IJTE, Vol. 13, No. 2, Autumn 2025, pp. 2173-2192
Research Paper

Enhancing Short-Term Traffic Flow Forecasting by Hybrid Deep Learning Architectures and Attention Mechanisms (Case Study: High-Density Karaj-Chalous Road, Iran)

Seyed Saber Naseralavi[1,*], Soheil Rezashoar[2], Akram Mazaheri[3], Saman Shafaati[1]

Received: 2025/01/15 *Accepted: 2025/09/07*

Abstract

The main tool to mitigate congestion and improve travel experiences effectively in intelligent traffic management is to predict the accurate and timely short-term traffic flow on high-volume roads. We present the performances of different deep learning models, such as LSTM, GRU, CNN, their hybrids CNN-LSTM and CNN-GRU, and versions with an attention mechanism for one-hour-ahead traffic flow prediction on mountainous and high-density Karaj-Chalous Road. The input data include the traffic data from two traffic counters. The cited data were derived for a period ranging from 01/01/1401 to 01/01/1403. Besides, the synoptic meteorological data were acquired within three-hour intervals, while the models are compared based on various quantitative accuracy and error metrics. The results showed that the CNN-LSTM model was the best among the rest, with an R^2 value of 0.83, because it captured complex traffic patterns and temporal dependencies effectively. The other models ranked next were LSTM, GRU, CNN-LSTM-GRU, and CNN-GRU, with R2 values of 0.82, 0.81, 0.80, and 0.80, respectively. While the weakest models, CNN and CNN-MultiHead-Attention, yielded an R^2 of 0.60 and 0.62, respectively, this is due to a lack of consideration in these models regarding the nature of traffic data as a time series. Employing attention mechanisms improved prediction accuracy in some model architectures. This effect was highly varied based on the model structure itself. The results depict that deep, hybrid models with the integration of attention mechanisms can give more reliable and valuable forecasts to the intelligent transportation management systems for better travel planning and congestion reduction in similar roadways.

Keywords: Short-term, traffic flow prediction, deep learning, attention mechanism, CNN-LSTM, Karaj-Chalous Road

* Corresponding author. E-mail: saber_naseralavi@uk.ac.ir
[1] Department of Civil Engineering, Shahid Bahonar University of Kerman, Kerman, Iran
[2] Department of Transportation Planning, Faculty of Engineering, Imam Khomeini International University, Qazvin, Iran
[3] Faculty of Civil and Environmental Engineering, Tarbiat Modares University, Tehran, Iran

1. Introduction

Traffic on high-density corridors is increasing, and planning and infrastructure management require short-term traffic flow prediction to be a top priority in transportation (Lv et al., 2015; Polson & Sokolov, 2016; E. I. Vlahogianni et al., 2014). Accurate and reliable short-term forecasts can help in decision-making, reduce congestion, improve safety, optimize traffic management systems, and overall provide a better travel experience (Wu et al., 2018; W. Zhang et al., 2019).

However, traffic prediction is a complex problem. The intrinsic characteristics of traffic systems—nonlinearity, spatio-temporal dependencies, and external factors (weather, incidents, etc.)—have exposed the limitations of traditional statistical methods (Ahmed & Cook, 1979; Williams & Hoel, 2003) and early machine learning approaches, especially when dealing with real-time dynamic traffic data (Jia et al., 2017; Ma et al., 2015; Zhao et al., 2017). These limitations highlight the gaps in the literature, and we need more advanced techniques to achieve higher accuracy and stability.

With the emergence of deep learning in recent years and its ability to extract features from large-scale heterogeneous data, traffic flow prediction has opened up new possibilities (Lv et al., 2015; Polson & Sokolov, 2016). Specifically, models based on LSTM and GRU, along with CNN, have enabled the understanding of long-term temporal patterns and spatial structures (Miglani & Kumar, 2019; Wu et al., 2018; W. Zhang et al., 2019). Attention mechanisms, by highlighting important and informative patterns, have improved the accuracy and efficiency of hybrid models like CNN-LSTM-Attention (Do et al., 2019; Liu et al., 2024; Xia et al., 2023).

But these advanced methods need to be tested in real-world complex scenarios. Karaj-Chalous Road, one of the busiest and toughest mountainous roads in Iran, is a good testbed for these models. The region's climate variations, seasonal travel patterns, and high traffic fluctuations create a realistic and challenging environment to evaluate the strengths and weaknesses of deep learning models. Traffic forecasting for this road has both scientific and practical value—developing intelligent traffic management systems, reducing delays, and improving road safety.

In this paper, we aim to fill the research gaps and introduce new aspects by evaluating several deep learning models, including LSTM, GRU, CNN, as well as CNN-LSTM and CNN-GRU hybrids, and their attention-enhanced variants for short-term traffic flow prediction on Karaj-Chalous Road. Using both traffic and meteorological data, we compare each model's performance using quantitative and qualitative metrics and identify the strengths and weaknesses of each method.

Following this introduction, Section 2 reviews the related work and gaps. Section 3 describes the study area and data. Section 4 explains the models and results. Section 5 discusses the findings and their implications. Section 6 concludes and suggests future work.

2. Literature Review

In the past couple of decades, short-term traffic flow prediction has emerged as one of the most important enablers in ITS for intelligent traffic management, efficient travel planning, reduction of waiting times, mitigation of congestion, and improvement of road safety (Lv et al., 2015; Ma et al., 2015; E. I. Vlahogianni et al., 2014). Advances in sensor technologies, IoT, big data analytics, and high-performance computing have made a vast amount of traffic data available to researchers. Therefore, traditional statistical methods, which are usually based on linear and stationary assumptions, have been replaced by novel machine learning and deep learning techniques (Tedjopurnomo et al., 2020; Wu et al., 2018; W. Zhang et al., 2019). However, despite the tremendous achievements, there are still considerable challenges, gaps, and shortcomings in the literature. The current research aims to fill these gaps by targeting high-density, mountainous roads like the Karaj-Chalous route and using state-of-the-art deep hybrid models, including CNN-LSTM architectures and attention mechanisms.

2.1. Historical Background and Classical Approachess

Early efforts in short-term traffic flow prediction were largely based on statistical models such as Box-Jenkins (ARIMA, SARIMA) models (Ahmed & Cook, 1979; Williams & Hoel, 2003). These models had the advantage of being interpretable and easy to implement and were able to model simple seasonal and weekly patterns. However, their inherent linearity was limiting in capturing real-world traffic data, which is complex, nonlinear, and volatile in nature (W. Zhang et al., 2019). They were gradually joined by more sophisticated statistical methods that included variants of Kalman filter-based models (J. Guo et al., 2014; Okutani & Stephenades, 1984), state-space representations (Stathopoulos & Karlaftis, 2003), and Bayesian networks (Sun et al., 2006). Still, these also tended to fail when noisy, non-stationary data and atypical conditions like accidents, special events, or extreme weather were involved.

2.2. Transition to Classical Machine Learning Methods

It was now possible to utilize methods such as Artificial Neural Networks (ANN), k-Nearest Neighbors (KNN), Support Vector Regression (SVR), and neuro-fuzzy models that took traffic flow prediction to another level and deeply began to shape the idea of machine learning (Chen & Grant-Muller, 2001; Dougherty & Cobbett, 1997; Jeong et al., 2013; Smith & Demetsky, 1997). These techniques outperformed classical linear models, which were limited to using simple relationships, as they were able to employ highly complex nonlinear boundaries (Park et al., 1998; Y. Zhang & Ye, 2008). However, they struggled with issues of effective feature extraction and feature configuration, as well as in making use of large-scale spatio-temporal properties of a large road network (Karlaftis & Vlahogianni, 2011; E. Vlahogianni et al., 2004).

2.3. The Emergence of Deep Learning: LSTM, GRU, CNN and GNN Architectures

Deep learning is a more recent model that is able to work exceptionally well with unstructured data, particularly with complex spatio-temporal data (Lv et al., 2015; Polson & Sokolov, 2016). Sequence memory architectures, such as LSTM and GRU units, have been successful in modeling long-range time dependencies (Ma et al., 2015; Shu et al., 2021; Zhao et al., 2017). Similarly, Convolutional Neural Networks, or CNNs, have performed well in modeling spatial features, especially on road structures that are network and graph-like (Wu et al., 2018; W. Zhang et al., 2019).

Hybrid architectures, such as CNN-LSTM or CNN-GRU, combine the strengths of convolution for the extraction of spatial features and recurrent memory units for temporal dependencies, which have brought significant improvements in predictive accuracy (Naheliya et al., 2023; Wu et al., 2018). In addition, attention mechanisms have been developed to place more emphasis on the most informative parts of the data (Do et al., 2019; Wang et al., 2020; Xia et al., 2023). Graph Convolutional Networks consider the topological structure of road networks and offer an enhanced way of modeling spatial correlations (Duan et al., 2019; Xia et al., 2023).

In recent years, GNNs have gained increasing traction in traffic forecasting due to their remarkable ability to handle non-Euclidean data structures and model complex spatial relationships inherent in transportation networks (Cirstea et al., 2022; Liu et al., 2023). These approaches have become foundational to modern traffic prediction systems, with STGCN (Yu et al., 2018) setting a significant precedent by effectively integrating graph convolution for spatio-temporal modeling. Subsequent advancements have furthered this paradigm: DCRNN (Y. Li et al., 2018) improved temporal modeling through recurrent networks with diffusion convolution, while GMAN (Zheng et al., 2020) achieved more refined feature representation using multi-head attention mechanisms.

The field has progressively moved toward more sophisticated modeling of dynamic spatio-temporal characteristics (Yuan & Li, 2021). Notable advancements include ASTGCN's (Luo et al., 2023) attention-enhanced framework for capturing fine-grained patterns, STSGCN's (Song et al., 2020) localized spatio-temporal graph learning approach, and STFGNN's (M. Li & Zhu, 2021) parallel feature fusion architecture. DSTAGNN (S. Guo et al., 2019) marked a conceptual leap by shifting from static to dynamic graph representations, better accommodating real-world traffic variability. Further research (Bui et al., 2022; Kong et al., 2024; Zhang et al., 2022) has advanced dynamic graph generation techniques to more effectively align with the continuously evolving nature of urban traffic systems.

However, current methodologies still face significant limitations. Many approaches rely on predetermined or fixed graph configurations, which restrict their ability to: (1) accurately capture the dynamic, multivariate nature of real-world traffic conditions; (2) effectively model complex spatio-temporal interrelationships; and (3) adapt to unexpected disruptions. For instance, while STSGCN (Song et al., 2020) excels at capturing spatial evolution, its temporal modeling remains underdeveloped. Similarly, AGCRN's (Bai et al., 2020) adaptive graph generation, combined with recurrent networks, enhances feature representation but falls short in incorporating broader contextual information.

2.4. External Factors and Multisource Data

A popular development in recent times includes the inclusion of exogenous variables like weather, events, construction, and vehicle type distributions into models that predict the flow of traffic (Jia et al., 2017; Liu et al., 2024; Xie et al., 2007). Some research has shown that the inclusion of meteorological data can drastically improve prediction accuracy for roads whose characteristics or conditions are highly dependent on weather, such as rain or snowfall (Jia et al., 2017). For example, LSTM and DBN models combined with weather information show superior performance compared to traffic-only-based models when applied to different weather conditions.

2.5. Critique, Gaps in the Literature, Significance, and Necessity of Present Study

Most of the research has targeted urban or relatively simple freeway conditions, with less attention to complex, mountainous, and high-volume roads where the pattern of traffic flow is highly volatile and strongly affected by weather conditions (Lippi et al., 2013; Miglani & Kumar, 2019; E. I. Vlahogianni et al., 2014). For example, the Karaj-Chalous Road, which experiences heavy traffic, seasonal and hourly fluctuations, and diverse climatic conditions, has been largely overlooked in deep and hybrid modeling. While some CNN-LSTM and CNN-GRU models with attention mechanisms have shown promise in complex environments (Wang et al., 2020; Wu et al., 2018), examination of the systematic effects of combined weather conditions with intricate spatiotemporal structures is limited (Jia et al., 2017). Existing literature often examines the impact of attention mechanisms in CNN-LSTM and CNN-GRU models only on simplified networks, without incorporating multisource data (e.g., historical traffic, real-time weather sensor data, and forecasted meteorological conditions). Thus, there is a need for models capable of integrating spatiotemporal structures and meteorological data in a hybrid optimization framework (e.g., CNN-LSTM plus Attention). From the above points, it is clear that there is an urgent need for a short-term traffic flow prediction model suitable for the Karaj-Chalous Road—a high-volume, mountainous, and weather-sensitive route. The contribution of this study, therefore, is the development.

3. Research Methodology

The methodology in this research involves a systematic, transparent, and reproducible application of state-of-the-art deep learning methods to the problem of short-term traffic flow prediction. Its primary objective is to leverage multisource data and evaluate a wide range of deep learning models, from classical architectures to hybrid structures with attention mechanisms, under the severe conditions characterizing the Karaj-Chalous Road. This route has unique geographic characteristics, heavy traffic flow, seasonal travel patterns, and varying climatic conditions, which constitute a realistic and complex setting for these models. We begin by reviewing the literature, identifying existing gaps in research and scientific needs.

Then, we consider the Karaj-Chalous Road as a case study and collect traffic and meteorological data. After cleaning, preparing, and integrating the datasets, the data is structured so that deep learning models can forecast one-hour-ahead traffic flow based on a five-hour historical window. Following the definition and implementation of different architectures, such as LSTM, GRU, CNN, and CNN-LSTM and CNN-GRU hybrids, possibly with attention mechanisms, the data is divided into training, validation, and test sets. Training includes hyperparameter tuning and model evaluation using statistical metrics: MSE, RMSE, MAE, and R^2. Finally, the results are analyzed, the best models are selected, and recommendations for further research are discussed. A flowchart of the main steps of the methodology, from left to right, is divided into four major phases.

The research method (Figure 1) is initiated in Phase 1, followed by a literature review to understand prior models, traditional methods, and deep learning approaches to identify research gaps. In Phase 2, the Karaj-Chalous Road is selected as the case study. Relevant traffic and weather data are gathered and prepared by cleaning, removing outliers, and integrating the traffic and meteorological information. Suitable time windows are defined for one-hour-ahead traffic flow prediction. In Phase 3, the deep learning model architectures are defined, including LSTM and GRU with their hybrid variants: CNN-LSTM and CNN-GRU. Data is divided into training, validation, and test sets. Training of these models, including hyperparameter tuning and performance evaluation using metrics like MSE, RMSE, MAE, and R^2, is performed. In Phase 4, the results of the models are analyzed: error metrics are considered, observed versus predicted plots and scatter plots are used to compare model performances, and the strengths and weaknesses of each model are identified. Lastly, conclusions are drawn, along with future directions to enhance the models and explore more advanced approaches.

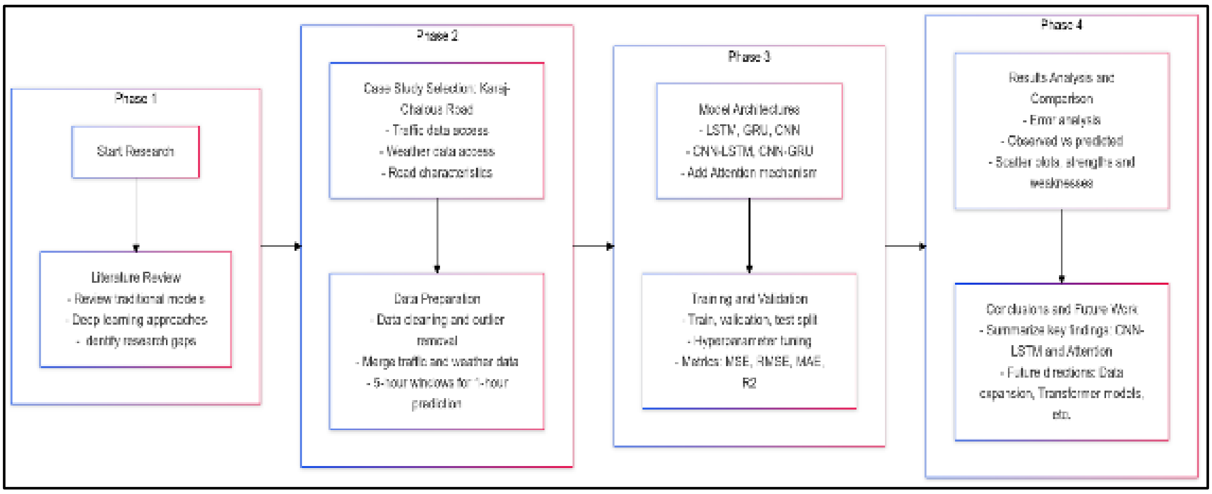

Figure 1. The Four Main Phases of the Research and Their Components

3.1. Traffic Data

Chalous Road (Road 59) is one of the most important arteries connecting central parts of Iran, including large cities such as Karaj and Tehran, with northern parts of the country along the Caspian Sea shoreline. It is renowned for its unspoiled natural beauty and picturesque views, making this mountainous route crowded on weekends and holidays while experiencing heavy traffic congestion due to tourist traffic. Traffic data were obtained from the Road Maintenance and Transportation Organization of Iran (Iran Road Maintenance and Transportation Organization, 2024), specifically from two traffic counters along the Karaj-Chalous corridor: one in the Karaj-to-Chalous direction (KC, code 183120) and one in the Chalous-to-Karaj direction (CK, code 183170). Since the road is two-way, traffic data from both directions were incorporated into the input dataset. The traffic data span from 01/01/1401 to 01/01/1403.

3.2. Meteorological Data

It is supplemented with meteorological data obtained from the Iran Meteorological Organization (Iran Meteorological Organization (IRIMO), 2024), recorded at the Siahbisheh synoptic station (station no. 40735), at an elevation of 1855.4 meters and approximately at 51.303°E, 36.231°N. Data collection occurred at three-hour intervals from 21/03/2022 to 20/03/2024 (01/01/1401 to 01/01/1403) and will serve as a valuable basis for understanding the weather conditions affecting the flow of traffic.

Figure 2 illustrates a satellite map highlighting the locations of traffic counters (stations 183120 and 183170) and the meteorological station (station no. 40735) along the Karaj-Chalous Road.

3.3. Variables Used in the Study

Table 1 lists the variables used in designing the algorithms and analyzing the data, totaling 33 parameters. Since traffic variables are considered for both the Karaj-to-Chalous and Chalous-to-Karaj directions, the initial 12 traffic-related variables per direction double to 24.

Figure 2. Satellite View of Chalous Road with Traffic and Meteorological Data Points

Table 1. Input Variables

	Input Feature	Description
Calendar variables	Date Time	-
	Hour	Hour of the day (1-24)
	Day of week	Day of the week (Saturday to Friday)
	Day of month	Day of the month (1-31)
	Holidays	official holidays and weekends based on the Persian calendar
	Operating time	Operating time of the counter (minutes)
Traffic Variables	Total number of vehicles	Total number of vehicles in the operating period
	Number of class 1 vehicles	Number of passenger cars and pickup trucks
	Number of class 2 vehicles	Number of small trucks, vans, and minibuses
	Number of class 3 vehicles	Number of regular trucks under 10m and three-axle vehicles
	Number of class 4 vehicles	Number of buses
	Number of class 5 vehicles	Number of trailers and heavy multi-axle vehicles
	Average speed	Average speed of traffic
	Speeding violation	Number of speed violations (vehicles exceeding the speed limit)
	Violation of following distance	Number of following distance violations (less than 2 seconds)
	Unauthorized overtaking violation	Number of unauthorized overtaking violations
	Estimated number	Estimated number of vehicles per hour
Meteorological	Horizontal visibility	Horizontal visibility (meters)
	Temperature	Temperature (°C)
	Relative humidity	Relative humidity (%)
	Cloud cover	Cloud cover (score 1-10)

3.4. Model Operation and Implementation

Based on data analysis, using the previous five hours of data to predict traffic volume for the next hour proved effective. As shown in Table

2, the model takes input from the preceding five hours to forecast the number of vehicles in the sixth hour. This process advances incrementally, predicting each subsequent hour based on the prior five hours. The implementation was carried out using Python in the Jupyter Notebook environment.

Table 2. Input-Output Structure for One-Hour-Ahead Traffic Flow Prediction

Input	Output
Data from hours 0 to 5	Predict number of vehicles in hour 6
Data from hours 1 to 6	Predict number of vehicles in hour 7
Data from hours 2 to 7	Predict number of vehicles in hour 8
...	...
Data from hours 23 to 4	Predict number of vehicles in hour 5

The preprocessing pipeline utilizes two key techniques to prepare data for modeling. First, the KNNImputer addresses missing values by identifying the three most similar data points (based on Euclidean distance) and imputing missing entries using their weighted average, thus preserving local data structures. Next, the StandardScaler standardizes all features by centering them on zero (subtracting the mean) and scaling to unit variance (dividing by the standard deviation), ensuring all variables are on a comparable scale and preventing features with larger magnitudes from dominating the model training process. Together, these preprocessing steps produce a more complete and normalized dataset, better suited for most machine learning algorithms.

The dataset was divided into training, validation, and test sets as follows: a training set with 12,034 records, a validation set with 1,000 records, and a test set with 3,000 records. This division allows the model to learn patterns from the training data, refine its parameters during validation, and evaluate its generalization capability on the test set. For model evaluation, the code employs Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2):

$$MSE = \frac{1}{T}\sum_{t=1}^{T}(\hat{y}_t - y_t)^2 \quad (1)$$

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{y}_t - y_t)^2} \quad (2)$$

$$R^2 = 1 - \frac{\sum_{t=1}^{T}(\hat{y}_t - y_t)^2}{\sum_{t=1}^{T}(y_t - \frac{1}{T}\sum_{t=1}^{T} y_t)^2} \quad (3)$$

In these equations, $\hat{y}_t$ represents the predicted value at time t, y_t the actual value at time t, and T is the total number of predictions. Lower MSE and RMSE values and an R^2 closer to one indicate better model performance.

3.5. Model Development

In this study, we apply a combination of advanced deep learning architectures for short-term traffic flow prediction. The ability of deep learning to capture complex features and temporal dependencies in data has attracted significant attention in traffic forecasting. Various models, integrating architectures such as LSTM, GRU, CNN, and attention mechanisms, have been proposed to improve predictive accuracy and performance.

The architecture of LSTMs and GRUs is renowned for capturing long-term temporal dependencies, making them well-suited for sequential data in traffic flow forecasting tasks (Ma et al., 2015; Polson & Sokolov, 2016; Zhao et al., 2017). In scenarios with highly variable conditions, LSTMs typically demonstrate greater robustness due to their more complex structure, while GRUs offer a simpler and faster training process, serving as an efficient, lightweight alternative (Lippi et al., 2013; Wu et al., 2018).

Incorporating an attention mechanism into the CNN-LSTM model (CNN-LSTM-Attention) accelerated error reduction and achieved lower validation errors. This model stands out for its higher efficiency and faster error convergence.

While the CNN-GRU model exhibited stable and consistent error reduction, its performance was less optimal compared to attention-equipped models. The gradual slope of its validation error reduction suggests room for further improvement.

The LSTM-MultiHead-Attention model, with multiple parallel attention mechanisms, achieved the most significant reduction in validation error, ranking among the top models in both speed and accuracy.

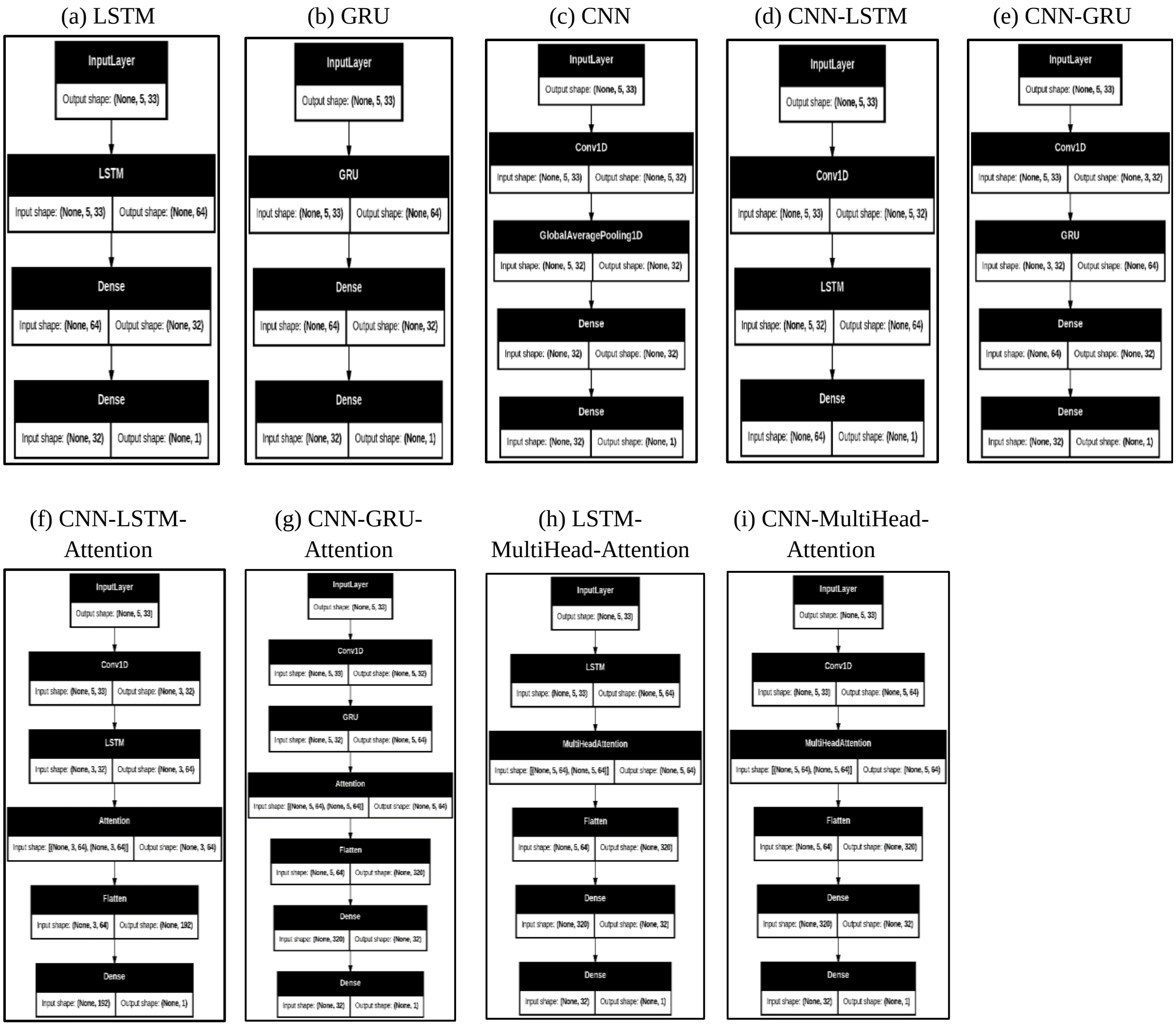

Figure 3. Architecture of the Developed Models

The integration of CNN and LSTM architectures, such as CNN-LSTM, leverages CNN's spatial feature extraction capabilities and LSTM's temporal modeling, resulting in improved accuracy and efficiency for predictions (Miglani & Kumar, 2019; Wu et al., 2018; W. Zhang et al., 2019). Incorporating an attention mechanism into these frameworks, such as CNN-LSTM-Attention, enables the model to focus on the most relevant patterns, further enhancing predictive performance (Do et al., 2019; Xia et al., 2023).

Furthermore, multi-attention-based models, such as multi-head attention in LSTM or CNN, can simultaneously analyze various spatiotemporal features, yielding significantly improved results for data with complex nonlinear conditions (Do et al., 2018; Wu et al., 2018). The GRU model combined with an attention mechanism, such as CNN-GRU-Attention, retains the simplicity and training speed of GRU while delivering performance comparable to LSTM, making it suitable for applications requiring rapid response and high accuracy (Ma et al., 2015; W. Zhang et al., 2019).

In general, hybrid architectures and advanced attention mechanisms provide a robust foundation for traffic flow analysis and prediction. The architecture, type, number of layers, their order, and the number of neurons for each developed model are illustrated in Figure 3. The selection of these architectures and settings was based on trial and error, repeated experiments, and evaluation of the error metrics for each architecture. Ultimately, the architecture yielding the lowest error was chosen as the superior model.

To ensure the transparency and reproducibility of this research, the detailed architectural configurations and hyperparameters for each of the nine developed models are provided in Table 3. The hyperparameter values, including the number of layers, the number of units per layer, and the kernel size for convolutional layers, were determined through a process of experimentation and tuning. The configuration that yielded the lowest error on the validation set was selected as the final architecture for each model.

For training all models, the Adam optimizer was utilized with an initial learning rate of 0.001 and a batch size of 32. An Early Stopping callback was employed to prevent overfitting, alongside a ReduceLROnPlateau scheduler to dynamically adjust the learning rate during training.

4. Results

Figure 4 illustrates the loss reduction during training for nine distinct models, all trained on data from the previous five hours. These curves track each model’s performance during the training and validation phases. Subsequent sections interpret individual results and provide a collective comparison.

The LSTM model exhibited stable performance, with a consistent decrease in both training and validation errors. An initial rapid decline in error stabilized over time, with minimal divergence between training and validation errors, indicating balanced learning without overfitting.

The GRU model displayed a similar pattern to the LSTM but achieved the minimum error more quickly (approximately within 25 iterations). This characteristic makes the GRU a faster and more efficient option for reaching optimal performance.

The CNN model showed a significant gap between training and validation errors, suggesting potential overfitting. Enhanced model tuning or techniques such as data augmentation could mitigate this issue (Zargari et al., 2025).

The hybrid CNN-LSTM model performed best, demonstrating a steady error reduction trend. The close alignment of training and validation errors reflects an optimal balance in the learning process.

Table 3. Input-Output Structure for One-Hour-Ahead Traffic Flow Prediction

Model	Layers & Activation Functions	Optimizer	Learning Rate	Batch Size
LSTM	Input (5,33), LSTM (64), Dense (32), Dense (1)	Adam	0.001	32
GRU	Input (5,33), GRU (64), Dense (32), Dense (1)	Adam	0.001	32
CNN	Input (5,33), Conv1D (32, k=1, ReLU), GlobalAveragePooling1D, Dense (32), Dense (1)	Adam	0.001	32
CNN-LSTM	Input (5,33), Conv1D (32, k=1, ReLU), LSTM (64), Dense (1)	Adam	0.001	32
CNN-GRU	Input (5,33), Conv1D (32, k=3, ReLU), GRU (64), Dense (32), Dense (1)	Adam	0.001	32
CNN-LSTM-Attention	Input (5,33), Conv1D (32, k=3, ReLU), LSTM (64, return_sequences=True), Attention, Flatten, Dense (1)	Adam	0.001	32
CNN-GRU-Attention	Input (5,33), Conv1D (32, k=1, ReLU), GRU (64, return_sequences=True), Attention, Flatten, Dense (32), Dense (1)	Adam	0.001	32
LSTM-MultiHead-Attention	Input (5,33), LSTM (64, return sequences=True), MultiHeadAttention (num_heads=6, key_dim=16), Flatten, Dense (32), Dense (1)	Adam	0.001	32
CNN-MultiHead-Attention	Input (5,33), Conv1D (64, k=1, ReLU), MultiHeadAttention (num_heads=2, key_dim=8), Flatten, Dense (32), Dense (1)	Adam	0.001	32

Incorporating an attention mechanism into the CNN-LSTM model (CNN-LSTM-Attention) accelerated error reduction and achieved lower validation errors. This model stands out for its higher efficiency and faster error convergence.

While the CNN-GRU model exhibited stable and consistent error reduction, its performance was less optimal compared to attention-equipped models. The gradual slope of its validation error reduction suggests room for further improvement.

The LSTM-MultiHead-Attention model, with multiple parallel attention mechanisms, achieved the most significant reduction in validation error, ranking among the top models in both speed and accuracy.

Despite utilizing multi-head attention, the CNN-MultiHead-Attention model experienced fluctuations in validation error, possibly due to high sensitivity to input data or the need for more precise hyperparameter tuning.

Finally, the CNN-GRU-Attention model performed comparably to the CNN-LSTM-Attention model but with a gentler validation error reduction slope, indicating potential for further optimization.

5. Discussion

Table 4 presents the performance evaluation results for traffic flow prediction, using the metrics MSE, RMSE, and R^2 across nine different models (including LSTM, GRU, CNN, CNN-GRU, CNN-LSTM, and their attention-equipped variants). Models with simpler architectures were compared alongside their corresponding attention-based versions. In addition to these quantitative metrics, model alignment with actual values was assessed using observed vs. predicted plots and scatter plots.

The results show that the CNN-LSTM and LSTM models achieved the highest accuracy, with the lowest MSE and RMSE and the highest R^2 values. The CNN-LSTM model, with an MSE of 13,487.7, RMSE of 116.14, and R^2 of 0.8356, was the most accurate, explaining over 83% of the data variance. This indicates a strong ability to align model outputs with observed data.

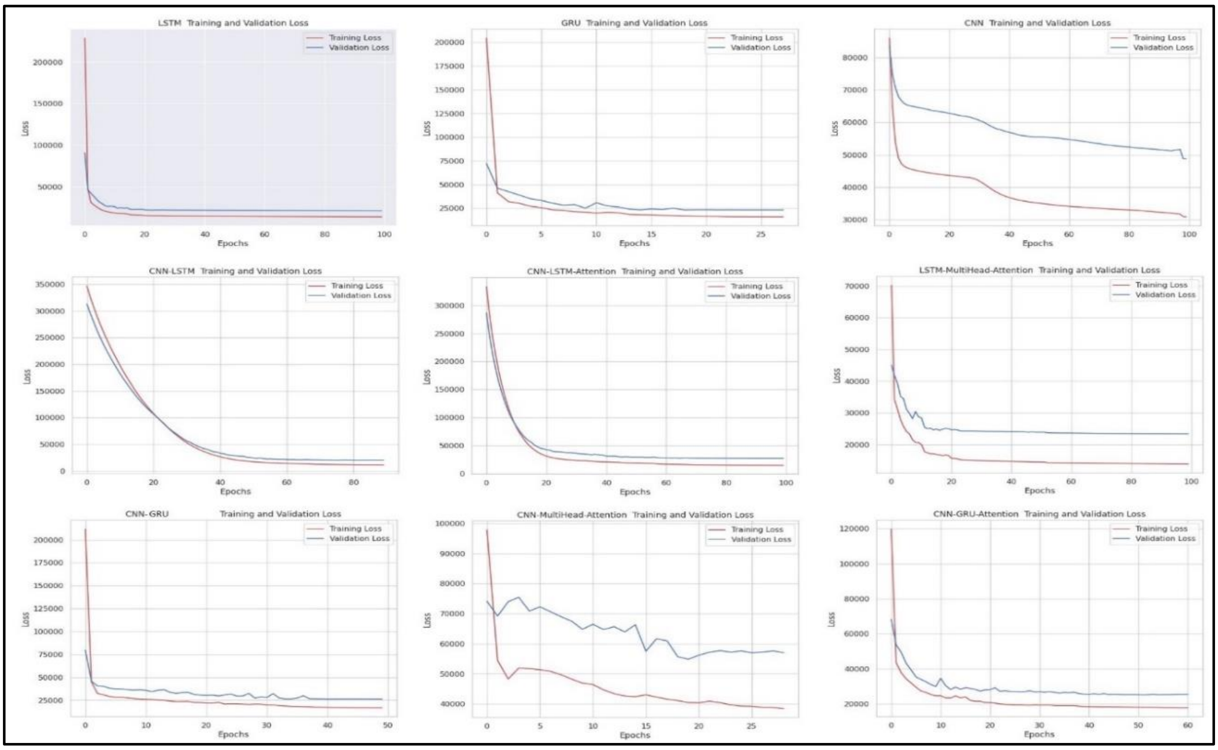

Figure 4. Performance Comparison of Deep Learning Models for Short-Term Traffic Flow Prediction

In contrast, the CNN and CNN-MultiHead-Attention models displayed poorer performance, with high MSE values of 30,369.6 and 32,076.2, respectively, high RMSE values, and low R^2 values of 0.6277 and 0.6068, respectively. This indicates that purely CNN-based architectures are less effective at capturing complex time-series patterns.

The analyses in this study demonstrate that hybrid deep architectures, such as CNN-LSTM, which combine CNN's spatial feature extraction with LSTM's temporal dependency modeling, significantly enhance traffic flow prediction accuracy. Among these, the CNN-LSTM model, adept at capturing temporal dynamics and managing long-term pattern dependencies, exhibited the best performance among the evaluated models.

To verify this, a Wilcoxon signed-rank test was conducted, with the results shown in Table 4. The p-values, all being less than 0.05, indicate that the superior performance of the CNN-LSTM model compared to all other models is statistically significant.

The model CNN-LSTM-Attention has a slight error decrease and better performance compared to the model CNN-LSTM. In all used models, however, the use of the attention mechanism has little influence, especially in weak models such as CNN-MultiHead Attention. This limited impact could be attributed to several factors. First, the inherent strength of the CNN-LSTM baseline may already capture the dominant spatio-temporal patterns, leaving marginal room for improvement by the attention layer. Second, the dataset, while spanning two years, might still be insufficient for training a more complex attention mechanism to its full potential. Finally, the specific single-head attention architecture used might be too simple to uncover more subtle dependencies in this highly volatile traffic environment.

Conversely, the generally poor performance of CNN-only models indicates that relying solely on spatial feature extraction, without accounting for temporal structures and long-term dynamics of time-series data, is insufficient for tasks like traffic prediction.

Additionally, attention mechanisms improved prediction accuracy in some models, with their effectiveness depending on the base model architecture.

Table 4. Input-Output Structure for One-Hour-Ahead Traffic Flow Prediction

Model	MSE	RMSE	R^2 (%)	Rank	p-value (vs. CNN-LSTM)
LSTM	14525.76	120.52	0.82	2	0.031
GRU	14979.58	122.39	0.81	3	0.025
CNN	30369.60	174.26	0.62	7	< 0.001
CNN-LSTM	13487.70	116.13	0.83	1	-
CNN-GRU	15956.22	126.31	0.80	5	0.018
CNN-LSTM-Attention	15726.59	125.40	0.80	4	0.045
CNN-GRU-Attention	16722.94	129.31	0.79	8	< 0.001
LSTM-MultiHead-Attention	15994.11	126.46	0.80	6	0.015
CNN-MultiHead-Attention	32076.18	179.09	0.60	9	< 0.001

Overall, the findings of this study highlight those robust architectures, such as CNN-LSTM, combined with consideration of temporal dependencies in real-world datasets, offer promising avenues for enhancing traffic flow forecasting performance. These results can inform the development of more effective models for practical applications, such as intelligent traffic management systems or journey planning.

Figure 5 displays the actual and forecasted traffic flow values for the nine models using the test dataset. The horizontal axis represents the number of test samples, and the vertical axis indicates the magnitude of traffic flow. The blue line represents observed values, while the orange line represents forecasts. The closer the orange line is to the blue line, the higher the model’s accuracy. Analysis of observed versus predicted plots revealed that all models can identify general data patterns, but discrepancies persist in predicting peak values. The best alignment between predicted and actual values was observed in the CNN-LSTM and LSTM models. Models incorporating attention mechanisms excelled at capturing sudden spikes in traffic (e.g., peak hours) and sharp drops (e.g., nighttime low traffic).

5.1. Analysis of Model Performance Under Extreme Conditions

To gain deeper insights, we analyzed model performance in specific traffic scenarios:

Performance on Holidays and Peak Traffic: During holidays and weekends, the Karaj-Chalous Road experiences sharp, sudden surges in traffic. We observed that hybrid models, especially CNN-LSTM, were more effective at capturing the onset of these peaks compared to simpler models. However, all models tended to slightly underestimate the maximum traffic volume during the most extreme peaks, suggesting a need for more data from such rare events.

Performance during Nighttime and Low Traffic: During late-night hours, traffic flow approaches zero. Models equipped with attention mechanisms (e.g., CNN-LSTM-Attention) demonstrated superior performance in this regime, accurately predicting near-zero values and avoiding the minor negative or positive biases seen in other models. This indicates their ability to effectively learn the "do-nothing" pattern from historical data when traffic cues are absent.

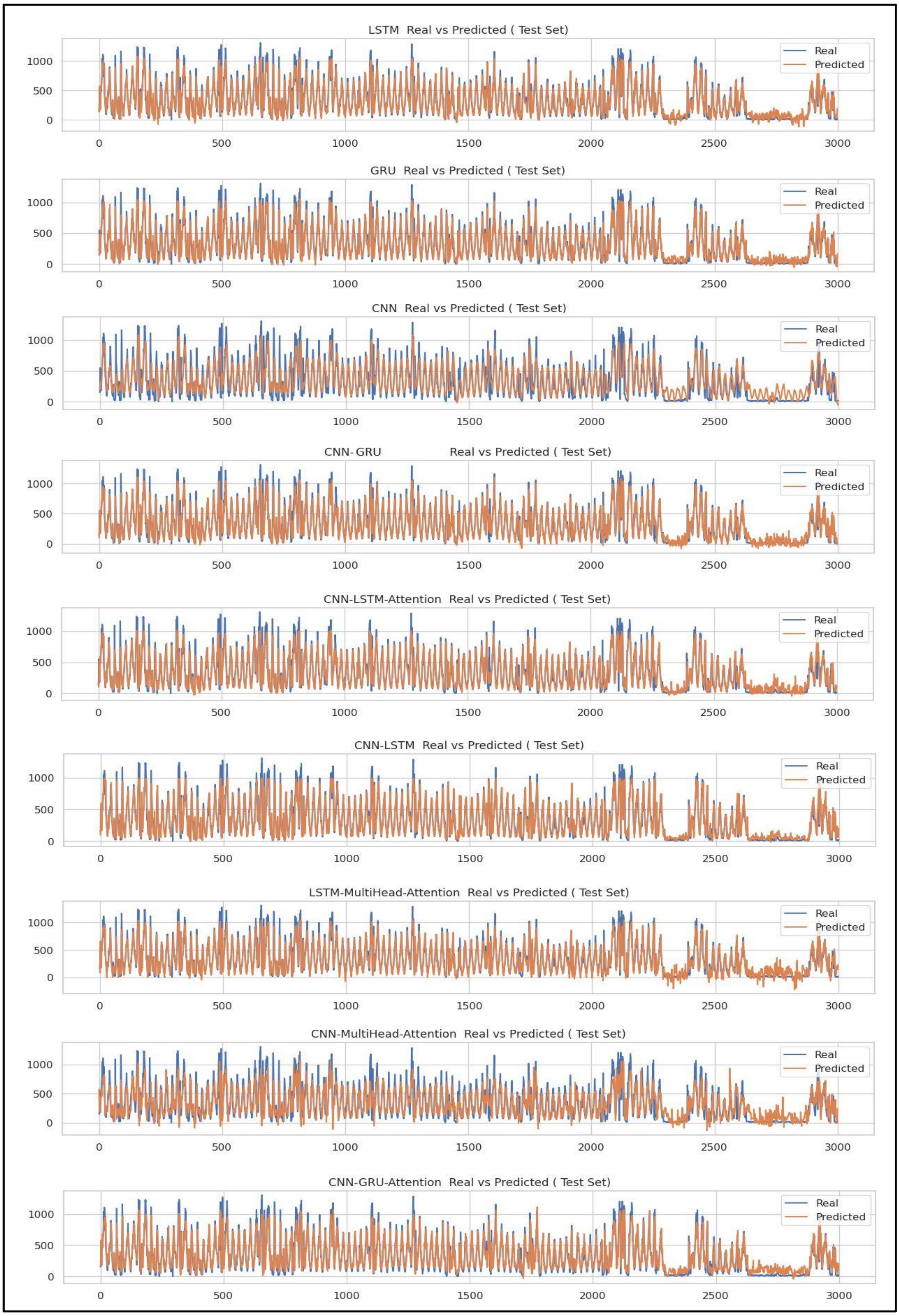

Figure 5. Actual vs. Predicted Traffic Flow for Nine Deep Learning Models on Test Dataset

In periods where observed values are near zero (e.g., low-traffic conditions like nighttime), predicted values should also approach zero. Attention-based models that have effectively learned from historical patterns can accurately forecast these fluctuations. Many intervals display oscillatory patterns (consecutive increases and decreases). Models that successfully capture these patterns generate more accurate forecasts. Deviations between predicted and observed patterns indicate weaker model performance in those segments.

Figure 6 presents scatter plots of actual versus predicted traffic flow values for each of the nine models. The dashed red line (y = x) represents perfect predictions, where predicted values equal observed values. Points closer to this line indicate higher model accuracy, with tighter clustering around the line reflecting better predictive performance. Models with blue points more closely clustered around y = x exhibit lower errors and provide the most accurate predictions.

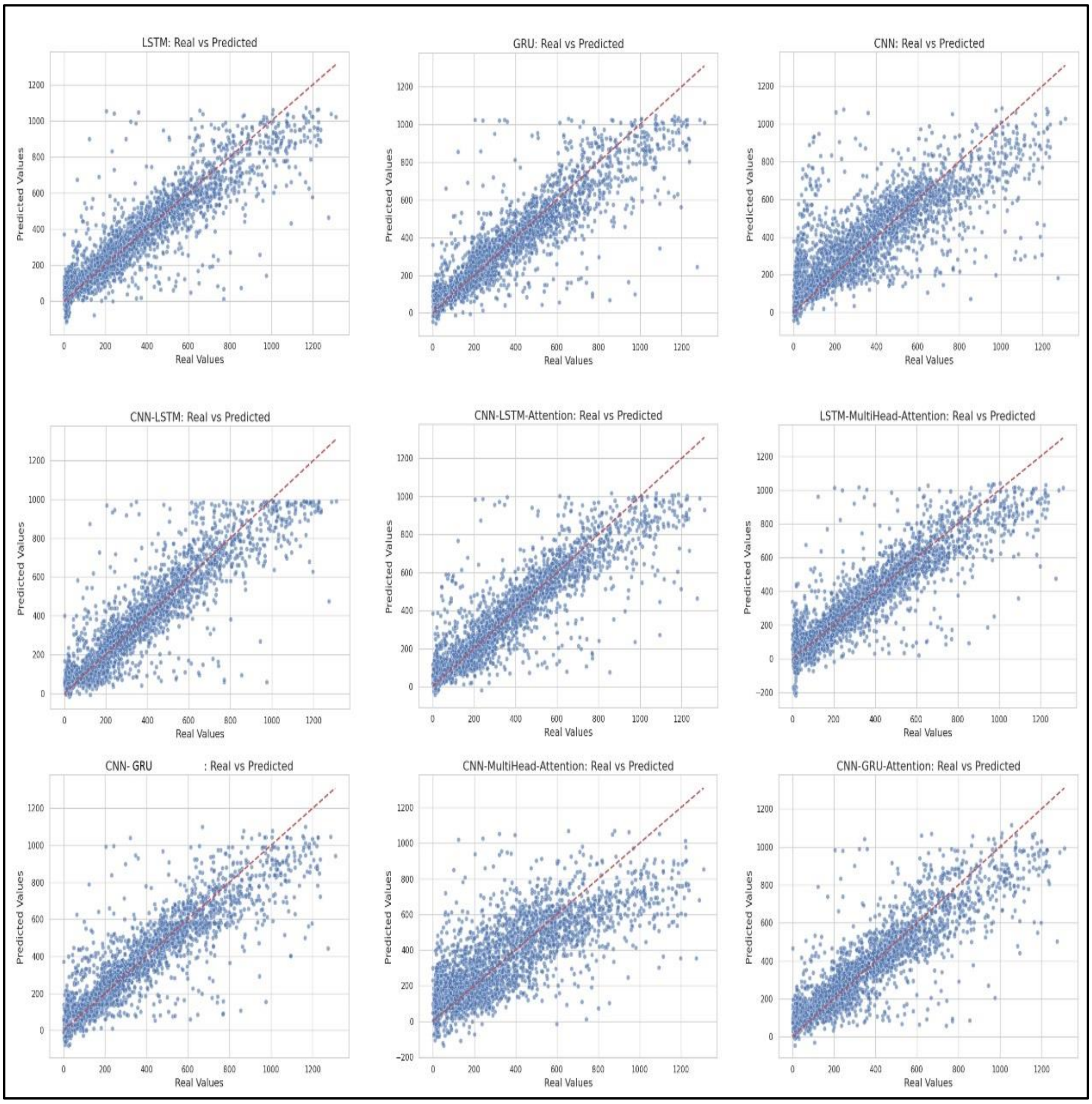

Figure 6. Scatter Plots of Actual vs. Predicted Traffic Flow for Deep Learning Models

6. Conclusions

This study explored the feasibility of short-term traffic flow prediction on the Karaj-Chalous Road using deep learning approaches. The primary innovation lies in the systematic comparison of nine deep learning architectures, ranging from basic to hybrid models, evaluated with comprehensive traffic and meteorological data. Our findings show that hybrid models, particularly CNN-LSTM with an R^2 of 0.83, outperform simpler architectures by effectively capturing complex spatiotemporal traffic patterns. Subsequent rankings placed LSTM (R^2=0.82), GRU (R^2=0.81), and CNN-GRU (R^2=0.80) as strong alternatives, while pure CNN (R^2=0.62) and CNN-MultiHead-Attention (R^2=0.60) models underperformed, likely due to their limited ability to model sequential dependencies. Although attention mechanisms provided marginal improvements in some architectures, their impact was limited by the base model's capacity. These results underscore the importance of thoughtful architectural design for accurate traffic state prediction, enabling intelligent decision-making systems for congested mountainous highways. Such models could support operational tools for congestion management, travel planning, and risk mitigation.

While this study provides promising insights into traffic prediction, its findings are subject to certain limitations, primarily the focus on a single road segment and a limited time period, which constrains the generalizability of our results. To address this, future research should proceed along several key avenues to develop more robust and scalable solutions. Architecturally, the implementation of Transformer-based models could better capture long-range spatiotemporal dependencies in traffic flow, while the exploration of Graph Neural Networks (GNNs) offers a powerful method for explicitly modeling the topological complexities of road networks. Concurrently, advancing optimization techniques is crucial; the use of metaheuristic algorithms, such as the Grey Wolf Optimizer (Khorshidi et al., 2025), for hyperparameter tuning and the development of hybrid optimization-deep learning pipelines could significantly enhance model efficiency and performance. Most critically, future work must prioritize data expansion and real-world validation. This includes incorporating diverse, multi-region datasets that integrate crucial socioeconomic and behavioral factors, thereby enriching the model's contextual understanding. The ultimate goal should be the validation of these advanced models within real-world traffic management systems, potentially through the development of embedded systems for real-time prediction and field trials to quantify their operational impacts. Such a comprehensive approach would not only overcome the limitations of the current study but also pave the way for more intelligent and adaptive transportation systems.

7. Conflict of Interest

No conflicts of interest are declared by the authors.

8. References

- Ahmed, M. S., & Cook, A. R. (1979). Analysis of Freeway Traffic Time-Series Data by Using Box-Jenkins Techniques.

- Bai, L., Yao, L., Li, C., Wang, X., & Wang, C. (2020). Adaptive graph convolutional recurrent network for traffic forecasting. Advances in Neural Information Processing Systems, 33, 17804–17815.

- Bui, K.-H. N., Cho, J., & Yi, H. (2022). Spatial-temporal graph neural network for traffic forecasting: An overview and open research issues. Applied Intelligence, 52(3), 2763–2774. https://doi.org/10.1007/s10489-021-02587-w

- Chen, H., & Grant-Muller, S. (2001). Use of sequential learning for short-term traffic flow forecasting. Transportation Research Part C-Emerging Technologies. https://doi.org/10.1016/S0968-090X(00)00039-5

- Cirstea, R.-G., Yang, B., Guo, C., Kieu, T., & Pan, S. (2022). Towards spatio-temporal aware traffic time series forecasting. 2022 IEEE 38th International Conference on Data Engineering (ICDE).

- Do, L. N. N., Taherifar, N., & Vu, H. (2018). Survey of neural network-based models for short-term traffic state prediction. Wiley Interdisciplinary Reviews Data Mining and Knowledge Discovery. https://doi.org/10.1002/WIDM.1285

- Do, L. N. N., Vu, H. L., Vo, B. Q., Liu, Z., & Phung, D. (2019). An effective spatial-temporal attention based neural network for traffic flow prediction. Transportation Research Part C: Emerging Technologies. https://doi.org/10.1016/J.TRC.2019.09.008

- Dougherty, M., & Cobbett, M. R. (1997). Short-term inter-urban traffic forecasts using neural networks. International Journal of Forecasting. https://doi.org/10.1016/S0169-2070(96)00697-8

- Duan, P., Mao, G., Liang, W., & Zhang, D. (2019). A Unified Spatio-Temporal Model for Short-Term Traffic Flow Prediction. IEEE Transactions on Intelligent Transportation Systems (Print). https://doi.org/10.1109/TITS.2018.2873137

- Dynamic Graph Convolutional Networks Based on Spatiotemporal Data Embedding for Traffic Flow Forecasting, 250 Knowledge-Based Systems 109028 (Elsevier 2022). https://www.sciencedirect.com/science/article/pii/S0950705122005032

- Guo, J., Huang, W., & Williams, B. M. (2014). Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification. Transportation Research Part C-Emerging Technologies. https://doi.org/10.1016/J.TRC.2014.02.006

- Guo, S., Lin, Y., Feng, N., Song, C., & Wan, H. (2019). Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. Proceedings of the AAAI Conference on Artificial Intelligence, 33(01), 922–929. https://ojs.aaai.org/index.php/AAAI/article/view/3881

- Iran Meteorological Organization (IRIMO). (2024). IRIMO Official Website. https://www.irimo.ir

- Iran Road Maintenance and Transportation Organization. (2024). RMTO Official Website. https://www.rmto.ir

- Jeong, Y., Byon, Y.-J., Castro-Neto, M., & Easa, S. (2013). Supervised Weighting-Online Learning Algorithm for Short-Term Traffic Flow Prediction. IEEE Transactions on Intelligent Transportation Systems (Print). https://doi.org/10.1109/TITS.2013.2267735

- Jia, Y., Wu, J., & Xu, M. (2017). Traffic Flow Prediction with Rainfall Impact Using a Deep Learning Method. https://doi.org/10.1155/2017/6575947

- Karlaftis, M. G., & Vlahogianni, E. I. (2011). Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. Transportation Research Part C-Emerging Technologies. https://doi.org/10.1016/J.TRC.2010.10.004

- Khorshidi, N., Zargari, S. A., Rezashoar, S., & Mirzahossein, H. (2025). Optimizing Travel Time Reliability with XAI: A Virginia Interstate Network Case Using Machine Learning and Meta-Heuristics. Machine Learning with Applications, 100709. https://doi.org/10.1016/j.mlwa.2025.100709

- Kong, J., Fan, X., Zuo, M., Deveci, M., Jin, X., & Zhong, K. (2024). ADCT-Net: Adaptive traffic forecasting neural network via dual-graphic cross-fused transformer. Information Fusion, 103, 102122.

- Li, M., & Zhu, Z. (2021). Spatial-temporal fusion graph neural networks for traffic flow forecasting. Proceedings of the AAAI Conference on Artificial Intelligence, 35(5), 4189–4196. https://ojs.aaai.org/index.php/AAAI/article/view/16542

- Li, Y., Yu, R., Shahabi, C., & Liu, Y. (2018). Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting (No. arXiv:1707.01926). arXiv. https://doi.org/10.48550/arXiv.1707.01926

- Liu, H., Zhu, C., Zhang, D., & Li, Q. (2023). Attention-based spatial-temporal graph convolutional recurrent networks for traffic forecasting. International conference on advanced data mining and applications.

- Lippi, M., Bertini, M., & Frasconi, P. (2013). Short-Term Traffic Flow Forecasting: An Experimental Comparison of Time-Series Analysis and Supervised Learning. IEEE Transactions on Intelligent Transportation Systems. https://doi.org/10.1109/TITS.2013.2247040

- Liu, T., Wang, Y., Zhou, H., Luo, J., & Deng, F. (2024). Distributed Short-Term Traffic Flow Prediction Based on Integrating Federated Learning and TCN. IEEE Access, 12, 148026–148036. IEEE Access. https://doi.org/10.1109/ACCESS.2024.3474300

- Luo, X., Zhu, C., Zhang, D., & Li, Q. (2023). Dynamic Graph Convolutional Network with Attention Fusion for Traffic Flow Prediction (No. arXiv:2302.12598). arXiv. https://doi.org/10.48550/arXiv.2302.12598

- Lv, Y., Duan, Y., Kang, W., Li, Z., & Wang, F. (2015). Traffic Flow Prediction With Big Data: A Deep Learning Approach. IEEE Transactions on Intelligent Transportation Systems (Print). https://doi.org/10.1109/TITS.2014.2345663

- Ma, X., Tao, Z., Wang, Y., Yu, H., & Wang, Y. (2015). Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. Transportation Research Part C-Emerging Technologies. https://doi.org/10.1016/J.TRC.2015.03.014

- Miglani, A., & Kumar, N. (2019). Deep learning models for traffic flow prediction in autonomous vehicles: A review, solutions, and challenges. Vehicular Communications. https://doi.org/10.1016/J.VEHCOM.2019.100184

- Naheliya, B., Redhu, P., & Kumar, K. (2023). A Hybrid Deep Learning Method for Short-Term Traffic Flow Forecasting: GSA-LSTM. SRS Journal. https://indjst.org/

- Okutani, I., & Stephenades, Y. J. (1984). Dynamic prediction of traffic volume through Kalman Filtering. https://doi.org/10.1016/0191-2615(84)90002-X

- Park, B., Messer, C., & Urbanik, T. (1998). Short-Term Freeway Traffic Volume Forecasting Using Radial Basis Function Neural Network. https://doi.org/10.3141/1651-06

- Polson, N. G., & Sokolov, V. O. (2016). Deep learning for short-term traffic flow prediction. https://doi.org/10.1016/J.TRC.2017.02.024

- Shu, W., Cai, K., & Xiong, N. (2021). A Short-Term Traffic Flow Prediction Model Based on an Improved Gate Recurrent Unit Neural Network. IEEE Transactions on Intelligent Transportation Systems (Print). https://doi.org/10.1109/TITS.2021.3094659

- Smith, B. L., & Demetsky, M. (1997). Traffic Flow Forecasting: Comparison of Modeling Approaches. https://doi.org/10.1061/(ASCE)0733-947X(1997)123:4(261)

- Song, C., Lin, Y., Guo, S., & Wan, H. (2020). Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting. Proceedings of the AAAI Conference on Artificial Intelligence, 34(01), 914–921. https://ojs.aaai.org/index.php/AAAI/article/view/5438

- Stathopoulos, A., & Karlaftis, M. G. (2003). A multivariate state space approach for urban traffic flow modeling and prediction. Transportation Research Part C-Emerging Technologies. https://doi.org/10.1016/S0968-090X(03)00004-4

- Sun, S., Zhang, C., & Yu, G. (2006). A bayesian network approach to traffic flow forecasting. IEEE Transactions on Intelligent Transportation Systems. https://doi.org/10.1109/TITS.2006.869623

- Tedjopurnomo, D. A., Bao, Z., Zheng, B., Choudhury, F., & Qin, A. K. (2020). A Survey on Modern Deep Neural Network for Traffic Prediction: Trends, Methods and Challenges. IEEE Transactions on Knowledge and Data Engineering. https://doi.org/10.1109/TKDE.2020.3001195

- Vlahogianni, E., Golias, J., & Karlaftis, M. (2004). Short-term traffic forecasting: Overview of objectives and methods. https://doi.org/10.1080/0144164042000195072

- Vlahogianni, E. I., Karlaftis, M. G., & Golias, J. (2014). Short-term traffic forecasting: Where we are and where we're going. Transportation Research Part C-Emerging Technologies. https://doi.org/10.1016/J.TRC.2014.01.005

- Wang, Y., Xu, S., & Di Feng. (2020). A New Method for Short-term Traffic Flow Prediction Based on Multi-segments Features. ICMLC. https://doi.org/10.1145/3383972.3384038

- Williams, B. M., & Hoel, L. (2003). Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results. Journal of Transportation Engineering. https://doi.org/10.1061/(ASCE)0733-947X(2003)129:6(664)

- Wu, Y., Tan, H., Qin, L., Bin Ran, & Jiang, Z. (2018). A hybrid deep learning based traffic flow prediction method and its understanding. Transportation Research Part C-Emerging Technologies. https://doi.org/10.1016/J.TRC.2018.03.001

- Xia, M., Jin, D., & Chen, J. (2023). Short-Term Traffic Flow Prediction Based on Graph Convolutional Networks and Federated Learning. IEEE Transactions on Intelligent Transportation Systems (Print). https://doi.org/10.1109/TITS.2022.3179391

- Xie, Y., Zhang, Y., & Ye, Z. (2007). Short-Term Traffic Volume Forecasting Using Kalman Filter with Discrete Wavelet Decomposition. Computer-Aided Civil and Infrastructure Engineering. https://doi.org/10.1111/J.1467-8667.2007.00489.X

- Yu, B., Yin, H., & Zhu, Z. (2018). Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting. https://doi.org/10.24963/ijcai.2018/505

- Yuan, H., & Li, G. (2021). A Survey of Traffic Prediction: From Spatio-Temporal Data to Intelligent Transportation. Data Science and Engineering, 6(1), 63–85. https://doi.org/10.1007/s41019-020-00151-z

- Zargari, S. A., Khorshidi, N., Mirzahossein, H., & Jin, X. (2025). Application of data augmentation techniques in predicting travel time reliability: evidence from England. Iranian Journal of Science and Technology, Transactions of Civil Engineering, 49(1), 921-933. https://doi.org/10.1007/s40996-024-01383-z

- Zhang, W., Yu, Y., Qi, Y., Shu, F., & Wang, Y. (2019). Short-term traffic flow prediction based on spatio-temporal analysis and CNN deep learning. Transportmetrica A: Transport Science. https://doi.org/10.1080/23249935.2019.1637966

- Zhang, Y., & Ye, Z. (2008). Short-Term Traffic Flow Forecasting Using Fuzzy Logic System Methods. J. Intell. Transp. Syst. https://doi.org/10.1080/15472450802262281

Zhao, Z., Chen, W., Wu, X., Chen, P. C. Y., & Liu, J. (2017). LSTM network: A deep learning approach for short-term traffic forecast. https://doi.org/10.1049/IET-ITS.2016.0208

- Zheng, C., Fan, X., Wang, C., & Qi, J. (2020). Gman: A graph multi-attention network for traffic prediction. Proceedings of the AAAI Conference on Artificial Intelligence, 34(01), 1234–1241. https://aaai.org/ojs/index.php/AAAI/article/view/5477