

Article

Multi-Factor Highway Freight Volume Prediction Based on Backpropagation Neural Network

Yanshuang Zhang ¹, Caixia Tian ^{2,*}, Baohua Guo ^{1,3} , Meixia Wang ¹, Zhezhe Zhang ¹ and Kgaugelo Morobeni ¹

¹ School of Energy Science and Engineering, Henan Polytechnic University, Jiaozuo 454000, China; 15236446375@163.com (Y.Z.); guobaohua@139.com (B.G.); 18403909837@163.com (M.W.); zhe991027@163.com (Z.Z.); kgaumorobeni404@gmail.com (K.M.)

² School of Resources and Environment, Henan Polytechnic University, Jiaozuo 454000, China

³ Jiaozuo Engineering Research Center of Road Traffic and Transportation, Henan Polytechnic University, Jiaozuo 454000, China

* Correspondence: tiancx@hpu.edu.cn

Abstract: With the development of the times, the traditional single-factor time series prediction cannot meet the needs of actual prediction, and it is necessary to comprehensively consider the influence of various variables on prediction results. Therefore, we use MATLAB R2022a to predict the multi-factor highway freight volume. According to the relevant data of highway freight volume in Chinese history, the BP neural network prediction model of highway freight volume is established, and the model is coded and calculated in the MATLAB software environment. Through repeated training of the data, the predicted value is finally obtained. The results show that the prediction accuracy of the BP neural network model based on multi-factor prediction is very high. Through the example analysis of China's highway freight volume, the original data are accurately fitted, and the validity of the highway freight volume prediction model based on BP neural network is proved. Through the prediction of freight volume, the investment in infrastructure construction is improved to promote the development of transportation industry and the progress of social economy.

Keywords: BP neural network; highway freight volume; multi-factor prediction



Citation: Zhang, Y.; Tian, C.; Guo, B.; Wang, M.; Zhang, Z.; Morobeni, K. Multi-Factor Highway Freight Volume Prediction Based on Backpropagation Neural Network. *Appl. Sci.* **2024**, *14*, 5948. <https://doi.org/10.3390/app14135948>

Received: 12 April 2024

Revised: 23 May 2024

Accepted: 30 May 2024

Published: 8 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Since the reform and opening up, the rapid development of the national economy has promoted the rapid development of various industries. For example, it has promoted the rapid development of road freight transportation. With the rise of online shopping, it has promoted the growth of China's logistics industry to an extent that has led to the development of China's freight transportation. The increase in the investment share of social fixed assets has promoted the development of trucks. In the 'Twelfth Five-Year Plan', the state clearly proposed to vigorously develop road freight transportation, improve the level of standardization and intensification, promote the development and improvement of freight transportation methods, and expand the radiation range of less-than-truckload and fast-operating business networks. With the completion of a well-off society in an all-round way, China's poverty alleviation has achieved great results. The economy has developed steadily, the income level of urban and rural residents has been continuously improved, and consumer demand has been increasing. In addition, China's highway network has been continuously improved. For instance, the structure of the highway has been continuously rationalized, the connection between urban and rural areas has been strengthened, and the highway network has been gradually built and improved. The progress of informatization technology will promote the development of highway transportation, improve the quality of transportation, and promote the transformation and upgrading of transportation modes while promoting the development of transportation industry [1].

As the most important mode of transportation, highway transportation occupies an important position in the whole of cargo transportation, and the connection between cargo transportation and the highway is inseparable. Highway freight volume is affected by many factors, such as the ownership of freight vehicles with highway diesel consumption as the evaluation index, as well as highway freight volume and freight turnover. It is also inseparable from the relationship between the population and the economic development level. The prediction of highway freight volume is of great significance. The fluctuation of highway freight volume can reflect the economic fluctuation to a certain extent, which is of great significance to the improvement of the national and international economic levels. In order to predict highway freight volume more accurately, we extract the key factors affecting freight volume from historical data and construct a multi-factor BP neural network model to achieve the accurate prediction of highway freight volume. Through the research of this paper, we expect to provide a more accurate and effective method for the prediction of highway freight volume. Through the prediction of highway freight volume, we can adjust the industrial structure of highway freight and the investment in infrastructure and provide strong support for the sustainable development of highway transportation [2–6].

2. Literature Review

Zhang [7] established a BP (backpropagation neural network, a multi-layer feedforward neural network trained according to error backpropagation algorithm) neural network model and analyzed the influencing factors of freight volume in Xinjiang. Through the training and fitting of the model, historical data were selected as the test samples of the network model, which proved that the prediction of the BP neural network was more accurate than other prediction methods. Taking the national railway freight volume as the research object, Luo [8] uses the methods of multiple linear regression and the BP neural network to obtain a more suitable prediction method for the national railway and conducts research on the prediction of national railway freight volume based on multiple factors. Practice shows that the two prediction methods have their own advantages and the combination of the two can improve the prediction accuracy. Chen [9] analyzed the characteristics of Shandong's freight volume while using several prediction models to make quantitative predictions under the influence of relevant factors and qualitatively corrected the prediction results to obtain the final comprehensive prediction results. Yang [10] established a PSO-SVM (particle swarm optimization–support vector machine) prediction model using a particle swarm support vector machine model and predicted the road freight volume under the influence of industrial transfer in the Yangtze River transport corridor to verify the effectiveness of the model. An [11] used the GM (1,1) model direct modeling method to improve and analyze the unbiased grey Verhulst model (single-sequence first-order nonlinear dynamic model) and the traditional grey Verhulst model. Through the simulation and prediction of Lanzhou highway freight volume, the superiority of the GM (1,1), traditional grey Verhulst, and unbiased grey Verhulst models are analyzed and the feasibility and the applicability of the improved unbiased grey Verhulst model are illustrated. On the basis of qualitative and quantitative analysis of the influencing factors of railway transportation, Wu [12] selected some factors affecting China's freight volume to carry out BP neural network prediction research based on multi-factor railway freight volume. Both prediction fitting and the algorithm convergence are good. The complex nonlinear mapping between the influencing factors and the railway freight volume is realized and the prediction results are excellent. The literature research and analysis are shown in Table 1 below.

Whether it is for the study of highway freight volume, or for the study of railway freight volume, we can receive inspiration from it so as to apply it to the study of highway freight volume. The application of research methods does not distinguish between fields but focuses on reliability. Based on the research of the above scholars, we can find that multiple regression and multi-factor prediction have a good effect on the prediction of freight volume. In summary, based on the comprehensive consideration of the impact indicators of freight volume, this paper selects relevant indicators and collects data. The

multi-factor prediction based on the BP neural network is used to predict China's freight volume and compare it with China's actual highway freight volume. The predicted results and reliability are obtained [13].

Table 1. Literature research and analysis.

Literature Serial Number	Research Method	Prediction Direction	Research Conclusions
[7]	BP neural network algorithm	Xinjiang freight volume	BP neural network prediction is more accurate than other prediction methods.
[8]		National railway freight volume.	The prediction accuracy of multiple linear regression method is higher.
[9]		Railway freight volume.	The algorithm has good convergence and good prediction results.
[10]	Combined model	Highway freight volume in the Yangtze River comprehensive transport corridor	The prediction results show the effectiveness of the model.
[11]		Freight volume in Shandong Province	The combination of qualitative prediction and quantitative prediction makes the prediction results more accurate.
[12]		Lanzhou to Zhongchuan highway freight volume	The improved model improves the accuracy of simulation prediction.

3. Neural Network Theory

3.1. MATLAB Multivariate Prediction Theory

In this paper, we mainly use the neural network regression prediction. MATLAB is a powerful tool for data analysis and numerical calculation. It can use a variety of functions and toolboxes for multivariate prediction. Multivariate prediction generally predicts the value of one or more output variables through a variety of input variables. The steps in multivariate prediction are shown in Figure 1. In MATLAB, we can use a variety of multivariate prediction algorithms to predict. For example, through the construction of multiple decision trees, their results are averaged or voted for by random forest regression prediction; to predict the relationship between independent variables and dependent variables and establish multiple linear regression prediction of linear model, the weighted average is performed according to the nearest neighbor weight to find the K-nearest neighbor regression prediction of the nearest sample point; support vector regression prediction uses the support vector machine idea by finding the optimal hyperplane; the principal component regression predicts the principal component analysis of the independent variables and the principal component regression analysis of the principal component; and the partial least squares regression prediction of the linear model is established after the dimension reduction of the independent variable and the dependent variable. The neural network model is used to establish the neural network regression prediction of the nonlinear mapping relationship between input and output [14–18].

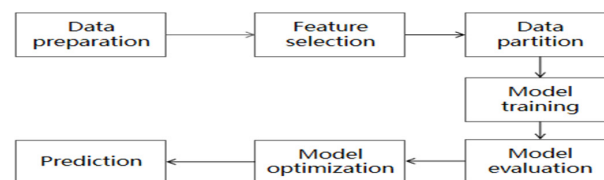


Figure 1. Multivariate flow chart.

- (1) Data preparation: Collect and organize input and output data. Ensure that the data are clean and complete and perform the necessary preprocessing, such as removing outliers, filling missing values, etc.
- (2) Feature selection: According to the characteristics of the problem and data, select the appropriate input variables as features. Correlation analysis, principal component analysis, and other methods can be used to assist selection.

- (3) Data partitioning: The data set is divided into training set and test set. Usually, most of the data are used to train the model, and the remaining part is used to evaluate the performance of the model.
- (4) Model training: Select the appropriate algorithm and use the training set to train the model. In MATLAB, functions such as fitlm (the model is usually created in the form of a table or an array of data sets according to the given data, which is used to fit the linear regression model), svmtrain (a powerful supervised learning algorithm widely used in classification and regression problems for training support vector machine (SVM) models), and trainlm (an optimization algorithm that combines Newton's method and the gradient descent method to improve training speed and accuracy) can be used to train the model.
- (5) Model evaluation: Use the test set to evaluate the performance of the model. Various indicators, such as mean square error and coefficient of determination, can be used to evaluate the accuracy and generalization ability of the model.
- (6) Model optimization: According to the evaluation results, the model is adjusted and optimized. Different parameter settings and feature combinations can be tried to improve the performance of the model.
- (7) Prediction: Use the trained model to predict. Input new data and obtain the corresponding output through the model.

3.2. BP Neural Network Theory

3.2.1. Algorithm Flow

The characteristic of the BP neural network is to improve signal propagation according to the reverse feedback of error. The topology of the BP neural network model is generally composed of three aspects: input layer, hidden layer, and output layer, as shown in Figure 2 below.

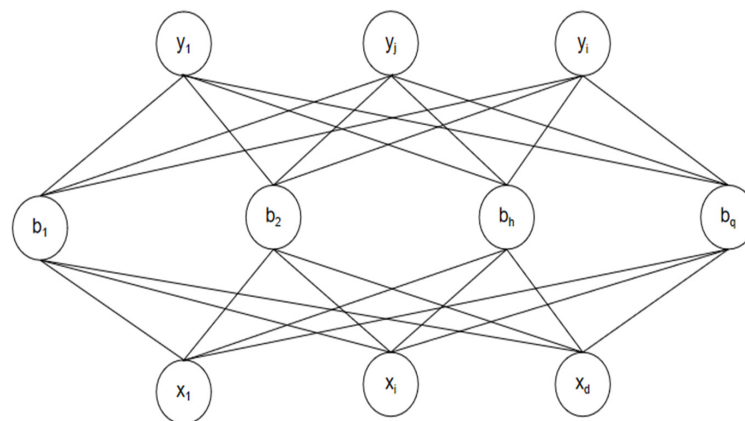


Figure 2. Topological structure of BP neural network model.

The output layers $y_1, \dots, y_j, \dots, y_i$ denote information output, that is, the target result.

The hidden layers $b_1, b_2, \dots, b_h, \dots, b_q$ denote data processing; you can set the number of layers.

The input layers $x_1, \dots, x_j, \dots, x_d$ denote information input and read data:

w is the black path between b and y .

w represents the weight from hidden layer to input layer.

v is the black path between b and x .

v represents the weight of the input layer to the hidden layer.

3.2.2. Working Process

In the working process, the weight and bias from the hidden layer to the output layer are adjusted in turn, and the weight and bias from the input layer to the hidden layer are adjusted in turn. The working process of the BP neural network is shown in Figure 3.

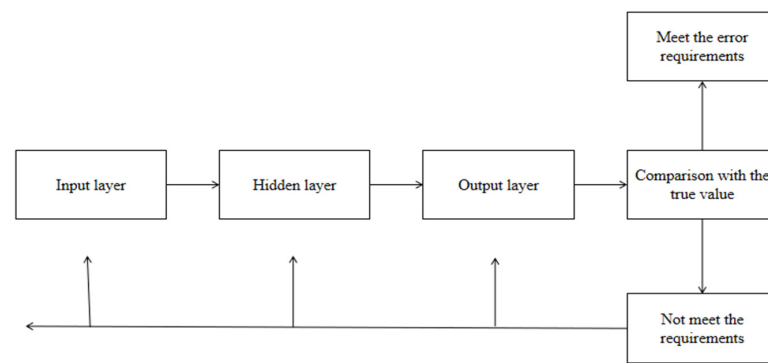


Figure 3. Working process diagram of BP neural network.

3.2.3. Propagation Process

(1) Forward propagation

The information enters the network from the input layer and passes through each layer in turn and enters the process of obtaining the final output layer result.

The calculation process mainly involves matrix multiplication, using the value of each layer multiplied by the corresponding weight + bias variable (activation function).

From the input layer to the hidden layer:

$$\alpha_h = \sum_{i=1}^d v_{ih} x_i \quad (1)$$

The activation function through the hidden layer is:

$$b_h = f(\alpha_h - \gamma_h) \quad (2)$$

Among them, the weight between the first neuron of the input layer and the first neuron of the hidden layer is the threshold of the first neuron of the hidden layer.

From the hidden layer to the output layer:

$$\beta_j = \sum_{h=1}^q w_{hj} b_h + \theta_j \quad (3)$$

Among them, the weight between the first neuron in the hidden layer and the first neuron in the output layer is the threshold of the first neuron in the output layer.

In the machine learning model, the parameters we start with are random, so the calculation results may have a large error with the real values. In order to reduce this error, we need to adjust the parameters according to the error, and this process requires the back propagation algorithm of the model.

(2) Backpropagation

The core idea of the backpropagation algorithm is to calculate the error between the output layer and the expected value, then propagate the error back to each layer of the network, and then adjust the parameters in the network. In this process, each neuron adjusts its own weight and bias according to the error so that the output of the network is closer to the expected value. This process is an iterative process, which needs to be repeated until the error reaches the minimum or meets the preset stop condition. The error is usually measured by calculating the square sum of the difference between the true value and the neural network training result. Of course, other calculation methods can also be selected. This error value will be used to guide the parameter adjustment in the back propagation process so that the network can better fit the real data.

4. Case Analysis of Highway Freight Volume Forecasting Based on Neural Network

4.1. Necessity Analysis

We analyzed the characteristics of the data, the limitations of resources and time, and the existing research. Therefore, considering the characteristics of the data studied in this paper and compared with the traditional prediction methods, the highway freight volume prediction based on BP neural network has the advantages of dealing with nonlinear relationships and complex data structures, and more accurately describing the relationship between highway freight volume and various influencing factors. Secondly, it has strong generalization ability and can adapt to different scenarios and conditions. The prediction can improve the prediction accuracy and stability through continuous learning and optimization and is very good in the accuracy of prediction. Before the research method of this paper is determined, we need to clearly define the problems we want to solve and the expected goals and consider the complexity and scale of the problem, as well as the type and quantity of data. For the deep learning algorithm that has been popular recently, it needs a lot of data support. However, the data of this study are limited (the reasons for the limited data are explained in detail below). Compared with the deep learning algorithm, the multi-factor analysis algorithm is more suitable for the case where the sample size is small and the variable relationship is more complex. The selection of multi-factor analysis or deep learning needs to be considered comprehensively according to the nature of the problem and the causal relationship between variables, and the multi-factor analysis algorithm is finally selected as the algorithm used in this paper.

For the prediction of the mode of transportation, we generally consider a variety of factors, usually considering the nature of the goods, such as whether they are perishable goods considering the time distance and cost of transportation. For the measurement of the volume of goods, we can directly measure the length, width, and height of the goods to calculate their volume. We can also use the water level method (put the goods in water, use the rise of the water level and the height to measure the volume of the goods) to measure the volume of the goods indirectly. For the prediction of cargo volume and transportation mode, a variety of factors need to be considered comprehensively. For the influencing factors of highway freight volume selected in this paper, highway freight turnover, highway freight volume, and highway diesel consumption are mainly measured by weight. The population is mainly measured by the number of people, and the gross domestic product is mainly measured by price. Their prediction can play an auxiliary role in the model data source and provide support for the data source.

4.2. Data Source

Through the analysis of relevant data over the years, as well as the review of the relevant literature, the final choices are road diesel consumption, population, gross domestic product, road freight turnover, road mileage, gross domestic product, road operating truck ownership and other indicators as influencing factors, road freight volume as the target, data from the China Statistical Yearbook, and the main query of the relevant data from 1997 to 2021. As for the data for diesel consumption, the data shown in the China Statistical Yearbook are not complete. The data for diesel consumption can only be queried as early as 1997. In addition, the data for some years may be missing. For the filling of missing values, we adopt the method of averaging the data for the adjacent two years. Due to the limited and missing data, the data in this paper mainly query the relevant data from 1997 to 2021. After querying the data, the data are summarized. The example analysis proves that the freight volume prediction model based on the BP neural network is effective, which provides a new prediction idea and method for highway freight volume prediction. The data over the years are shown in Table 2 [19–23].

Table 2. Data.

Particular Year	Road Diesel Consumption (Ten Thousand Tons)	Population (Ten Thousand People)	Gross Domestic Product (Billion Yuan)	Road Freight Turnover (Billions of Tons of Kilometers)	Highway Mileage (Ten Thousand Kilometers)	Gross Domestic Product of Highway (Billion Yuan)	The Number of Road Operating Trucks (Ten Thousand)	Volume of Road Haulage (Ten Thousand Tons)
1997	1380	123,626	78,803	5272	123	4149	601	976,536
1998	1902	124,761	83,818	5483	128	4662	628	976,004
1999	2222	125,768	89,367	5724	135	5176	409	990,444
2000	2544	126,743	99,066	6129	168	6162	486	1,038,813
2001	2754	127,627	109,276	6330	170	6871	509	1,056,312
2002	2965	128,453	120,480	6783	177	7494	537	1,116,324
2003	34,485	129,227	136,576	7100	181	7915	592	1,159,957
2004	4182	129,988	161,415	7841	187	9307	648	1,244,990
2005	4965	130,756	185,999	8693	335	10,669	605	1,341,778
2006	5747	131,448	219,029	9754	346	12,186	641	1,466,347
2007	6794	132,129	270,704	11,355	358	14,605	684	1,639,432
2008	7343	132,802	321,230	32,868	373	16,368	761	1,916,759
2009	7892	133,450	347,935	37,189	386	16,522	907	2,127,834
2010	8519	134,091	410,354	43,390	401	18,784	1050	2,448,052
2011	9485	134,916	483,393	51,375	411	21,842	1179	2,820,100
2012	10,727	135,922	537,329	59,535	424	23,763	1253	3,188,475
2013	10,921	136,726	588,141	55,738	436	26,043	1419	3,076,648
2014	11,043	137,646	644,380	56,847	446	28,534	1453	3,113,334
2015	11,163	138,326	685,571	57,956	458	30,520	1389	3,150,019
2016	11,068	139,232	742,694	61,080	470	33,029	1352	3,341,259
2017	11,254	140,011	830,946	66,772	477	37,122	1369	3,686,858
2018	11,167	140,541	915,244	71,249	485	40,337	1356	3,956,871
2019	9867	141,008	983,751	59,636	501	42,466	1088	3,435,480
2020	9532	141,212	1,005,451	60,172	520	40,583	1110	3,426,413
2021	9984	141,260	1,141,231	69,088	528	48,424	1173	3,913,889

4.3. Multicollinearity Test

(1) Multicollinearity test

Before any modeling work, a more thorough investigation should be conducted on the relationship between the input features and how they affect the output variance to avoid the selected variables being independent and having the same distribution, which may lead to wrong predictions due to problems such as multicollinearity or heteroscedasticity.

The VIF (variance inflation factor) is an indicator used to measure the severity of multicollinearity between independent variables. When the VIF value is greater than 10, we usually think that there is a serious multicollinearity problem. Multicollinearity means that there is a high degree of correlation between independent variables, which may lead to inaccurate parameter estimation of the regression model, thus affecting the predictive ability and explanatory ability of the model.

The variance inflation factor of the regression coefficient can be expressed as:

$$VIF_i = 1 / (1 - R_i^2) \quad (4)$$

The square indicates the coefficient of determination obtained by fitting the regression equation with the first variable as the dependent variable and the remaining independent variables.

The larger the VIF value, the stronger the correlation between the variable and the other independent variables.

Tolerance and VIF are two important indicators in collinearity diagnosis, and they are reciprocal relationships. Tolerance greater than 0.1 is usually indicates no collinearity problem. Tolerance and VIF can be selected to judge collinearity in the study, but in general, describing the VIF value is more common because it is more intuitive, and when the VIF value is large, the tolerance value will be small and not easy to observe. They are generally used to calculate the correlation between the independent variables in the multiple linear regression model, so it is necessary to ensure the applicability and accuracy of the model.

Using SPSS for linear regression analysis, the data can be obtained and are presented in Table 3.

Table 3. Collinearity statistics.

Model	Collinearity Statistics	
(constant)	allowance	VIF
Road diesel consumption	0.704	1.42
Population	0.016	61.654
Highway freight volume turnover	0.03	32.914
Gross domestic product	0.003	369.396
Highway mileage	0.05	20.078
Gross domestic product by road	0.002	454.474
Number of road operating trucks	0.071	14.067

The purpose of multicollinearity regression analysis is to predict dependent variables through independent variables. From the table, we can see that except for the tolerance of road diesel consumption greater than 0.1 and VIF less than 10, other independent variables do not meet the requirements, and the collinearity between independent variables is more serious.

The DW (Durbin-Watson) test is often used to detect whether the residual has autocorrelation. The D-W value is between 0–4, and the closer it is to 2, the greater the possibility that the observed values are independent of each other. Using SPSS for analysis, the following Table 4 can be obtained.

Table 4. DW test.

Input Variable	Durbin-Watson
Road diesel consumption	1.762
Population	
Highway freight volume turnover	
Gross domestic product	
Highway mileage	
Gross domestic product by road	
Number of road operating trucks	

From the table, we can see that the value of DW is 1.762, which is very close to 2, and the observation values are more likely to be independent of each other.

(2) Reduction in multicollinearity

By gradually excluding the factors with strong influence of multicollinearity, and then conducting multiple multicollinearity tests through SPSS, the collinearity Table 5 and independent test Table 6 can finally be obtained.

Table 5. Multicollinearity reduction.

Colinearity Statistics		
(constant)	allowance	VIF
Road diesel consumption	0.884	1.131
Highway mileage	0.147	6.812
Number of road operating trucks	0.239	4.19
Gross domestic product	0.173	5.78

Table 6. Independent test table.

Input Variable	Durbin-Watson
Road diesel consumption	1.383
Highway mileage	
Number of road operating trucks	
Gross domestic product	

Finally, we choose five independent variables: road diesel consumption, road mileage, road operating truck ownership, gross domestic product, and Chinese freight volume as the dependent variable.

4.4. Data Determination

After the previous data query, through the multicollinearity test, the problem of reducing the multicollinearity or heteroscedasticity leads to wrong predictions and finally retains the following data, as shown in Table 7.

Table 7. Data determination table.

Particular Year	Road Diesel Consumption (Ten Thousand Tons)	Gross Domestic Product (Billion Yuan)	Highway Mileage (Ten Thousand Kilometers)	The Number of Road Operating Trucks (Ten Thousand)	Volume of Road Haulage (Ten Thousand Tons)
1997	1380	78,803	123	601	976,536
1998	1902	83,818	128	628	976,004
1999	2222	89,367	135	409	990,444
2000	2544	99,066	168	486	1,038,813
2001	2754	109,276	170	509	1,056,312

Table 7. Cont.

Particular Year	Road Diesel Consumption (Ten Thousand Tons)	Gross Domestic Product (Billion Yuan)	Highway Mileage (Ten Thousand Kilometers)	The Number of Road Operating Trucks (Ten Thousand)	Volume of Road Haulage (Ten Thousand Tons)
2002	2965	120,480	177	537	1,116,324
2003	34,485	136,576	181	592	1,159,957
2004	4182	161,415	187	648	1,244,990
2005	4965	185,999	335	605	1,341,778
2006	5747	219,029	346	641	1,466,347
2007	6794	270,704	358	684	1,639,432
2008	7343	321,230	373	761	1,916,759
2009	7892	347,935	386	907	2,127,834
2010	8519	410,354	401	1050	2,448,052
2011	9485	483,393	411	1179	2,820,100
2012	10,727	537,329	424	1253	3,188,475
2013	10,921	588,141	436	1419	3,076,648
2014	11,043	644,380	446	1453	3,113,334
2015	11,163	685,571	458	1389	3,150,019
2016	11,068	742,694	470	1352	3,341,259
2017	11,254	830,946	477	1369	3,686,858
2018	11,167	915,244	485	1356	3,956,871
2019	9867	983,751	501	1088	3,435,480
2020	9532	1,005,451	520	1110	3,426,413
2021	9984	1,141,231	528	1173	3,913,889

4.5. Model Construction

According to the strong nonlinear mapping characteristics of the neural network, only the input and output data of the object are given, and the characteristics of the mapping relationship between input and output can be achieved through the learning function of the network itself. By using the strong nonlinear mapping characteristics of the neural network, that is, through the input and output data of a given object, the mapping relationship between input and output can be achieved through the learning function of the network itself. The BP neural network includes input layer, hidden layer, and output layer. The input layer contains four factors, namely, road diesel consumption, road mileage, road operating truck ownership, and gross domestic product. For independent variables, the output layer is road freight volume, and the number of hidden layers is five. The effect is better. This is obtained through multiple tests. Finally, a 4-5-1 network is determined, and the structure is shown in Figure 4. The maximum number of iterations of the training is set to 1000 times, and the error threshold of the training is set to 10^{-6} . Three sets of data are selected from the collated data to test the model. The first 17 sets of data, namely the data from 1997 to 2013, are selected as the divided training set, and the last five sets of data, namely the data from 2014 to 2018, are selected as the divided test set. In general, we use 70% to 80% of the data as the training set and the remaining data are used for the division of the test set. Therefore, the ratio of the divided training set to the test set is 7:2, and the size of the data set is divided according to the amount of data and the data studied. After determining the network structure and data grouping, Matlab R2022a software is used for network training and analysis [24–28].

The BP neural network model consists of input layer, hidden layer, and output layer. Each layer contains one or more neuron nodes. Based on the multi-layer feedforward neural network trained by the error back propagation algorithm, the mapping relationship between the input value and the output value is established by continuously training and learning the data. The number of hidden layers is determined by Formula (5).

$$l = \sqrt{m + n} + a \quad (5)$$

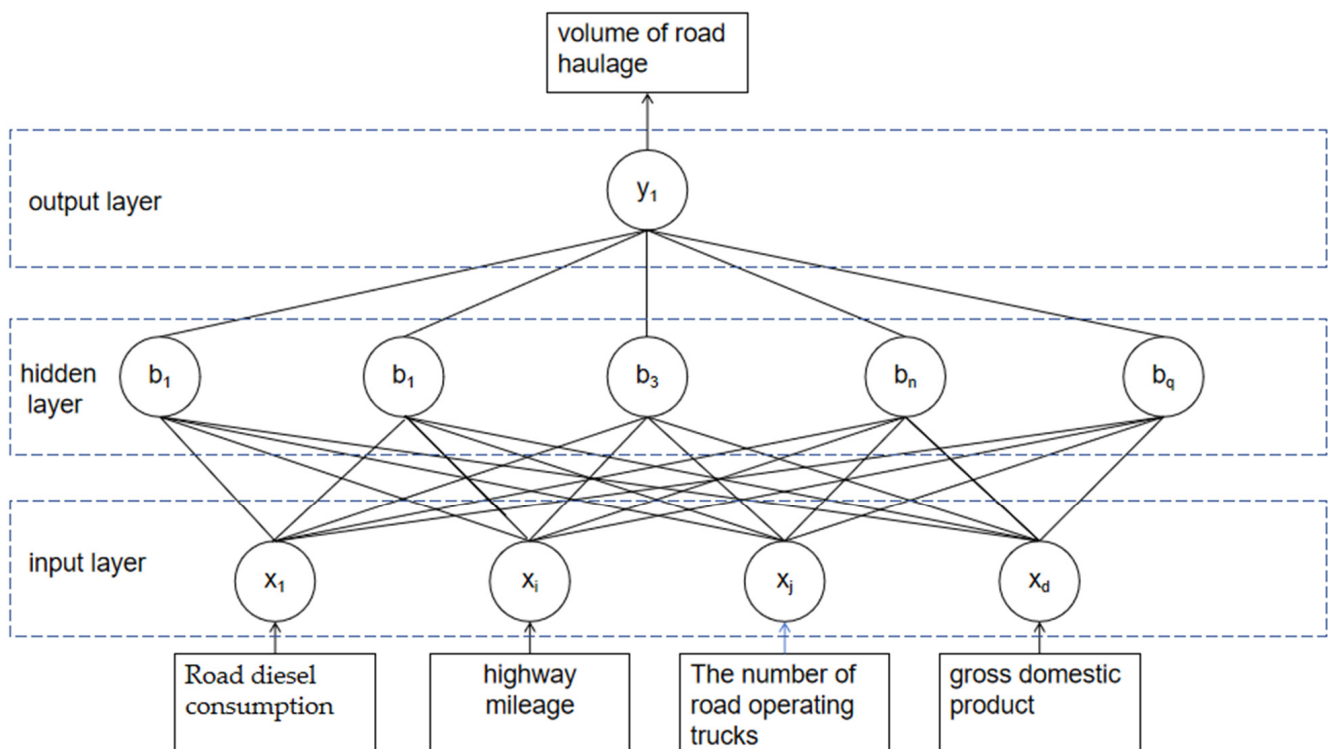


Figure 4. Model structure.

In the formula: m is the number of neurons in the output layer; n is the number of neurons in the input layer; a is a constant between 1 and 12; and l is the number of neurons in the hidden layer, $l \in [4, 15]$.

The reference activation function is sigmoid (logistic), also known as the S-shaped growth curve. The input signal from negative infinity to positive infinity can be transformed into output between 0 and 1. The activation function is shown in the following Figure 5.

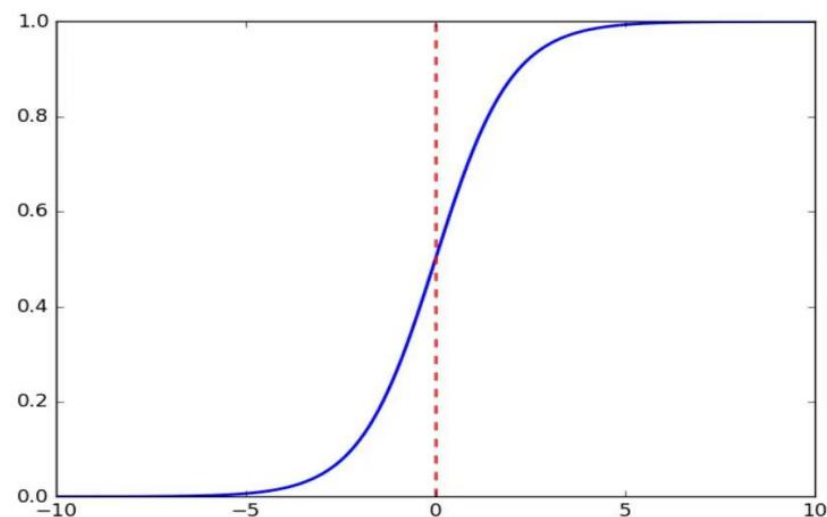


Figure 5. Activation function.

The properties of the activation function are very good:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (6)$$

$$f'(x) = f(x)(1 - f(x)) \quad (7)$$

The BP neural network model was evaluated by the coefficient of determination (R^2), the mean absolute error (MAE), the mean deviation (MBE), and the root mean square error (RMSE) between the estimated value and the true value. The calculation formula is as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (x_{sim} - x_i)^2}{\sum_{i=1}^n (\bar{x} - x_i)^2} \quad (8)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |a_i - b_i| \quad (9)$$

$$MBE = \frac{1}{n} \sum_{i=1}^n (a_i - b_i) \quad (10)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (a_i - b_i)^2} \quad (11)$$

In the formula:

x_i is the experimental value;

\bar{x}_i is the average value of the test;

x_{sim} is for the predicted value;

a_i and b_i represent the predicted value and the actual value, respectively.

n represents the number of samples.

By calculating R^2 , the correlation between the predicted value and the real value of the model is observed. The higher the correlation is, the better the prediction effect is. The smaller the average absolute error is, the smaller the gap between the predicted value and the real value is. The closer the average deviation is to 0, the better the model fitting effect is.

4.6. MATLAB Running Process

Based on the analysis of historical data and the influencing factors of the highway freight volume, the BP neural network is used to predict the highway freight volume in China. The neural network view is shown in Figure 6 below, which mainly includes the hidden layer, the input layer, and the output layer. The specific operation process is shown in Figure 7 below.

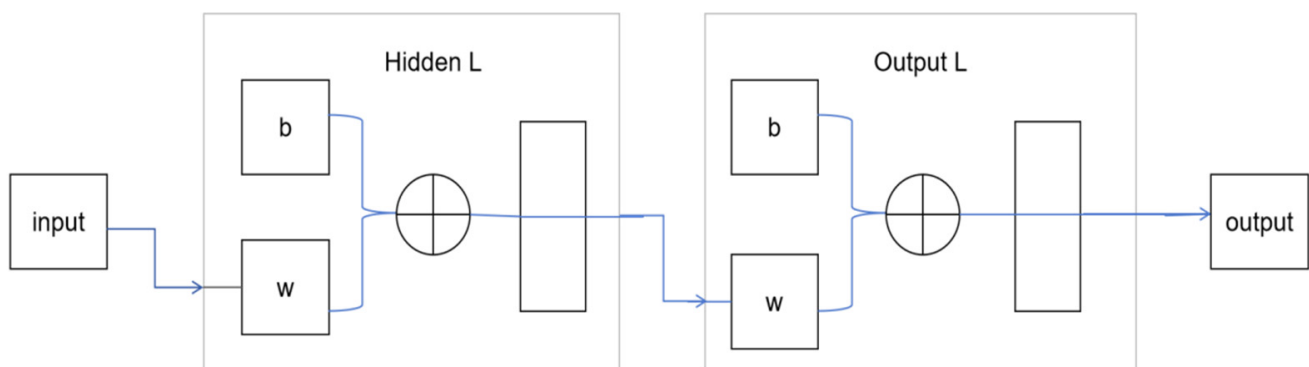


Figure 6. Neural network view.

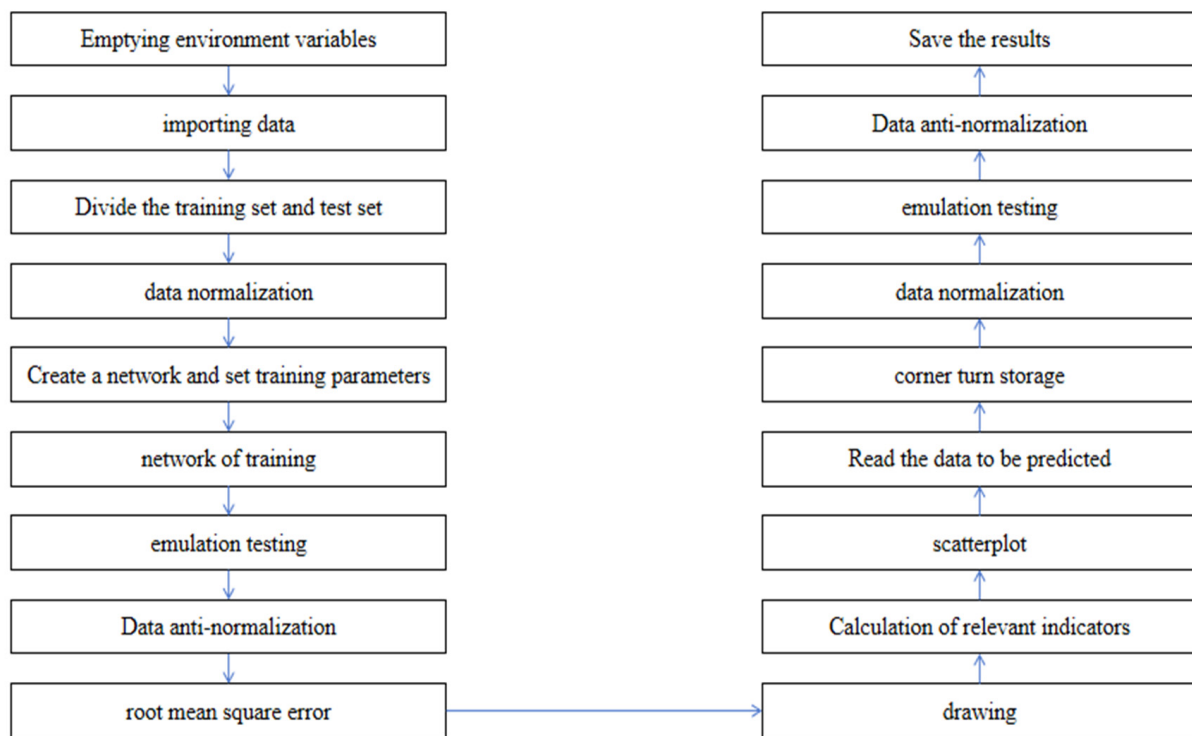


Figure 7. Operation flow chart.

- (1) Emptying the environment variables. Delete the defined variables and start building our code again.
- (2) Import data. The data of five indicators such as road diesel consumption, road mileage, road operating truck ownership, gross domestic product, and road freight volume from 1997 to 2018 were selected as the imported data, a total of 22 sets of data, and the specific data were derived from the National Statistical Yearbook.
- (3) Divide the training set and test set. Partitioning the training set and test set is an important task in machine learning and data analysis, which is used to evaluate the performance and generalization ability of the model. Here, we select the first 17 sets of data, that is, the data from 1997 to 2013 as the divided training set; select the last 5 sets of data, that is, the data from 2014 to 2018 as the divided test set; and select the data from the last 3 years, 2019–2021, as the verification set.

The training set is mainly used to construct the model. It refers to the data set used to train the model and algorithm. The classifier is established by training and fitting some parameters. By learning and summarizing the data patterns, trends, and rules in the training set, the model predicts, classifies, or infers unknown data. The test set is used to evaluate the final model. It is mainly used to evaluate the generalization ability of the final selected model after the entire training process is completed. The purpose is to provide an unbiased estimate of the model performance.

Validation set: used for model selection, that is, to perform the final optimization and determination of the model and to assist in the construction of our model.

The test set plays a vital role in the construction of machine learning models. The test set is a data set independent of the training set and the validation set, which is used to evaluate the performance of the trained model on unknown data.

- (4) Data normalization. The data are scaled to a certain range according to a certain proportion to improve the accuracy and efficiency of subsequent data processing or analysis, accelerate the speed of gradient descent to find the optimal solution, and improve the training efficiency of the model.

- (5) Create a network. In MATLAB, a new feedforward neural network is created, where p_train is the training input data, t_train is the training target data, and 5 represents the number of neurons in the hidden layer of the network.
- (6) Set training parameters. The maximum number of iterations of training is set to 1000 times. The training algorithm will try to minimize the error by adjusting the weight and deviation of the network and perform up to 1000 iterations. If the set error threshold is reached within this number of iterations or other stop conditions are met, the training may end early. The training error threshold is set to 10^{-6} . The training algorithm will try to make the output error of the network less than or equal to this threshold. When the threshold is reached or is lower than this threshold, the training will stop, even if the maximum number of iterations has not been reached. The learning rate is set to 0.01, which determines the magnitude of the network weight and bias update in each iteration. A high learning rate may lead to fast or unstable training, and a low learning rate may lead to more stable training.
- (7) Training network. The train function is called to train the neural network net using the input data p_train and the target data t_train .
- (8) Simulation test. The trained neural network net is used to simulate (predict) the training data p_train and the test data p_test , and the simulated outputs t_sim1 and t_sim2 are obtained, respectively.
- (9) Data denormalization. The mapminmax (data preprocessing function) function is used to normalize the input data P_train and P_test and the target data T_train and T_test and scales the data to a specific range (between 0 and 1) for better neural network training.
- (10) Root mean square error. Calculate the error between the neural network prediction result and the actual target value. The calculation method of root mean square error (RMSE) is used.
- (11) Drawing. The comparison diagram between the neural network prediction results and the actual target values is drawn, and the root mean square error (RMSE) of each data set is displayed. The comparison diagram of the training set prediction results and the comparison diagram of the test set prediction results can be obtained.
- (12) Calculation of relevant indicators. Several key indicators of neural network prediction results, R2 (coefficient of determination), MAE (mean absolute error), MBE (mean deviation), and RMSE (root mean square error) are calculated.
- (13) Draw scatter plot. By drawing the scatter plot to visualize the relationship between the predicted value and the true value of the training set and the test set, the training set scatter plot and the test set scatter plot are drawn.
- (14) Read the predicted data. Read the data from the Excel file, and read the data of the five indicators, namely road diesel consumption, road mileage, road operating truck ownership, gross domestic product, and road freight volume from 2019 to 2021.
- (15) Data transpose. The kes matrix is transposed, and the transposed matrix is re-assigned to kes.
- (16) Data normalization. Each value in the original data are mapped to the specified range by linear transformation.
- (17) Simulation test. Perform neural network simulation tests.
- (18) Data denormalization. The denormalized data are stored in T_sim3 and are directly compared with the original output data to evaluate the performance of the model.
- (19) Save the results. Save the data as an Excel file, and open and view these results when needed.

4.7. Analysis of Output Results

- (1) The training set prediction results are compared with the test set prediction results, as shown in Figure 8.

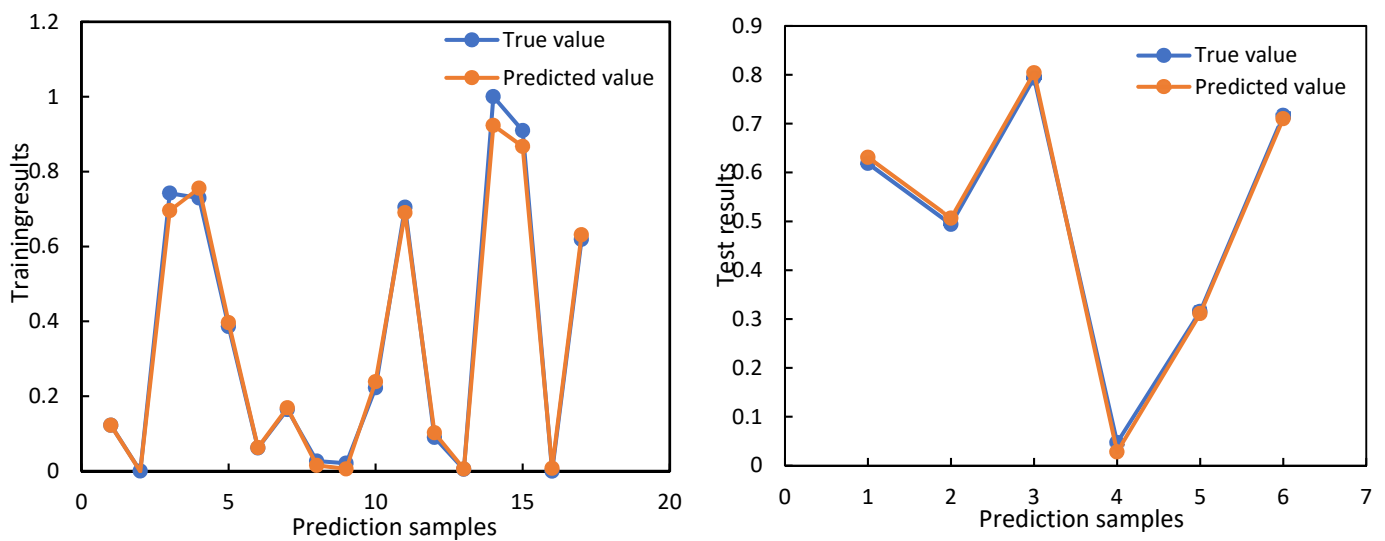


Figure 8. Comparison of training set prediction results and test set prediction results.

The comparison of the prediction results of the training set is as follows: $RMSE = 78,755.4264$. The comparison of the prediction results of the test set is as follows: $RMSE = 35,812.9792$. We can see that the difference between the two is large. The data show that the model performs better in the training set and performs relatively poorly in the test set.

- (2) The analysis and comparison of training set and prediction set fitting graph is as shown in Figure 9.

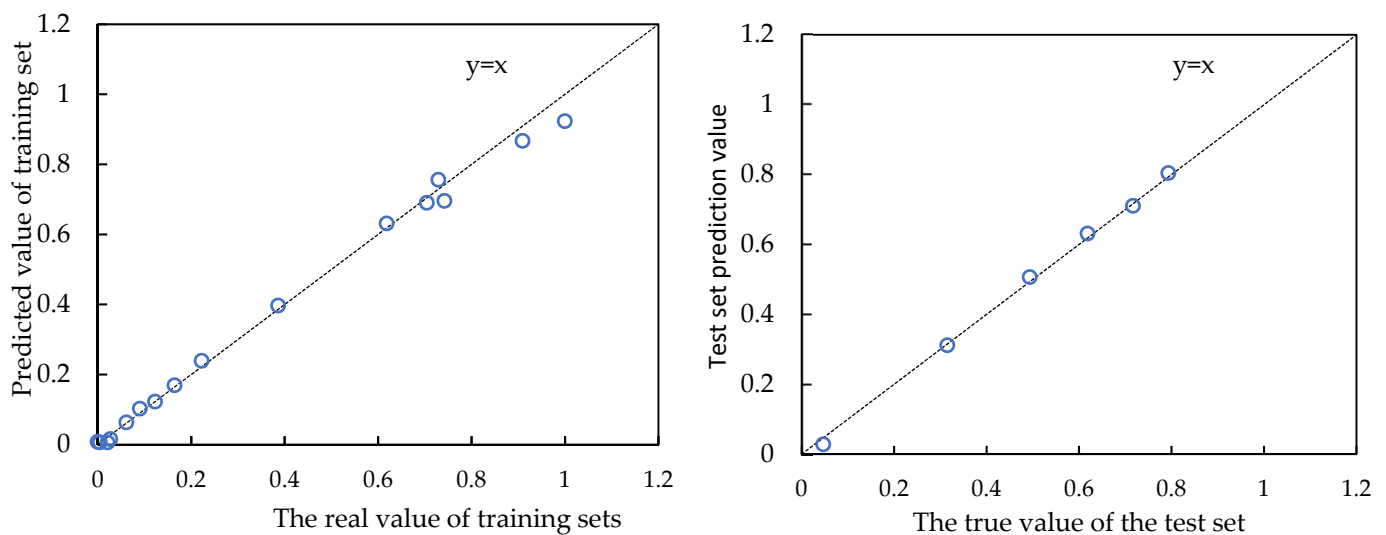


Figure 9. Comparison diagram of training set and prediction set fitting.

From the figure, we can see that in the predicted value image of the training set and the prediction set, in the fitting relationship between the scatter points of the predicted value and the real value of the training set, and the prediction set and the linear fitting line, we can compare and observe that the fitting effect of the training set and the test set is very significant and the scatter points are in or very close to the linear fitting line, which indicates that the model is very accurate in the prediction results of the training set.

- (3) Output analysis of the prediction results of the training set and test set. The prediction results of the training set and the test set are shown in Table 8.

Table 8. Test results of training set and prediction set.

	R^2	MAE	MBE	RMSE
training sets	0.99424	52,617.2825	−20,266.4305	78,755.4264
testing set	0.99776	32,936.508	2942.7318	35,812.9792

The training set R^2 is 0.99424, and the test set R^2 is 0.99776, which are very close to 1, indicating that the model fits well in training. The MAE of the training set is 52,617.2825, and the MAE of the test set is 32,936.508. The MAE of the test set is lower, indicating that the average absolute value of the prediction error is smaller and the prediction is more accurate. The MBE of the training set is −20,266.4305, indicating that the average predicted value of the model on the training set is smaller than the true value. The MBE of the test set is 2942.7318, indicating that the average predicted value of the model on the test set is larger than the true value. The RMSE of the training set is 78,755.4264, and the RMSE of the test set is 35,812.9792. The RMSE of the test set is much smaller than that of the training set, which again indicates that the model has better prediction performance on the test set. Based on the above analysis, the following conclusions can be drawn: the model performs well on the training set but performs better on the test set.

- (4) Analysis and forecast of freight volume in 2019–2021. The forecast results of China's freight volume in 2019–2021 are shown in Table 9.

Table 9. Forecast results of China's freight volume from 2019 to 2021.

	2019	2020	2021
Forecast (ten thousand tons)	3,837,847.547	3,827,709.298	3,968,614.472
Actual (ten thousand tons)	3,435,480	3,426,413	3,913,889

We use the data from 1997 to 2013 as the training set, and select the data from 2014 to 2018 as the test set to predict the freight volume in China from 2019 to 2021. It is predicted that the freight volume of China's highways in 2019 will be 38,378.47547 million tons. In 2020, the freight volume of China's highways will be 3,827,709.298 million tons. In 2021, the freight volume of China's highways will be 3,968,614.472 million tons. The actual freight volume of China's highways will be 3,435,480 million tons in 2019. In 2020, China's road freight volume will be 3,426,413 million tons, and in 2021, China's freight volume will be 3,913,889 million tons. It can be seen from the comparison that the predicted Chinese road freight volume from 2019 to 2021 is higher than the actual Chinese road freight volume.

5. Discussion

The reason for the phenomenon that the forecast of China's road freight volume is higher than China's actual road freight volume is due to the outbreak of the new coronavirus epidemic in 2019; the mobility of personnel is reduced, people voluntarily reduce travel, reduce online shopping, or reduce the type of goods purchased to reduce the indirect infection caused by freight transportation. National policies during the epidemic affected road freight volume. During the epidemic, domestic economic activities have been seriously affected. Domestic production capacity has recovered slowly, and investment and consumption have been in a downturn. Many enterprises have experienced problems such as supply chain disruption and employee shortages during the epidemic, and they cannot maintain a normal production rhythm. This has directly reduced the number of goods that need to be transported. Enterprises are on the verge of bankruptcy due to the serious impact of the epidemic, which has affected the volume of road freight transportation. At the same time, transportation cost is also a factor worth considering. In recent years, due to the increase in fuel price and labor cost, the proportion of road freight transportation has decreased.

In addition, road freight has certain limitations. Compared with other modes of transportation, road freight has the disadvantages of less cargo capacity and high transportation costs and is not suitable for long-distance transportation. Compared with other freight modes that are substitutes for each other, when the substitutes for road freight are more cost-effective, out of the choice of maximum value, shippers often choose the substitutes for road freight, which is also a reason for the decline in road freight volume. Finally, at the national level, the state strongly advocates the development of multimodal transport in 'road to rail' and 'road to water', resulting in a slow downward trend in the proportion of road freight volume.

6. Comparative Analysis

Based on the analysis of historical data and the influencing factors of highway freight volume, the BP neural network is used to establish the time series prediction model of highway freight volume in Nanning, and the future highway freight volume in Nanning is predicted. The historical data of highway freight volume in Nanning from 2002 to 2013 is used as the research object, and the data are divided into two parts. The data from 2002 to 2012 are used as the training sample of the neural network, and the data in 2013 are used as the test sample to complete the training and fitting of the model. The historical data from 2003 to 2013 are selected as the input sample to predict the highway freight volume in 2014. After multiple evaluations, the moving average method is used to select the predicted data. The predicted data in 2014 are compiled into the input data to predict the highway freight volume in 2015, and the polynomial is used to fit the trend of the original data. Through the analysis of the development trend of social freight volume in Nanning's historical years, the prediction model expression is established to determine the development trend of freight volume in Nanning's future years. The correlation coefficient between highway freight volume and prediction years is 0.9779, while the correlation coefficient of the training set in this paper is 0.99424 and the correlation coefficient of the test set is 0.99776. From this point of view, the fitting effect of the BP neural network based on multiple factors is more accurate and the prediction accuracy is higher [29].

7. Conclusions

The innovation of this study lies in the establishment of a multi-layer feedforward neural network to build a road freight volume prediction model based on the BP neural network. A variety of influencing factors such as road diesel consumption, population, gross domestic product, road freight turnover, road freight volume, etc., are used as input variables, and road freight volume is used as the output variable. Using the learning and simulation capabilities of the neural network, a complex mapping relationship between input and output is established. Multi-factor prediction based on the BP neural network can identify the interaction and correlation between different factors so as to predict the output more accurately. It has advantages in dealing with complex multi-factor prediction problems. It can not only achieve good performance on the training set but also effectively predict new and unseen data. In this study, through the establishment of BP neural network, the multi-factor prediction method is used to accurately fit the original data, and finally the predicted value of China's highway freight volume is obtained. An effective prediction of China's freight volume has been made, and a high prediction accuracy has been achieved. The model has a small error, meets the prediction requirements for future freight volume, meets the needs of economic development, meets the needs of regional and social development, and provides a certain reference for correctly predicting the development trend of future freight volume.

The significance of highway freight volume prediction involves many aspects. It is conducive to enterprises to grasp the development trend of the market, understand the changes in demand, and correctly formulate development strategies to cope with market changes, reduce operating costs, improve transportation efficiency, and gain a foothold in the market. Through the prediction of highway freight volume, it is conducive to the

rational planning of highway operation lines, strengthening the planning and management of highway vehicles and rational planning of traffic flow. For the case analysis of road freight volume in this study, we can improve from the following aspects, such as increasing the vertical number of data and horizontal influencing factors and increasing the sample size of the predicted data. While selecting the influencing factors, we should consider the correlation with freight volume. In the future, we can try to enrich the influencing factors of freight volume and increase the spatial dimension characteristics so as to enrich the influencing factors of China's highway freight volume. In addition, we can also adjust the number of training rounds and select the appropriate model.

Author Contributions: Methodology, M.W.; software, C.T.; writing—original draft preparation, Y.Z.; writing—review, Z.Z.; writing—editing, B.G.; writing—polishing, K.M. All authors have read and agreed to the published version of the manuscript.

Funding: This study did not receive any external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in the article.

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

References

1. Liang, R.; Wu, S. Problems and Improvement of Trial Scheme for Highway Freight Volume Statistics. *J. Highw. Transp. Res. Dev.* **2018**, *12*, 103–110. [[CrossRef](#)]
2. Muhin, M.; Luburić, G.; Rajsman, M. Dynamics and trends of the development of transport relations in road freight transport. *Teh. Vjesn.* **2017**, *24*, 635–642.
3. Alises, A.; Vassallo, M.J.; Guzmán, F.A. Road freight transport decoupling: A comparative analysis between the United Kingdom and Spain. *Transp. Policy* **2014**, *32*, 186–193. [[CrossRef](#)]
4. Zhang, Y. Highway freight volume forecasting based on wavelet kernel-based support vector machine. *J. Conver. Inf. Technol.* **2013**, *8*, 132–138.
5. Ivaković, Č.; Stanković, R.; Šafran, M. Transport Exchange as a Factor in Optimising International Road Freight Traffic. *Promet-Traffic Transp.* **2012**, *10*, 303–308.
6. Kveiborg, O.; Fosgerau, M. Decomposing the decoupling of Danish road freight traffic growth and economic growth. *Transp. Policy* **2006**, *14*, 39–48. [[CrossRef](#)]
7. Zhang, Y.Y.; Eli, I. Freight Volume Prediction in Xinjiang Based on BP Neural Network. *Transp. Res.* **2011**, *17*, 144–147.
8. Luo, P.; Zhu, L. A multifactor-based forecasting study of national railroad freight volume. *China High-Tech. Enterp.* **2007**, *10*, 29–30.
9. Chen, S. The Research of Forecast Method and Application of Cargo. Ph.D. Thesis, Wuhan University of Technology, Wuhan, China, 2008.
10. Yang, Q.Q. The Research on the Prediction of the Highway Traffic Volume in the Integrated Transportation Channel of the Yangtze River. Ph.D. Thesis, Chongqing Jiaotong University, Chongqing, China, 2016.
11. An, Y.E. Highway Freight Volume Forecasting Based on Rough Set and Grey Theory. Ph.D. Thesis, Lanzhou Jiaotong University, Lanzhou, China, 2017.
12. Wu, X.L.; Fu, Z. Research on BP Neural Network Prediction of Railway Freight Transportation Volume Based on Multiple Factors. *Railw. Freight Transp.* **2009**, *10*, 33–36+5.
13. Ji, Y.Z.; Feng, Y.H.; Wang, S.Z. Forecast of road freight volume in Changchun. *J. Chang. Univ. Technol.* **1997**, *3*, 27–31.
14. Zhao, J.Y.; Zhou, S.F.; Cui, X.J.; Wag, G.Q. Predictive method of highway freight volume based on fuzzy linear regression model. *J. Traffic Transp. Eng.* **2012**, *12*, 80–85.
15. Wang, Y.; Shao, C.F. Research on the Forecasting of Road Freight Volume Based on Support Vector Machine. *Logist. Technol.* **2010**, *29*, 142–144+15.
16. Chen, D.Q.; Chen, Y.C. Research on Combined Forecast of Highway Freight Transportation Volume in Fujian Province. *Sci. Technol. Inf.* **2008**, *26*, 187–188.
17. Meng, G.M.; Hu, Y.J.; He, Y.M. Forecast and analysis of road freight transportation volume in Mudanjiang Nongken sub-bureau. *Commun. Sci. Technol. Heilongjiang* **2008**, *06*, 94–95.
18. Gai, C.Y.; Pei, Y.L. Study of the gray model MARKOV chain model on highway freight forecast. *China J. Highw. Transp.* **2003**, *03*, 114–117.
19. An, R.; Hua, G.; Dong, N. Freight Volume Forecasting in Nanning Based on BP Neural Network. *Transp. Res.* **2015**, *1*, 58–64.
20. Zheng, C.W.; Lin, L.H.; Tian, R. Study on Forecasting of Highway Freight Transportation Volume of Hohhot. *Logist. Technol.* **2015**, *34*, 195–197.

21. Zhou, Y. Research on the Influential Factors of Urban Road Freight Volume Based on Variable Precision Rough Set-Taking Chongqing as an Example. Ph.D. Thesis, Chongqing Jiaotong University, Chongqing, China, 2014.
22. Fan, Z.M.; Zhao, L.W.; Wang, Y.P. Modeling and Forecasting of Road Freight Transportation in China. *Times Financ.* **2014**, *03*, 320–321.
23. Cao, Y.G.; Wu, F.P. Application of Combinatorial Forecasting to Highway Freight Volume in Suzhou. *Logist. Technol.* **2013**, *32*, 264–265+388.
24. Wu, H.Y.; Liu, X.; Liu, Y.; Lin, R.P. Research on Highway Freight Volume Prediction under the Background of “Double Carbon” in the Post-COVID-19 Epidemic. *Logist. Sci.-Tech.* **2022**, *45*, 88–93.
25. Xue, F.; Su, R.F.; Yang, S.; Yao, Y.Z.; Zhang, J. Prediction of highway freight transport volume based on regression analysis. *Automob. Appl. Technol.* **2019**, *15*, 65–69.
26. Chen, B. Research on the Application of Novel Machine Learning Methods in Urban Highway Freight Transportation Volume Forecasting. Ph.D. Thesis, Shenzhen University, Shenzhen, China, 2019.
27. Gong, D.F.; Tian, Q.M. Design and Application of Estimating Measures for Highway Freight Volume in Zhejiang Province via Large Data. *J. Wenzhou Polytech.* **2018**, *18*, 46–50.
28. Yang, T. Research on Highway Freight Volume Forecast in Yunnan Province Based on Weighted Combination Model. Ph.D. Thesis, Chongqing Jiaotong University, Chongqing, China, 2016.
29. Jin, H.; Zhang, J.; Wang, Z.S. Demand Forecasting Empirical Study of Highway Freight Traffic Volume Based on Combined Forecasting—Taking Tangshan City as an Example. *Stat. Manag.* **2016**, *1*, 52–55.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.