffled Frog Leaping Algorithm [51], Ocean Predator Algorithm [77], Bacterial Foraging Optimization Algorithm [52], Whale Optimization Algorithm [78], sparrow search algorithm [57], Gaussian perturbation bat algorithm [54], Butterfly Optimization Algorithm [19], Salp Swarm Algorithm [79]
LSTM	Bilstm [80], Attention-LSTM [81], sparrow search algorithm [31]
RF	The performance mainly depends on the initial settings of the model parameters, such as the values of nodesize, ntree and mtry [35,36,63,66]. At present, there is no other optimization method for the random forest algorithm.
ELM	Particle swarm optimization algorithm [15], genetic algorithm [64,65], Flame optimization [66], Manta foraging optimization [41], sparrow search algorithm [20], Gaussian perturbation bat algorithm [67]

The primary evaluation criteria for the accuracy of each carbon emission prediction model are the variance (R^2), mean square error (RMSE) and mean absolute percentage error (MAPE), as shown in Equations (7)–(9):

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

where m is the number of actual (predicted) values, y_i is the ith actual value and $\hat{y}_i$ is the ith predicted value.

$$R^2 = \frac{1}{n}\sum_{1}^{n}(x_i - \overline{x})^2 \quad (8)$$

where n is the number of actual (predicted) values, x_i is the ith predicted value and $\overline{x}$ is the average value of the predicted values.

$$MAPE = \frac{100\%}{n}\sum_{1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (9)$$

where n is the number of actual (predicted) values, y_i is the ith actual value and $\hat{y}_i$ is the ith predicted value.

5. Summary

1. In comparison to traditional carbon emission prediction methods, intelligent prediction methods offer significant advantages in terms of efficiency and accuracy, particularly when dealing with large and complex data samples. Perhaps the most prominent benefit is that it can self-learn through vast amounts of data, which reduces the workload of humans and yields even better prediction results than traditional models. Numerous scholars have employed machine learning techniques such as BPNN, SVM, LSTM, RF and ELM for carbon emission prediction applications. While a few have used other methods such as decision trees, linear regression and Markov models to achieve their prediction goals, the five categories of the machine learning predictive models discussed remain the most widely used approaches in the field of carbon emission prediction.
2. Whether it is a conventional or intelligent model, the data it analyzes are basically derived from official statistical reports. The corresponding carbon emissions are computed through energy consumption data by using the IPCC carbon emission calculation method. The input variable represents the influencing factors of carbon emissions, while the output variable is the actual amount of carbon emissions.
3. In terms of prediction accuracy, the SVM- and ELM-based carbon emission models demonstrate the better performance with a lower mean square error or average absolute percentage error than other algorithms. Following closely are the BPNN and LSTM models. At present, the reference research on carbon emission prediction models based on random forest is not very sufficient, but as shown by the existing research, its accuracy can also reach a relatively high level, but its overall prediction level is uneven, so we think that the carbon emission prediction model based on random forest still has a lot of room for development.
4. Researchers have made significant progress in improving carbon emission prediction models by using machine learning. Currently, researchers are primarily focused on three major areas in their ideas for developing these models: collecting and calculating carbon emission data, processing and optimizing carbon emission data and selecting and improving model algorithms (Figure 10).

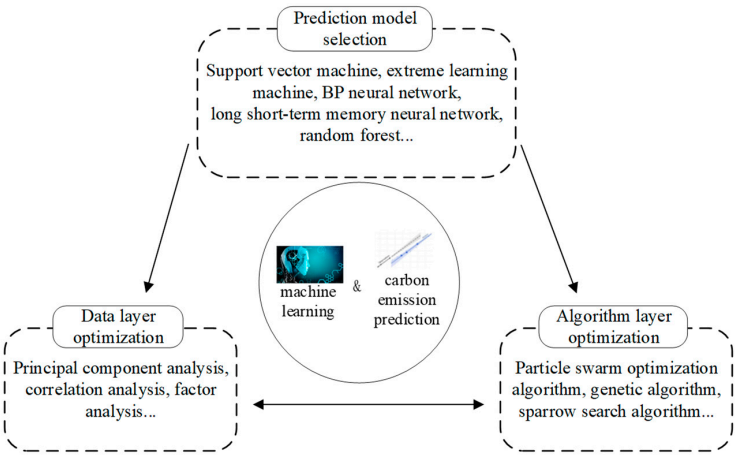

Figure 10. Optimal path of neural network prediction model. Own elaboration.

6. Prospect

Currently, most data collection methods rely on official releases. However, with the development of smart cities, more regions are implementing their own energy consumption monitoring systems. The data obtained from these systems are more accurate and reliable than the data collected from human statistics. Researchers in areas with well-established monitoring systems can extract information from these systems to perform corresponding carbon emission calculations, which can improve the accuracy of the data. Combining official statistics with monitoring systems can further enhance the accuracy of the data. However, in cities and regions without widely available monitoring systems, researchers still primarily rely on official and relevant organization survey data.

Based on post-2020 research, the number of publications related to LSTM- and SVM-based carbon emission prediction models surpasses that of other prediction models. From an application standpoint, SVMs and ELMs show a more prominent performance, with overall better results compared to other methods. This phenomenon may be due to LSTM's natural suitability for time-series data in other fields, and since most of the data related to carbon emissions are also time-series data, researchers tend to prioritize LSTM, but the results may not always meet expectations. SVMs can adapt well to large amounts of data such as data on carbon emissions, while the effect of ELMs may be ignored by researchers. The principle behind it is worth studying.

However, the development of machine learning is quite rapid, and the elimination rate is also very high. While studying the field of carbon emission predictions, researchers should also pay attention to the relevant developments in the field of machine learning, including the latest improvement and optimization methods of these five types of algorithms and the emerging prediction methods that perform well in other industries.

Generally, machine-learning-based carbon emission prediction models offer significant advantages in terms of efficiency and accuracy over traditional artificial-theory-based models. However, due to their inherent complexity, these models are not easily modifiable, which can result in certain drawbacks, such as the limited interpretability of their results. To overcome this, future carbon emission prediction models based on machine learning should integrate traditional theories, not just by borrowing some general machine learning prediction algorithms, but also by incorporating domain knowledge. By doing so, these models can improve both their prediction accuracy and efficiency, while also making their predicted results more interpretable.

Moreover, recent studies have shown that many researchers are utilizing the latest algorithm called "transfer learning" from the field of artificial intelligence to predict energy consumption [82–84]. However, there is limited research on applying transfer learning to carbon emission prediction, despite the high correlation between carbon emission and energy consumption. Hence, it is highly likely that the development of transfer learning in the field of carbon emission prediction will become a growing trend. Transfer learning involves the development of a new model by using the model developed by task A as the initial point of task B (Figure 11). Currently, the data available for carbon emission prediction are huge and extremely complex, with only a few regions having popularized carbon emission monitoring systems. Most regions lack sufficient data to develop carbon emission prediction models, and the application of transfer learning can greatly reduce the amount of data required. Although many carbon emission prediction models exist, there are still limited studies on the mutual reference and optimization of these models. In the future, transfer learning is expected to play a significant role in the integration of carbon emission models.

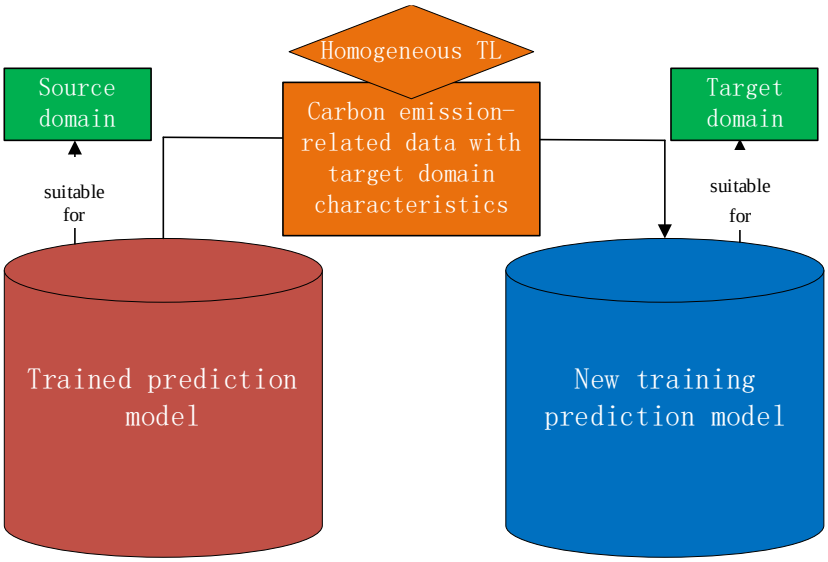

Figure 11. Transfer learning process of prediction model. Own elaboration.

Author Contributions: Conceptualizing, review and editing, Y.Z.; Data collection, writing, drafting-Original draft, R.L.; Supervision, review and funding acquisition, Z.L.; Supervision, review and editing, L.L.; Supervision, J.W.; Data collection, W.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Seneviratne, S.I.; Donat, M.G.; Pitman, A.J.; Knutti, R.; Wilby, R.L. Allowable CO_2 emissions based on regional and impact-related climate targets. *Nature* **2016**, *529*, 477–483. [CrossRef]
2. Reichstein, M.; Bahn, M.; Ciais, P.; Frank, D.; Mahecha, M.D.; Seneviratne, S.I.; Zscheischler, J.; Beer, C.; Buchmann, N.; Frank, D.C.; et al. Climate extremes and the carbon cycle. *Nature* **2013**, *500*, 287–295. [CrossRef]
3. Kelley, C.P.; Mohtadi, S.; Cane, M.A.; Seager, R.; Kushnir, Y. Climate change in the Fertile Crescent and implications of the recent Syrian drought. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 3241–3246. [CrossRef] [PubMed]
4. Orlowsky, B.; Hoekstra, A.Y.; Gudmundsson, L.; Seneviratne, S.I. Today's virtual water consumption and trade under future water scarcity. *Environ. Res. Lett.* **2014**, *9*, 074007. [CrossRef]
5. Hunt, A.S.P.; Wilby, R.L.; Dale, N.; Sura, K.; Watkiss, P. Embodied water imports to the UK under climate change. *Clim. Res.* **2014**, *59*, 89–101. [CrossRef]
6. Zhang, C.; Su, B.; Zhou, K.; Yang, S. Decomposition analysis of China's CO_2 emissions (2000–2016) and scenario analysis of its carbon intensity targets in 2020 and 2030. *Sci. Total Environ.* **2019**, *668*, 432–442. [CrossRef]
7. Song, Y.; Zhang, Y.; Zhang, Y. Economic and environmental influences of resource tax: Firm-level evidence from China. *Resour. Policy* **2022**, *77*, 102751. [CrossRef]
8. Sun, W.; Cai, J.; Mao, H.; Guan, D. Change in Carbon Dioxide (CO_2) Emissions from Energy Use in China's Iron and Steel Industry. *J. Iron Steel Res. Int.* **2011**, *18*, 31–36. [CrossRef]
9. Thepkhun, P.; Limmeechokchai, B.; Fujimori, S.; Masui, T.; Shrestha, R.M. Thailand's Low-Carbon Scenario 2050: The AIM/CGE analyses of CO_2 mitigation measures. *Energy Policy* **2013**, *62*, 561–572. [CrossRef]
10. Wang, Q.; Wang, Y.; Hang, Y.; Zhou, P. An improved production-theoretical approach to decomposing carbon dioxide emissions. *J. Environ. Manag.* **2019**, *252*, 109577. [CrossRef]
11. Ang, B.W. The LMDI approach to decomposition analysis: A practical guide. *Energy Policy* **2005**, *33*, 867–871. [CrossRef]
12. Song, W. Research on Carbon Emission Prediction Model ofConstruction Industry Based on Machine Learning. Master's Thesis, Xi'an University of Architecture and Technology, Xi'an, China, 2020.
13. Chen, C.; Song, M. Visualizing a field of research: A methodology of systematic scientometric reviews. *PLoS ONE* **2019**, *14*, e0223994. [CrossRef] [PubMed]

14. Xi, J. An important speech at the Climate Ambition Summit by Xi Jinping. *World Aff.* **2021**, *6*.
15. Sun, W.; Wang, C.; Zhang, C. Factor analysis and forecasting of CO_2 emissions in Hebei, using extreme learning machine based on particle swarm optimization. *J. Clean. Prod.* **2017**, *162*, 1095–1101. [CrossRef]
16. Zhang, S.; Huo, Z.; Zhai, C. Building Carbon Emission Scenario Prediction Using STIRPAT and GA-BP Neural Network Model. *Sustainability* **2022**, *14*, 9369. [CrossRef]
17. Ameyaw, B.; Yao, L.; Oppong, A.; Agyeman, J.K. Investigating, forecasting and proposing emission mitigation pathways for CO2 emissions from fossil fuel combustion only: A case study of selected countries. *Energy Policy* **2019**, *130*, 7–21. [CrossRef]
18. Wen, L.; Cao, Y. Influencing factors analysis and forecasting of residential energy-related CO_2 emissions utilizing optimized support vector machine. *J. Clean. Prod.* **2020**, *250*, 119492. [CrossRef]
19. Wen, L.; Cao, Y. A hybrid intelligent predicting model for exploring household CO_2 emissions mitigation strategies derived from butterfly optimization algorithm. *Sci. Total Environ.* **2020**, *727*, 138572. [CrossRef] [PubMed]
20. Kong, F.; Song, J.; Yang, Z. A daily carbon emission prediction model combining two-stage feature selection and optimized extreme learning machine. *Environ. Sci. Pollut. Res.* **2022**, *29*, 87983–87997. [CrossRef]
21. Hu, Z.; Gong, X.; Liu, H. Prediction of household consumption carbon emission in western cities Based on BP model: Case of Xi'an city. *J. Arid. Land Resour. Environ.* **2020**, *34*, 82–89. [CrossRef]
22. Lu, C.; Li, W.; Gao, S. Driving determinants and prospective prediction simulations on carbon emissions peak for China's heavy chemical industry. *J. Clean. Prod.* **2020**, *251*, 119642. [CrossRef]
23. Sun, W.; Ren, C. Short-term prediction of carbon emissions based on the EEMD-PSOBP model. *Environ. Sci. Pollut. Res.* **2021**, *28*, 56580–56594. [CrossRef]
24. Zhang, D.; Wang, T.; Zhi, J. Carbon emissions prediction based on IPSO-BP neural network model and eco-economic analysis of Shandong province. *Ecol. Sci.* **2022**, *41*, 149–158. [CrossRef]
25. Zhang, X.; Yan, F.; Liu, H.; Qiao, Z. Towards low carbon cities: A machine learning method for predicting urban blocks carbon emissions (UBCE) based on built environment factors (BEF) in Changxing City, China. *Sustain. Cities Soc.* **2021**, *69*, 102875. [CrossRef]
26. Mladenovic, I.; Sokolov-Mladenovic, S.; Milovancevic, M.; Markovic, D.; Simeunovic, N. Management and estimation of thermal comfort, carbon dioxide emission and economic growth by support vector machine. *Renew. Sustain. Energy Rev.* **2016**, *64*, 466–476. [CrossRef]
27. Gou, G. SVR-based prediction of carbon emissions from energy consumption in Henan Province. In Proceedings of the 3rd International Conference on Advances in Energy Resources and Environment Engineering, Harbin, China, 28–30 May 2018.
28. Xu, Y.; Song, W. Carbon Emission Prediction of Construction Industry Based on FCS-SVM. *Ecol. Econ.* **2019**, *35*, 37–41.
29. Sun, W.; Jin, H.; Wang, X. Predicting and Analyzing CO_2 Emissions Based on an Improved Least Squares Support Vector Machine. *Pol. J. Environ. Stud.* **2019**, *28*, 4391–4401. [CrossRef]
30. Zuo, Z.; Guo, H.; Cheng, J. An LSTM-STRIPAT model analysis of China's 2030 CO_2 emissions peak. *Carbon Manag.* **2020**, *11*, 577–592. [CrossRef]
31. Zhao, J.; Kou, L.; Wang, H.; He, X.; Xiong, Z.; Liu, C.; Cui, H. Carbon Emission Prediction Model and Analysis in the Yellow River Basin Based on a Machine Learning Method. *Sustainability* **2022**, *14*, 6153. [CrossRef]
32. Li, Y. Forecasting Chinese carbon emissions based on a novel time series prediction method. *Energy Sci. Eng.* **2020**, *8*, 2274–2285. [CrossRef]
33. Yu, Y.; Sun, R.; Sun, Y.; Shu, Y. Integrated Carbon Emission Estimation Method and Energy Conservation Analysis: The Port of Los Angles Case Study. *J. Mar. Sci. Eng.* **2022**, *10*, 717. [CrossRef]
34. Wang, Y.; Watanabe, D.; Hirata, E.; Toriumi, S. Real-Time Management of Vessel Carbon Dioxide Emissions Based on Automatic Identification System Database Using Deep Learning. *J. Mar. Sci. Eng.* **2021**, *9*, 871. [CrossRef]
35. Lin, J.; Lu, S.; He, X.; Wang, F. Analyzing the impact of three-dimensional building structure on CO_2 emissions based on random forest regression. *Energy* **2021**, *236*, 121502. [CrossRef]
36. Fang, Y.; Lu, X.; Li, H. A random forest-based model for the prediction of construction-stage carbon emissions at the early design stage. *J. Clean. Prod.* **2021**, *328*, 129657. [CrossRef]
37. Huang, G.B.; Zhu, Q.Y.; Siew, C.K. Extreme learning machine: A new learning scheme of feedforward neural networks. In Proceedings of the 2004 IEEE International Joint Conference on Neural Networks, Budapest, Hungary, 25–29 July 2004; Volume 1–4, pp. 985–990.
38. Tang, J.; Deng, C.; Huang, G. Extreme Learning Machine for Multilayer Perceptron. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *27*, 809–821. [CrossRef]
39. Huang, G.; Huang, G.; Song, S.; You, K. Trends in extreme learning machines: A review. *Neural Netw.* **2015**, *61*, 32–48. [CrossRef]
40. Li, Y.; Hu, H. Influential Factor Analysis and Projection of Industrial CO_2 Emissions in China Based on Extreme Learning Machine Improved by Genetic Algorithm. *Pol. J. Environ. Stud.* **2020**, *29*, 2259–2271. [CrossRef]
41. Wang, W.; Wang, J. Determinants investigation and peak prediction of CO_2 emissions in China's transport sector utilizing bio-inspired extreme learning machine. *Environ. Sci. Pollut. Res.* **2021**, *28*, 55535–55553. [CrossRef]
42. Wen, L.; Yuan, X. Forecasting CO_2 emissions in Chinas commercial department, through BP neural network based on random forest and PSO. *Sci. Total Environ.* **2020**, *718*, 137194. [CrossRef] [PubMed]

43. Qiu, G.; Cai, Z. Research on Carbon Emission Prediction in Shaanxi Province Based on Rough Set and Neural Network Method. *Ecol. Econ.* **2019**, *35*, 25–30.
44. Zhang, G.; Zhang, Z.; Liu, P.; Liu, M. The Running Mechanism and Prediction of the Growth Rate of China's Carbon Emissions. *Chin. J. Manag. Sci.* **2015**, *23*, 86–93. [CrossRef]
45. Yan, F.; Liu, S.; Zhang, X. Prediction of Carbon Emission for Land Use Based on PCA-BP Neural Network. *J. Hum. Settl. West China* **2021**, *36*, 1–7. [CrossRef]
46. Tursun, M.; Ding, W.; Xie, J. Prediction and Impact Factor Analysis of Agricultural Carbon Emission Based on Neural Network. *Environ. Eng.* **2017**, *35*, 156–160. [CrossRef]
47. Dai, D.; Li, K.; Zhao, S.; Zhou, B. Research on prediction and realization path of carbon peak of construction industry based on EGM-BP model. *Front. Energy Res.* **2022**, *10*, 981097. [CrossRef]
48. Zhou, J.; Du, S.; Shi, J.; Guang, F. Carbon Emissions Scenario Prediction of the Thermal Power Industry in the Beijing-Tianjin-Hebei Region Based on a Back Propagation Neural Network Optimized by an Improved Particle Swarm Optimization Algorithm. *Pol. J. Environ. Stud.* **2017**, *26*, 1895–1904. [CrossRef]
49. Wen, L.; Liu, Y. A Research About Beijing's Carbon Emissions Based on the IPSO-BP Model. *Environ. Prog. Sustain. Energy* **2017**, *36*, 428–434. [CrossRef]
50. Zhou, J.; Jin, B.; Du, S.; Zhang, P. Scenario Analysis of Carbon Emissions of Beijing-Tianjin-Hebei. *Energies* **2018**, *11*, 1489. [CrossRef]
51. Dai, S.; Niu, D.; Han, Y. Forecasting of Energy-Related CO_2 Emissions in China Based on GM(1,1) and Least Squares Support Vector Machine Optimized by Modified Shuffled Frog Leaping Algorithm for Sustainability. *Sustainability* **2018**, *10*, 958. [CrossRef]
52. Sun, W.; Zhang, J. Analysis influence factors and forecast energy-related CO_2 emissions: Evidence from Hebei. *Environ. Monit. Assess.* **2020**, *192*, 1–17. [CrossRef]
53. Wei, S.; Wang, T.; Li, Y. Influencing factors and prediction of carbon dioxide emissions using factor analysis and optimized least squares support vector machine. *Environ. Eng. Res.* **2017**, *22*, 175–185. [CrossRef]
54. Wu, Q.; Meng, F. Prediction of energy-related CO_2 emissions in multiple scenarios using a least square support vector machine optimized by improved bat algorithm: A case study of China. *Greenh. Gases Sci. Technol.* **2020**, *10*, 160–175. [CrossRef]
55. Li, J.; Zhang, B.; Shi, J. Combining a Genetic Algorithm and Support Vector Machine to Study the Factors Influencing CO_2 Emissions in Beijing with Scenario Analysis. *Energies* **2017**, *10*, 1520. [CrossRef]
56. Sun, W.; Liu, M. Prediction and analysis of the three major industries and residential consumption CO_2 emissions based on least squares support vector machine in China. *J. Clean. Prod.* **2016**, *122*, 144–153. [CrossRef]
57. Zhao, J.; Kou, L.; Jiang, Z.; Lu, N.; Wang, B.; Li, Q. A novel evaluation model for carbon dioxide emission in the slurry shield tunnelling. *Tunn. Undergr. Space Technol.* **2022**, *130*, 104757. [CrossRef]
58. Xue, Y.; Ren, J.; Bi, X. Impact of Influencing Factors on CO_2 Emissions in the Yangtze River Delta during Urbanization. *Sustainability* **2019**, *11*, 4183. [CrossRef]
59. Zeng, S.; Su, B.; Zhang, M.; Gao, Y.; Liu, J.; Luo, S.; Tao, Q. Analysis and forecast of China's energy consumption structure. *Energy Policy* **2021**, *159*, 112630. [CrossRef]
60. Huang, Y.; Shen, L.; Liu, H. Grey relational analysis, principal component analysis and forecasting of carbon emissions based on long short-term memory in China. *J. Clean. Prod.* **2019**, *209*, 415–423. [CrossRef]
61. Kong, F.; Song, J.; Yang, Z. A novel short-term carbon emission prediction model based on secondary decomposition method and long short-term memory network. *Environ. Sci. Pollut. Res.* **2022**, *29*, 64983–64998. [CrossRef] [PubMed]
62. Wen, Y.; Wu, R.; Zhou, Z.; Zhang, S.; Yang, S.; Wallington, T.J.; Shen, W.; Tan, Q.; Deng, Y.; Wu, Y. A data-driven method of traffic emissions mapping with land use random forest models. *Appl. Energy* **2022**, *305*, 117916. [CrossRef]
63. Wang, G.; Han, Q.; de Vries, B. Assessment of the relation between land use and carbon emission in Eindhoven, the Netherlands. *J. Environ. Manag.* **2019**, *247*, 413–424. [CrossRef]
64. Silitonga, A.S.; Masjuki, H.H.; Ong, H.C.; Sebayang, A.H.; Dharma, S.; Kusumo, F.; Siswantoro, J.; Milano, J.; Daud, K.; Mahlia, T.M.I.; et al. Evaluation of the engine performance and exhaust emissions of biodiesel-bioethanol-diesel blends using kernel-based extreme learning machine. *Energy* **2018**, *159*, 1075–1087. [CrossRef]
65. Li, Y.; Dong, H.; Lu, S. Research on application of a hybrid heuristic algorithm in transportation carbon emission. *Environ. Sci. Pollut. Res.* **2021**, *28*, 48610–48627. [CrossRef]
66. Sun, W.; Wang, Y.; Zhang, C. Forecasting CO_2 emissions in Hebei, China, through moth-flame optimization based on the random forest and extreme learning machine. *Environ. Sci. Pollut. Res.* **2018**, *25*, 28985–28997. [CrossRef]
67. Li, W.; Zhang, S.; Lu, C. Research on the driving factors and carbon emission reduction pathways of China's iron and steel industry under the vision of carbon neutrality. *J. Clean. Prod.* **2022**, *361*, 132237. [CrossRef]
68. Sun, W.; Sun, J. Prediction of carbon dioxide emissions based on principal component analysis with regularized extreme learning machine: The case of China. *Environ. Eng. Res.* **2017**, *22*, 302–311. [CrossRef]
69. Chai, Z.; Yan, Y.; Simayi, Z.; Yang, S.; Abulimiti, M.; Wang, Y. Carbon emissions index decomposition and carbon emissions prediction in Xinjiang from the perspective of population-related factors, based on the combination of STIRPAT model and neural network. *Environ. Sci. Pollut. Res.* **2022**, *29*, 31781–31796. [CrossRef]
70. Li, J.; Shi, J.; Li, J. Exploring Reduction Potential of Carbon Intensity Based on Back Propagation Neural Network and Scenario Analysis: A Case of Beijing, China. *Energies* **2016**, *9*, 615. [CrossRef]

71. Sun, W.; Xu, Y. Using a back propagation neural network based on improved particle swarm optimization to study the influential factors of carbon dioxide emissions in Hebei Province, China. *J. Clean. Prod.* **2016**, *112*, 1282–1291. [CrossRef]
72. Chu, X.; Zhao, R. A building carbon emission prediction model by PSO-SVR method under multi-criteria evaluation. *J. Intell. Fuzzy Syst.* **2021**, *41*, 7473–7484. [CrossRef]
73. Wang, J.; Yang, F.; Zhang, X. Analysis of the Influence Mechanism of Energy-Related Carbon Emissions with a Novel Hybrid Support Vector Machine Algorithm in Hebei, China. *Pol. J. Environ. Stud.* **2019**, *28*, 3475–3487. [CrossRef]
74. Wang, L.; Xue, X.; Zhao, Z.; Wang, Y.; Zeng, Z. Finding the de-carbonization potentials in the transport sector: Application of scenario analysis with a hybrid prediction model. *Environ. Sci. Pollut. Res.* **2020**, *27*, 21762–21776. [CrossRef] [PubMed]
75. Qiao, W.; Lu, H.; Zhou, G.; Azimi, M.; Yang, Q.; Tian, W. A hybrid algorithm for carbon dioxide emissions forecasting based on improved lion swarm optimizer. *J. Clean. Prod.* **2020**, *244*, 118612. [CrossRef]
76. Wang, J.; Yang, F.; Chen, K. Regional carbon emission evolution mechanism and its prediction approach: A case study of Hebei, China. *Environ. Sci. Pollut. Res.* **2019**, *26*, 28884–28897. [CrossRef] [PubMed]
77. Qin, X.; Zhang, S.; Dong, X.; Zhan, Y.; Wang, R.; Xu, D. China's carbon dioxide emission forecast based on improved marine predator algorithm and multi-kernel support vector regression. *Environ. Sci. Pollut. Res.* **2022**, *30*, 5730–5748. [CrossRef]
78. Zhao, H.; Guo, S.; Zhao, H. Energy-Related CO_2 Emissions Forecasting Using an Improved LSSVM Model Optimized by Whale Optimization Algorithm. *Energies* **2017**, *10*, 874. [CrossRef]
79. Zhao, H.; Huang, G.; Yan, N. Forecasting Energy-Related CO2 Emissions Employing a Novel SSA-LSSVM Model: Considering Structural Factors in China. *Energies* **2018**, *11*, 781. [CrossRef]
80. Ameyaw, B.; Li, Y.; Annan, A.; Agyeman, J.K. West Africa's CO_2 emissions: Investigating the economic indicators, forecasting, and proposing pathways to reduce carbon emission levels. *Environ. Sci. Pollut. Res.* **2020**, *27*, 13276–13300. [CrossRef]
81. Lin, X.; Zhu, X.; Feng, M.; Han, Y.; Geng, Z. Economy and carbon emissions optimization of different countries or areas in the world using an improved Attention mechanism based long short-term memory neural network. *Sci. Total Environ.* **2021**, *792*, 148444. [CrossRef]
82. Fu, Q.; Liu, Q.; Gao, Z.; Wu, H.; Fu, B.; Chen, J. A Building Energy Consumption Prediction Method Based on Integration of a Deep Neural Network and Transfer Reinforcement Learning. *Int. J. Pattern Recognit. Artif. Intell.* **2020**, *34*, 2052005. [CrossRef]
83. Tuong, L.; Minh, T.V.; Tung, K.; Hwang, E.; Rho, S.; Baik, S.W. Multiple Electric Energy Consumption Forecasting Using a Cluster-Based Strategy for Transfer Learning in Smart Building. *Sensors* **2020**, *20*, 2668. [CrossRef]
84. Qian, F.; Gao, W.; Yang, Y.; Yu, D. Potential analysis of the transfer learning model in short and medium-term forecasting of building HVAC energy consumption. *Energy* **2020**, *193*, 315–324. [CrossRef]