ISSN 2616-7433 Vol. 6, Issue 9: 52-58, DOI: 10.25236/FSST.2024.060909

Research on Carbon Emission Prediction of Qinghai Province Based on Machine Learning

Shasha Xiong[1,a], Xinyi Zhang[1,b], Bo Yang[1,c]

[1]*College of Economics and Management, Qinghai Normal University, Xining City, Qinghai Province, 810008, China*
[a]*18723728914@163.com,* [b]*1211513982@qq.com,* [c]*bo9818@qq.com*

Abstract: *This study aims to predict carbon emissions in Qinghai Province using advanced machine learning techniques to aid in achieving sustainable development goals. A "top-down" carbon emission calculation model was developed, revealing that the carbon emissions from Qinghai's transportation sector increased significantly from 2004 to 2022. To further analyze the influencing factors, an improved STIRPAT model was constructed, taking into account variables such as population, economic growth, energy consumption, and transportation development. The study also employed a Long Short-Term Memory (LSTM) neural network for prediction, comparing its performance with other models such as BP neural network and Support Vector Machines (SVM). The results show that the LSTM model achieved the best predictive accuracy, indicating that Qinghai's carbon emissions will peak around 2036. This research provides valuable insights into carbon emission dynamics and offers practical guidance for policy-making and environmental management in the region.*

Keywords: *Carbon Emissions, Machine Learning, STIRPAT Model, LSTM, Qinghai Province*

1. Introduction

With the increasing urgency of addressing climate change, accurately predicting carbon emissions has become a pivotal task for policymakers and researchers alike. Qinghai Province, located in northwestern China, presents unique challenges and opportunities in this context. The region's economic development, heavily reliant on resource extraction and transportation, has led to a steady increase in carbon emissions over the past two decades. Understanding and forecasting these emissions is crucial for formulating effective mitigation strategies and ensuring sustainable growth. Despite the progress made in recent years, existing models often fail to capture the complex interplay between economic, environmental, and social factors that influence carbon emissions in Qinghai[1, 2].

The application of machine learning techniques in carbon emission prediction offers a novel approach to addressing these complexities. Traditional statistical models, while useful, are limited by their linear assumptions and inability to handle large, multidimensional datasets. Machine learning models, particularly neural networks such as Long Short-Term Memory (LSTM), have shown great promise in capturing non-linear relationships and temporal dependencies within data[3, 4]. This study leverages the strengths of machine learning to develop a robust predictive model tailored to the unique characteristics of Qinghai's carbon emissions, considering factors such as population growth, economic activity, energy consumption, and infrastructure development.

This research not only aims to provide an accurate prediction of future carbon emissions in Qinghai Province but also seeks to identify the primary drivers behind these emissions. By integrating socio-economic and environmental variables into a comprehensive model, the study contributes to a deeper understanding of the region's carbon dynamics[5, 6]. The findings are expected to support policymakers in designing targeted strategies to achieve carbon neutrality and align with China's broader goals of peak carbon and carbon neutrality by 2060. This study also lays the groundwork for future research in applying advanced data science techniques to environmental management and policy evaluation.

ISSN 2616-7433 Vol. 6, Issue 9: 52-58, DOI: 10.25236/FSST.2024.060909

2. Model Design

2.1. "Top-Down" Carbon Emission Calculation Model

This study investigates the trend of transportation carbon emissions in Qinghai Province using a "top-down" approach, combined with fuel consumption to calculate the transportation carbon emissions in Qinghai Province. The research considers that the carbon emissions from energy consumption in the transportation industry mainly originate from the combustion of fossil fuels. It is assumed that all the energy is completely burned; therefore, the physical consumption data of major fossil fuels are used to estimate carbon emissions. Since electricity and thermal energy do not produce carbon dioxide in the end-use consumption of the transportation industry, the carbon emissions from electricity and thermal energy are excluded when calculating the transportation carbon emissions of Qinghai Province.

Based on the carbon emission coefficients provided in the "IPCC 2006 Guidelines for National Greenhouse Gas Inventories" and the average values calculated by Professor Xu Guoquan, the carbon emission coefficients for various energy types are summarized in Table 1.

Table 1: Conversion Coefficients of Various Energy Sources and Carbon Emission Factors

Energy Type	Unit	Standard Coal Conversion Coefficient	Carbon Emission Factor
Crude Oil	10,000 tons	1.4286	0.5857
Natural Gas	billion m^3	1.33	0.4483
Diesel	10,000 tons	1.457	0.592
Gasoline	10,000 tons	1.471	0.553
Coke	10,000 tons	0.9714	0.855
Kerosene	10,000 tons	1.471	0.571
Coal	10,000 tons	0.7143	0.7559

2.2. Influencing Factors Selection Based on the Improved STIRPAT Model

The improved STIRPAT model, derived from the IPAT model, allows for the inclusion of multiple independent variables related to scale, structure, and technology in the model. The improved STIRPAT model provides a more flexible framework for exploring and quantifying the specific contributions of these factors to environmental impacts.

By using the STIRPAT model and selecting its influencing factors, it is possible to better understand and analyze the multiple driving forces of environmental impacts, providing a scientific basis for the formulation of effective environmental policies and measures.

2.3. Carbon Emission Prediction Model Based on LSTM Neural Network

The Long Short-Term Memory (LSTM) network is a deep learning model for time-series data based on artificial intelligence. The LSTM neural network is an improved form of gated RNN, which constructs special memory storage units. Through backpropagation through time, it ensures that the gradients do not vanish or explode as the information is transmitted along the time-series path. This model can change at every moment, and by accumulating sequential information, it selectively removes unnecessary conditions, solving the problems of gradient vanishing and long-term dependency in traditional neural networks. This makes it more effective in time-series prediction. The specific structure is shown in Figure 1.

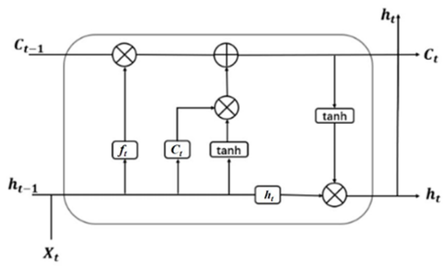

Figure 1: LSTM Model Structure

ISSN 2616-7433 Vol. 6, Issue 9: 52-58, DOI: 10.25236/FSST.2024.060909

Through the above process, the LSTM neural network can achieve better long-term dependency capabilities compared to the RNN neural network, allowing it to handle time-series data more effectively. The LSTM-based carbon emission prediction model can capture the long-term dependency relationships within carbon emission data and its influencing factors. It can adapt to different types of data variables, including nonlinear relationships and complex time dynamics, providing high flexibility. The LSTM model can also be trained in an online learning mode, enabling real-time prediction, which is crucial for carbon emission monitoring and forecasting as environmental policies and market conditions may change rapidly, requiring models that can respond quickly.

To build an effective prediction model, it is essential to preprocess the data set properly. This involves normalizing the data, handling missing values, and dividing the data into training, validation, and test sets. By doing so, the model's accuracy and generalization capabilities can be significantly improved. In this study, 70% of the data is used for training, 15% for validation, and 15% for testing.

3. Carbon Emission Calculation Method and Selection of Influencing Factors

3.1. Carbon Emission Calculation Method

A "top-down" carbon dioxide emission calculation model was established to calculate the annual carbon dioxide emissions from the transportation industry in Qinghai Province from 2004 to 2022, as shown in Table 2.

Table 2: Carbon Dioxide Emissions in Qinghai Province (2004-2013) (100 million tons)

Year	*CO_2 Emissions (100 million tons)*
2004	*0.239*
2005	*0.255*
2006	*0.321*
2007	*0.361*
2008	*0.436*
2009	*0.423*
2010	*0.453*
2011	*0.531*
2012	*0.604*
2013	*0.658*
2014	*0.614*
2015	*0.567*
2016	*0.652*
2017	*0.62*
2018	*0.597*
2019	*0.597*
2020	*0.724*
2021	*0.751*
2022	*0.778*

As shown in Table 2, the carbon dioxide emissions in Qinghai Province increased from 23.9 million tons in 2004 to 77.8 million tons in 2022, representing a growth of approximately 226% over the 18-year period. This significant increase in carbon dioxide emissions from the transportation industry in Qinghai Province indicates that the province faces challenges in environmental policy and management, necessitating more effective policy and technological measures to address environmental challenges and promote sustainable development.

3.2. Selection of Influencing Factors for Carbon Emissions

This study selected nine variables from four dimensions—population, economy, environment, and transportation—as the main factors influencing carbon emissions in the transportation industry and expanded the STIRPAT model accordingly. The expanded variables are as follows: Population dimension includes total year-end population and urbanization rate; economic dimension includes per capita GDP and added value of the tertiary industry; environmental dimension includes green coverage rate and urban green area; transportation dimension includes motor vehicle ownership, passenger turnover, and freight turnover.

ISSN 2616-7433 Vol. 6, Issue 9: 52-58, DOI: 10.25236/FSST.2024.060909

The total population directly affects energy consumption and carbon emissions. A larger population means higher demand for basic needs such as housing, food, and transportation, leading to increased energy use and carbon emissions. A high urbanization rate indicates more people moving from rural areas to cities. Urbanization is often accompanied by industrialization and modernization, which increase energy consumption and carbon emissions. Urban lifestyles typically rely more on fossil fuels, such as cars and electricity, than rural lifestyles. The changes in population and urbanization rate in Qinghai Province from 2004 to 2022 are shown in Figure 2 and Figure 3.

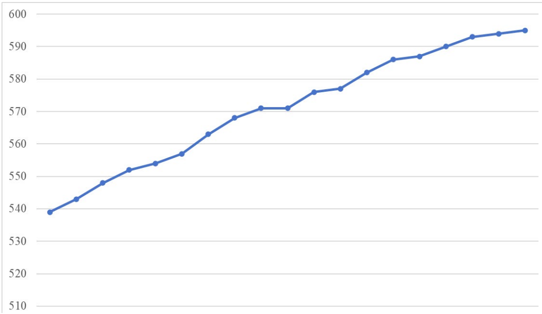

Figure 2: Population Change

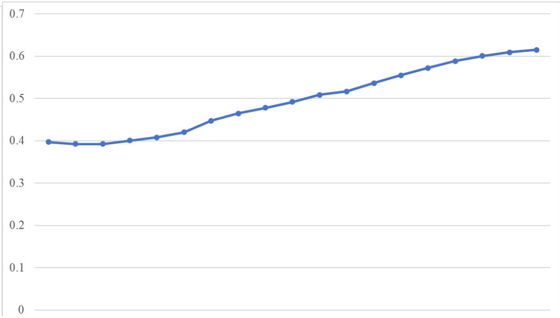

Figure 3: Urbanization Rate Change

Per capita GDP is an important indicator of economic activity. Increases in economic activity are usually accompanied by increased energy consumption, especially in industrial production and the service sector. Therefore, the growth in per capita GDP is often related to an increase in carbon emissions. The changes in per capita GDP and added value of the tertiary industry in Qinghai Province from 2004 to 2022 are shown in Figure 4 and Figure 5 below.

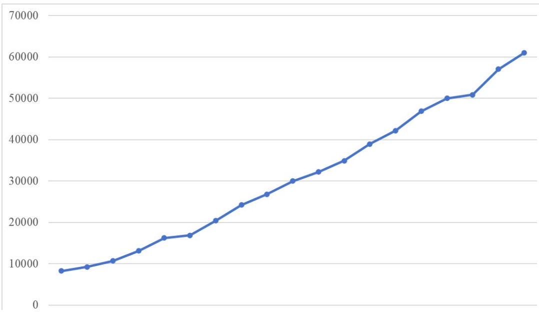

Figure 4: Per Capita GDP Change

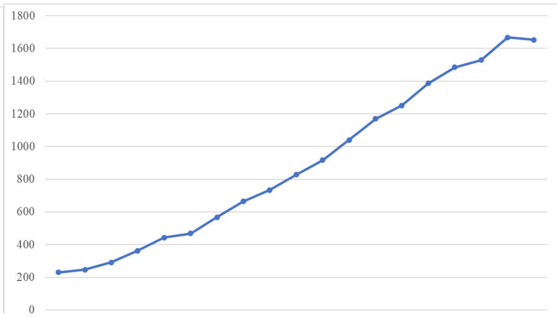

Figure 5: Added Value of the Tertiary Industry

Regions with a high green coverage rate can absorb more carbon dioxide through photosynthesis, thereby helping to reduce atmospheric carbon content. Moreover, good greenery helps to improve the urban microclimate and reduce the urban heat island effect, indirectly reducing the use of cooling equipment such as air conditioners, thereby reducing energy consumption and carbon emissions. Similar to the green coverage rate, increasing urban green area also helps in carbon absorption and reducing the urban heat island effect, which is positive for reducing carbon emissions. The changes in green coverage rate and urban green area in Qinghai Province from 2004 to 2022 are shown in Figure 6 and Figure 7 below.

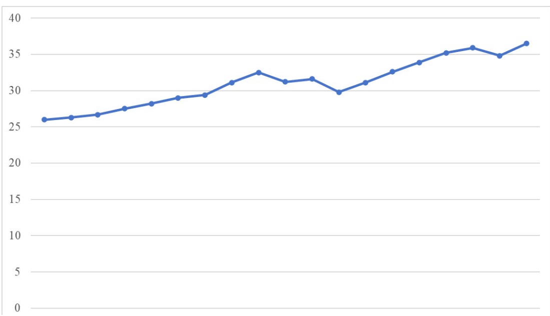

Figure 6: Green Coverage Rate in Built-up Areas

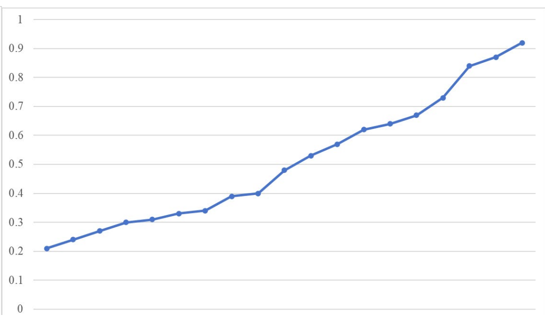

Figure 7: Urban Green Area

Motor vehicles are one of the main sources of urban carbon emissions. An increase in motor vehicle ownership directly leads to increased fuel consumption, thus increasing carbon emissions. Passenger

ISSN 2616-7433 Vol. 6, Issue 9: 52-58, DOI: 10.25236/FSST.2024.060909

turnover and freight turnover reflect the scale of transportation activities. The larger the volume of passengers and goods transported, the higher the required energy consumption and corresponding carbon emissions, especially for fossil fuel-dependent modes of transport such as cars, airplanes, and ships. The changes in motor vehicle ownership, freight turnover, and passenger turnover in Qinghai Province from 2004 to 2022 are shown in Figure 8 and Figure 9 below.

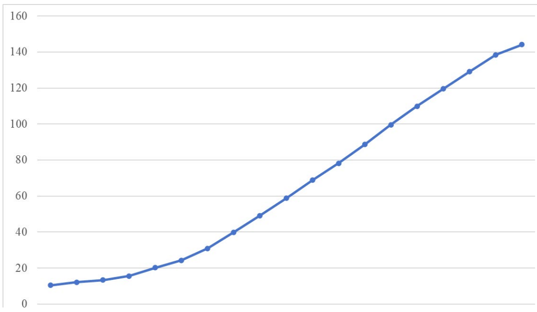

Figure 8: Change in Motor Vehicle Ownership

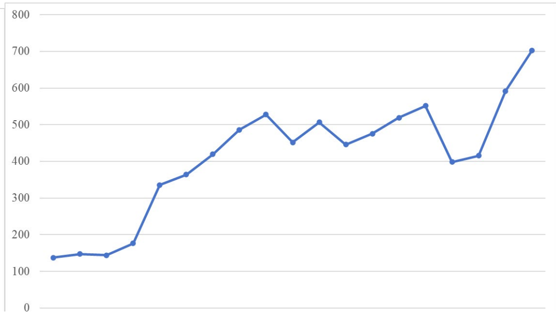

Figure 9: Change in Freight Turnover

By comprehensively analyzing the impact of economic, population, environmental, and transportation dimensions on carbon emissions, the study helps quantify the specific contributions of each factor to carbon emissions, thereby enabling the formulation of more effective targeted emission reduction strategies.

3.3. Correlation Analysis Model

Correlation analysis was conducted on the variables selected in this study to describe the relationship between various linear variables using the Pearson correlation coefficient. The calculation method of the Pearson correlation coefficient is as follows:

$$r=\frac{\sum(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum(x_i-\overline{x})^2}\sqrt{\sum(y_i-\overline{y})^2}}$$

The magnitude of the correlation coefficient reflects the strength of the correlation between different variables. The range of the correlation coefficient is [-1, 1]. The closer the absolute value of the correlation coefficient is to 1, the stronger the correlation between the variables; the closer it is to 0, the weaker the correlation between the variables. The positive or negative sign of the correlation coefficient indicates whether the relationship between the variables is positive or negative.

The results of the correlation analysis show that the correlation coefficients between carbon emissions and year-end permanent population, urbanization rate, added value of the tertiary industry, per capita GDP, green coverage rate of built-up areas, urban green area, motor vehicle ownership, and freight turnover are all greater than 0.8, indicating a high correlation between carbon emissions and these influencing factors. The correlation coefficient between carbon emissions and passenger turnover is 0.5, indicating a moderate correlation. Among them, economic and population indicators are highly correlated with carbon emissions, suggesting that as economic development and urbanization progress, urban transportation carbon emissions have increased significantly. The correlation analysis results for the transportation dimension show that the correlation between freight turnover and carbon emissions is also high (correlation coefficient of 0.905), indicating that an increase in freight activities also leads to an increase in carbon emissions, while the correlation between passenger turnover and carbon emissions is relatively low (correlation coefficient of 0.5). This suggests that the direct impact of passenger transportation on carbon emissions may not be as significant as freight transportation, but it still has a considerable impact on carbon emissions.

In summary, all variables have reached high levels of correlation coefficients, showing strong statistical evidence supporting the existence of correlations among these variables. This indicates that the influencing factors of transportation carbon emissions selected based on the improved STIRPAT model are reasonable and can be further studied.

3.4. Traffic Carbon Emission Prediction Model Based on LSTM

First, the selected nine influencing variables and carbon emission time series data are normalized to remove dimensional differences. Then, the matrix composed of the values from the previous T periods

ISSN 2616-7433 Vol. 6, Issue 9: 52-58, DOI: 10.25236/FSST.2024.060909

of the processed sequence data is input into the LSTM model for training, and the T+1 period data is used as the output error correction for comparison, making it consistent with a batch of supervised learning data sets of the LSTM model, where T is the training step (time_step).

In this study, data from 2004 to 2016 were used as the training set, and data from 2017 to 2022 were used as the validation set. Considering the small sample size, the number of hidden layers (layer_size) of the LSTM model was set to 1, the training step T was set to 3, the number of training epochs (epoch) was set to 500, the optimizer (optimizer) was selected as Adam, and the loss function was selected as MAE (Mean Absolute Error). To achieve better prediction results, the number of hidden nodes (hidden_size) and batch size (batch_size) of the LSTM model were adjusted in sequence, and each group of hidden nodes and batch size was trained 10 times. The model with the best performance was selected based on model evaluation indicators. The selection range of hidden nodes and batch size was [1, 6].

To evaluate the model's prediction ability and accuracy, RMSE (Root Mean Square Error), MAE (Mean Absolute Error), and MAPE (Mean Absolute Percentage Error) were used to analyze the prediction model. The evaluation indicators of the LSTM model are shown in Table 3.

Table 3: Evaluation Indicators of the LSTM Model

Sample	Indicator	Value
Training Set	R^2	0.984
	RMSE	0.02
	MAE	0.016
	MAPE	0.001
Test Set	R^2	0.892
	RMSE	0.042
	MAE	0.036
	MAPE	0.011

As shown in Table 3, the LSTM model performs very well in fitting the training set, with RMSE = 0.020 in the training set and RMSE = 0.042 in the test set. The training set indicators are R^2 = 0.984, MAE = 0.016, and MAPE = 0.001, while the test set indicators are R^2 = 0.892, MAE = 0.036, and MAPE = 0.011. This shows that the LSTM model performs excellently on both the training set and the test set. Although the model's performance decreases slightly in the test set compared to the training set, which is mainly due to model generalization, the model still shows good predictive fitting ability, making it suitable for reasonable prediction of traffic carbon emissions.

To verify the prediction effect of different models on the CO2 emissions of the transportation industry in Qinghai Province, BP neural network, support vector regression (SVR) model, and convolutional neural network (CNN) were constructed for comparison with the LSTM model. The comparison results of each evaluation indicator are shown in Table 4.

Table 4: Comparison of Prediction Accuracy of Different Models

Sample	Indicator	BP	SVR	CNN	LSTM
Training Set	R^2	0.483	0.839	0.908	0.984
	RMSE	0.104	0.057	0.043	0.02
	MAE	0.076	0.045	0.034	0.016
	MAPE	17.555	10.286	7.856	0.001
Test Set	R^2	0.483	0.882	-0.611	0.892
	RMSE	0.823	0.103	0.096	0.042
	MAE	0.542	0.095	0.074	0.036
	MAPE	42.55	12.365	11.892	0.011

As shown in Table 4, the R^2 values of BP neural network, SVR, CNN, and LSTM in the test set are 0.483, 0.882, -0.611, and 0.892, respectively, with the LSTM model having the highest R^2 value in the test set. Meanwhile, the R^2 of LSTM in the training set is also the highest at 0.984, indicating that LSTM has the best fitting effect on both the test set and the training set. In terms of RMSE, the values of the LSTM model in the training set and test set are the smallest, at 0.020 and 0.042, respectively. In terms of MAE, the LSTM model has the lowest error in both the training set and the test set, at 0.016 and 0.036, respectively. In terms of MAPE, the LSTM model also has the lowest error, with the MAPE of the training set being 0.001 and the MAPE of the test set being 0.011. In summary, the LSTM model performs the best in both the training set and the test set, with the highest coefficient of determination and the lowest error indicators, making it the most stable and accurate model among the four models.

ISSN 2616-7433 Vol. 6, Issue 9: 52-58, DOI: 10.25236/FSST.2024.060909

4. Conclusions

The study shows that from 2004 to 2022, the carbon emissions in Qinghai Province have shown an overall upward trend, which has had a significant impact on economic development and living environment. Qinghai Province faces significant pressure to reduce emissions and save energy.

This paper first analyzes the current status of carbon emission research and adopts a "top-down" carbon emission calculation method for the transportation industry to estimate the carbon emissions of Qinghai Province from 2004 to 2022. Based on this, an improved STIRPAT model was established to analyze the influencing factors of Qinghai Province's transportation industry from four dimensions: population, economy, environment, and transportation. Then, based on the improved STIRPAT-LSTM model, the carbon emissions of the transportation industry in Qinghai Province from 2023 to 2040 were predicted. The results show that the carbon emissions of the transportation industry in Qinghai Province are expected to reach a "carbon peak" in 2036.

References

[1] Wang X. N., Gu K. P. "Research on the Current Situation of Carbon Source Emission Estimation Methods in China" [J]. Environmental Science and Management, 2006, 31(4): 78-80.

[2] Liu Z., Guan D., Wei W., et al. "Reduced Carbon Emission Estimates from Fossil Fuel Combustion and Cement Production in China" [J]. Nature, 2015, 524(7565): 335-338.

[3] Wang M., Zhou Z. X., Feng M. M., et al. "CO2 Emission Monitoring Model and Accuracy Verification of Thermal Power Units Using Direct Measurement Method" [J]. Coal Chemical Industry, 2022, 50(02): 18-21+33.

[4] Qin Z., Zhang J. E., Luo S. M., Ye Y. Q. "System Dynamics Prediction of Energy Consumption and CO2 Emissions in China" [J]. Chinese Journal of Eco-Agriculture, 2008, 66(04): 1043-1047.

[5] Chen B., Yang W. S. "Research on the Carbon Emission Accounting Method of Industrial Parks" [J]. China Population, Resources and Environment, 2017, 27(03): 1-10.

[6] Deng G. Y., Ma R., Zhang Z. J. "Decomposition Study on the Carbon Emission Effect of Energy Consumption in Various Industries in China" [J]. Statistics and Decision, 2018, 34(15): 124-127.