Journal of Computer Science and Electrical Engineering
Print ISSN: 2663-1938
Online ISSN: 2663-1946
DOI: https://doi.org/10.61784/jcsee3092

SHORT-TERM TRAFFIC PREDICTION BASED ON A BI-GRU-ATTEN- ARIMA RESIDUAL FUSION MODEL

YiYang Jiao
School of Economics, Jinan University, Guangzhou 510632, Guangdong, China.
Corresponding Email: jiaoyiyang0804@163.com

Abstract: Intelligent Transportation Systems (ITS) mitigate traffic congestion through real-time planning and management, where short-term traffic forecasting is crucial. Because traffic time-series data are highly nonlinear, intricate, and history-dependent, we blend spatiotemporally correlated road-traffic datasets with exogenous factors such as holidays and major events. On this basis, we propose a residual-fusion model, Bi-GRU-Atten-ARIMA, which couples the nonlinear feature-learning capacity of a bidirectional gated recurrent unit (Bi-GRU) with an attention mechanism for adaptive feature weighting, while also exploiting the linear autocorrelation strengths of an ARIMA model. By jointly capturing nonlinear and linear patterns, the model significantly enhances forecasting accuracy. Two empirical studies on major Hong Kong region roadways—covering one-month and fifty-day datasets, respectively—validate its effectiveness, showing that the Bi-GRU-Atten-ARIMA residual-fusion model outperforms competing approaches in short-term urban-traffic prediction. Leveraging these precise forecasts, we further implement a congestion-warning module that quickly flags anomalous conditions within the traffic system.
Keywords: Short-term traffic prediction; Bidirectional gated recurrent unit; Residual fusion; Hybrid model

1 INTRODUCTION

Intelligent Transportation Systems (ITS)—real-time, precise, and highly efficient platforms for integrated transport management—play a pivotal role in easing congestion, optimising routes and networks, and pinpointing the best travel paths. Short-term traffic forecasting sits at the heart of ITS. A scientifically robust prediction model is essential for enhancing our understanding of and response to real-time traffic condition changes, thereby facilitating smart mobility and promoting the sustainable and healthy development of urban traffic management.

Firstly, single models may underperform in certain prediction tasks due to the inherent complexity of traffic data, the coexistence of linear and nonlinear structures, or the excessive intricacy of the model architecture. In response to these limitations, this study proposes a hybrid model that integrates traditional statistical approaches with deep learning techniques through an attention mechanism. Moreover, in the domain of short-term traffic prediction, deep learning methods commonly treat deep neural networks as the final stage of data processing. However, the residual sequence produced by these models is often not subjected to white noise testing to determine whether it still contains meaningful information. As a result, it is highly probable that traffic-related features remain embedded in the residuals.

To address these limitations, this paper proposes a residual fusion model based on Bi-GRU-Atten-ARIMA. The main contributions of this paper are as follows:

• This study integrates the deep learning model Bi-GRU with the traditional statistical model ARIMA. The Bi-GRU component is employed to learn and predict complex nonlinear patterns in historical traffic data, while the ARIMA model is subsequently used to extract linear features. This innovative hybrid approach leverages the respective strengths of both modeling paradigms, thereby addressing the limitations of single-model frameworks and enhancing overall prediction accuracy.

• In traffic flow prediction, correlations often exist between traffic volumes at different spatial locations. To effectively exploit this spatial correlation during the feature selection stage, this paper evaluates and rank the importance of various location detectors in predicting the traffic volume of a target detector. The most representative detector is then selected as the input for the hybrid prediction model, which not only reduces computational complexity but also improves the model's accuracy and stability.

• The incorporation of an attention mechanism further enhances the model's ability to focus on critical information across temporal positions and influencing factors within the sequence. This design enables the model to more accurately capture key spatiotemporal features in traffic flow data, reduce the impact of redundant variables, and ultimately improve overall predictive performance.

2 RELATED WORK

Over the past three decades, a wide range of methods have been developed to predict macroscopic traffic conditions. These approaches stem from diverse disciplines, including statistics, control theory, artificial intelligence, and applied mathematics. While various classification schemes have been proposed in the literature, this paper adopts a broad categorization that divides existing methodologies into two main groups: time series prediction models grounded in traditional statistical theory and deep learning models based on neural networks.

2.1 Traffic Flow Prediction Using Statistical Models

In the early stages of traffic-flow prediction research, traditional statistical models were widely adopted because of their well-established theoretical foundations, high computational efficiency, and strong interpretability. Representative approaches include the Historical Average (HA) model, Kalman Filtering, and the Autoregressive Integrated Moving Average (ARIMA) model. ARIMA, a classical time-series model that captures sequential trends, effectively handles temporal dependencies and non-stationarity in data and therefore occupies an important role in short-term traffic forecasting. M. Ahmed et al. were the first to apply the ARIMA model to highway traffic-flow prediction[1]. In order to enhance the adaptability of the model and further consider the impact of exogenous variables on traffic flow, the ARIMAX model was applied in subsequent studies. B. Williams used data from all relevant upstream sections as input to predict traffic flow on selected downstream sections[2]. The results showed that the ARIMAX model that takes spatial correlation into account is superior to the ARIMA model. Furthermore, Y. Kamarianakis and P. Prastacos proposed the Spatiotemporal ARIMA (STARIMA) model[3], which integrates spatial correlations between adjacent locations with temporal autocorrelations in a unified framework. Many traditional statistical time series models have already been applied to traffic condition forecasting. Benchmark models such as ARIMA have been widely used for single-location traffic prediction and are frequently selected by researchers as a baseline for comparison with other models. However, most of these are linear models that perform well in extracting linear features but have limitations in capturing nonlinear fluctuations.

2.2 Traffic Flow Prediction Using Deep Learning

In recent years, with the development of traffic big data and computing power, deep learning methods have gradually become the mainstream of traffic forecasting research. Traffic flow forecasting is essentially a time series forecasting problem. To capture the temporal correlations in the data, recurrent neural networks (RNNs) and their variants are widely used for traffic time series modeling. X. Huang et al. applied the long short-term memory (LSTM) network to traffic-flow prediction and verified that it outperforms traditional recurrent neural networks (RNNs) in capturing long-term temporal dependencies[4]. Subsequently, B. Li et al. incorporated multivariate auxiliary information on top of the LSTM framework and developed a multivariate LSTM model to improve highway-traffic forecasting accuracy[5]. To further highlight critical time segments, E. Sherafat et al. proposed an LSTM-Attention model that uses an attention mechanism to adaptively weight key temporal features[6]. As temporal modeling capabilities advanced, bidirectional structures became a new research focus. P. Redhu and K. Kumar and B. Naheliya et al. successively introduced particle-swarm-optimized (PSO) and moth-flame-optimized (MFOA) bidirectional LSTM (Bi-LSTM) models that leverage both historical and future information to enhance predictive accuracy[7,8]. Considering spatial correlations among different road segments, J. Wang and C. Susanto and F. Ma et al. proposed hybrid CNN-LSTM models[9,10], in which convolutional layers extract spatial features while LSTM layers capture temporal dynamics, thereby achieving joint spatiotemporal modeling. Going a step further, N. Singh et al. developed an attention-based spatiotemporal LSTM model that excels at capturing complex spatiotemporal dependencies[11]. In parallel with the LSTM family, gated recurrent unit (GRU) networks and their variants have also gained attention. H. Ding et al. and R. Li et al. demonstrated the effectiveness of GRU in traffic-flow and travel-time prediction[12,13], while G. Shi and L. Luo reported that GRU can outperform LSTM in certain urban-rail scenarios[14]. To strengthen long-sequence dependency modeling, N. Chauhan et al. proposed a bidirectional GRU (Bi-GRU) with a confined-attention mechanism[15], further improving prediction under complex traffic conditions. X. Sun et al. built a shared-weight spatiotemporal GRU model to efficiently capture spatial features[16].

A review of the literature on both categories of prediction methods reveals that, while statistical models are grounded in well-established theoretical frameworks, their overly rigid assumptions often constrain predictive accuracy, resulting in performance bottlenecks and limited model effectiveness. In contrast, deep learning–based approaches have demonstrated strong potential due to the powerful feature extraction and fitting capabilities of neural networks. Nevertheless, these models also exhibit notable limitations, including the need for large-scale training data, challenges in selecting appropriate network architectures, dependence on empirical tuning of hyperparameters, lack of guaranteed convergence to an optimal solution, and relatively slow training processes. In real-world scenarios, time series data are often complex, characterized by a mixture of linear trends and nonlinear, non-stationary fluctuations. As a result, single-model approaches to time series prediction frequently fall short of delivering satisfactory performance. To address these gaps, this study proposes a Bi-GRU-Atten-ARIMA residual-fusion model. The design follows a "division-of-labor" strategy: a bidirectional GRU equipped with an attention mechanism first extracts nonlinear spatiotemporal features and assigns adaptive weights to key temporal inputs, while an ARIMA module models the residual sequence to capture remaining linear autocorrelations. By integrating the strengths of both paradigms, the hybrid framework enhances predictive accuracy, robustness, and interpretability, offering a more comprehensive solution for complex urban traffic-flow forecasting.

3 METHODOLOGY

The architecture of the proposed Bi-GRU- Atten-ARIMA model comprises two primary components: a deep learning module and a traditional statistical module. The deep learning component is designed to extract nonlinear features, while the statistical model focuses on capturing linear relationships. Additionally, the spatial characteristics and temporal dependencies of the time series data are incorporated at the input stage, enabling a more holistic and accurate prediction of time-dependent patterns. The nonlinear feature extraction module consists of an input layer, Bi-GRU layers, an attention mechanism layer, and an output layer. For the linear component, the ARIMA model is employed to further model the residuals from the nonlinear predictions. The model construction process is outlined as follows:

• Input Layer: The input time series data, along with selected features and the target prediction sequence, are formatted into a supervised learning structure compatible with neural networks, organized according to time steps.

• Bi-GRU Layer: This layer is composed of two Bi-GRU layers, each with a different number of neurons, forming a hierarchical feature extractor. The deeper hidden layers are tasked with capturing more complex patterns, whereas the shallower layers focus on finer-grained features. This hierarchical configuration improves the model's flexibility and enables better adaptation to varying data characteristics. The two Bi-GRU layers perform bidirectional training on the preliminary feature vectors extracted from the preceding layer, capturing deeper temporal dependencies in the load data. All outputs from these layers are subsequently fed into the Attention layer.

• Attention Layer: This layer assigns different weights to the hidden states output by the Bi-GRU layer, emphasizing the influence of key features on the prediction results.

• Output Layer: A fully connected layer is used to connect with the Attention layer. The Sigmoid function is adopted as the activation function, followed by a denormalization process to obtain the final nonlinear prediction output $\hat{N}$. The structure of the neural network section is shown in Fig. 1.

• Residual Series: The residual sequence is obtained by subtracting the denormalized prediction values from the true values $\hat{L}$ of the original time series.

• ARIMA Model: The residual sequence is further analyzed using the ARIMA model to generate predicted values for the residuals on the test set, representing the linear component of the prediction $\hat{L}$ .Finally, the overall prediction results $\hat{Y}$ of the hybrid model are obtained by adding the nonlinear prediction output from the neural network and the predicted values of the residual sequence. The calculation formula is shown in Eq. (1).

$$\hat{Y} = \hat{N} + \hat{L} \tag{1}$$

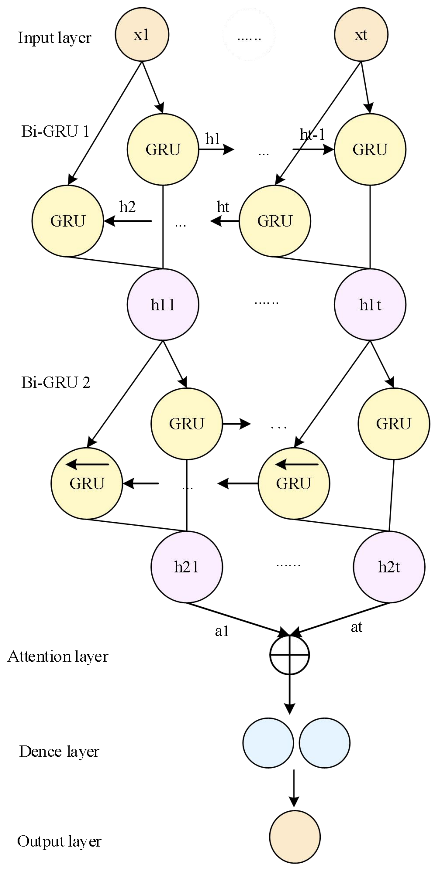


Figure 1 Structure of the Attention-Based Bi-GRU Model

The model first capitalizes on the strength of nonlinear feature extraction to effectively capture both local and global patterns in time series traffic flow data, thereby improving its capacity to model temporal dynamics. In the prediction phase, an attention mechanism is introduced to account for the varying influence of different feature states on the final output. By assigning adaptive weights to the outputs of each hidden layer, the model achieves more precise nonlinear feature prediction. In addition, the model incorporates a linear feature extraction component that is capable of identifying and modeling trends and periodicities within the time series. Through the integration of the nonlinear and linear prediction results, the proposed model delivers enhanced accuracy in short-term traffic flow forecasting. Fig. 2 shows the flow chart of the proposed approach.

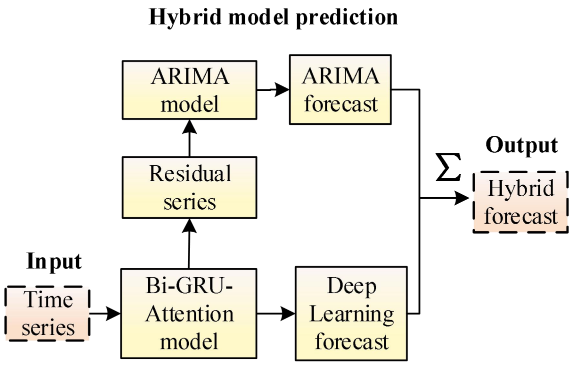


Figure 2 Flow Chart of the Proposed Approach

4 CASE DATA

4.1 Dataset Description

The dataset employed in this study comprises three key traffic indicators: traffic volume, average vehicle speed, and road occupancy. These data were obtained from the Transport Department of the Hong Kong Special Administrative Region and collected via real-time traffic monitoring devices, including major road detectors and smart lampposts distributed throughout Hong Kong region. Two empirical traffic sequence datasets were utilized for analysis. The first dataset spans from March 1 to March 31, 2024, encompassing all three indicators. To enhance forecasting robustness and improve training accuracy through an expanded sample size, a second dataset extends the observation period to 50 days, covering February 11 to March 31, 2024. Both datasets were acquired from the official government open data platform, DATA.GOV.HK.

4.2 Data Selection and Integration

The raw data comprise 30-second interval records of traffic flow, vehicle speed, and road occupancy, continuously collected throughout each day from all major road detectors and smart lampposts across the city. However, this study focuses specifically on representative arterial roads in densely populated central urban districts. In selecting the roads, considerations included urban planning structure, traffic density, and spatial connectivity. Ultimately, seven interconnected primary roads were chosen as the study area: West Kowloon Corridor, Gascoigne Road, Gascoigne Road Flyover, Princess Margaret Road, Hung Hom Road, Salisbury Road, and Nathan Road.
Based on the specified detector and smart lamp post IDs along seven designated roads and the availability of corresponding data, a total of 47 detectors and smart lamp posts were selected to construct the dataset. The raw traffic time series data, initially recorded at one-minute intervals, were subsequently aggregated (for traffic flow) or averaged (for average speed and occupancy rate) to produce time series data at 15-minute intervals. Accordingly, in the one-month dataset, each detector comprises three traffic indicator time series, each containing 2,976 observations. Likewise, in the 50-day dataset, each detector includes three traffic indicator time series, each with 4,800 observations.

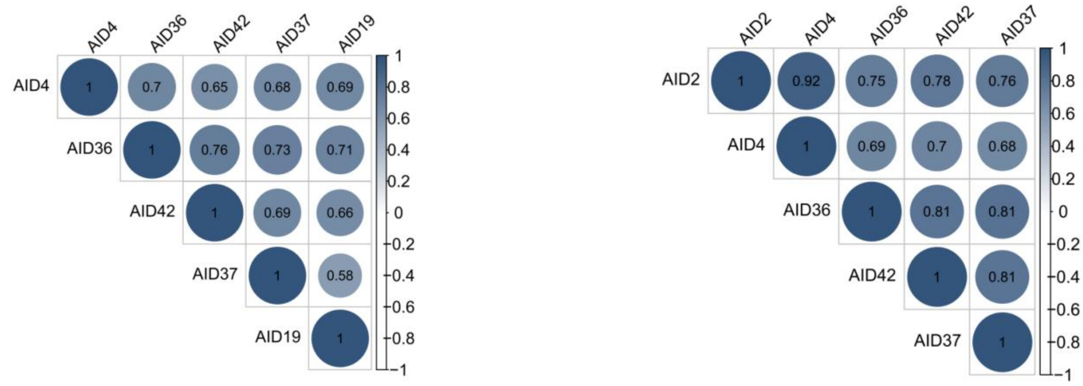
4.3 Data Preprocessing

Before model training and prediction, data preprocessing is needed, including removing duplicates, filling missing values, and correcting outliers. This study uses forward-backward filling to ensure data integrity and continuity, making the time series more realistic for subsequent model analysis and interpretation.
The dataset was partitioned into three non-overlapping subsets: a training set, a validation set, and a test set. The training set was used for model fitting, the validation set for estimating generalization error, and the test set for evaluating predictive performance. For both datasets, the training, validation, and test sets were allocated in a 6:2:2 ratio to support model training and assessment.
To reduce the impact of partial dimensions between data and the absolute magnitude of traffic flow eigenvalues on prediction results, and improve the model's training effect and generalization ability, min-max normalization is used to scale the original data into dimensionless data within the range of [0,1].

4.4 Feature Selection

To comprehensively capture the spatial correlations inherent in traffic flow, we analysed the interrelationships among detector data to determine how the location of each detector correlates with others. For each detector, the five detectors with the highest correlation were ultimately selected as representative influencing features. These selected features were then used as input variables for the primary traffic flow prediction model, thereby enhancing its predictive accuracy and stability.

For instance, in the case of the detector AID01111, the five most correlated detectors in terms of average speed were identified as AID4, AID36, AID42, AID37, and AID19. Regarding traffic volume, the top five most relevant detectors were AID2, AID4, AID36, AID42, and AID37. Fig. 3(a) shows the top five correlation coefficients of traffic speed time series, and Fig. 3(b) shows the top five correlation coefficients of traffic volume time series.



(a) (b)

Figure 3 Correlograms of Network-Wide Traffic Flows: (a) Speed and (b) Volume

To comprehensively account for external factors affecting traffic prediction, two binary variables—holiday and major event indicators—were incorporated into the processed traffic time series datasets. The corresponding holiday and major event data for the one-month and 50-day periods were merged into the traffic datasets, resulting in two final datasets that include influencing factor information: Dataset 1 (one month) and Dataset 2 (50 days), which were used in the empirical analysis.

5 EXPERIMENTAL SETTINGS

5.1 Experimental Setup

The deep learning framework used for model construction in this study is TensorFlow, with Python 3.11 as the programming language. The working environment is Windows 11.

In the experiments, all models are trained with a learning rate of 0.001, a batch size of 512, and 100 epochs. The Adam optimization algorithm is employed for model training, and L1 regularization is applied to prevent overfitting.

5.2 Model Performance Evaluation Metrics

Four evaluation metrics are used in this study to assess the accuracy of the prediction results: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and the Coefficient of Determination (R^2)The corresponding formulas are defined as Eq. (2)- Eq. (5):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right| \tag{2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{3}$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{4}$$

$$R^2 = \left(1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}\right) \times 100\% \tag{5}$$

Where: $\hat{y}_i$ denotes the predicted value by the model, y_i represents the actual value, $\overline{y}$ is the mean of the actual values, n is the number of predictions. A smaller error indicates better prediction performance of the model; a higher R^2 value suggests a better model fit.

6 RESULTS

6.1 Residual Analysis

The trained neural network was used to predict the traffic speed data on the test sets, producing the nonlinear component of the forecasts, denoted as $\hat{N}$. By subtracting the nonlinear predictions from the original time series, the residual sequences for the two datasets were obtained. These residuals were further analysed using the ARIMA model. Initial tests were conducted to assess whether the residual sequences constituted white noise, followed by an evaluation of autocorrelation. The Ljung–Box test statistics were 80.32 for Dataset 1 and 296.77 for Dataset 2, with p-values exceeding 0.05. Thus, at the 5% significance level, the null hypothesis of independence could not be rejected, indicating that the residuals do not represent white noise. This suggests that the residuals still contain linear patterns, which can be further modelled using the ARIMA approach. The specific ARIMA configurations were determined using the auto.arima function, and the resulting models produced the linear component of the final prediction. The overall forecast from the Bi-GRU-ARIMA model with attention is derived by summing the nonlinear and linear components. This hybrid model successfully captures both the nonlinear and linear dynamics inherent in the original traffic time series, thereby enhancing forecasting accuracy.

6.2 Evaluation

To provide a more intuitive comparison and validation of the prediction performance of the proposed attention-based Bi-GRU-ARIMA hybrid model on traffic indicators, Tables 2 and 3 present the evaluation results for six models—including the hybrid model, ARIMA, GRU, Bi-GRU, CNN-Bi-GRU, and Bi-GRU-ARIMA—on both Dataset 1 and Dataset 2, focusing on average speed and traffic volume. It is worth noting that the (p,d,q) parameters set in ARIMA differ between Dataset 1 and Dataset 2.

Table 1 Model Comparison and Evaluation Results on Dataset 1 (Speed)

Model	MAE	RMSE	MAPE(%)	R^2
ARIMA(1,0,1)	2.36	3.17	-	78.49
GRU	2.41	3.46	3.69	74.67
Bi-GRU	2.38	3.35	3.63	76.2
CNN-Bi-GRU	2.38	3.32	3.64	76.59
Bi-GRU-ARIMA	1.56	2.14	2.39	90.24
Bi-GRU-Attention-ARIMA	1.20	1.66	1.83	94.05

Table 2 Model Comparison and Evaluation Results on Dataset 1 (Volume)

Model	MAE	RMSE	MAPE(%)	R^2
ARIMA(1,0,2)	29.69	39.52	-	80.78
GRU	28.13	37.40	19.85	83.05
Bi-GRU	27.99	37.09	18.50	83.32
CNN-Bi-GRU	27.13	36.91	19.14	83.49
Bi-GRU-ARIMA	23.34	31.15	17.03	88.12
Bi-GRU-Attention-ARIMA	21.42	29.36	15.28	89.37

A comparison of the four evaluation-metrics presented in Tables 1 and 2 reveals the following insights:

- Compared with the GRU, Bi-GRU, and ARIMA, as a traditional time-series statistical method, demonstrates better flexibility and adaptability when dealing with data exhibiting clear periodicity and seasonality—such as the average traffic speed shown in Table 1.
- Model Bi-GRU, which incorporates an additional backpropagation layer compared to Model GRU, shows a certain improvement in prediction performance: MAE、RMSE and MAPE all show a degree of reduction. Specifically, the average speed metric improved by approximately 1.53%, while the improvement in the traffic volume metric R^2 was relatively modest, at only 0.27%.
- After the introduction of CNN, the stacked neural network model CNN-Bi-GRU, which extracts both spatial correlation and temporal dependency from the time series, achieved slight improvements in predicting both average speed and traffic volume. However, the performance gains are limited, likely due to the homogeneity in the extraction of nonlinear features, which restricts the overall enhancement of the model's predictive capability.
- After the Bi-GRU model extracted the nonlinear features of the time series, the residuals were further modeled using the ARIMA model to capture the remaining linear components. This two-stage approach significantly improved the model's predictive performance and proved more effective than simply incorporating a convolutional neural network. Compared with the single Bi-GRU model: the average speed metric saw a reduction of 0.82 in MAE, 1.21 in RMSE,

and 1.24% in MAPE, while R^2 increased by approximately 14.04%. For traffic volume, MAE decreased by 4.65, RMSE by 5.94, MAPE by 1.47%, and R^2 increased by about 4.8%.

• On top of the extraction of nonlinear and linear information, the Bi-GRU-Atten-ARIMA model introduces an attention mechanism, enabling the model to focus more effectively on task-relevant features. The inclusion of this mechanism further enhanced prediction accuracy, with the average speed metric increasing by approximately 3.81%, while the improvement for traffic volume was relatively smaller at 1.25%. Overall, the Bi-GRU-Atten-ARIMA model outperformed both standalone models and stacked neural network models across different indicators.

Table 3 Model Comparison and Evaluation Results on Dataset 2 (Speed)

Model	MAE	RMSE	MAPE(%)	R^2
ARIMA(1,0,1)	2.40	3.22	-	77.85
GRU	2.07	2.74	3.02	77.12
Bi-GRU	1.96	2.54	2.83	80.22
CNN-Bi-GRU	1.98	2.57	2.85	79.75
Bi-GRU-ARIMA	1.60	2.08	2.30	86.82
Bi-GRU-Attention-ARIMA	1.48	1.87	2.19	90.03

Table 4 Model comparison and evaluation results on dataset 2 (volume)

Model	MAE	RMSE	MAPE(%)	R^2
ARIMA(2,1,3)	22.65	30.98	-	78.67
GRU	24.90	32.77	20.64	76.45
Bi-GRU	23.90	31.31	20.10	78.49
CNN-Bi-GRU	22.11	30.51	16.52	79.25
Bi-GRU-ARIMA	19.19	25.41	15.97	85.86
Bi-GRU-Attention-ARIMA	18.10	23.97	14.78	87.42

Based on the values of the four evaluations metrics presented in Tables 3 and 4, the following observations can be made:

The traditional statistical model ARIMA achieved a R^2 of only 77.85% when evaluating the average speed indicator, a result not significantly different from the fitting performance of model GRU. Its performance on traffic volume was similarly limited. However, model Bi-GRU showed a clear improvement over GRU, indicating that Bi-GRU is more effective at capturing dependencies in long-sequence time series data, thereby enhancing model accuracy. Meanwhile, the stacked deep learning model (CNN-Bi-GRU) failed to deliver significant improvements in predictive accuracy. In contrast, combining traditional statistical models with deep learning substantially improved prediction performance. Compared to the single Bi-GRU model, Bi-GRU-ARIMA achieved an improvement of 6.60% in speed prediction. For traffic volume, MAE decreased by 4.71, RMSE by 5.90, MAPE by 4.13%, and R^2 increased by approximately 7.37%. On this basis, the addition of an attention mechanism further enhanced model performance, with Bi-GRU-Attention-ARIMA improving R^2 by 3.21% and 1.56%, respectively. Ultimately, the hybrid model achieved an R^2 of over 85% for both speed and traffic volume, with more than a 5% improvement compared to both standalone models and the stacked neural network model.

7 DISCUSSION

7.1 Dataset1 vs. Dataset 2

A comparison of the evaluation results between the two datasets reveals the following insights:

• For the average speed indicator, the traditional statistical model ARIMA, being relatively simple, performs well on smaller datasets. It tends to be more effective when applied to compact datasets. In contrast, for larger datasets, deep learning models are better equipped to capture long-term dependencies within complex time series structures, thereby offering superior predictive performance.

• In Dataset 1, the predictive performance of the neural network's nonlinear component is weaker than that observed in Dataset 2, which benefits from a larger data volume. However, after incorporating ARIMA to extract linear features, the overall performance of the hybrid model on Dataset 1 surpasses that of Dataset 2, despite the stronger neural network performance in the latter. This may be attributed to the neural network in Dataset 2 already capturing most of the time series information during training, leaving limited room for further enhancement through ARIMA. In contrast, when the neural network's performance is less optimal—as in Dataset 1—the ARIMA model can make more effective use of the remaining information in the residuals. Moreover, due to the pronounced periodicity in real-time traffic data, the ARIMA model's contribution tends to be more significant when more extractable information is available, leading to better overall results.

• Based on the evaluation results for both average speed and traffic volume, it is evident that as the model's ability to extract time series features improves, so does its predictive performance. The proposed hybrid model consistently delivers the best results across different indicators. However, the degree of improvement varies by indicator. The model performs particularly well in predicting average speed, suggesting that its neural network structure is well-suited to capturing the complex patterns and dependencies in speed data. In contrast, the improvement in predicting traffic volume is relatively modest, indicating that the model's feature-matching capabilities for traffic volume data may be less effective.

7.2 Forecast

The prediction analysis focuses on two key macroscopic traffic variables: speed and traffic volume. In addition to the numerical results presented in the four tables above, the prediction outcomes are also visualized. Using one-month and 50-day traffic time series data collected by AID01111, prediction graphs were generated for speed series and volume series . Fig. 4- Fig. 7 shows the prediction graphs of actual vs. predicted values by proposed model.

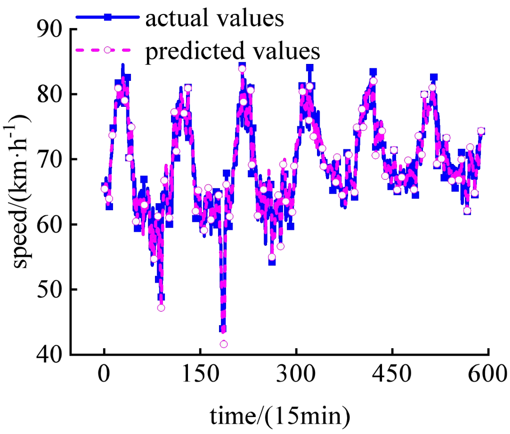


Figure 4 Actual vs. Predicted Speed by Proposed Model for Dataset 1

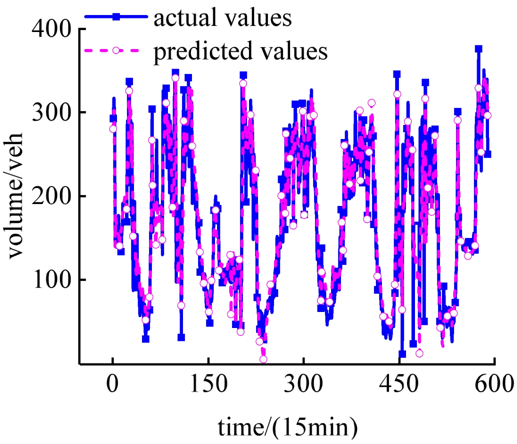


Figure 5 Actual vs. Predicted Volume by Proposed Model for Dataset 1

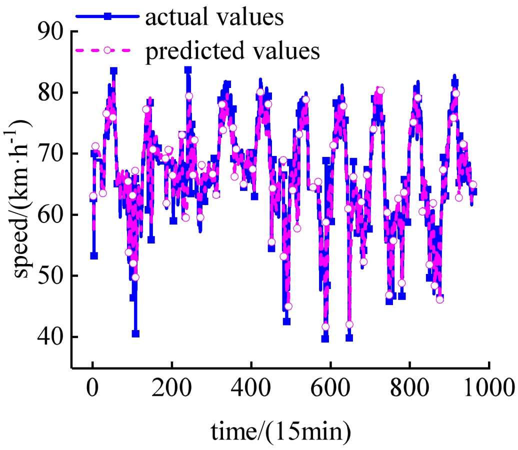


Figure 6 Actual vs. Predicted Speed by Proposed Model for Dataset 2

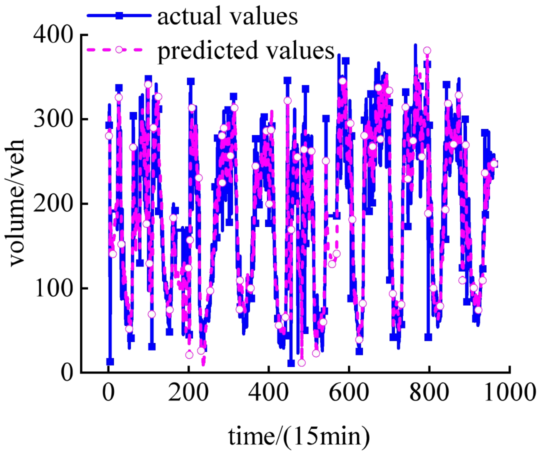


Figure 7 Actual vs. Predicted Volume by Proposed Model for Dataset 2

7.3 Warning Window

Building upon the predictions generated by the hybrid model, early warnings for traffic congestion or anomalies can also be issued. In the subsequent application of time series forecasting, and based on the speed limits and occupancy distribution characteristics of expressways in Hong Kong region, we define abnormal or congested conditions as those where the average speed falls below 40 km/h and occupancy exceeds the upper quartile of the original data series. The prediction performance of the proposed model on the occupancy rate indicator is shown in Table 5. When such conditions are identified, an alert is triggered. This early warning mechanism enables traffic authorities to proactively monitor and manage congestion and related risks. It supports the implementation of intelligent traffic systems and contributes to more efficient urban traffic management. Ultimately, this approach can improve road network efficiency, reduce economic losses caused by delays, minimize fuel consumption and environmental impact, and enhance the overall travel experience for urban residents.

Table 5 Evaluation Results of Bi-GRU-Attention-ARIMA Model for Occupancy Rate Indicator

Dataset	MAE	RMSE	MAPE(%)	R^2
1	0.62	0.87	12.16	95.7
2	0.57	0.42	12.07	96.98

8 CONCLUSION

This study began by preprocessing the raw traffic indicator data, converting it into short-term time series with 15-minute intervals. Time series plots of the processed data revealed notable characteristics such as periodicity, nonlinearity, and volatility. In response to these temporal patterns, the processed sequences were fed into a neural network model for prediction. An attention mechanism was incorporated to reassign weights to the most relevant information, thereby generating a prediction sequence that captures the nonlinear component. Subsequently, empirical analysis using autocorrelation function (ACF) plots was conducted to examine the residuals from the neural network predictions. The presence of linear autocorrelation in the residuals suggested that additional linear information could still be extracted. A separate model was therefore employed to predict the residual sequence, and the final prediction of the hybrid model was obtained by summing the predicted residuals with the neural network's nonlinear output. Based on a comparative evaluation with other models, the following key conclusions were drawn:

- The proposed hybrid model exhibits superior predictive performance in short-term traffic flow forecasting compared to existing approaches. It performs consistently well across both long-term and short-term datasets and across different traffic indicators, including average speed and traffic volume. For instance, in predicting average speed using Dataset 1, the hybrid model achieved improvements of 15.56%, 19.38%, 17.85%, 17.46%, and 3.81% in R^2 over Models ARIMA, GRU, Bi-GRU, CNN-Bi-GRU, and Bi-GRU-ARIMA, respectively. Unlike traditional statistical models or standalone deep learning models, the hybrid model offers more comprehensive feature extraction by jointly considering linear and nonlinear components. Furthermore, the integration of an attention mechanism enables the model to effectively distinguish the varying importance of input features, thereby further enhancing its predictive accuracy.
- The accuracy of traffic flow prediction is affected by the characteristics of the dataset used. A comparison of the neural network prediction results between Dataset 1 and Dataset 2 shows that deep learning models perform more effectively on larger and more complex datasets. Such models are better equipped to capture long-term dependencies within time series data, resulting in more effective training and greater improvements over baseline models. However, analysis of the overall predictive framework reveals that as deep learning extracts a larger portion of information, the amount of remaining information in the residuals decreases. Consequently, the model yields better overall predictive performance for the one-month dataset than for the 50-day dataset. Furthermore, under the same model architecture, performance improvements vary depending on the complexity and underlying patterns of different indicators in the datasets. The better the model structure aligns with the specific characteristics of a traffic indicator, the more significant the performance enhancement.

CONFLICTS OF INTEREST

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Ahmed M, Cook A. Analysis of Freeway Traffic Time-Series Data by Using Box-Jenkins Techniques. Transportation Research Record. 1979. https://www.semanticscholar.org/paper/ANALYSIS-OF-FREEWAY-TRAFFIC-TIME-SERIES-DATA-BY-Ahmed-Cook/c6fc010c45d2bd96b82b5696c997d3050d997095.

[2] Williams B. Multivariate Vehicular Traffic Flow Prediction: Evaluation of ARIMAX Modeling. Transportation Research Record, 2001, 1776(1): 194-200.

[3] Kamarianakis Y, Prastacos P. Space–Time Modeling of Traffic Flow. Computers & Geosciences, 2005, 31(2): 119-133.

[4] Huang X, Li X, Wang W. Temporal Data-Driven Short-Term Traffic Prediction: Application and Analysis of LSTM Model. Theoretical and Natural Science, 2023, 14, 205-211.

[5] Li B, Xiong J, Wan F, et al. Incorporating Multivariate Auxiliary Information for Traffic Prediction on Highways. Sensors, 2023, 23(7): 3631.

[6] Sherafat E, Farooq B, Karbasi A, et al. Attention-LSTM for Multivariate Traffic State Prediction on Rural Roads. arXiv Preprint. 2023, arXiv: 2301.02731. DOI: https://doi.org/10.48550/arXiv.2301.02731.

[7] Bharti, Redhu P, Kumar K. Short-Term Traffic Flow Prediction Based on Optimized Deep Learning Neural Network: PSO-Bi-LSTM. Physica A: Statistical Mechanics and its Applications, 2023, 625, 129001. DOI: https://doi.org/10.1016/j.physa.2023.129001.

[8] Naheliya B, Redhu P, Kumar K. MFOA-Bi-LSTM: An Optimized Bidirectional Long Short-Term Memory Model for Short-Term Traffic Flow Prediction. Physica A: Statistical Mechanics and its Applications, 2024, 634, 129448. DOI: https://doi.org/10.1016/j.physa.2023.129448.

[9] Wang J, Susanto C. Traffic Flow Prediction with Heterogenous Data Using a Hybrid CNN-LSTM Model. Computers, Materials & Continua, 2023, 76(3): 3097-3112. DOI: https://doi.org/10.32604/cmc.2023.040914.

[10] Ma F, Deng S, Mei S. A Short-Term Highway Traffic Flow Forecasting Model Based on CNN-LSTM with an Attention Mechanism. Proc. of Journal of Physics: Conference Series, IOP Publishing, 2023, 2491(1): 012008. DOI 10.1088/1742-6596/2491/1/012008.

[11] Singh N, Kumar K, Pokhriyal B. Attention Based Spatiotemporal Model for Short-Term Traffic Flow Prediction. International Journal of System Assurance Engineering and Management, 2025, 16(4): 1517-1531.

[12] Ding H, Li Z, Su N. Traffic Prediction Based on the GRU Neural Network. Applied and Computational Engineering, 2023, 8: 287-291.

[13] Li R, Hao Z, Yang X, et al. Urban Road Travel Time Prediction Based on Gated Recurrent Unit Using Internet Data. IET Intelligent Transport Systems, 2023, 17(12): 2396-2409.

[14] Shi G, Luo L. Prediction and Impact Analysis of Passenger Flow in Urban Rail Transit in the Postpandemic Era. Journal of Advanced Transportation, 2023, 1, 448864.

[15] Chauhan N, Kumar N, Eskandarian A. A Novel Confined Attention Mechanism Driven Bi-GRU Model for Traffic Flow Prediction. IEEE Transactions on Intelligent Transportation Systems, 2024, 25(8): 9181-9191.

[16] Sun X, Chen F, Wang Y, et al. Short-Term Traffic Flow Prediction Model Based on a Shared Weight Gate Recurrent Unit Neural Network. Physica A: Statistical Mechanics and its Applications, 2023, 618, 128650. DOI: https://doi.org/10.1016/j.physa.2023.128650.