ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

Research on Multivariable Carbon Emission Prediction Technology for Road Sections Based on GAN-KAN Hybrid Model

Yuan Chai[1,2,*], *Jiaqi Liu*[3] *and Yu Song*[4]

[1] Guangxi Communications Investment Group Co., Ltd., 530022, Guangxi, China
[2] Collage of Transportation, Tongji University, 201804, Shanghai, China
[3] School of Civil and Architectural Engineering, Nanning University of Science and Technology, 541000, Guangxi, China
[4] School of Civil Engineering, Guilin University of Technology, 541004 Guangxi, China

SUMMARY: *Based on the accurate prediction of carbon emissions from road traffic, intelligent transportation systems and green mobility will be used for data-driven decision-making. Given the shortcomings of the old way of dealing with complex multi-variate non-linear relationships and missing data, we propose a new GAN-KAN hybrid model in this paper. Fill in the missing values of the time-series traffic data in the model, and use a function-structure decomposition method to model multivariate carbon emissions more accurately and understandably. Based on the experimental results, the enhanced GAN increased the number of samples from 18,200 to 45,600; a variable coverage rate of 98.7% was achieved; the KL divergence was reduced from 0.138 to 0.025; and the error was corrected to 4.91 gCO2·km-1. According to the above experiment, the hybrid model's predicted carbon emissions for expressways, arterial roads and local roads are 146.8 gCO2·km-1, 159.3 gCO2·km-1 and 176.7 gCO2·km-1, respectively; at the same time, the root mean square error is 6.48 gCO2·km-1 and the mean absolute error is 5.21 gCO2·km-1. The carbon emission prediction error is less than 5% and thus superior to that of the other model. The proposed hybrid model can address the two problems of poor data quality and difficulty in relationship modelling for carbon emission prediction, thus demonstrating good practicality and engineering application value.*

KEYWORDS: *Intelligent transportation; Carbon emission prediction; GAN-KAN hybrid model; Road traffic; Data augmentation; Nonlinear modeling*

1 Introduction

With the deep development of the "dual-carbon" strategy, control over carbon emissions from the transportation sector has gradually achieved some results and thus become one of the main directions for promoting urban green and low-carbon development [1]. As one of the primary sources of carbon emissions, road traffic emissions are influenced by many factors simultaneously, such as vehicle speed, traffic volume, road grade and weather conditions, and therefore exhibit strong non-linearity, multidimensional coupling and spatiotemporal heterogeneity [2]. Therefore, to build a high-precision and highly adaptable traffic carbon emission prediction model, intelligent management of urban transportation systems can be achieved, and green mobility policies optimized [3]. Most of the existing research has used

*chaiyuan@tongji.edu.cn
https://doi.org/10.65102/is2026932 Published: 30 April, 2026

conventional machine learning or deep neural networks and is still facing many difficulties when dealing with actual traffic data [4]. On the one hand, raw traffic data suffer from extensive missing values, discontinuous data collection and class imbalance; thus, model training efficiency and generalisation ability are severely constrained [5]. On the other hand, the mapping relationship among carbon emissions and traffic-related factors is highly irregular, and traditional models are not ideal for expressing these non-linear reasons; thus, prediction accuracy is reduced under conditions of small samples or long-tail categories [6]. In addition, the current method lacks integrated optimisation of data-quality improvement and model-structure interpretability, thus failing to meet the stability and accuracy requirements of actual applications [7].

Many studies have focused on enhancing the prediction accuracy and model generalisation ability for traffic carbon emission prediction in recent years by improving the handling of complex relationships among multiple variables and time-series data. Huo and others have dealt with the problem of predicting transportation-related carbon emissions under uncertainty by constructing a support vector machine regression model and Optimizing it with a genetic algorithm. Optimizing the penalty factor and kernel function parameters of the support vector machine has improved prediction accuracy and identified periods of high carbon emissions successfully [8]. Khajavi and Rastgoo have proposed a hybrid model of random forests and metaheuristic algorithms, such as the Harris Hawks' optimisation algorithm, to address the complex relationship between road traffic characteristics and CO_2 emissions more accurately in their research. Application experiments in the first few cities of China have achieved excellent predictive results [9]. Sun and others have put forward a parallel emission regulation system to address the problems in traffic emission supervision of intelligent transportation systems and smart cities. Through low-cost computational experiments and parallel execution, the framework improved emission-aware intelligent scheduling capabilities and offered new ideas for the development of future low-carbon mobility policies [10]. Ojadi and others have employed big data analytics and artificial intelligence technology to put forward a system for improving the sustainability of the transportation sector's supply chain and reducing greenhouse gas emissions; they have achieved good results in carbon emission control and green logistics management [11]. The former has increased the accuracy of carbon emission forecasts, adjusted traffic pollution control measures, etc., to promote all-around sustainable development.

In recent years, GANs and KANs have emerged as powerful generative models and data-modeling tools, respectively, and are now frequently applied to various complex data-generation and prediction problems in traffic carbon emission prediction. To solve the problems of limited handling of imbalanced data in traditional GANs, Huang and Jafari put forward an improved balanced generative adversarial network with gradient penalty (BAGAN-GP). By adding an autoencoder structure and a gradient penalty function, these models have shown improved stability and higher-quality generation for minority-class samples; thus, GANs can be effectively applied to imbalanced-data problems [12]. Yin and others proposed an improved distribution-matching distillation method that adds GAN loss and multi-step sampling to enhance the quality and speed of generation for images. This way solved the training instability problem in image generation and improved the visual quality substantially; thus, GANs have been shown to be highly effective for fast image generation [13]. On the other hand, Koenig et al. integrated KAN with neural ordinary differential equations to propose the KAN-ODE framework and achieved good results in dynamic system modelling by utilising the high modelling accuracy and low parameter count of KAN. The above works provide theoretical support for the integration of GAN and KAN, and show that KAN can be applied to model time-dependent and complex data [14]. Saravani and others have also used KAN to predict chlorophyll-a concentration in a lake and shown that, compared with other traditional neural network models, KAN performs better in dealing with nonlinear relationships and long-term dependence in time-series data, thereby verifying the strong nonlinear modeling capability of KAN [15].

In short, the existing studies on traffic carbon emission prediction have used a wide variety of models and methods, such as genetic algorithm-optimized support vector machine regression models, hybrid approaches combining random forests with metaheuristic algorithms, and deep learning techniques like LSTM and GAN, and have achieved good results. The above ways have improved the accuracy of the predictions and solved the problem of imbalanced data; thus, various traffic-emission-reduction measures have been optimized. However, the current research still faces problems such as a lack of data, complex spatiotemporal non-linear relationships, and insufficient model generalisation ability; thus, for large-scale multivariate traffic data, stability and interpretability of the model are still serious issues. Therefore, this study proposes a hybrid model based on an improved GAN and KAN to achieve high-quality traffic data augmentation and missing value repair through GAN, and at the same time, use the powerful nonlinear modelling ability of KAN to enhance the representation of relationships among carbon emissions and various other traffic-related variables.

2 Prediction framework design for GAN-KAN hybrid models

2.1 Traffic Data Augmentation and Missing Data Repair Method Based on Improved GAN

Raw road-segment data in traffic carbon emission prediction tasks often have problems such as discontinuous data collection, missing values, and class imbalance, which reduce the accuracy and stability of the following model [16, 17]. Given the problems mentioned above, a hybrid prediction framework has been proposed that combines GAN and KAN, and a two-stage collaborative optimisation strategy is used to improve both the generality and representation power of the prediction model. Within this framework, the GAN will be used to augment traffic time-series data and fill in missing values; high-quality, structurally complete training samples will thus be obtained for the KAN model to learn complex non-linear relationships among multiple variables accurately. As the front-end data generation module of the hybrid model, the GAN can learn the underlying distribution of the data from a small number of original samples and synthesize large-scale and diverse traffic datasets with variable correlations and structural consistency. This can help to solve the problems of modelling induced by sample scarcity. The entire structure of the new system is shown in Figure 1.

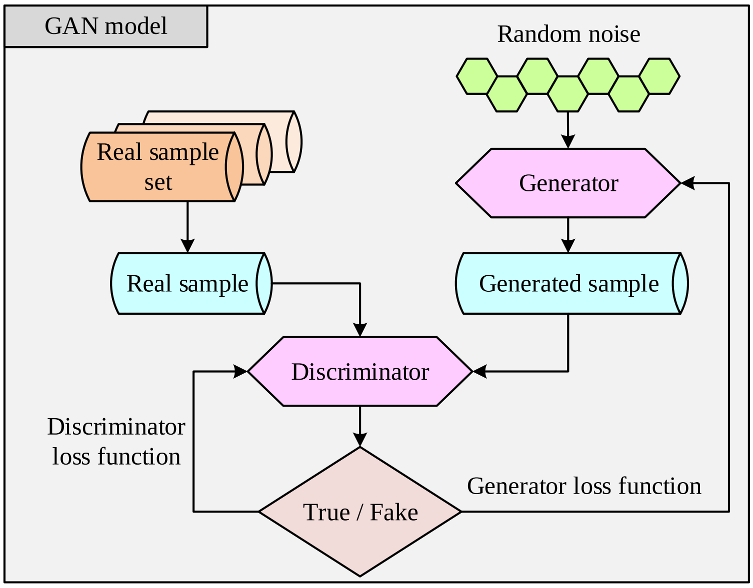


Figure1: Schematic Diagram of the GAN Model

As shown in Fig. 1, GAN consists of two parts: a generator and a discriminator. These two parts continuously optimize during the adversarial game, thereby achieving an approximate model of the real data distribution. GAN first P_{data} extracts samples from the input set of real samples, and simultaneously x samples input z from a random noise distribution P_z to the generator G to generate pseudo-samples $x'=G(z)$. The discriminator D's task is to determine whether the input sample comes from real data (label 1) or generated data (label 0), and based on its judgment , it feeds back gradient information to update the parameters of GAN G and GAN respectively D. The entire adversarial training process can be formalized as a minimax optimization problem as shown in Equation (1).

$$\min_G \max_D E_{x\sim P_{diata}}[logD(x)]+E_{z\sim P_z}[log(1\text{-}D(G(z)))] \tag{1}$$

In equation (1), $x \sim P_{diata}$ represents sampling from the real data distribution, $z \sim P_z$ represents sampling from the noise distribution (usually Gaussian or uniform distribution), $D(x)$and and $D(G(z))$represent the probability that the discriminator outputs "true" for the real sample and the generated sample, respectively. However, when used for multivariate road segment carbon emission prediction, traditional GANs have difficulty handling the mixed structure of continuous and discrete variables in traffic data, the generated samples are prone to distortion, and the generator cannot effectively learn the distribution when facing a few classes of scenarios such as extreme weather and low-frequency road segments, resulting in insufficient sample diversity [18]. Therefore, the study improves GANs, and the specific measures are shown in Fig. 2.

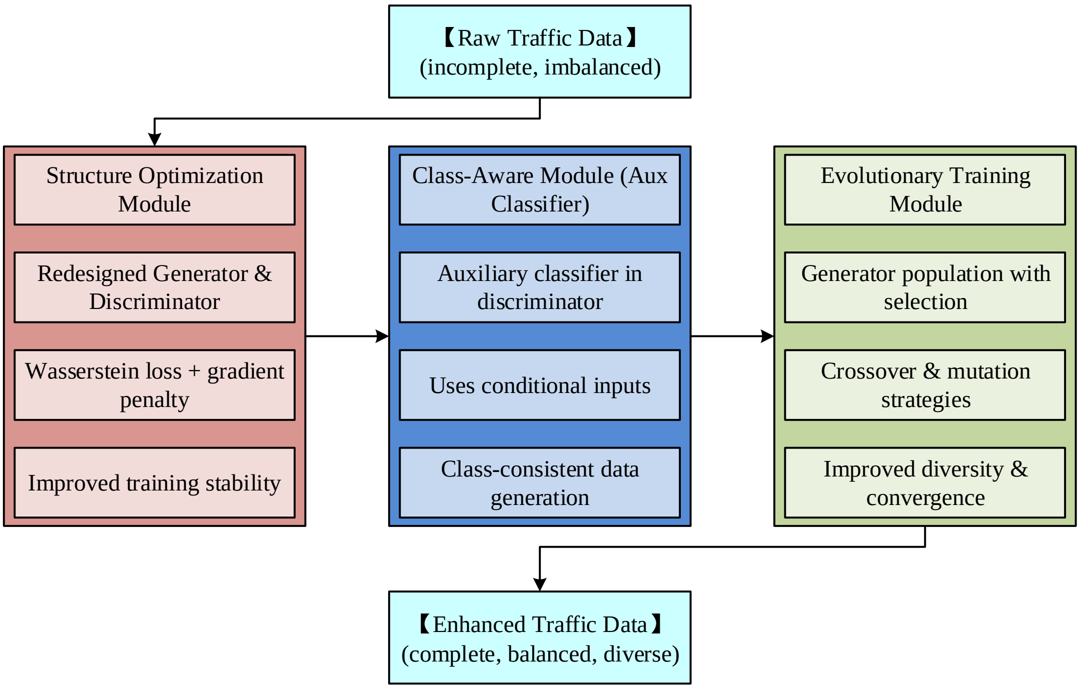


Figure 2: Improvements to the GAN Architecture

As shown in Figure 2, the three improvements of this study over traditional GANs are structural optimisation, category awareness and a training mechanism. First, in terms of structure optimisation, the basic architecture of the generator and discriminator has been rebuilt. The Generator is a multi-layer residual network that can handle temporally continuous data well. To enhance the discrimination capability of the discriminator for sample authenticity and subtle variations, LeakyReLU activation and layer normalisation have been introduced at different stages. After structural optimisation, in order to further reduce the training instability and gradient vanishing problems, the Wasserstein distance is used instead of the JS divergence of traditional GANs as the adversarial objective function, defined as in Equation (2).

$$\min_{G}\max_{D\in F}[D(x)] - E_{z\sim P_z}[\log\ D(G(z))] \tag{2}$$

In Fequation (2), is the function space that satisfies the 1-Lipschitz condition. To ensure the Lipschitz continuity of the discriminator function, a gradient penalty term is further introduced , and the final loss function of the discriminator is constructed as shown in equation (3).

$$L_D=E_{\tilde{x}\sim P_g}[D(\tilde{x})]-E_{x\sim P_{data}}[D(x)]+\lambda\cdot E_{\hat{x}\sim P_{\hat{x}}}[(\|\nabla_{\hat{x}}D(\hat{x})\|_2-1)^2] \tag{3}$$

In equation (3), $\tilde{x}$is the generated sample, is $\hat{x}$the sample obtained by linear interpolation between and , $\nabla_{\hat{x}}D(\hat{x})$represents $x\sim P_{data}$the gradient $\tilde{x}\sim P_g$of the discriminator pair $\hat{x}$, λ and is the gradient penalty weight coefficient, which is usually taken as 10. This loss function constrains the gradient magnitude of the discriminator to be close to 1, thereby improving the stability and convergence efficiency of model training.

Secondly, regarding category perception enhancement, this study investigates embedding an auxiliary classifier into the discriminator structure to achieve supervised learning of the generated sample category labels, addressing various discrete variables in traffic data such as road type, time period, and weather conditions. The classifier outputs a C vector of length , representing the probability distribution of each category. Its training objective is to minimize the cross-entropy loss between the true sample category label and the predicted probability, as shown in Equation (4).

$$L_{AC}=-E_{(x,c)\sim P_{data}}\sum_{i=1}^{C}1_{[i=c]}\,logP(c=i\mid x) \tag{4}$$

In equation (4), $c$represents the true class label, $C$is the total number of all classes, $1_{[i=c]}$is the indicator function, $P(c=i\mid x)$and is the predicted probability that the sample belongs to class i. The final comprehensive loss function of the discriminator is shown in equation (5).

$$L_D^{total}=L_D+\alpha\cdot L_{AC} \tag{5}$$

In Equation (5), αrepresents the weighting factor of the classification loss. The introduction of the auxiliary classifier enables the generator to receive more explicit category information feedback during training, generating data that better matches the actual category distribution and improving the sample diversity and reliability under a few categories, such as extreme weather and nighttime periods. Finally, in terms of the training mechanism, the study adopts an evolutionary optimization strategy to enhance the diversity and global search capability of the generator. Specifically, a generator population is constructed, and iterative optimization is performed using a genetic algorithm (GA). In each generation, individual generators are evaluated according to the fitness function, as shown in Equation (6).

$$F(G_i)=E_{x\sim P_{data}}[D(x)]-E_{z\sim P_z}[D(G_i(z))] \tag{6}$$

In Equation (6), G_i represents the i-th i generator, and represents $D(G_i(z))$ the discriminator output of its generated samples. $F(G_i)$A larger value indicates that the samples generated by this generator are closer to the true distribution. In the selection operation, individuals with high fitness are retained. In the crossover operation, the parameter vectors of two individuals are partially swapped to generate a new individual. In the mutation operation, Gaussian noise is added to the individual parameters. This evolutionary training process improves the generator's search capability in the high-dimensional traffic data feature space, effectively avoiding getting trapped in local optima and enhancing the diversity and structural rationality of the generated data. Therefore, based on the improved GAN, the steps of the traffic data augmentation and missing data repair method are shown in Fig. 3.

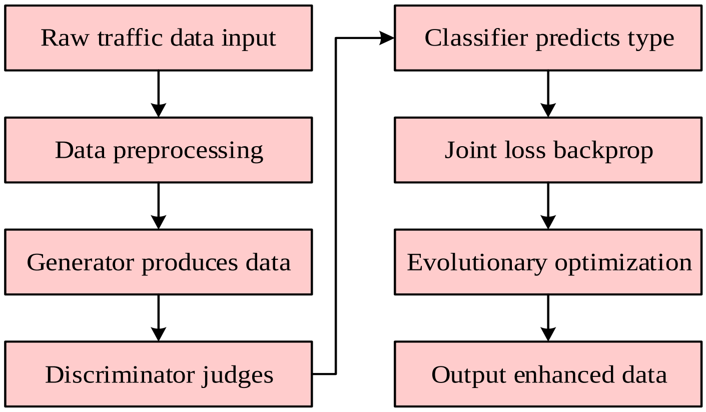


Figure 3: Traffic Data Augmentation and Missing Value Imputation Based on Improved GAN.

As shown in Figure 3, the eight steps of the proposed method are as follows. First, the raw traffic data are imported, and they generally have problems such as missing values and class imbalance. Second, data preprocessing is carried out to unify the input format by time alignment, normalisation and missing-value imputation. The third is the structurally optimised generator, and it outputs a new sample after receiving input data. The fourth step is to have the discriminator determine how real the samples are and thus keep them in the distribution. In the fifth step, a classifier identifies the category of each sample and enhances category-consistency learning. In the sixth step, a combined loss function of adversarial loss and classification loss is used for backpropagation to optimize the parameters of both the generator and the discriminator. At the seventh step, an evolutionary strategy is used to find the fit of several generators, and then crossover and mutation operations are performed to increase the diversity and overall performance of the samples. Finally, high-quality augmented data with a full structure and balanced categories have been generated to provide strong support for the following carbon emission prediction model.

2.2 A Multivariate Carbon Emission Function Modeling Method Based on KAN

The well-structured and class-balanced traffic time-series data produced by the improved GAN have effectively solved the issues of missingness and imbalance in the model. To increase the accuracy of the non-linear mapping function for traffic indicators and carbon emissions further, and to improve prediction performance, a KAN-based multivariate carbon emission function model has been used. Kolmogorov-Arnold Networks (KANs) are new neural network architectures based on the Kolmogorov-Arnold Representation Theorem. Replace the original activation function with a learnable one-dimensional function to enhance the non-linearity of representation and improve function interpretability in KAN. KANs have shown excellent modelling results for low-sample and high-dimensional data compared with fully connected neural networks. The structure of the KAN network is shown in Figure 4.

As shown in Figure 4, the overall KAN architecture has several "learnable one-dimensional function modules" and "weighted summation units". Each input feature at the bottom layer is first processed by a set of independent one-dimensional nonlinear transformation units. The above functions can be in the form of parameterized spline functions, interpolation functions or higher-order polynomials, as shown in Equation (7).

$$\Phi_{ij}(x_i) \sum_{k=1}^{M} c_{ijk} \mathrm{B}_k(x_i) \tag{7}$$

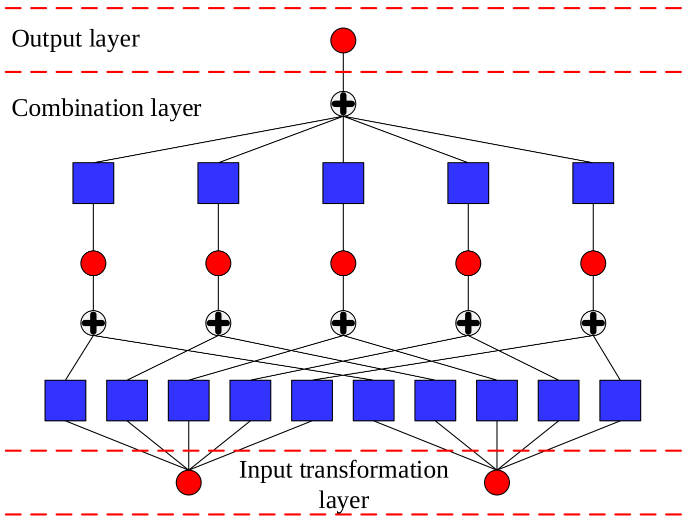


Figure 4: Structure of the KAN Network

In equation (7), x_i represents the i-th i traffic input variable, $B_k(x_i)$ is the i-th k basis function, c_{ijk} is the learnable spline coefficient, M is the number of basis functions for each activation function, $\Phi_{ij}(x_i)$ and is the nonlinear response of the i-th i input to the i-th j hidden unit. Subsequently, all transformation results are input into the hidden layer neurons through a weighted summation method, as shown in equation (8).

$$h_j = \sum_{i=1}^{n} \alpha_{ij} \cdot \Phi_{ij}(x_i) \tag{8}$$

In equation (8), α_{ij} is the connection weight, represents k the contribution of the i-th h_j input to the i-th hidden unit, j and is j the output of the i-th hidden unit. Finally, all the outputs of the hidden units are combined through the second layer to generate the final carbon emission prediction value, as shown in equation (9).

$$\hat{y} = \sum_{j=1}^{m} w_j \cdot h_j \tag{9}$$

In Equation (9), $\hat{y}$ represents the predicted carbon emission value output by the model, and w_j represents the weights from the hidden layer to the output layer. To achieve accurate learning of the complex mapping relationship between traffic multivariates and carbon emission intensity, the KAN model uses the minimum mean squared error (MSE) as the main training objective function. This loss function measures the magnitude of the error between the model's predicted value and the actual observed value, and its expression is shown in Equation (10).

$$L_{MSE}=\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2 \tag{10}$$

In equation (10), N represents the total number of training samples, y_i is i the actual carbon emission value of the i-th sample, and the smaller the loss value, the closer the model prediction is to the actual distribution [19, 20]. During the training process, a batch gradient descent strategy is adopted, and the Adam optimizer is introduced for parameter updates. The multivariate traffic carbon emission modeling process based on KAN is shown in Fig. 5.

As shown in Figure 5, the improved GAN module is first used to generate structurally complete and class-balanced traffic time-series data that serve as the input data source for the

KAN model. Next, the input variable vectors are sequentially processed by a one-dimensional nonlinear function transformation layer to extract the nonlinear relationships among these variables and the target. Finally, the extracted features are given weights and the final result of carbon emission prediction is obtained. Finally, the model parameters are updated by minimising the loss function in cooperation with the Adam optimizer and backpropagation, and thus the relationships among multiple carbon emission maps can be learned effectively. Based on this, after carrying out data augmentation with the improved GAN and KAN-based modeling procedures, the entire workflow of multivariate road-segment carbon emission prediction technology based on the GAN-KAN hybrid model has been established, as shown in Figure 6.

As shown in Figure 6, the four modules of the proposed prediction technology are Data acquisition and pre-processing, Data augmentation, Function modelling, and Prediction output, which work together to achieve this. In the data acquisition and preprocessing stage, the raw traffic data from urban road monitoring systems are collected; this includes attributes such as traffic flow, average speed, road grade and weather type, and is then time-aligned, normalized and missing-value imputed. In the data augmentation stage, the preprocessed data are fed into the improved GAN module to generate high-quality and class-balanced traffic datasets via structural optimisation, category-aware learning and evolutionary training mechanisms, thus increasing the sample size and improving distribution rationality. In the function modelling stage, the augmented data are fed into the KAN network to learn the non-linear mapping function of traffic features to carbon emission intensity. A Mean Squared Error loss function is used to optimise the parameters of the model iteratively through the Adam optimizer for improved prediction accuracy. Finally, the generated carbon emission prediction results are provided in the output stage to support the planning of green mobility strategies and low-carbon policies for subsequent intelligent traffic management systems and carbon management platforms.

3 Experiment and Results Analysis

3.1 Evaluation of the data augmentation and restoration effects of improved GANs

A few experiments have been carried out to determine whether the enhanced GAN model for traffic data augmentation and missing value imputation has achieved its goal. The Experimental Environment and Parameter Settings are shown in Table 1.

Based on the configurations shown in Table 1, traffic monitoring data for a few main arterial roads in a city in China in 2022 were selected as the experimental dataset. The multi-source data in the dataset includes vehicle speed, traffic flow, road grade, time period and weather information, with a total of 18,200 records. After cleaning and normalising the data, nine input variables and a carbon emission output variable were kept. Then, an 8:2 split was made of the dataset to construct a training set and a test set for evaluation of the data augmentation and missing value replacement capabilities of the improved GAN. Traditional GAN, Wasserstein generative adversarial network with gradient penalty (WGAN-GP), auxiliary classifier generative adversarial network (AC-GAN) and variational autoencoder (VAE) were selected as the comparative methods. First, the performance of all models in traffic time-series data augmentation and missing-value imputation was tested. A uniform input sample size of 18,200 records was used, and all generative models were trained for the same number of training epochs and parameter settings. The results are as follows: Figure 7.

As shown in Figure 7(a), the number of samples produced by each model increases with an increase in the number of training epochs. The improved GAN consistently shows the best sample generation performance at all epochs and has reached 45,600 samples by the 300th epoch, which is 2.5 times the original number of samples. The above performances are higher than those of traditional GAN (31,800 samples), WGAN-GP (37,100 samples), AC-GAN (34,400 samples) and VAE (33,000 samples). The above results show that structural optimisation, category-aware enhancement and evolutionary training strategies have improved the data generation capability. As shown in Figure 7(b), the improved GAN has a coverage rate of 94.5% at the beginning of training and reaches a maximum of 98.7% at the 300th epoch; this is higher than that of the

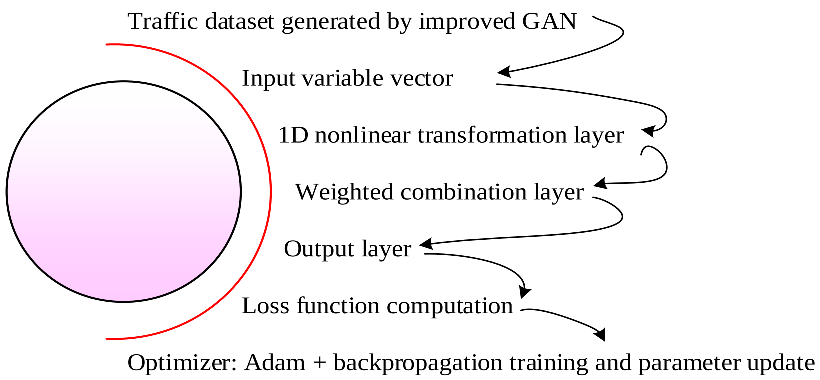


Figure 5: Model Process of Traffic Carbon Emission Routes Based on KAN.

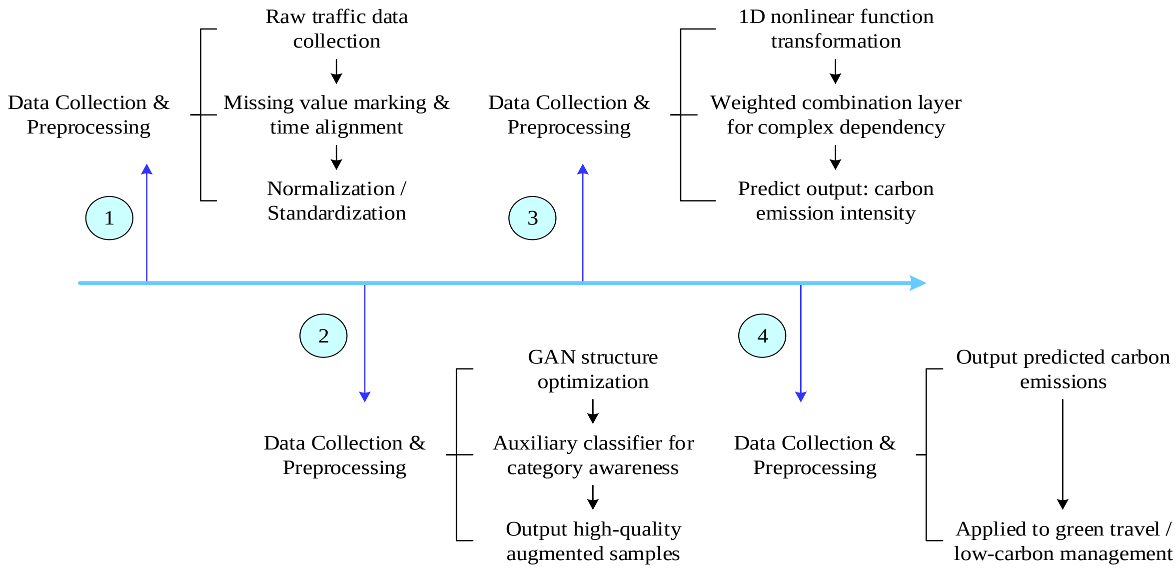
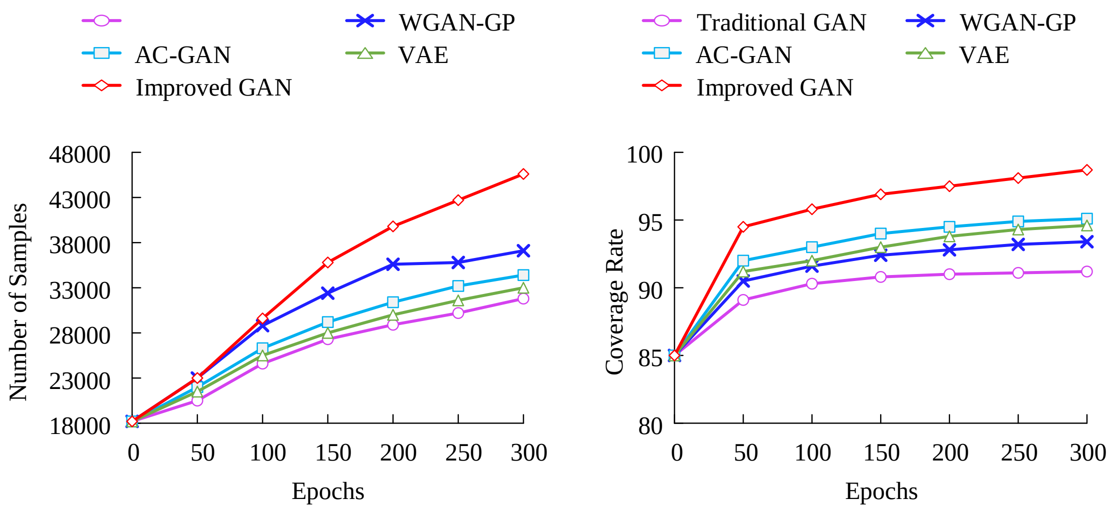


Figure 6: Multivariate R oad Segment C arbon E mission P rediction B ased on GAN- KAN H ybrid M odel

Traditional GAN (91.2%), WGAN-GP (93.4%), AC-GAN (95.1%) and VAE (94.6%). This shows that the improved GAN has a stronger model for maintaining multivariate structural consistency and can address problems of variable missingness and structural imbalance in traffic data more effectively.

Table 1: Experimental Environment and Parameter Settings

Classification	project	Configuration/Description
Hardware	processor	Intel Core i7-12700H
	Graphics Processing Unit	NVIDIA RTX 4060 (8GB VRAM)
	Memory	32 GB DDR5
Software	operating system	Windows 11 64-bit
	programming language	Python 3.10
	Deep learning framework	PyTorch 2.0 + CUDA 11.7
	Driver version	NVIDIA Driver 531.41
	Key library	numpy, pandas, matplotlib, scikit-learn, seaborn, etc.
Training settings	Optimizer	Adam
	Initial learning rate	0.0002
	Batch size	128
	Number of training rounds	300
Evolutionary strategy	Population size	10
	Iterative Algebra	20



(a) Comparison of Sample Size (b) Comparison of Coverage Rate

Figure 7: Comparison of D ata A ugmentation and V ariable C overage A mong Diffrent M odels

To assess the impact of improvement on the distribution of generated data and repair accuracy of missing values by the improved GAN, Kullback–Leibler (KL) divergence and mean repair error were chosen as the main indices. The corresponding results are as follows: Table 2. As shown in Table 2, the proposed improved GAN model has achieved the best results for traffic data missing-value repair. The KL divergence is as low as 0.025, and compared to the original GAN, there has been an 81.9% reduction; thus, the distribution of the generated data is now much closer to that of real data. At the same time, the average repair error has been reduced to 4.91 gCO2·km^{-1}, a decrease of 37.4% compared with the previous method; it also has a higher accuracy for reconstructing missing data. Compared with advanced models such as WGAN-GP, AC-GAN and VAE, the improved GAN has consistently outperformed them in all main evaluation indicators and thus verified the efficacy of structural optimisation, category-aware learning, and evolutionary training strategies for improving data quality.

3.2 Performance Analysis of Carbon Emission Prediction Using the Hybrid Model

To evaluate the performance of the GAN-KAN hybrid model in multivariate road-segment carbon emission prediction, a multi-layer perceptron (MLP), long short-term memory network (LSTM), gated recurrent unit (GRU), and Transformer model were selected as comparative methods. First, the Carbon Emission Prediction Performance Under Small Sample Size was evaluated. A total of 18,200 original traffic time-series records served as the foundation for the dataset. All models were trained for the same number of training epochs, loss functions and hyperparameters, and predictions were performed on the same test set. The results are as follows: Figure 8.

As shown in Figure 8, the GAN-KAN hybrid model performs better than all other models in carbon emission prediction under small-sample conditions. The mean absolute error (MAE), root mean square error (RMSE) and mean absolute percentage error (MAPE) of this model are 5.21 gCO2·km^{-1}, 6.48 gCO2·km^{-1} and 4.87% respectively, which are the lowest among the other models; therefore, it has good predictive accuracy and stability. On the other hand, traditional neural network models such as MLP, LSTM, GRU, and Transformer perform poorly across all indicators; thus, it can be concluded that the proposed GAN-KAN model has stronger modelling and generalisation capabilities in data-scarce situations. An expanded group of tests was also performed with the addition of a new one. Apply the improved GAN to extend the original 18,200 traffic records, thereby expanding the training set to 45,600 structurally complete and class-balanced samples. This augmented dataset was then employed to train the GAN-KAN, MLP, LSTM, GRU, and Transformer models separately, and their carbon emission prediction performance was evaluated on the same test set. The results are as follows: Figure 9.

As shown in Figure 9(a), the GAN–KAN model has the highest prediction accuracy in all rounds, and the average accuracy across all comparison models is 97.33%. The peak accuracy of the GAN-KAN model in the second round is 98.13%, and it is both highly accurate and stable. As shown in Figure 9(b), the GAN-KAN model has the lowest error in all experiments and, with prediction errors less than 5% and an average error of 3.92%, outperforms the other models. The traditional MLP and GRU models show a larger error fluctuation, and the maximum error reaches 9.31%; thus, they are less stable when processing complex multivariate traffic data. Although LSTM and Transformer are good in some rounds, their overall errors are still relatively high compared to those of GAN-KAN, and some rounds exceed 6%. The above results show that the GAN-KAN model has better generalisation and robustness in the presence of extended data, thus improving the accuracy of carbon emission forecasting. Finally, predictions for the different kinds of roads have been made and are shown in Table 3.

As shown in Table 3, among the three categories of roads, the GAN–KAN hybrid model performs the best in carbon emission prediction. The predicted value for expressways is 146.8 gCO2·km^{-1}, and the error with the actual value of 146.1 gCO2·km^{-1} is only 0.7 gCO2·km^{-1}; among all the models, this is the most accurate. The predicted values for arterial roads and local roads are 159.3 gCO2·km-1 and 176.7 gCO2·km-1, respectively, with errors kept at 1.0 gCO2·km-1, and they are better than those of MLP and GRU models. The above results show that the GAN-KAN model has maintained high prediction accuracy and is also relatively generalisable to multiple road types.

Table 2: Comparison of KL D ivergence and I mputation E rror A mong D ifferent M odels

Model	KL divergence	Correction error (gCO_2·km-1)	KL decreased (%)	Error reduction (%)
Traditional GAN	0.138 ± 0.009	7.84 ± 0.31	/	/
WGAN-GP	0.096 ± 0.007	6.52 ± 0.28	30.4	16.8
AC-GAN	0.083 ± 0.006	5.97 ± 0.22	39.9	23.8
VAE	0.058 ± 0.005	5.22 ± 0.19	58.0	33.5
Improved GAN	0.025 ± 0.004	4.91 ± 0.15	81.9	37.4

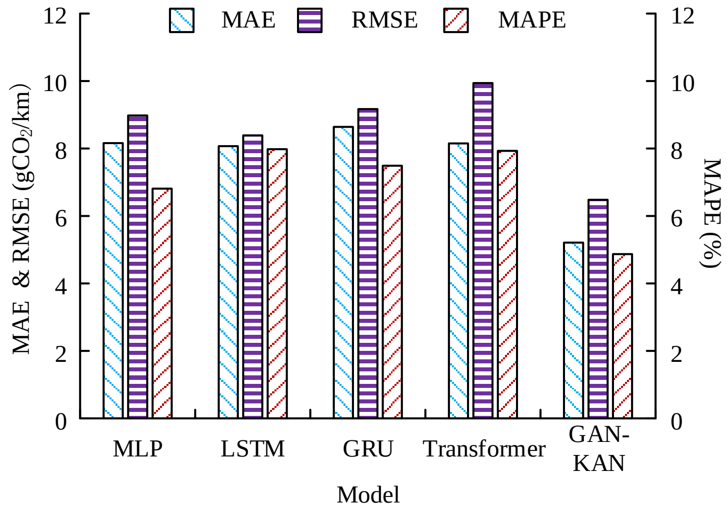


Figure 8: Carbon-E Mission P Rediction R Results under Small-to-Medium Conditions

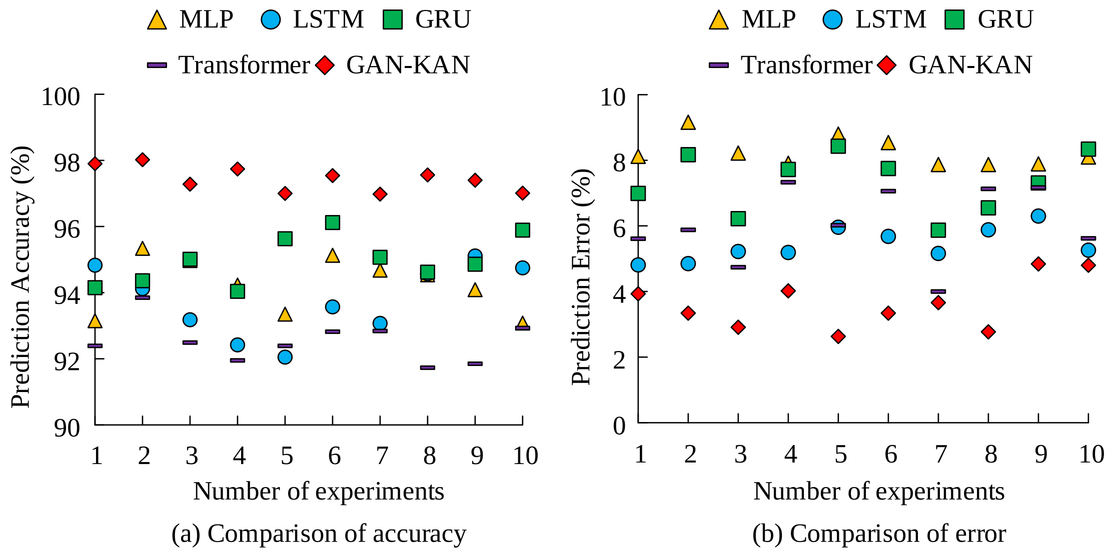


Figure 9: Carbon-E Mission Prediction Results Under Large Sample Conditions

Table 3: Carbon-E Mission P Rediction R results: Cross-different Road Types

Road type	True value	MLP	LSTM	GRU	Transformer	GAN-KAN
Expressway	146.1	153.2	148.5	151.6	149.3	146.8
Main road	158.9	164.9	157.1	161.3	158.4	159.3
Branch road	176.2	181.6	172.4	178.9	174.8	176.7

4 Conclusion

A hybrid carbon emission prediction model has been proposed in this paper, which integrates an improved GAN and KAN to address problems such as missing values, data imbalance and strong non-linear mapping in multivariate traffic data. Based on the above experiments, the following conclusions have been drawn.

(1) The improved GAN effectively performs traffic data augmentation and missing-value repair, constructing a high-quality sample set with complete structure and balanced categories, thereby ensuring the reliability of subsequent carbon emission prediction.

(2) The introduction of the KAN model enables accurate characterization of the complex nonlinear relationships between carbon emissions and multiple traffic variables, significantly enhancing prediction accuracy.

(3) Under small-sample conditions, the proposed model achieves substantially lower prediction errors than the comparative models. Under large-sample conditions, the prediction accuracy reaches 97.33%, with the error controlled within 3.92%, demonstrating strong stability and generalization capability.

Overall, the proposed GAN-KAN hybrid model has shown stable performance and high adaptability in carbon emission prediction, and multi-road-type prediction tasks are particularly outstanding. However, the model's performance in the event of extreme circumstances is still not ideal. In the future, multi-modal data and transfer learning can be used to improve the model's response speed to sudden changes and expand its cross-regional generalisation ability.

Funding

Open Fund of Guangxi Key Laboratory of Karst Dynamics: Migration Laws and Solidification-remediation Mechanisms of Heavy Metal Contaminated Soil in Karst Areas (Grant No. RZ2300151383)

About the Author

Yuan Chai, male, born in Fuxin, Liaoning, China in 1989. He is a doctor who studied at Guangxi University. At present, he is an engineer at Guangxi Communications Investment Group Co., Ltd. and a postdoctoral researcher at Tongji University. The first is green and low-carbon expressway construction.

Jiaqi Liu was born in Fushun, Liaoning Province, China in 1989. She graduated from Liaoning Technical University with a master's degree and is now at Nanning College of Technology. Her first subject of study is Highway Engineering.

Yu Song is a woman born in November 1981 who is also a doctor and professor from Dalian, Liaoning Province. She is now a senior-level academic and works in the School of Civil Engineering at Guilin University of Technology. Research direction: environmental geotechnical engineering (bio-geotechnical engineering). Presided over the National Natural Science Foundation of China 1 and provincial and ministerial projects 5; forty academic papers were published as the first author/corresponding author, of which some are SCI, EI, etc.

References

[1] Liu, Y., & Liu, H. (2022). Characteristics, influence factors, and prediction of agricultural carbon emissions in Shandong Province. Chinese Journal of Eco-Agriculture, 30(4), 558-569.

[2] Wang, G., Yan, C., et al. (2025). Highway construction carbon emission assessment and low-carbon measures and carbon compensation for geotechnical engineering: A review. China Journal of Highway and Transport, 38(01), 46-72.

[3] Tang, X., Liu, S., Wang, Y., et al. (2024). Study on carbon emission reduction countermeasures based on carbon emission influencing factors and trends. Environmental Science and Pollution Research, 31(9), 14003-14022.

[4] Wong, E. Y. C., Ling, K. K. T., Tai, A. H., et al. (2024). Two-stage multilateral trade-based prediction model for freight transport carbon emission of Belt and Road countries along Eurasian Landbridges. International Journal of Sustainable Transportation, 18(8), 633-650.

[5] Wang, L., Chen, R., Dai, L., et al. (2024). Multi-objective maintenance optimization strategy for bridge networks oriented to low-carbon. China Journal of Highway and Transport, 37(04), 188-200.

[6] Li, J., Yang, Z., Jiang, W., et al. (2025). Carbon emission prediction and sensitivity evaluation of virtual power plants based on big data and multiscale analysis. Scalable Computing: Practice and Experience, 26(2), 837-848.

[7] Wang, Y., & Si, Q. (2024). Forecasting Chinese carbon emission intensity based on the interactive effect GM(1,N) power model. Gray Systems: Theory and Application, 14(1), 21-48.

[8] Huo, Z., Zha, X., Lu, M., et al. (2023). Prediction of carbon emission of the transportation sector in Jiangsu Province: Regression prediction model based on GA-SVM. Sustainability, 15(4), 3631.

[9] Khajavi, H., & Rastgoo, A. (2023). Predicting the carbon dioxide emission caused by road transport using a random forest model combined by meta-heuristic algorithms. Sustainable Cities and Society, 93, 104503.

[10] Sun, Y., Hu, Y., Zhang, H., et al. (2023). A parallel emission regulatory framework for intelligent transportation systems and smart cities. IEEE Transactions on Intelligent Vehicles, 8(2), 1017-1020.

[11] Ojadi, J. O., Odionu, C., Onukwulu, E., et al. (2024). Big data analytics and AI for optimizing supply chain sustainability and reducing greenhouse gas emissions in logistics and transportation. International Journal of Multidisciplinary Research and Growth Evaluation, 5(1), 1536-1548.

[12] Huang, G., & Jafari, A. H. (2023). Enhanced balancing GAN: Minority-class image generation. Neural Computing and Applications, 35(7), 5145-5154.

[13] Yin, T., Gharbi, M., Park, T., et al. (2024). Improved distribution matching distillation for fast image synthesis. Advances in Neural Information Processing Systems, 37(1), 47455-47487.

[14] Koenig, B. C., Kim, S., & Deng, S. (2024). KAN-ODEs: Kolmogorov–Arnold network ordinary differential equations for learning dynamical systems and hidden physics. Computer Methods in Applied Mechanics and Engineering, 432, 117397.

[15] Saravani, M. J., Noori, R., Jun, C., et al. (2025). Predicting chlorophyll-a concentrations in the world's largest lakes using Kolmogorov-Arnold networks. Environmental Science & Technology, 59(3), 1801-1810.

[16] Senanayake, R. D., Yao, X., Froehlich, C. E., et al. (2022). Machine learning-assisted carbon dot synthesis: Prediction of emission color and wavelength. Journal of Chemical Information and Modeling, 62(23), 5918-5928.

[17] Rao, C., Huang, Q., Chen, L., et al. (2023). Forecasting the carbon emissions in Hubei Province under the background of carbon neutrality: A novel STIRPAT extended model with ridge regression and scenario analysis. Environmental Science and Pollution Research, 30(20), 57460-57480.

[18] Ojadi, J. O., Onukwulu, E., Odionu, C., et al. (2023). AI-driven predictive analytics for carbon emission reduction in industrial manufacturing: A machine learning approach to sustainable production. International Journal of Multidisciplinary Research and Growth Evaluation, 4(1), 948-960.

[19] Ahmed, M., Shuai, C., & Ahmed, M. (2022). Influencing factors of carbon emissions and their trends in China and India: A machine learning method. Environmental Science and Pollution Research, 29(32), 48424-48437.

[20] Shao, Y., Fan, G., Feng, Z., et al. (2023). Prediction of forest fire occurrence in China under climate change scenarios. Journal of Forestry Research, 34(5), 1217-1228.