

Research on a Carbon Emission Prediction Model for Deep Foundation Pit Construction Based on the BP-GA Algorithm

Yang Song

*Department of Energy and Power Engineering, Henan Agricultural University, Zhengzhou, China
19838023092@163.com*

Abstract: To address the challenges posed by the multi-source heterogeneous characteristics of carbon emissions in deep foundation pit construction and the insufficient generalization capability of traditional prediction models, this study proposes a hybrid model based on a genetic algorithm-optimized BP (Backpropagation) neural network. The model employs an elite retention strategy and a single-point crossover mechanism to optimize the initial weights of the network. Additionally, it utilizes a two-layer neural network structure with 16 and 8 nodes, respectively, to achieve a nonlinear mapping from input features to carbon emissions. Validation using a dataset of over 1,000 engineering cases demonstrates the model's strong predictive performance, achieving a Mean Squared Error (MSE) of 2856.34 and a Mean Absolute Error (MAE) of 43.11. These results highlight the model's reliability as a prediction tool for carbon emissions in deep foundation pit projects, offering a promising approach for improving the accuracy of carbon footprint assessments in construction engineering.

Keywords: Carbon Emission, BP-GA Algorithm, Deep Foundation Pit

1. Introduction

In the face of the severe challenges of global climate change, the “dual carbon” strategy has become a key goal for economic development and environmental protection worldwide. As the world's largest developing country, China, upholding its responsibility as a major nation, has also proposed the “dual carbon” goals—striving to peak carbon dioxide emissions by 2030 and achieving carbon neutrality by 2060. According to the China Building Energy Consumption Research Report (2022), in 2020, China's building and construction energy consumption and carbon emissions reached 2.27 billion tons of standard coal equivalent (tce) and 5.08 billion tons, accounting for 45.5% and 50.9% of the national total, respectively [1]. The construction industry has become a major contributor to China's energy consumption and carbon emissions. Therefore, promoting the low-carbon transition of the construction sector is an essential pathway to achieving carbon peaking and carbon neutrality.

With the continuous advancement of urbanization, the utilization of urban underground space has received increasing attention, making deep foundation pit engineering an increasingly important part of urban construction. However, its high energy consumption and large carbon emissions have made it a focal point in the context of green development. Qin Ao et al., taking a station of Chengdu Metro Line 3 as an example, found that the carbon emissions from the materialization process of metro stations account for a significantly higher proportion of their total lifecycle emissions [2]. When the station's service life is set at 50 years, these emissions can reach up to 49% of the total. Therefore, promoting low-carbon development in deep foundation pit projects is imperative. However, different

engineering projects require different carbon emission prediction methods. Li Dezhi et al. reviewed the research status of building lifecycle carbon emission measurement from three aspects: measurement stages, measurement methods, and measurement tools [3]. Pan Yiqun et al. summarized and categorized the major building carbon emission calculation models used domestically and internationally based on fundamental modeling methods, which is of great significance for understanding and analyzing global carbon reduction efforts [4]. Zeng Xudong et al. proposed that low-carbon building design should integrate BIM (Building Information Modeling) technology with carbon emission analysis from the planning stage, which can facilitate the realization of low-carbon architecture [5]. Luo Zhixing et al. developed a multiple regression model to calculate embodied carbon emissions in building conceptual design, providing a new approach to carbon reduction from the design phase [6]. Guo Chun et al. established a predictive model for tunnel construction carbon emissions using linear regression analysis, offering a new method for regional carbon emission forecasting [7]. However, carbon emissions during construction are influenced by multiple factors, including construction equipment energy consumption, material usage, transportation, and construction methods, exhibiting highly nonlinear characteristics. Traditional prediction methods, such as linear models or empirical formula-based analysis, often fail to accurately capture these complex relationships, leading to significant prediction deviations and limiting their practical application in engineering projects. To address this, this paper proposes a carbon emission prediction model based on a BP neural network and a genetic algorithm. By combining the strong nonlinear mapping capability of the BP neural network with the global optimization ability of the genetic algorithm, the model enhances prediction accuracy and robustness, providing valuable insights for carbon emission forecasting and low-carbon development in deep foundation pit engineering.

2. Machine learning algorithm

2.1. Comparative analysis of algorithms

Although the traditional BP neural network is widely used in the field of machine learning, it has significant limitations in practical applications. Due to its gradient descent-based optimization mechanism, the training process is prone to getting stuck in local optima and is highly sensitive to the selection of initial network weights. Additionally, the traditional BP algorithm has a slow convergence speed and often exhibits oscillations during training. This instability becomes more pronounced when dealing with large-scale network structures, making the training process unreliable and susceptible to overfitting. For complex nonlinear problems such as carbon emissions in deep foundation pit construction, relying solely on gradient descent makes it difficult to find the global optimal solution.

While the genetic algorithm (GA) alone has strong global search capabilities, it also faces inherent limitations. First, when dealing with large-scale neural networks, GA has high computational complexity and requires significant time. Second, although it excels at global search, its performance in local optimization is relatively weak. In the later stages of convergence, the optimization efficiency declines significantly, making it difficult to achieve fine-tuned optimization in high-dimensional continuous parameter spaces, which ultimately affects the prediction accuracy.

To address the shortcomings of the two algorithms mentioned above, this study proposes a BP-GA hybrid algorithm. This method combines the global search capability of the genetic algorithm (GA) with the local optimization features of the BP algorithm. By using GA to optimize the initial network weights, the risk of getting trapped in local optima is effectively reduced. This hybrid strategy significantly decreases sensitivity to initial parameter selection, accelerates model convergence, and improves training stability. More importantly, while maintaining strong exploration capabilities, the BP-GA hybrid algorithm ensures optimization accuracy, making it particularly

suitable for handling complex nonlinear prediction problems such as carbon emissions in deep foundation pit construction.

2.2. Genetic algorithm optimization design

To enhance the global optimization capability of the BP neural network, this paper uses a genetic algorithm to pre-optimize the network weights. The algorithm adopts a real-valued encoding method, where the parameters to be optimized are encoded into chromosomes containing weight matrices $W1(5 \times 16)$, $W2(16 \times 8)$, $W3(8 \times 1)$, and their corresponding biases. The population size is set to 100, and the number of evolutionary generations is set to 200. The fitness function is defined as the reciprocal of the mean squared error (MSE): $f = 1/(MSE + 1e-10)$, where a small constant of $1e-10$ is added to avoid division by zero errors. In the selection process, an elite retention strategy is used, where the top 10 individuals with the highest fitness are directly retained for the next generation. The crossover operation randomly selects from the top 20 high-quality parents, and the offspring are generated by taking the arithmetic mean of the parents' genes. To maintain population diversity, a mutation operation is applied to individuals with a probability of 0.1. Gene mutation is achieved by adding Gaussian noise with a mean of 0 and a standard deviation of 0.1.

2.3. Neural network structure design

Considering the complexity of predicting carbon emissions in deep foundation pit construction, a feedforward neural network structure with two hidden layers is designed. The input layer of the network contains 5 feature nodes, corresponding to key influencing factors such as equipment energy consumption and material usage. The first hidden layer is configured with 16 neurons, and the second hidden layer with 8 neurons, both using the ReLU activation function to enhance the network's nonlinear expression capability. The output layer uses a single neuron with a linear activation function to directly output the predicted carbon emissions.

2.4. Model training strategy

The model is trained using the Adam optimizer, with a learning rate set to 0.001 to balance convergence speed and stability. To account for computational efficiency and memory usage, the batch size is set to 32. The BP model is trained for 500 epochs, with an early stopping strategy (patience=20) to prevent overfitting; the BP-GA hybrid model is trained for 200 epochs. The selection of these hyperparameters is based on multiple experimental comparisons, ensuring efficient convergence during training while maintaining model performance.

3. Carbon emission prediction model construction

3.1. Data Description

The raw data collected in this study includes five main factors affecting carbon emissions as input features:

1. Construction equipment energy consumption (unit: kWh): Reflects the electricity consumption during the operation of machinery.
2. Material usage (unit: tons): Includes the amount of major construction materials such as concrete and steel reinforcement.
3. Transportation distance (unit: kilometers): The cumulative distance of material and equipment transportation.
4. Construction method (categorical variable): 0 for open excavation method, 1 for top-down construction method.

5. Weather conditions (categorical variable): 0 for sunny days, 1 for rainy days.

The target output variable of the model is the carbon emissions per unit phase (unit: kg CO₂), which is obtained based on the engineering monitoring data.

Table 1: Carbon emission and energy consumption statistics of deep foundation pits.

Number	Equipment energy consumption	Material usage	Transportation distance	Construction method	Weather condition	Carbon Emission Per unit stage
1	1061.81	133.31	59.72	0	1	1155.71
2	1926.07	293.86	56.93	0	1	1987.61
3	1597.99	442.83	182.19	0	0	1870.55
4	1397.99	379.50	57.41	1	0	1725.29
5	734.03	412.95	61.67	0	0	1090.65
6	733.99	346.45	154.29	1	0	1145.35
7	587.13	361.52	95.45	1	1	1043.54
8	1799.26	432.14	157.58	0	0	1949.17
9	1401.67	162.35	22.42	0	1	1357.87
.....						
993	1542.04	121.51	162.00	1	0	1572.62
994	908.22	324.60	115.01	0	0	1170.58
995	1385.35	110.91	85.35	0	0	1340.43
996	1041.46	388.12	35.03	0	1	1370.73
997	637.37	345.63	174.41	1	0	1074.51
998	1875.97	480.48	39.88	1	1	2296.80
999	705.23	81.03	68.86	0	1	734.99
1000	1925.36	75.67	65.11	0	0	1689.98
1001	1169.01	176.98	175.57	0	0	1211.50

3.2. Data preprocessing

MinMaxScaler is used to normalize the input features and target values to the [0,1] range to eliminate the influence of dimensions. The dataset is randomly divided into a training set and a testing set in an 8:2 ratio, using a fixed random seed (42) to ensure the reproducibility of the experiment.

3.3. Model training process

In the genetic algorithm phase, the best fitness is recorded every 10 generations to monitor the population's evolutionary progress. During the BP network training process, the loss values and MAE metrics for both the training and validation sets are recorded and saved as training_history.xlsx for analyzing the model's convergence performance. The comparison between the predicted results and the actual values is saved in prediction_results.xlsx for subsequent analysis.

4. Result Analysis

This study conducted a systematic comparison of three models—BP neural network, genetic algorithm, and BP-GA hybrid algorithm—for carbon emission prediction in deep foundation pit construction. The experimental results indicate that the BP-GA hybrid algorithm significantly outperforms the individual algorithms across all evaluation metrics. In terms of prediction accuracy,

Fig. 1 shows that although the BP-GA algorithm exhibits relatively large errors in the initial epochs, both the training set and validation set errors rapidly decrease as the number of epochs increases and then stabilize. This trend is also reflected in Fig. 2, which illustrates the variation of the mean absolute error (MAE) during the training process, further indicating that the BP-GA algorithm achieves small errors. The BP-GA hybrid algorithm has a mean squared error (MSE) of 2856.34 kg CO₂, showing significant improvement compared to the BP neural network (3107.84 kg CO₂) and genetic algorithm (4932.46 kg CO₂), with reductions of 8.1% and 42.1%, respectively. A similar trend is observed in the root mean squared error (RMSE) comparison, where the BP-GA model has an RMSE of 53.44 kg CO₂, while the BP and GA models have RMSE values of 55.75 kg CO₂ and 70.23 kg CO₂, respectively. This indicates that the BP-GA hybrid algorithm effectively reduces the absolute error level of the predictions. In terms of prediction stability, as shown in Fig. 3 and 4, all data points are approximately distributed around the dashed line of equality (predicted values equal to true values), and small errors occur with a significantly higher frequency, indicating the algorithm's high prediction accuracy. The BP-GA hybrid algorithm has a MAE of 43.11 kg CO₂, which is noticeably lower than the BP model (43.35 kg CO₂) and the GA model (54.12 kg CO₂). More importantly, in terms of the mean absolute percentage error (MAPE), the BP-GA model achieves 3.29%, while the BP model and GA model have MAPE values of 3.26% and 4.26%, respectively. This indicates that the BP-GA hybrid algorithm maintains a high level of prediction accuracy across different prediction intervals.

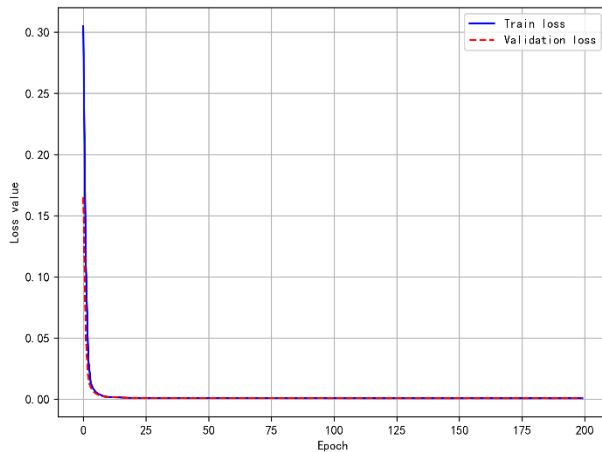


Figure 1: Loss curves during training process.

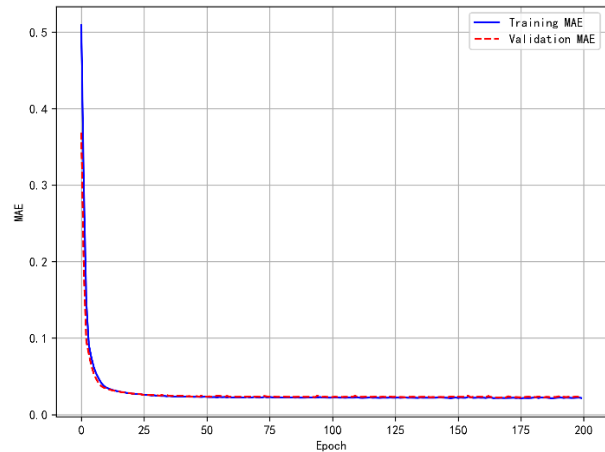


Figure 2: MAE variation curves.

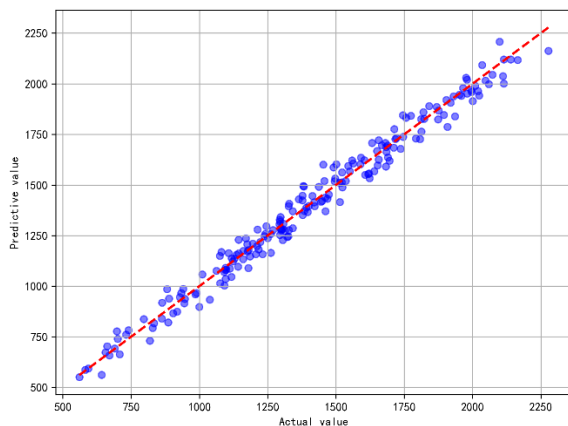


Figure 3: Scatter plot of predicted value vs actual value.

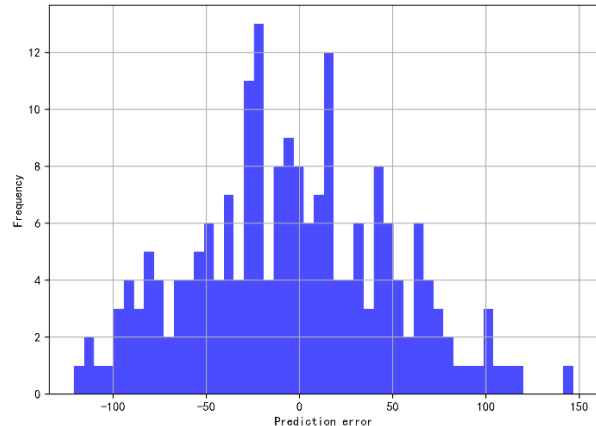


Figure 4: Histogram of prediction error distribution.

In terms of model goodness of fit, the coefficient of determination (R^2) of the BP-GA hybrid algorithm reaches 0.98, which is on par with the BP model and better than the GA model's 0.97. It can be seen from Fig.5 that the prediction points of the BP-GA model (green) are more closely clustered around the ideal prediction line, whereas the prediction points of the BP model (blue) and GA model (red) exhibit greater dispersion, particularly in high carbon emission areas, where prediction deviations are more pronounced. This indicates that the BP-GA hybrid model has better generalization capability when handling extreme values.

Error distribution analysis shows that the prediction errors of all three models exhibit characteristics of an approximately normal distribution, but with noticeable differences. From the error histogram, it can be observed that the BP-GA hybrid algorithm (green) has a more symmetric error distribution concentrated near the zero error line, as shown in Fig.6. Although the peak is not the highest, the overall distribution is more compact, with the error range mainly concentrated between ± 100 kg CO₂. In contrast, the BP model (blue) has a higher peak near the zero error line, but its overall error distribution is flatter. The GA model (red) has the most dispersed error distribution, with many outliers within the ± 200 kg CO₂ range, showing greater prediction uncertainty. From Fig.7 to Fig.9, it can be further observed that the residuals of the BP-GA model remain stable within the ± 100 kg CO₂ range across the entire prediction interval, whereas the residuals of the BP model and GA model reach ± 250 kg CO₂ and ± 200 kg CO₂, respectively. This visually confirms that the BP-GA model exhibits more stable prediction performance. Particularly in the tail of the error distribution, the individual algorithms show more extreme prediction deviations, reflecting their limitations in handling complex nonlinear relationships. Comparison chart of predicted carbon emissions and actual carbon emissions of three models.

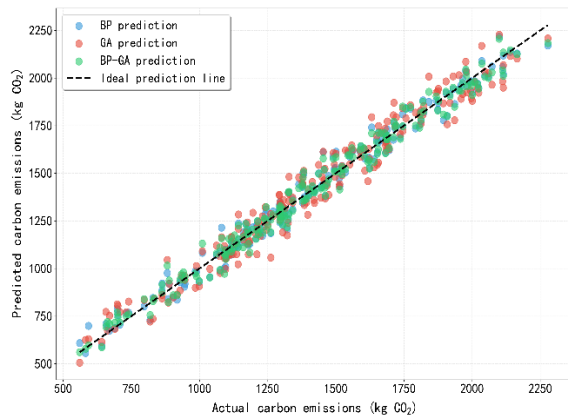


Figure 5: Comparison chart of predicted carbon emissions and actual carbon emissions of three models.

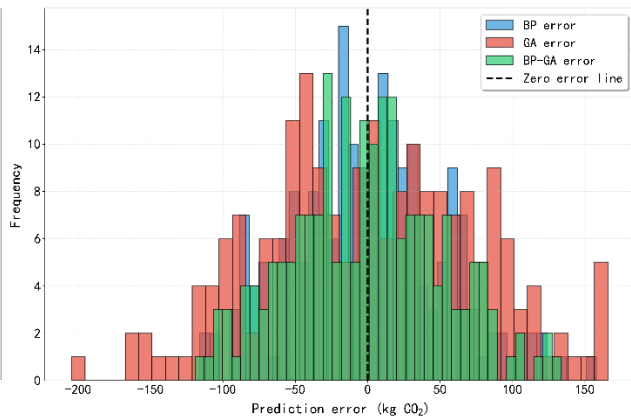


Figure 6: Histogram of error distribution in predicting carbon emissions for three models.

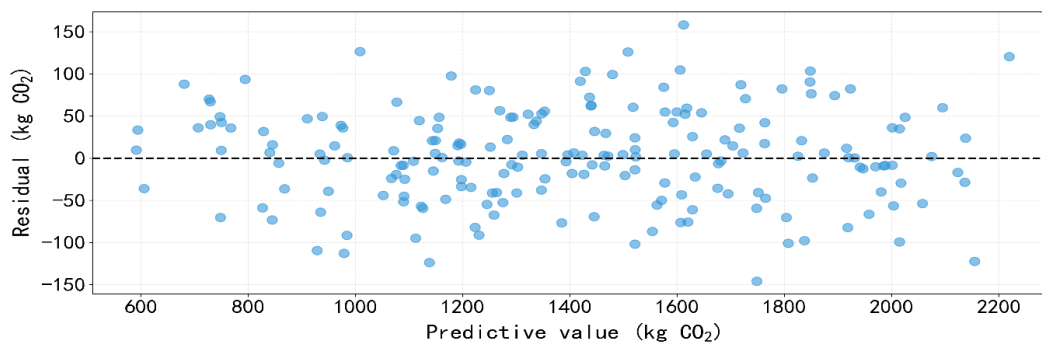


Figure 7: Residual distribution of BP model.

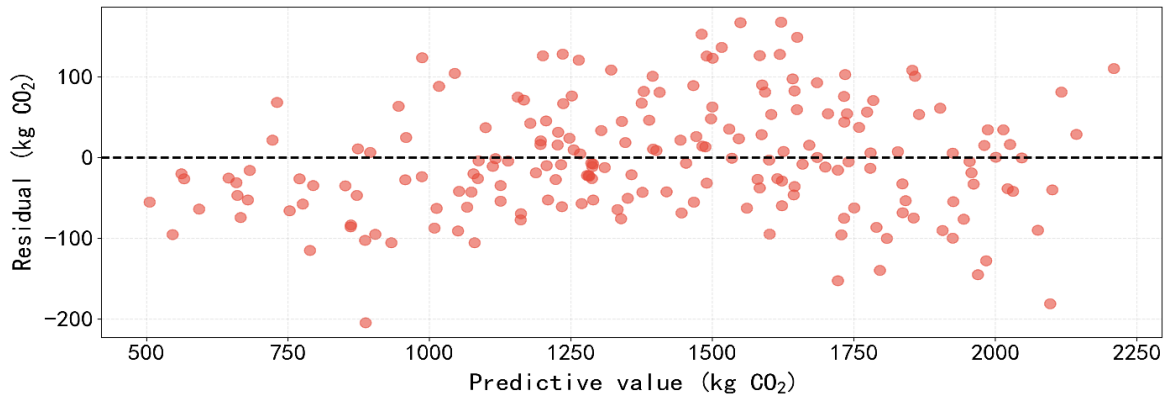


Figure 8: Residual distribution of GA model.

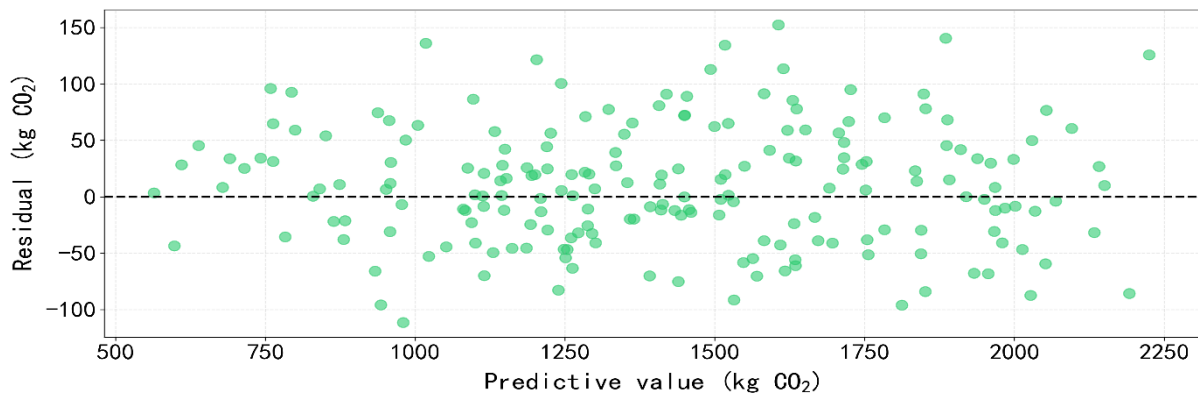


Figure 9: Residual distribution of BP-GA hybrid model.

In-depth analysis of the prediction errors reveals that the BP-GA hybrid algorithm not only performs excellently in overall prediction accuracy but also demonstrates significant advantages in the reliability and stability of the prediction results. The algorithm successfully combines the local optimization capability of the BP neural network with the global search capability of the genetic algorithm, effectively overcoming the inherent limitations of single algorithms when dealing with complex nonlinear issues such as carbon emissions in deep foundation pit construction. The experimental results fully demonstrate that the BP-GA hybrid algorithm can provide a more reliable and accurate technical solution for predicting carbon emissions in deep foundation pit construction.

5. Conclusion

(1) The GA-BP hybrid model developed in this study achieves accurate prediction of carbon emissions in deep foundation pit construction by optimizing the initial weights using the genetic algorithm (population size of 100, 200 iterations) and combining it with a neural network (16-8-1 structure).

(2) Compared to the traditional BP and GA single algorithms, the hybrid model reduces the MSE by more than 8.1%, with a higher correlation between the predicted and actual values and a more concentrated error distribution. The training strategy using the Adam optimizer (learning rate of 0.001) and batch training (batch_size=32) effectively balances computational efficiency and convergence stability.

(3) The current model exhibits larger prediction errors under extreme conditions. It is recommended to further improve the model's performance by expanding the dataset, optimizing the network structure, and refining feature engineering.

References

- [1] China Building Energy Efficiency Association, Professional Committee of Building Energy Consumption and Carbon Emission Data. 2022 China Building Energy Consumption and Carbon Emission Research Report . Chongqing: China Building Energy Efficiency Association, Professional Committee of Building Energy Consumption and Carbon Emission Data, 2022.
- [2] Qin,A.,Yuan,Y.P.,Jiang,F.J. (2020) Research on Life Cycle Carbon Emissions of Metro Stations: Taking a Station on Chengdu Line 3 as an Example. *Construction Economy*,41,329-334.
- [3] Li,D.Z.,Cui,P.,Ou,X.X. (2014) Progress of Evaluating the Life-cycle Carbon Emission of Buildings at Home and Abroad. *Modern Management Science*,2014(7),109-111.
- [4] Pan,Y.Q.,Liang,Y.M.,Zhu,M.Y. (2021) Review of Building Carbon Emission Calculation Models in Context of Carbon Neutrality. *Heating Ventilating & Air Conditioning*,51(7),37-48.
- [5] Zeng,X.D.,Qin,Y.Y. (2010) Strategy of Low-carbon Building Design in the Initial Design Stage.*New Architecture*,2010(4),114-117.
- [6] Luo,Z.X.,Cang,Y.J.,Yang,L.,Wang,Y.Q. (2021) Study on the Design-oriented Calculation Methods of Building Embodied Carbon Emissions,2021(12),1-7.
- [7] Guo,C.,Xu,J.F.,Zhang,J.P. (2020) Calculation Methods and Prediction Models of Carbon Emission of Tunnel Construction,2020(8),1140-1146.