DOI: 10.54254/2753-8818/2026.CH30484

Sustainability Efficiency Prediction of Smart Cities Based on Machine Learning Algorithms

Bowen Hu[1*], Dawei Chen[1], Yongbo Chen[1], Hongyan Ge[1], Ang Li[1], Ruian Yan[2]

[1]*School of Artificial Intelligence and Big Data, Henan University of Technology, Zhengzhou, China*
[2]*IFLYTEK CO.LTD, Hefei, China*
**Corresponding Author. Email: 2639973982@qq.com*

Abstract. In the current era of deep integration of globalization and digitalization, the construction of smart and sustainable cities has become the core direction of urban development. Urban infrastructure planning, as a key carrier, directly affects the operational efficiency and ecological sustainability of cities. Although existing machine learning algorithms have certain advantages in complex data processing, the design governance data of urban infrastructure planning has both temporal nature and multi-agent correlation. A single regression model often has difficulty simultaneously achieving long-distance variable dependency capture and bidirectional temporal feature extraction. To this end, this paper proposes the Transformer-BiGRU regression algorithm, and first conducts correlation analysis and violin plot analysis. Then compare its regression performance with algorithms such as Random Forest, Decision tree, ExtraTrees, GBDT, AdaBoost, and XGBoost. The results show that this algorithm has a strong correlation with the above-mentioned algorithms in the performance evaluation dimension. Both demonstrate their effects by reducing error indicators such as MSE, RMSE, MAE, and MAPE, and improving the goodness of fit of R^2. Among them, its MSE 9.185, RMSE 3.031, and MAE 2.556 are lower than those of three algorithms such as Random Forest and close to GBDT. MAPE 6.335 is slightly higher than GBDT but significantly lower than the other four algorithms. R^2 0.766 is the highest among all algorithms. This not only confirms its consistency with the performance evaluation logic of traditional ensemble learning algorithms, but also demonstrates the superior fitting ability and prediction accuracy after integrating deep learning, providing technical support for the scientific decision-making of smart and sustainable urban infrastructure planning.

Keywords: Smart sustainable city construction, Transformer, BiGRU, Correlation analysis

1. Introduction

Against the backdrop of the deep integration of globalization and digitalization, the construction of smart and sustainable cities has become the core direction of urban development. As a key carrier, urban infrastructure planning directly affects the operational efficiency and ecological sustainability of cities [1]. Design governance, as a crucial link connecting multiple stakeholders, plays a key role in coordinating the demands of various entities such as the government, enterprises, the public, and

DOI: 10.54254/2753-8818/2026.CH30484

technology companies, and promoting the implementation of sustainable solutions. However, in a multi-stakeholder collaborative scenario, the sustainability effectiveness of design managers in promoting plans is influenced by multiple dimensions such as policy support, depth of cooperation, public participation, and technology adaptation, with complex and dynamically changing relationships [2]. Traditional assessment methods rely on empirical judgment and single-indicator analysis, making it difficult to quantify the nonlinear correlations among multiple factors and accurately capture the key influencing mechanisms in the design governance process. This leads to a lag and one-sidedness in the prediction of sustainability effectiveness, which restricts the refined advancement of smart sustainable city design governance. There is an urgent need for more efficient analytical tools to solve this problem [3].

Machine learning algorithms, with their powerful data mining and nonlinear fitting capabilities, provide an effective path for solving the prediction problems of multi-dimensional complex systems. In the design governance scenario of urban infrastructure planning, machine learning regression algorithms can integrate multi-stakeholder collaborative data, design manager capability characteristics, institutional environment and other multi-source variables. By automatically learning the hidden correlations among variables, an accurate performance prediction model can be established [4]. Compared with traditional methods, machine learning algorithms do not require the presetting of fixed association rules, can flexibly adapt to the dynamic changes of multi-stakeholder collaboration, efficiently process high-dimensional and heterogeneous urban governance data, and significantly improve the accuracy and efficiency of sustainability effectiveness prediction. Meanwhile, the algorithm can quantify the significance of each influencing factor, providing data support for design managers to optimize collaborative strategies and improve governance mechanisms, facilitating the transformation of urban infrastructure planning from experience-driven to data-driven, and laying a technical foundation for the scientific advancement of smart and sustainable urban design and governance [5].

Although existing machine learning algorithms have shown certain advantages in complex data processing, the design governance data of urban infrastructure planning has both temporal nature and multi-agent correlation. A single regression model often fails to simultaneously capture long-distance variable dependencies and extract bidirectional temporal features. For this purpose, this paper proposes the Transformer-BiGRU regression algorithm. This algorithm integrates the multi-head attention mechanism of the Transformer model with the bidirectional time series modeling capability of BiGRU. It can not only accurately capture the long-distance correlation features in the multi-stakeholder collaboration process, but also fully explore the bidirectional time series evolution laws of design governance data, effectively adapting to the multi-dimensional and dynamic data source characteristics in urban infrastructure planning.

2. Data sources

The dataset used in this article is an open-source dataset, containing a total of 817 valid data entries. The dataset takes sustainability effectiveness as the core predictor variable and is used to comprehensively measure the environmental sustainability, resource utilization efficiency and long-term operation and maintenance adaptability of the promotion plans of design managers in urban infrastructure projects. The independent variable system includes multi-stakeholder collaboration dimensions (government participation, depth of enterprise cooperation, public participation, and collaboration degree of technology companies), design manager capability dimensions (project management experience), system and resource dimensions (resource integration capability, system completeness), and project characteristic dimensions (project type) The smart drainage project has a

DOI: 10.54254/2753-8818/2026.CH30484

positive adjustment for sustainability efficiency and the dimension of technology application (the level of smart technology application).

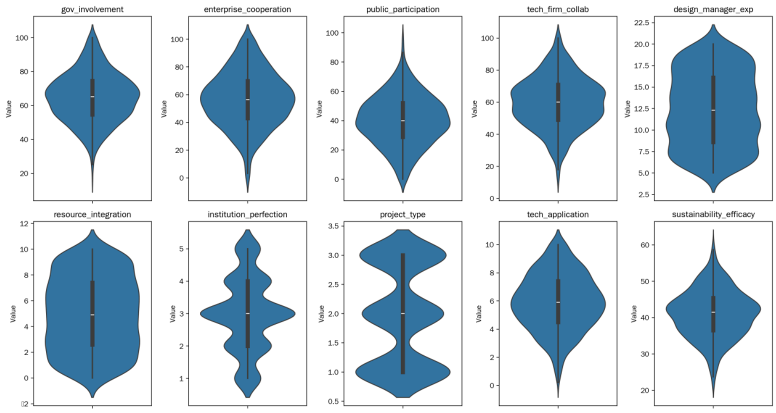

Figure 1. Violin diagrams of each variable

Calculate the correlations among various variables and draw a correlation heat map, as shown in Figure 2. Correlation analysis of each variable revealed that the correlation coefficient between tech_firm_collab and sustainability_efficacy was 0.49, indicating a relatively high correlation. The correlation coefficients of gov_involvement and enterprise_cooperation with sustainability_efficacy were 0.44 and 0.42 respectively, and there was also a certain degree of positive correlation.

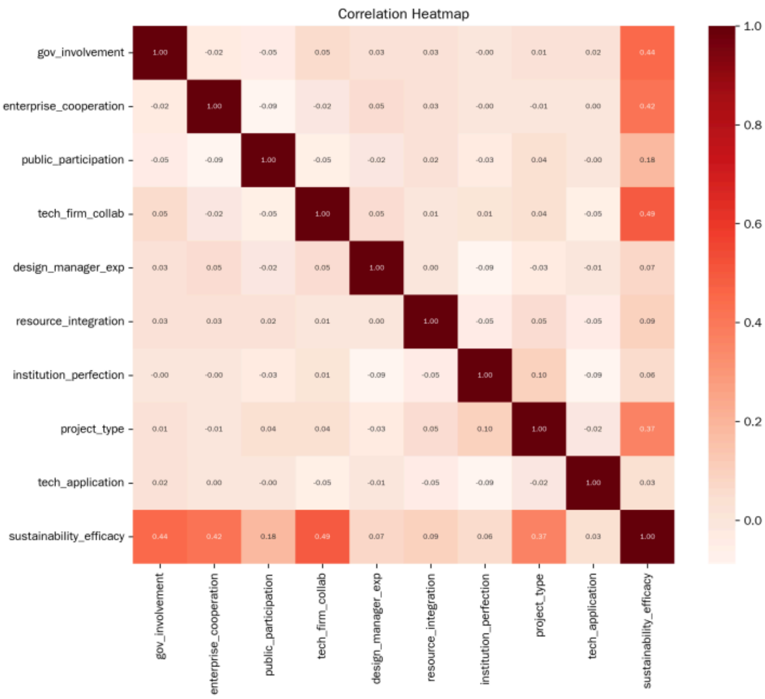

Figure 2. The correlation heat map

DOI: 10.54254/2753-8818/2026.CH30484

3. Method

3.1. Transformer

The Transformer algorithm is a deep learning model based on the self-attention mechanism. Its core advantage lies in its ability to efficiently capture long-distance dependencies in data and completely break away from the traditional recurrent neural network's reliance on sequence order. Its core component is the multi-head attention mechanism, which allocates weights to different dimensions and positions of the input data through parallel computing to accurately identify the implicit associations between variables. At the same time, it combines position encoding to supplement time series information, ensuring the model's sequential sensitivity to sequential data [6]. In the scenario of smart sustainable urban design and governance, the Transformer algorithm can effectively handle multi-source heterogeneous data generated by the collaboration of multiple stakeholders, such as the complex correlations of variables like government participation, enterprise cooperation depth, and public participation. It can automatically mine the global dependencies among variables without relying on manual feature engineering It provides powerful feature extraction capabilities for sustainability effectiveness prediction, especially suitable for processing high-dimensional and multi-correlation urban governance data [7]. The network structure of the Transformer is shown in Figure 3.

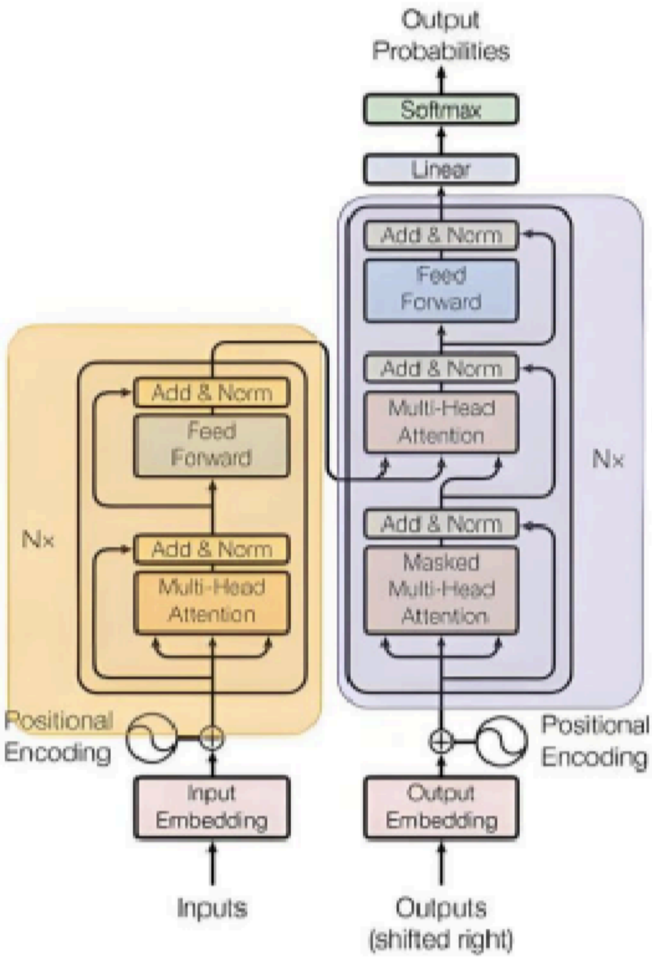

Figure 3. The network structure of the Transformer

DOI: 10.54254/2753-8818/2026.CH30484

3.2. BiGRU

BiGRU is a time series modeling algorithm optimized on the basis of GRU. By integrating the outputs of forward GRU and backward GRU, it comprehensively captures the bidirectional evolution characteristics of time series data [8]. This algorithm retains the update gate and reset gate structure of GRU, which can adaptively adjust the forgetting and retention of information, effectively alleviating the vanishing gradient problem of traditional recurrent neural networks. Meanwhile, its bidirectional structure enables it to capture the evolution patterns of historical data as well as perceive the potential impact of future data [9]. In the design governance scenario of multi-stakeholder collaboration, the process by which design managers promote solutions has significant temporal characteristics, such as the iteration of technology adaptation and the dynamic adjustment of stakeholder collaboration. The BiGRU algorithm can accurately mine the dependency relationships of these time series data, quantify the impact of variable changes at different stages on sustainability performance, and provide reliable time series feature extraction support for time series performance prediction [10].

3.3. Transformer-BiGRU

The Transformer-BigRU regression algorithm is a hybrid model that integrates the core advantages of Transformer and BiGRU, specifically designed for data scenarios with dual features of multivariable association and time series evolution. This algorithm adopts a three-stage structure of feature extraction + time series modeling + regression prediction: Firstly, through the multi-head attention mechanism of Transformer, global correlation features are extracted from multi-source variables to accurately capture the complex dependencies in multi-stakeholder collaboration; Subsequently, the extracted features are input into the BiGRU layer, and the temporal evolution law of the data is mined through a bidirectional gating structure, taking into account both historical influences and future trends. Finally, a regression prediction head is constructed through a fully connected layer to output continuous predicted values of sustainability performance. This fusion model overcomes the deficiency of a single Transformer in capturing temporal details and simultaneously compensates for the limitations of BiGRU in long-distance variable association mining. It perfectly ADAPTS to the dual data characteristics of multi-variable association and temporal dynamics of scheme promotion in the design and governance of smart sustainable cities through multi-stakeholder collaboration. It can provide more comprehensive feature support for the prediction of sustainable performance, significantly improving the prediction accuracy and generalization ability.

4. Result

In terms of parameter Settings, the proportion of the training set to the total samples is 0.7, and the output dimension is 1. In terms of model structure, the number of input channels of the Transformer is consistent with the dimension of the input features. The maximum position encoding length is 512, the number of self-attention heads is 4, the total number of key channels is 128, and it includes two self-attention layers. The number of hidden units in the forward GRU of BiGRU is 6, and the number of hidden units in the reverse GRU is 10. There are also splicing layers, ReLU activation layers, Dropout layers with a discard probability of 0.01, and fully connected layers with an output dimension of 1. The training uses the Adam optimizer, with a maximum of 200 training rounds and a batch size of 256. Before each round of training, the data is shuffled. The initial learning rate is 0.01,

DOI: 10.54254/2753-8818/2026.CH30484

the learning rate decline factor is 0.1 and decreases once every 80 rounds, the L2 regularization coefficient is 0.001, and the gradient threshold is 10.

For the comparative models, this paper conducts comparative experiments using Random Forest, Decision tree, ExtraTrees, GBDT, AdaBoost and XGBoost. The experimental results are shown in Table 2. Draw a line chart comparing the indicators of each variable as shown in Figure 4.

Table 1. The results of the comparative experiment

Model	MSE	RMSE	MAE	MAPE	R^2
Random Forest	20.617	4.541	3.593	8.982	0.522
Decision tree	23.149	4.811	3.763	9.167	0.509
ExtraTrees	13.329	3.651	2.889	7.135	0.699
GBDT	9.872	3.142	2.573	6.309	0.736
AdaBoost	10.847	3.293	2.666	6.469	0.731
XGBoost	11.307	3.363	2.804	7.093	0.751
Transformer-BiGRU	9.185	3.031	2.556	6.335	0.766

The Transformer-BiGRU regression algorithm proposed in this paper has a strong correlation with algorithms such as Random Forest, Decision tree, ExtraTrees, GBDT, AdaBoost, and XGBoost in the dimension of regression performance evaluation. The regression effect is reflected through the reduction of error indicators such as MSE, RMSE, MAE, and MAPE, as well as the improvement of R^2 goodness-of-fit. Among them, the MSE 9.185, RMSE 3.031, and MAE 2.556 of Transformer-BiGRU were all lower than those of Random Forest, Decision tree, and ExtraTrees, and were only relatively close to the corresponding indicators of GBDT. MAPE 6.335 is slightly higher than GBDT's 6.309 but significantly lower than the other four algorithms, while R^2 0.766 is the highest among all algorithms. This not only reflects its consistency with traditional ensemble learning algorithms in performance evaluation logic but also demonstrates its superior fitting ability and prediction accuracy after integrating deep learning features compared to other algorithms.

DOI: 10.54254/2753-8818/2026.CH30484

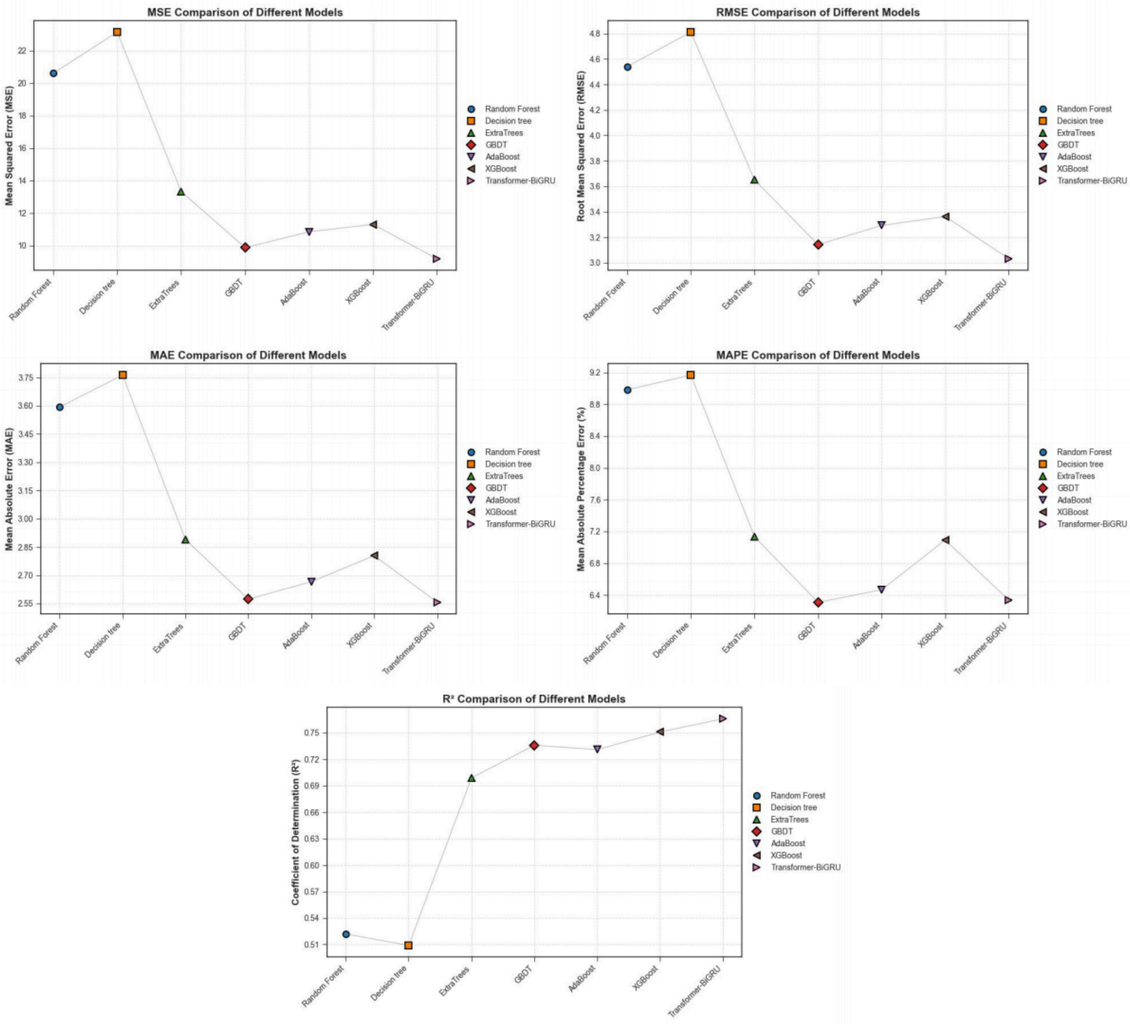

Figure 4. The line chart comparing the indicators of each variable

Output the line graph of the predicted and actual values of the Transformer-BiGRU test set, as shown in Figure 5.

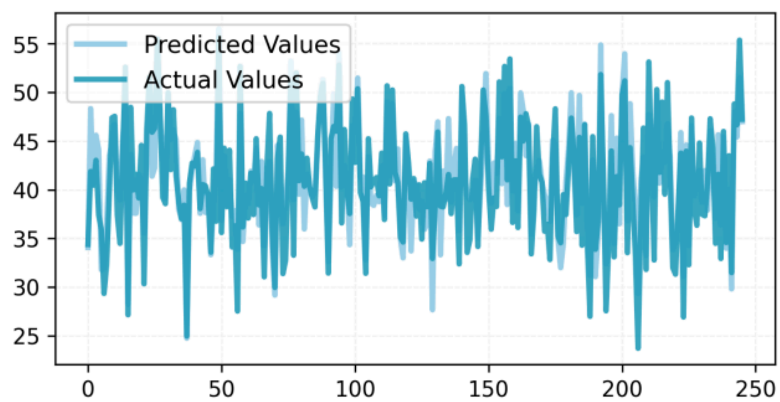

Figure 5. The line graph of the predicted and actual values of the Transformer-BiGRU test set

DOI: 10.54254/2753-8818/2026.CH30484

5. Conclusion

In the current era where globalization and digitalization are deeply intertwined, the construction of smart and sustainable cities has become the core path for urban development. Urban infrastructure planning, as the core carrier, is directly related to the operational efficiency and ecological sustainability of cities. Although existing machine learning algorithms can handle complex data, the design governance data of urban infrastructure planning has both temporal nature and multi-agent correlation. A single regression model often struggles to simultaneously capture long-distance variable dependencies and extract bidirectional temporal features. To this end, this paper proposes the Transformer-BiGRU regression algorithm. Firstly, correlation analysis and violin plot analysis are carried out, and then compared with various algorithms such as Random Forest and Decision tree. This algorithm is strongly correlated with these algorithms in the regression performance evaluation, and the effect is reflected by reducing error indicators such as MSE and RMSE and improving the goodness of fit of R^2 : Its MSE 9.185, RMSE 3.031, and MAE 2.556 are lower than those of three algorithms such as Random Forest, and only close to GBDT. MAPE was 6.335, slightly higher than GBDT's 6.309, but significantly lower than the other four. R^2 0.766 is the highest among all algorithms. This not only confirms its consistency with the performance evaluation logic of traditional ensemble learning algorithms, but also highlights the better fitting and prediction accuracy after integrating deep learning, providing key technical support for the scientific decision-making of smart and sustainable urban infrastructure planning.

References

[1] Salemi, Alireza, and Hamed Zamani. "Evaluating retrieval quality in retrieval-augmented generation." Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2024.
[2] Leng, Quinn, et al. "Long context rag performance of large language models." arXiv preprint arXiv: 2411.03538 (2024).
[3] Shi, Yunxiao, et al. "Enhancing retrieval and managing retrieval: A four-module synergy for improved quality and efficiency in rag systems." arXiv preprint arXiv: 2407.10670 (2024).
[4] Zhao, Shengming, et al. "Towards understanding retrieval accuracy and prompt quality in rag systems." arXiv preprint arXiv: 2411.19463 (2024).
[5] Jin, Bowen, et al. "Long-context llms meet rag: Overcoming challenges for long inputs in rag." arXiv preprint arXiv: 2410.05983 (2024).
[6] Mengmeng, Su, et al. "An Effective Retrieval Method to Improve RAG Performance." 2024 7th International Conference on Data Science and Information Technology (DSIT). IEEE, 2024.
[7] Tang, Yixuan, and Yi Yang. "Multihop-rag: Benchmarking retrieval-augmented generation for multi-hop queries." arXiv preprint arXiv: 2401.15391 (2024).
[8] Jiang, Wenqi, et al. "Rago: Systematic performance optimization for retrieval-augmented generation serving." Proceedings of the 52nd Annual International Symposium on Computer Architecture. 2025.
[9] Oche, Agada Joseph, et al. "A systematic review of key retrieval-augmented generation (rag) systems: Progress, gaps, and future directions." arXiv preprint arXiv: 2507.18910 (2025).
[10] Luo, Kun, et al. "Does RAG Really Perform Bad For Long-Context Processing?." arXiv preprint arXiv: 2502.11444 (2025).