Traffic & Transportation

Urban Traffic Level of Service Prediction Method Based on DBO-Transformer Multi-Source Information Fusion

Chen LIJING[1], Li YONG[2]

Original Scientific Paper
Submitted: 3 Mar 2025
Accepted: 12 May 2025
Published: 30 Mar 2026

[1] fjchenlijing@163.com, Department of Public Security, Fujian Police College, Fuzhou, China
[2] Corresponding author, yl112@fjpsc.edu.cn, Department of Public Security, Fujian Police College, Fuzhou, China

Publisher:
Faculty of Transport and Traffic Sciences, University of Zagreb

ABSTRACT
This study proposes an advanced framework for scientifically predicting the traffic level of service (LOS) that overcomes the constraints of conventional methodologies dependent primarily on real-time vehicle speed and traffic flow metrics. Recognising the empirical challenges in transformer model parameter optimisation, which is typically set based on engineering experience, we develop a novel dung beetle optimiser (DBO)-optimised transformer architecture. The model utilises traffic flow data from Fuzhou urban expressways to systematically optimise two pivotal hyperparameters: the number of self-attention heads and the learning rate. The model's performance is evaluated using mean squared error as the primary metric, supplemented by experimental data visualisation for a comprehensive assessment. The study investigates the model's predictive accuracy for the next moment's level of service under various influencing factors, including vehicle speed, traffic flow, time, weather, holidays, temperature and traffic conditions. The results indicate that the DBO-transformer model, which integrates multi-source information fusion, achieves exceptional performance in traffic level of service prediction, with an accuracy of 98.0495%. This performance surpasses that of the standard transformer model, the DBO-optimised BP neural network model and the LSTM neural network model.

KEYWORDS
traffic level of service; transformer; dung beetle optimisation; multi-source information.

1. INTRODUCTION

Traffic congestion in large cities is a significant issue, not only leading to low transportation efficiency but also causing considerable inconvenience to residents. Vehicles moving at low speeds in congested conditions substantially increase fuel consumption and exhaust emissions, to some extent hindering social and economic development. Therefore, accurately predicting urban traffic congestion is of great practical significance for alleviating traffic pressure and creating a comfortable and convenient travel environment [1].

Traffic congestion prediction is an essential research topic in the field of traffic state identification [2]. It estimates the future traffic state based on historical and current traffic flow data using specific models or theoretical methods. Accurate traffic congestion prediction is a key component of modern intelligent transport systems. Current research mainly focuses on mathematical statistical parametric models, simulation-based traffic state models and big data-driven artificial intelligence algorithms. Models such as the autoregressive integrated moving average differential [3], historical average method [4] and incremental Bayesian [5] primarily quantify the distribution characteristics of collected traffic data, but they exhibit significant shortcomings in detailed fluctuations fitting. For instance, Rajalakshmi V et al. [6] developed an ARIMA-MLP model for traffic flow prediction, achieving a squared correlation coefficient of 0.936763 during peak hours, while the prediction accuracy dropped to 0.87638 during off-peak hours. Traffic state models mainly

simulate traffic flow conditions by setting relevant parameters using traffic simulation software like VISSIM [7] and PARAMICS [8]. However, this software, based on simplified assumptions, often fails to accurately simulate complex traffic behaviours and unexpected situations, leading to distorted traffic congestion predictions due to data biases. With the advent of the big data era, deep learning technology has increasingly shown advantages in short-term traffic flow prediction [9]. Deep learning not only provides stable prediction results independent of model hyperparameters but also effectively uncovers hidden features in traffic flow data, demonstrating significant advantages in modelling complex traffic flow data. For instance, Chen Yue et al. [10] combined CNN's ability to process spatial information with BiLSTM's ability to handle temporal information, proposing a CS-BiLSTM network prediction model that effectively improves the accuracy of traffic congestion prediction. Zhang Shengrui et al. [11] proposed an FCM-RBF short-term traffic congestion prediction model with a prediction accuracy of 95%. Li Shuai et al. [12] proposed a CS-transformer urban traffic congestion prediction model that more effectively extracts spatiotemporal features of traffic data, showing high predictive performance. Jianqiang et al. [13] integrated graph convolutional networks with the transformer architecture, proposing a ConstFormer congestion prediction model that accurately predicts traffic congestion.

Previous studies have predominantly evaluated congestion levels based on real-time indicators such as occupancy and speed, with limited consideration of external factors. Moreover, there has been little focus on predicting the traffic level of service, and research on traffic level of service prediction often lacks the support of big data methodologies. In predictive systems, it is essential to continuously monitor changes in influencing factors, such as weather, time of day, vehicle speed and holidays [14], and with the help of historical data, with big data techniques to achieve more accurate predictions of urban traffic level of service. For example, Qinyang Wang et al. [15] advanced traffic flow prediction by considering the diversity of vehicles and utilising visual data, providing valuable insights for intelligent traffic systems. The transformer model, with its self-attention mechanism, is capable of capturing long-haul dependencies in time series data, making it well-suited for expressing the complex spatiotemporal dynamics of traffic flow. Therefore, this study adopts the widely used transformer model and employs intelligent optimisation methods to select optimal parameters for predicting the traffic level of service on urban expressways.

2. METHODOLOGY

2.1 Transformer-based traffic level of service evaluation model

In the context of traffic level of service prediction, temporal dependencies are not merely sequential and local; capturing global temporal dependencies is of significant importance. The transformer model, as depicted in *Figure 1*, comprises six key components: self-attention mechanism, multi-head attention, multi-layer perceptron, positional encoding, residual connections and layer normalisation [16]. The self-attention mechanism enables the transformer architecture to capture long-range dependencies when processing long sequences of traffic data, thereby accurately analysing the patterns of traffic level of service variations. The self-attention mechanism in the transformer is extended to multiple attention heads, each of which can learn different attention weights to better capture diverse types of relationships. The incorporation of the multi-layer perceptron enhances the transformer model's capability for non-linear feature extraction, significantly improving both training and prediction speeds. This allows the model to handle multi-source heterogeneous data, integrating information such as traffic flow, weather conditions and vehicle speed for comprehensive prediction. The design of residual connections facilitates the fitting of the mapping function as the number of model layers increases, effectively mitigating issues such as gradient vanishing and gradient explosion that arise with deeper models. Layer normalisation accelerates the convergence speed of the model by normalising the features. Within the transformer model, the most critical components primarily encompass the following three aspects.

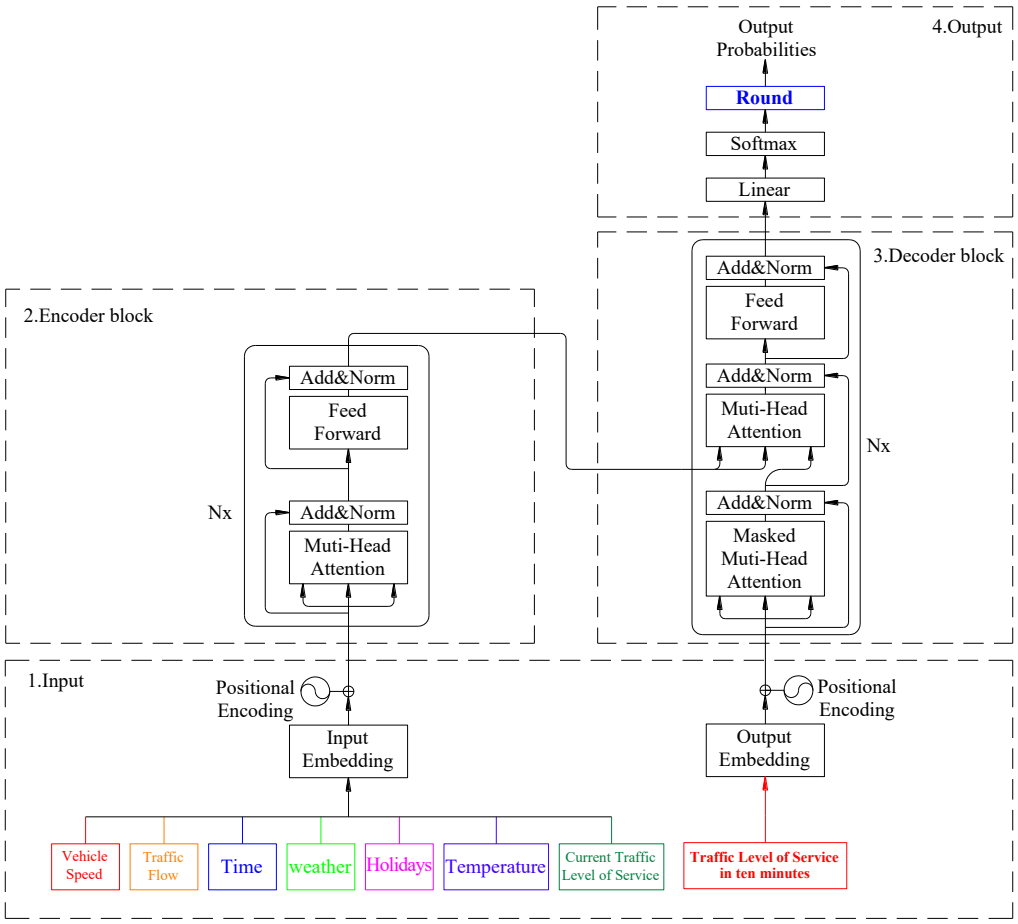


Figure 1 – Transformer-based level of service prediction

1) Positional Encoding

Positional encoding is an essential technical mechanism in transformer models that explicitly incorporates positional information to compensate for the self-attention mechanism's inability to inherently capture sequential ordering in input sequences. While the transformer architecture achieves superior computational efficiency through parallel processing of all input elements, this design inherently results in permutation invariance – the inability to discern the original ordering of elements. The integration of positional encoding provides the model with crucial sequential awareness, enabling it to effectively learn and utilise positional relationships within the data. The encoding process is represented by *Equation 1.*

$$E_t = \begin{cases} PE(pos, 2i) = \sin(pos/10000^{2i/d_{model}}) \\ PE(pos, 2i+1) = \cos(pos/10000^{2i/d_{model}}) \end{cases} \quad (1)$$

In this equation, *pos* represents the time index within the input traffic flow segment, which allows the network to learn the importance of different time points for the prediction task and *dmodel* denotes the encoding dimension.

The positional encoding output is additively combined with the transformer's input features $\boldsymbol{H}_t$ to form the input matrix $\boldsymbol{X}^S$ for the multi-head attention layer, as expressed by *Equation 2*.

$$X^S = E_t + H_t \quad (2)$$

2) Multi-Head Attention Mechanism

Single-head attention can learn a spatial representation by mapping the input features to a learnable, high-dimensional subspace. Each node in this subspace can be divided into three subspaces, represented by *Equations 3* to *5*. Here $W_q^S \in R^{d_{model} \times d_k^S}$, $W_k^S \in R^{d_{model} \times d_k^S}$, $W_v^S \in R^{d_{model} \times d_k^S}$, d_k^S represent the dimensions of the value vectors in each attention head.

$$\boldsymbol{Q}^S = X^S \boldsymbol{W}_q^S \tag{3}$$

$$\boldsymbol{K}^S = X^S \boldsymbol{W}_k^S \tag{4}$$

$$\boldsymbol{V}^S = X^S \boldsymbol{W}_v^S \tag{5}$$

After obtaining these three high-dimensional subspaces, the single-head attention output $\boldsymbol{M}^S$ can be calculated using *Equation 6*. When the input dimension is high, the function may enter a saturation zone during the calculation of attention weights $(d_k^S)^{0.5}$. Therefore, a scaling factor is introduced to keep the input within a reasonable range.

$$\boldsymbol{M}^S = Softmax\left(\frac{\boldsymbol{Q}^S \boldsymbol{K}^{S^T}}{\sqrt{d_k^S}} \boldsymbol{V}^S\right) \tag{6}$$

To address the limitation of classical attention mechanisms, which cannot extract information from different subspaces, researchers in [16] proposed the multi-head attention mechanism. This mechanism allows the model to learn multiple subspaces and capture different dependencies within these subspaces. The outputs of multiple single-head attentions are concatenated to achieve multi-head attention, enhancing the model's representation capability, as shown in *Equation 7*.

$$MultiHead(\boldsymbol{Q}^S, \boldsymbol{K}^S, \boldsymbol{V}^S) = Concat(\boldsymbol{M}_1^S, \boldsymbol{M}_2^S, \dots, \boldsymbol{M}_h^S)\boldsymbol{W}^O \tag{7}$$

Here $\boldsymbol{W}^O \in \boldsymbol{R}^{hd_k^S \times d_{model}}$, h represents the number of attention heads in the multi-head attention mechanism.

Following the multi-head self-attention layer is the feed-forward neural network layer, which consists of two fully connected layers. The first layer uses the ReLU activation function, while the second layer uses a linear activation function. The core role of the feed-forward neural network layer is to further transform the output of the multi-head self-attention layer through non-linear transformations, extracting more complex features.

3) Training and Optimisation of the Transformer Model

The transformer model uses mean squared error as the loss function, as shown in *Equation 8*. During training, the Adam optimiser is employed to optimise the model. The Adam optimiser combines the advantages of momentum and RMSProp, automatically and flexibly adjusting the learning rate based on the training process. This effectively improves the training efficiency and convergence speed, helping the model to reach an optimal solution more quickly and accurately.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - o_i)^2 \tag{8}$$

In the equation, y_i represents the predicted traffic level of service, and o_i represents the actual traffic level of service.

2.2 Dung beetle optimisation algorithm

The original transformer model requires manual parameter setting, typically determined based on engineering experience, which is labour-intensive and cumbersome, making it challenging to identify the optimal parameter combination. To enhance the simulation capability of the transformer model, this paper constructs a dung beetle optimiser-based transformer (DBO-transformer) model. This model enables rapid identification of the optimal positions for two hyperparameters – the number of self-attention heads and the learning rate – during local search, thereby providing better parameters for transformer training. The inspiration for the dung beetle optimisation algorithm primarily stems from the behaviours of dung beetles, such as ball rolling, dancing, foraging, stealing and reproduction [17]. By simulating these behaviours, the algorithm seeks the optimal solution to optimisation problems.

1) Rolling Behaviour

When rolling dung balls without obstacles, dung beetles use the sun for navigation, allowing the dung ball to roll in a straight line. This process is represented by *Equations 9* and *10*:

$$x_i(t+1) = x_i(t) + a \times k \times x_i(t-1) + b \times \Delta x \quad (9)$$

$$\Delta x = |x_i(t) - X_w| \quad (10)$$

In the equation, t represents the current iteration, $x_i(t)$ represents the position of the i-th dung beetle at the t-th iteration, α is a natural coefficient (taking values of -1 or 1), $k \in (0,0.2]$ is the deflection coefficient, X_w represents the global worst position, $b \in (0,1)$ is a constant, $|x_i(t) - X_w|$ represents the light intensity, where higher values indicate weaker light.

2) Dancing Behaviour

When encountering obstacles that prevent forward movement, dung beetles can change direction to obtain a new path. This process is represented by *Equation 11*:

$$x_i(t+1) = x_i(t) + tan(\theta) \times |x_i(t) - x_i(t-1)| \quad (11)$$

In the equation, θ is the deflection angle, which takes values within the interval [0,π]. When θ is 0, π/2 or π, the position of the dung beetle remains unchanged. $|x_i(t) - x_i(t-1)|$ represents the position offset of the i-th dung beetle between iterations t and t-1.

3) Reproduction Behaviour

The study employs a boundary selection strategy to simulate the egg-laying area of female dung beetles, dynamically adjusting the egg-laying zone. This process is represented by *Equations 12* and *13*:

$$L_b^* = max(X^* \times 1 - R, L_b) \quad (12)$$

$$U_b^* = min(X^* \times 1 + R, U_b) \quad (13)$$

In the equation, X^* represents the current local optimal position, L_b^* and U_b^* are the lower and upper bounds of the egg-laying zone, $R = (1-t)/T_{max}$, T_{max} is the maximum number of iterations, and L_b and U_b are the lower and upper bounds of the optimisation problem.

Mature dung beetles emerge from the ground to forage, with the foraging area dynamically updated. This process is represented by *Equations 14* and *15*:

$$L_{bb} = max(X_b \times 1 - R, L_b) \quad (14)$$

$$U_{bb} = min(X_b \times 1 + R, U_b) \quad (15)$$

In the equation, X_b represents the global optimal position, L_{bb} and U_{bb} are the lower and upper bounds of the foraging zone. The position update formula for young dung beetles is:

$$x_i(t+1) = x_i(t) + C_1 \times (x_i t - L_{bb}) + C_2 \times (x_i t - U_{bb}) \quad (16)$$

In the equation, C_1 is a random number following a normal distribution and $C_2 \in (0,1)$ is a random vector.

4) Stealing Behaviour

Stealing as a competitive behaviour is common among dung beetles. This process is represented by *Equation 17*.

$$x_i(t+1) = X_b + S \times g \times |x_i(t) - X^*| + |x_i(t) - X_b| \quad (17)$$

In the equation, X_b represents the optimal position for food competition, $x_i(t)$ is the position of the i-th young dung beetle at iteration t, g is a random vector following a normal distribution with size $1 \times D$, S is a constant value.

In the DBO algorithm, the first step involves initialising relevant parameters, including the initial data size, the number of independent variables, and the upper and lower bounds of the independent variables, followed by calculating the initial fitness value. The fitness value represents the adaptability of the parameter scheme after optimisation. During the iteration process, each parameter is updated. Subsequently, the parameters are optimised, and the fitness value after optimisation is calculated. Finally, it is determined whether the parameters and fitness values meet the termination conditions. If satisfied, the process exits and outputs the results; otherwise, the above steps are repeated, as illustrated in *Figure 2*.

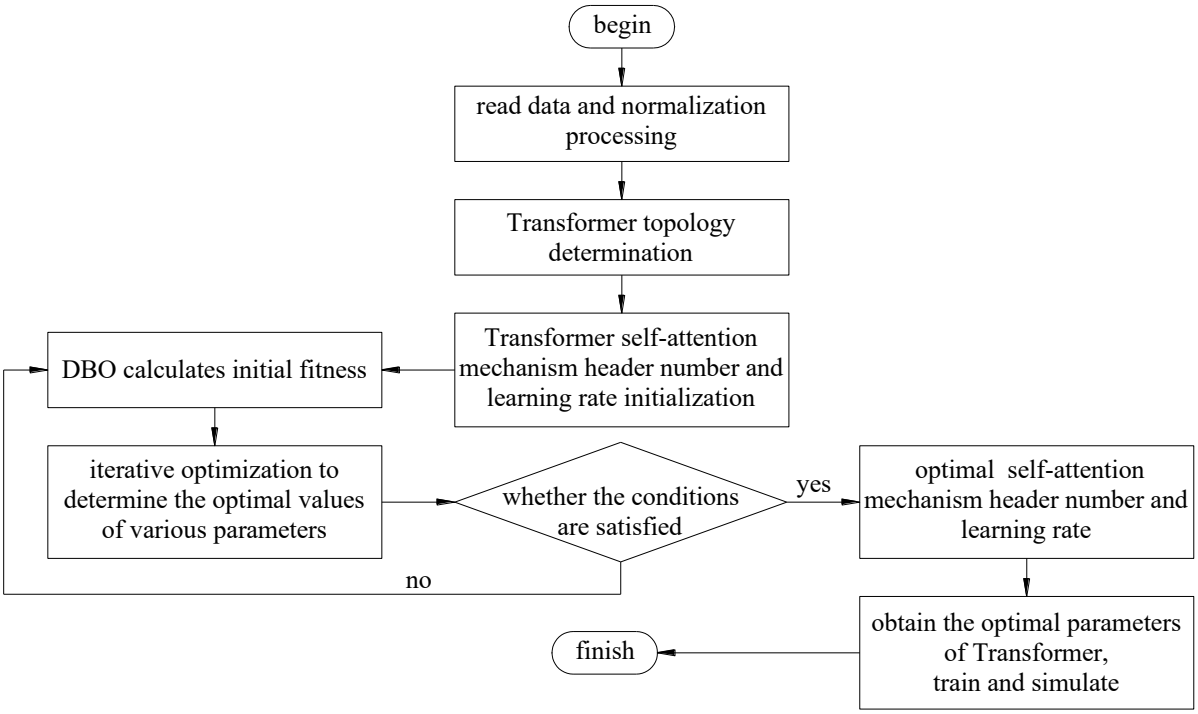


Figure 2 –DBO-transformer neural network model computation

The DBO-transformer model is initialised with a learning rate of 0.001, a maximum training epoch of 200, a dung beetle population size of 10, an evolution count of 20 and individual value bounds of [-5,5]. Additionally, during the selection of optimal parameters, the minimum learning rate is set to 10^-10, and the number of self-attention heads is constrained to positive integers.

3. EXPERIMENTAL VALIDATION

3.1 Classification of traffic level of service

The urban road traffic performance evaluation system provides detailed criteria for classifying the states of expressways, arterial roads, secondary roads and branch roads in urban areas, as shown in *Table 1* [18]. In this study, the average travel speed of each road segment in the network was calculated at 10-minute intervals (not exceeding 15 minutes) to determine the traffic level of service for each segment. In this study, the traffic level of service is quantitatively classified using a five-level ordinal scale, where 1 to 5 respectively indicate smooth, basically smooth, slow, moderate congestion and severe congestion.

Table 1 – Classification of traffic level of service (km/h)

Road class	Smooth	Basically smooth	Slow	Moderate congestion	Severe congestion
Expressway	>65	(50,65]	(35,50]	(20,35]	≤20
Arterial road	>40	(30,40]	(20,30]	(15,20]	≤15
Secondary & branch roads	>35	(25,35]	(15,25]	(10,15]	≤10

3.2 Data sources

This study utilises a traffic dataset from an expressway on a bridge in Fuzhou, collected in 2021 with a sampling interval of 1 minute. The dataset contains a large volume of data, and although there are minor instances of missing data, they do not significantly impact the experimental results. The dataset accurately reflects the temporal characteristics of vehicle speed variations at the observation points.

Factors influencing the traffic level of service

Traffic volume and speed are direct factors contributing to the traffic level of service, particularly during peak morning and evening hours when traffic flow concentrates. Increased vehicle density on roads leads to reduced average speeds and limited safe distances between vehicles. When road capacity is insufficient to meet demand, traffic congestion occurs, potentially spreading across the network. The level of service is influenced by a combination of factors, and its prediction requires the use of big data techniques. In general, apart from vehicle density and speed, the traffic level of service is also affected by factors such as weather, weekends, holidays and peak travel periods [19]. The use of multi-source data fusion techniques can provide more comprehensive information and more accurate results [20]. This study predicts the next moment's traffic level of service based on the previous ten minutes' vehicle speed, traffic volume, time, weather, holidays, temperature and traffic level of service, aiming to contribute to the assessment and improvement of traffic level of service. Weather conditions are classified into six levels based on the type and intensity of precipitation: clear, cloudy, light rain, moderate rain, heavy rain and storm. Holidays are categorised into four levels: the day before a holiday, the last day of a holiday, the holiday itself and other days.

Correlation analysis of influencing factors

To validate the correlation between various influencing factors and the traffic level of service, analysis of variance was used for quantitative data, and the chi-square test was applied for categorical data. These tests were employed to examine the differences in traffic level of service based on the previous time step's vehicle speed (x_1), traffic volume (x_2), time (x_3), weather (x_4), holidays(x_5), temperature (x_6) and traffic level of service (x_7). The results are presented in *Table 2*. As shown in the table, vehicle speed, traffic volume, time, weather, holidays, temperature and traffic level of service from the previous time step all exhibit significant correlations with traffic level of service ($p < 0.05$).

Table 2 – Correlation analysis between influencing factors and traffic level of service at the previous time step

Factor	Vehicle speed	Traffic flow	Time	Weather	Holidays	Temperature	Current traffic level of service
χ^2	—	—	—	829.131	306.217	—	104751.510
F	19463.747	15014.185	108.215	—	—	108.215	—
p	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Based on multi-source information, traffic data from 23 days were selected, including 15 weekdays, 4 holidays, 2 days at the end of holidays and 2 days before holidays. The weather conditions included 4 days of clear weather, 5 days of cloudy weather, 4 days of light rain, 3 days of moderate rain, 3 days of heavy rain and 4 days of storms, covering the months of January, April, June, August, September and October. A total of 32,571 time points were extracted from the dataset, with 6,170 classified as smooth traffic, 8,022 as basically smooth, 7,733 as slow, 10,105 as moderate congestion and 541 as severe congestion.

3.3 Data preprocessing

To eliminate the differences in units and value ranges among various influencing factors and to enhance the model's generalisation capability, the data related to the traffic level of service and its influencing factors were normalised using *Equation 18*:

$$\bar{x} = (x - x_{min})/(x_{max} - x_{min}) \tag{18}$$

In the equation, $\bar{x}$ represents the normalised data, x denotes the original data, x_{min} is the minimum value in the original dataset and x_{max} is the maximum value in the original dataset. The normalisation process transforms the data into a standardised range, ensuring uniformity across all features. A subset of the normalised data is presented in *Table 3*.

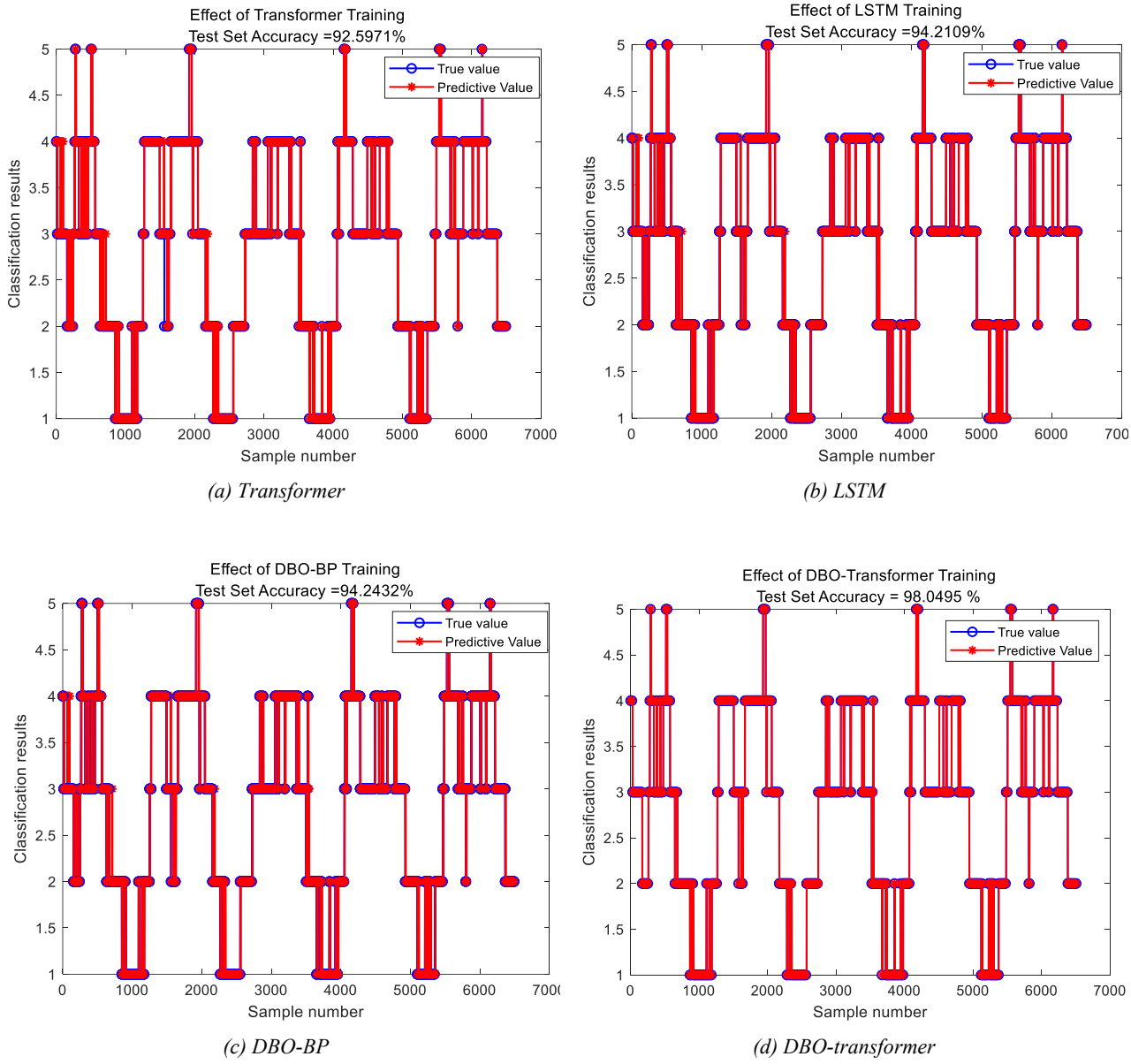
Table 3 – 20 sets of normalised data

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	y
1	0.638	0.106	0.065	0.600	0.000	1.000	0.000	0.000
2	0.585	0.265	0.013	0.400	0.333	0.694	0.250	0.000
3	0.658	0.168	0.021	0.200	0.667	0.449	0.000	0.000
4	0.558	0.106	0.009	0.200	1.000	0.755	0.000	0.000
5	0.538	0.398	0.008	0.000	0.333	0.082	0.250	0.250
6	0.502	0.062	0.017	0.800	0.000	0.939	0.250	0.250
7	0.568	0.088	0.040	1.000	0.000	0.980	0.000	0.250
8	0.528	0.027	0.014	0.400	0.000	0.000	0.000	0.250
9	0.449	0.460	0.340	0.200	0.667	0.449	0.250	0.500
10	0.422	0.487	0.320	0.400	0.333	0.694	0.500	0.500
11	0.492	0.487	0.347	0.800	0.333	0.755	0.250	0.500
12	0.548	0.451	0.284	1.000	0.000	0.980	0.250	0.500
13	0.296	0.673	0.298	1.000	0.000	0.980	0.500	0.750
14	0.262	0.504	0.418	0.800	0.333	0.755	0.750	0.750
15	0.243	0.611	0.404	0.200	0.667	0.449	0.750	0.750
16	0.309	0.540	0.403	0.400	0.333	0.694	0.500	0.750
17	0.123	0.292	0.900	0.000	0.333	0.082	1.000	1.000
18	0.213	0.327	0.750	0.000	0.000	0.224	0.750	1.000
19	0.083	0.389	0.580	0.800	0.333	0.755	1.000	1.000
20	0.179	0.310	0.780	0.800	0.000	0.939	1.000	1.000

The dataset was divided into 80% for training and 20% for testing. Input features included vehicle speed, traffic flow, time, weather, holidays, temperature and current traffic level of service, while the output was the traffic level of service 10 minutes later. The model was trained using these inputs, and the resulting predictions were rounded to produce the final output.

3.4 Experimental results comparison

To further validate the superiority of the proposed method, four prediction models – transformer, LSTM, DBO-BP and DBO-transformer – were developed to predict the traffic level of service. For the transformer model during the training phase, the input channel was set to 27, the number of self-attention heads was 2 with 64 units per head, the number of training iterations was 200, and the learning rate was 0.0000001. After optimisation using the dung beetle optimiser, the number of self-attention heads was adjusted to 4, and the learning rate was set to 0.0024. Implemented in MATLAB R2023b, the comparative prediction results and accuracy rates of the four models for the traffic level of service are illustrated in *Figures 3(a)* to *3(d)*.



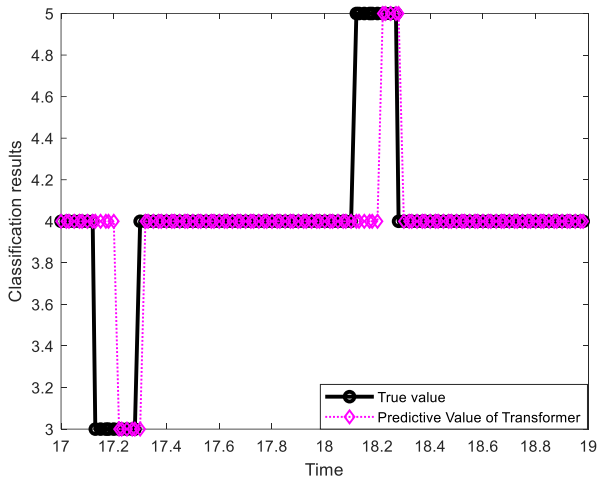
(a) Transformer *(b) LSTM*

(c) DBO-BP *(d) DBO-transformer*

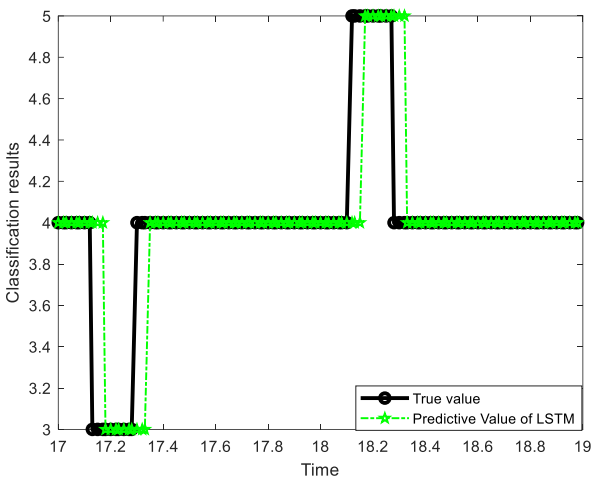
Figure 3 – Prediction results of different models

As shown in *Figure 3*, all four models achieve relatively accurate predictions of traffic level of service with accuracy rates exceeding 90% after integrating multi-source information. Among them, the DBO-transformer model demonstrates the best predictive performance, achieving an accuracy rate of 98.0495%.

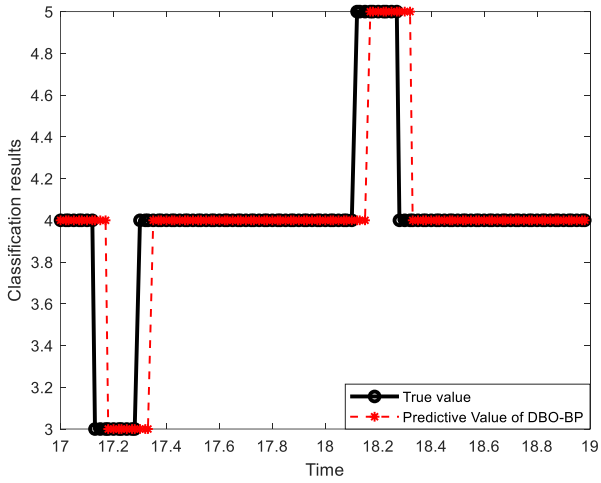
To provide a clearer comparison of the four models, the prediction results and accuracy rates of the traffic level of service for the four comparative models during a specific day from 17:00 to 19:00 are illustrated in *Figures 4(a)* to *4(d)*. As shown in *Figure 4*, the transformer, LSTM and DBO-BP models all underperform compared to the DBO-transformer model. The DBO-transformer model demonstrates superior stability and accuracy.



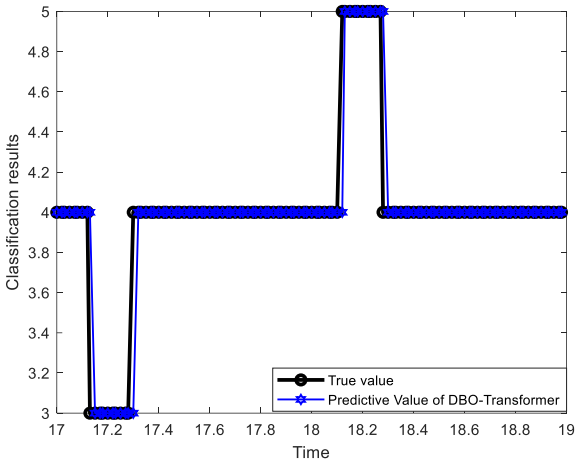
(a) Transformer



(b) LSTM



(c) DBO-BP



(d) DBO-transformer

Figure 4 – Prediction results of different models from 17:00 to 19:00

To more comprehensively evaluate the effectiveness of the proposed method, the coefficient of determination (R^2), mean absolute error (*MAE*), root mean square error *(RMSE*) and mean absolute percentage error (*MAPE*) were selected to analyse the four prediction models.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - o_i)^2}{\sum_{i=1}^{n}(\bar{y}_i - o_i)^2} \tag{19}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - o_i| \tag{20}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - o_i)^2} \tag{21}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - o_i|}{o_i} \tag{22}$$

The calculation results of the four evaluation metrics for the four models are presented in *Table 4*. As shown in *Table 4*, the DBO-transformer model, which employs the optimal parameters identified by the dung beetle optimiser, demonstrates superior performance. It significantly enhances the accuracy of traffic level of service prediction, outperforming the LSTM and DBO-BP neural network models.

Table 4 – Comparison of test set results

	R^2	*MAE*	*RMSE*	*MAPE*
Transformer	0.9353	0.0743	0.2738	3.37%
LSTM	0.9488	0.0584	0.2435	2.59%
DBO-BP	0.9491	0.0580	0.2428	2.58%
DBO-transformer	0.9828	0.0197	0.1414	0.88%

3.5 Further analysis and discussion

The prediction results of the four models – transformer, LSTM, DBO-BP and DBO-transformer were compared with the actual values. The counts of discrepancies between predicted and actual values are presented in *Table 5*. From the data in the table, it can be observed that the discrepancies between actual and predicted values for the four models are limited to three cases: -2, -1 and 1. Notably, the number of discrepancies with a value of -2 is extremely small, almost negligible.

Table 5 – Counts of discrepancies between actual and predicted values

	-2 (severely underestimated LOS classification)	**-1 (slightly underestimated LOS classification)**	**1 (slightly overestimated LOS classification)**
Transformer	2	296	184
LSTM	3	188	185
DBO-BP	3	186	186
DBO-transformer	1	63	63

Note: Deviation value = Actual LOS grade - Predicted LOS grade

Specifically, the transformer model exhibits a significantly higher frequency of discrepancies with a value of -1, indicating that its predictions are more dispersed at this deviation level. The LSTM and DBO-BP models show very similar performance across all discrepancy values, and both fall short of the DBO-transformer model in terms of overall prediction stability and accuracy, although the distinction between them is not pronounced.

In summary, the DBO-transformer model demonstrates lower frequencies of discrepancies, whether -2, -1 or 1, compared to the other three models. This indicates that the predictions of the DBO-transformer model are more closely distributed around the actual values, showcasing superior prediction stability and accuracy. In scenarios requiring stringent prediction accuracy, the DBO-transformer model exhibits more prominent application potential.

4. CONCLUSIONS

This study, based on the vehicle speed and traffic flow data monitored from an expressway in Fuzhou, innovatively proposes a traffic operation status evaluation model. This model comprehensively considers multiple factors such as vehicle speed, traffic flow, time, weather conditions, whether it is a holiday, temperature, and the current traffic operation status. It can directly predict the traffic level of service at the next moment, breaking the reliance of traditional prediction methods on current vehicle speed and flow data. Using the dung beetle optimiser algorithm to find the optimal parameters for the transformer and establishing the DBO-transformer traffic level of service prediction performance provides a new approach and method for urban traffic level of service prediction. Furthermore, this offers strong support for the further improvement of the intelligent traffic control system.

Numerous studies have attempted to use traditional time series prediction models and neural network models to predict the short-term traffic level of service. However, most of these studies are based on single data sources or simple indicators, lacking sufficient consideration for the complexity of actual traffic congestion situations, especially the comparative analysis based on multi-source information fusion. To more accurately predict urban traffic level of service, this study delves into the impact of multiple factors such as vehicle speed, traffic flow, time, weather, holidays, temperature and the previous traffic operation status on traffic level of service. This is achieved by utilising key factors from the previous moment to explore a method for predicting the traffic level of service by effectively using historical data without referencing real-time data.

By employing the dung beetle optimiser algorithm to identify the optimal learning rate and number of self-attention heads for the transformer model, a DBO-transformer-based traffic level of service evaluation model was established. This model was compared and analysed against the DBO-optimised BP neural network, LSTM neural network and the standard transformer model. The results of the experiment demonstrate that all four models exhibit strong predictive performance, with accuracy rates exceeding 90%. Further in-depth comparative analysis reveals that the transformer model holds significant advantages in feature extraction, parallel computation and handling long sequences. The DBO-transformer model, enhanced by intelligent optimisation, achieves the best prediction performance.

This study primarily focuses on core factors such as traffic flow, speed, weather and date. Due to challenges in data acquisition, spatial factors were not incorporated into the analysis of the traffic level of service. Future work aims at collecting more comprehensive data, integrating both temporal and spatial factors to refine the analysis of the traffic level of service.

ACKNOWLEDGEMENTS

This work was supported in part by the National Natural Science Foundation of China under Grant 62301159 and in part by the Natural Science Foundation of Fujian Province under Grant 2023J01229.

REFERENCES

[1] Li Yong, Liu Yulan, Zou Kai. Research on the critical value of traffic congestion propagation based on coordination game. *Procedia Engineering*. 2016;137:754-761. DOI: 10.1016/j.proeng.2016.01.313.

[2] Wang Y, et al. Real-time joint traffic state and model parameter estimation on freeways with fixed sensors and connected vehicles: State-of-the-art overview, methods, and case studies. *Transportation Research Part C: Emerging Technologies*. 2022;134:103444. DOI: 10.1016/j.trc.2021.103444.

[3] Trinh XS, et al. Incremental unscented kalman filter for real-time traffic estimation on motorways using multi-sourc e data. *Transportmetrica A: Transport Science*. 2022;18(3):1127-1153. DOI: 10.1080/23249935.2021.1931548.

[4] Zhang W, et al. Queue length estimation and accuracy assessment method for intersections based on trajectory data. *China Journal of Highway and Transport.* 2022;35(3): 216-225. DOI: 10.19721/j.cnki.1001-7372.2022.03.018.

[5] Wang D, Li M, Lu S. Short-term traffic volume prediction for scenic area highways based on Bayes-ARIMA. *Highway*. 2024;69(04):225-234.

[6] Rajalakshmi V, Ganesh Vaidyanathan S. Hybrid time-series forecasting models for traffic flow prediction. *Promet – Traffic&Transportation*. 2022;34(4):537-549. DOI: 10.7307/ptt.v34i4.3998.

[7] Yu X, et al. Prediction of traffic congestion diffusion and dissipation on highways based on VISSIM simulation. *Journal of Shandong Institute of Technology*. 2023;31(03):47-52.

[8] Moradi H, et al. The contribution of connected vehicles to network traffic control: A hierarchical approach. *Transportation Research Part C: Emerging Technologies*. 2022;139:103644. DOI: 10.1016/j.trc.2022.103644.

[9] Xiong Z, et al. A review of short-term traffic flow prediction based on deep learning. *Computer Engineering and Applications*. 2025;1-19. [Online] Available: http://kns.cnki.net/kcms/detail/11.2127.TP.20241210.1603.006.html.

[10] Chen Y, et al. A traffic congestion prediction algorithm based on Softmax-enhanced convolutional neural network and bidirectional long short-term memory network framework. *Science Technology and Engineering*. 2022;22(29):12917-12926. DOI: 10.3969/j.issn.1671-1815.2022.29.027.

[11] Zhang S, et al. A short-term traffic congestion prediction model based on FCM-RBF. *Journal of Chongqing University of Technology (Natural Science)*. 2023;37(03):12-21. DOI: 10.3969/j.issn.1674-8425(z).2023.03.002.

[12] Li S, Yang L, Zhao X. A short-term urban traffic congestion prediction algorithm based on deep learning. *Science Technology and Engineering*. 2023;23(25):10866-10878. DOI: 10.3969/j.issn.1671-1815.2023.25.028.

[13] Yan J, et al. A study on urban traffic congestion prediction: A case of Xi'an. *Computer Engineering and Applications*. 2025;61(01):330-340. DOI: 10.3778/j.issn.1002-8331.2407-0100.

[14] Zhou Y, et al. Traffic prediction technology driven by graph neural networks: Exploration and challenges. *Journal of Internet of Things*. 2021;5(04):1-16. DOI: 10.11959/j.issn.2096-3750.2021.00235.

[15] Qinyang W, et al. Fusing visual quantified features for heterogeneous traffic flow prediction. *Promet – Traffic&Transportation.* 2024;36(6):1068-1077. DOI: 10.7307/ptt.v36i6.667.

[16] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *31st Conference on Neural Information Processing Systems.* 2017;6000-6010. DOI: 10.48550/arXiv.1706.03762.

[17] Jiankai X, Bo S. Dung beetle optimizer: a new meta-heuristic algorithm for global optimization. *The Journal of Supercomputing*. 2023;79(7):7305-7336. DOI: 10.1007/s11227-022-04959-6.

[18] Beijing Bureau of Quality and Technical Supervision. DB11/T 785-2011. Urban Road Traffic Performance Index. Standards Press of China, Beijing, 2011.

[19] Wei Q, et al. A study on the prediction of urban road traffic congestion index considering multiple factors. *Journal of Transportation Systems Engineering and Information Technology*. 2017;17(01):74-81. DOI: 10.16097/j.cnki.1009-6744.2017.01.012.

[20] Liang Q, et al. Mining and prediction of public transport travel behavior features driven by multi-source data. *Science Technology and Engineering.* 2021;21(28)2:11. DOI: 10.3969/j.issn.1671-1815.2021.28.004.