

Article

Advancing Urban Planning with Deep Learning: Intelligent Traffic Flow Prediction and Optimization for Smart Cities

Fatema A. Albalooshi 

College of Information Technology, University of Bahrain, Sakhir P.O. Box 32038, Bahrain; falbalooshi@uob.edu.bh

Abstract

The accelerating pace of urbanization has significantly complicated traffic management systems, leading to mounting challenges, such as persistent congestion, increased travel delays, and heightened environmental impacts. In response to these challenges, this study presents a novel deep learning framework designed to enhance short-term traffic flow prediction and support intelligent transportation systems within the context of smart cities. The proposed model integrates Gated Recurrent Units (GRUs) and Long Short-Term Memory (LSTM) networks, augmented by an attention mechanism that dynamically emphasizes relevant temporal patterns. The model was rigorously evaluated using the publicly available datasets and demonstrated substantial improvements over current state-of-the-art methods. Specifically, the proposed framework achieves a 3.75% reduction in the Mean Absolute Error (MAE), a 2.00% reduction in the Root Mean Squared Error (RMSE), and a 4.17% reduction in the Mean Absolute Percentage Error (MAPE) compared to the baseline models. The enhanced predictive accuracy and computational efficiency offer significant benefits for intelligent traffic control, dynamic route planning, and proactive congestion management, thereby contributing to the development of more sustainable and efficient urban mobility systems.

Keywords: urban planning; smart cities; traffic flow prediction; deep learning; smart transportation; GRU; LSTM; attention mechanism; PEMS datasets



Academic Editor: Zhihong Yao

Received: 8 August 2025

Revised: 6 September 2025

Accepted: 11 September 2025

Published: 2 October 2025

Citation: Albalooshi, F.A. Advancing Urban Planning with Deep Learning: Intelligent Traffic Flow Prediction and Optimization for Smart Cities. *Future Transp.* **2025**, *5*, 133. <https://doi.org/10.3390/futuretransp5040133>

Copyright: © 2025 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Rapid urbanization and increased vehicle ownership present unprecedented challenges for transportation systems. Traffic congestion, unpredictable travel times, and elevated emissions adversely affect the efficiency of transportation systems and, in turn, hinder economic productivity and the quality of life in urban environments [1,2]. In particular, persistent traffic congestion constitutes a critical challenge for metropolitan areas, leading to economic inefficiencies, excessive fuel consumption, and increased greenhouse gas emissions that deteriorate both environmental quality and public health standards [3,4]. Traditional traffic management systems—predominantly reliant on static signal timings or reactive interventions—are increasingly inadequate for addressing the nonlinear and highly dynamic characteristics of modern traffic flows [5].

The emergence of smart city paradigms, enabled by advances in sensor technologies, Internet of Things (IoT) infrastructure, and artificial intelligence (AI), has created opportunities for the development of adaptive and intelligent traffic control strategies [6,7]. Deep learning has, in particular, proven to be a transformative approach [8], demonstrating significant potential to model complex spatiotemporal dependencies and to extract latent patterns from high-dimensional traffic datasets [9,10].

Accurate short-term traffic flow forecasting extends beyond the immediate needs of intelligent transportation systems; it directly informs critical aspects of urban planning and sustainable city design. Reliable predictions can guide dynamic road capacity allocation, such as adaptive signal timing or reversible lane strategies, which enable transportation networks to respond flexibly to fluctuating demand. Forecasting also provides an evidence base for infrastructure investment planning, allowing urban planners to identify persistent congestion hotspots and prioritize road expansions or new public transit corridors where they will yield the greatest long-term benefit.

Equally important, predictive traffic analytics play a pivotal role in advancing sustainability goals. By mitigating congestion, optimizing routing, and reducing idling, forecasting models support the reduction of vehicle emissions and improvements in energy efficiency. These outcomes are directly aligned with the United Nations Sustainable Development Goal (SDG) 11: (sustainable cities and communities) [11], which emphasizes the need for resilient, efficient, and environmentally responsible urban transport systems. Thus, the proposed methodological contributions to enhance computational performance and to provide tangible tools for planners and policy makers to design safer, more livable, and sustainable cities.

Performing accurate and timely traffic flow prediction is the cornerstone of intelligent transportation systems (ITS). Reliable forecasts support applications such as dynamic signal control, congestion mitigation, proactive route planning, and efficient emergency response allocation [12,13]. Recent research has confirmed the effectiveness of recurrent neural networks (RNNs), particularly Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) architectures, in capturing sequential dependencies in traffic data [14,15]. However, standalone models often struggle to balance predictive accuracy with computational efficiency, especially under real-time constraints [16].

To overcome these limitations, this study proposes a hybrid deep learning architecture that integrates GRU and LSTM units, augmented by a temporal attention mechanism. This design enhances the model's ability to capture both long-term dependencies and contextually relevant temporal features, thereby advancing predictive accuracy and operational efficiency in traffic flow forecasting. The key contributions of this work are as follows:

- A novel hybrid deep learning architecture is introduced, combining GRU and LSTM units with an attention mechanism to improve the extraction of complex temporal patterns inherent in traffic data.
- Comprehensive evaluation on the publicly available PEMS-03, PEMS-04, PEMS-07, and PEMS-08 datasets [17] demonstrates that the proposed model outperforms established baselines and recent deep learning methods, achieving measurable error reductions.
- Extensive ablation studies are conducted to assess the contribution of each architectural component and to verify the practical feasibility of real-world deployment.

By jointly addressing predictive performance and computational efficiency, this research delivers a robust, scalable, and efficient solution for urban traffic forecasting. The proposed model can serve as a foundational component of intelligent traffic management systems, thereby supporting the broader vision of sustainable, efficient, and adaptive smart cities.

The remainder of this paper is organized as follows: Section 2 reviews related works. Section 3 details the proposed architecture. Section 4 presents the experimental setup, results, and comparative analysis. Finally, Section 5 concludes the paper and outlines directions for future research.

2. Literature Review

Traffic forecasting has undergone a remarkable transformation over the last decade, evolving from statistical modeling to advanced deep learning solutions. Early predictive models were grounded in classical time-series algorithms, such as ARIMA and Kalman filters [18,19], offering simplicity and interpretability but limited in handling nonlinear spatial–temporal dependencies inherent in complex road networks.

Razali et al. [20] review traffic flow prediction using machine learning and deep learning, highlighting existing gaps, techniques, and evaluation methods. They analyze various predictive models, discussing their strengths and limitations in handling traffic data complexity and variability. The study emphasizes the need for more robust and adaptive prediction frameworks to improve traffic management and decision-making in dynamic urban environments.

As data availability increased, machine learning approaches—including Support Vector Regression (SVR), Random Forests (RF), and k-Nearest Neighbors (k-NN)—were shown to surpass classical methods in forecasting traffic speed and volume, particularly under non-linear and dynamic traffic conditions [21,22]. Although these methods improved the regression accuracy, they often fell short at modeling dependencies that stretch both in space and time. Moreover, they required careful feature engineering and struggled when scaling to large, heterogeneous datasets.

The maturation of deep learning introduced hybrid architectures combining Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) [23], particularly LSTMs. CNNs offer powerful spatial feature extraction from structured sensor arrays, while LSTMs capture temporal sequence dynamics, a combination that has demonstrated superior performance in traffic forecasting tasks [24,25]. The alliance of these two components enabled superior short-term predictions and paved the way for real-time traffic forecasting applications.

Recent advancements in Graph Neural Networks (GNNs) have significantly improved the accuracy and robustness of traffic flow prediction by explicitly modeling the spatial structure of road networks. Traditional temporal models often fail to capture complex inter-sensor dependencies, whereas GNN-based models integrate topological and temporal features in a unified framework.

Theodoropoulos et al. [26] proposed the WEST GCN-LSTM architecture, which introduces weighted stacked spatiotemporal graph convolutions combined with LSTM units. This model incorporates domain-specific policies, such as shared borders and adjustable hop distances, to better capture regional traffic patterns. Similarly, Jin et al. [27] introduced the STGNPP model, a spatiotemporal graph neural point process framework designed specifically for predicting traffic congestion events, modeling them as stochastic processes over time and graph space. Although the GSTPRN framework introduces a promising integration of spatial–temporal modeling components [28], computational efficiency and scalability to larger networks are essential. Reliance on self-attention and personalized propagation introduces significant overhead in real-time deployment scenarios.

Advancements in transformer and attention-based models have further revolutionized traffic forecasting by tailoring self-attention mechanisms to spatial and temporal dynamics. Prabowo et al. [29] introduced Graph Self-attention WaveNet (G-SWaN), which adapts self-attention across sensor pairs by learning unique dynamics for each sensor and sensor pair, yielding superior performance on multiple traffic datasets. Liu et al. [30] proposed STAEformer, a vanilla transformer integrated with Spatio-Temporal Adaptive Embedding, which captures intrinsic traffic patterns and achieves state-of-the-art results on five real-world datasets. Li et al. [31] enhanced existing deep forecasting frameworks with a Dynamic Regression module, explicitly modeling structured residual noise across sensors

and time, thereby improving both interpretability and robustness. Together, these models demonstrate the transformative potential of attention-based transformer architectures in capturing complex spatiotemporal dependencies for intelligent transportation systems.

Moreover, federated learning paradigms have increasingly emerged in traffic-forecasting applications, offering promising solutions for privacy-preserving, decentralized model training across distributed traffic sensors. Liu et al. [32] proposed FedOSTC, an on-line spatiotemporal federated learning framework that combines local GRU encoders with graph-attention-based spatial aggregation, enabling privacy-preserving, adaptive forecasting across distributed sensors. Similarly, Alqubaysi et al. [33] introduced a federated graph neural network approach for urban traffic prediction that improves model generalization while safeguarding data privacy. More recently, Wang et al. [34] developed an adaptive federated learning strategy that dynamically adjusts client contributions to enhance the forecasting accuracy in heterogeneous traffic networks.

Hong et al. [35] propose a resilience recovery method for complex traffic networks using trend forecasting. Their approach predicts network stress trends to guide adaptive recovery strategies, improving both robustness and post-disruption recovery efficiency. Simulation results demonstrate its effectiveness over conventional methods in maintaining traffic network stability.

Moreover, Abduljabbar et al. [36] surveyed machine learning models for sustainable traffic prediction, highlighting the trade-offs between accuracy and resource consumption. Likewise, Mystakidis et al.'s systematic review [13] emphasized the challenge of balancing predictive power with model compactness when applied to urban environments.

Chen et al. [37] proposed an attention-augmented LSTM framework that selectively emphasizes critical time steps, resulting in enhanced prediction accuracy during peak hours. Similarly, Qaffas [38] demonstrated the effectiveness of transformer-based attention modules in capturing dynamic traffic correlations across sensors in an IoT-enabled environment.

Despite these advancements, several challenges persist. Most current models are limited in their ability to generalize across different urban settings due to overfitting on location-specific datasets [39,40]. In light of these findings, the proposed research builds on recent hybrid architectures by incorporating both GRU and LSTM layers, enhanced with an attention mechanism. This aims to improve the temporal learning efficiency and adaptability under dynamic traffic scenarios, contributing to the ongoing effort to make ITS infrastructures more intelligent, responsive, and sustainable.

3. Materials and Methods

This section outlines the proposed deep learning methodology for intelligent traffic flow prediction and optimization. First, the dataset used in this study is described. Then, the preprocessing steps applied to ensure the quality and suitability of the data are presented. After that, the hybrid model architecture is detailed, followed by a description of the training and evaluation procedures.

3.1. Dataset Description

The PEMS-03, PEMS-04, PEMS-07, and PEMS-08 publicly available datasets [17] are selected for this research due to their comprehensive coverage and widespread use in traffic forecasting studies. Collected by the California Department of Transportation (Caltrans) through the Performance Measurement System (PEMS), these datasets offer high-resolution real-time traffic information from multiple detectors deployed across the freeway system.

The datasets encompass comprehensive traffic information, including performance metrics, vehicle volume, and recorded crash incidents. For example, Figure 1 presents a

comparative analysis of travel times at different hours of three days for Mainline Lanes, effectively highlighting peak hour congestion patterns.

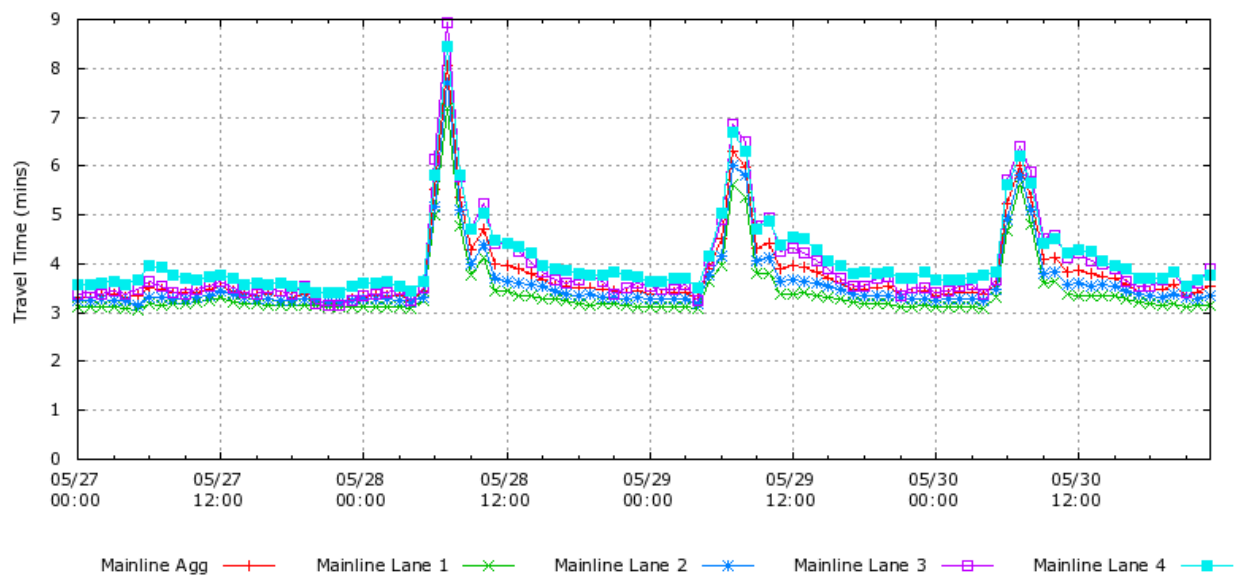


Figure 1. Sample data from PEMS-04 showing the travel time comparison throughout three days, captured on Mainlane Lanes.

To ensure reproducibility and transparency, a detailed summary of the datasets employed in this study is provided in Table 1. Four widely used benchmark subsets from the Caltrans Performance Measurement System (PeMS)—namely PeMS-03, PeMS-04, PeMS-07, and PeMS-08—were selected due to their extensive coverage of freeway networks and prior use in the traffic-forecasting literature. Each dataset consists of high-resolution measurements collected at five-minute intervals, including traffic speed, flow, and occupancy from fixed loop detectors.

Table 1. Summary of the PeMS datasets used in this study.

Dataset	Sampling Interval	Sensors	Time Span	Missing Rate (%)	Variables	Train/Val/Test Splits
PeMS-03	5 min	358	January–February 2018 (2 months)	3.2	Speed, Flow, and Occupancy	1 January–31 January / 1 February–10 February / 11 February–28 February
PeMS-04	5 min	307	January–February 2018 (2 months)	2.7	Speed, Flow, and Occupancy	1 January–31 January / 1 February–10 February / 11 February–28 February
PeMS-07	5 min	883	May–June 2017 (2 months)	4.1	Speed, Flow, and Occupancy	1 May–31 May / 1 June–10 June / 11 June–30 June
PeMS-08	5 min	170	July–August 2016 (2 months)	3.5	Speed, Flow, and Occupancy	1 July–31 July / 1 August–7 August / 8 August–31 August

The table reports the number of sensors deployed in each district, the temporal coverage of the datasets, the proportion of missing values, and the variables used in this study. To facilitate a fair evaluation, the datasets were partitioned into training, validation, and testing sets following community practice, with exact date ranges specified rather than approximate percentage splits. This explicit reporting allows for direct comparison with prior works and eliminates ambiguity regarding data usage.

Normalization was performed using the training set only, and the resulting parameters were applied consistently to the validation and testing sets. This procedure ensures that no information leakage occurred across partitions. By providing detailed dataset statistics, including missing rates and split protocols, we aim to support reproducibility and enable other researchers to replicate our experiments under identical conditions.

3.2. Data Preprocessing

Preprocessing plays a critical role in preparing raw traffic data for deep learning applications. The following steps were undertaken to enhance the data quality and ensure optimal model performance:

1. **Missing Value Imputation:** Real-world traffic datasets often contain missing values due to sensor outages, data corruption, or transmission errors. To ensure data integrity and consistency, two imputation strategies were employed based on the size of the missing segment. For short-term missing values (typically spanning a few time steps), linear interpolation is applied. Given two observed values, x_{t_1} and x_{t_2} , at time steps t_1 and t_2 , the missing value x_t for $t_1 < t < t_2$ is computed as follows:

$$x_t = x_{t_1} + \frac{t - t_1}{t_2 - t_1}(x_{t_2} - x_{t_1}) \quad (1)$$

This method ensures temporal smoothness and continuity in the time series data. Moreover, for longer gaps, a spatial nearest-neighbor imputation strategy was used. Let $x_t^{(i)}$ denote the missing value at detector i and time t . The imputed value is estimated using the values from spatially adjacent detectors $j \in \mathcal{N}(i)$ as follows:

$$x_t^{(i)} = \frac{1}{|\mathcal{N}(i)|} \sum_{j \in \mathcal{N}(i)} x_t^{(j)} \quad (2)$$

where $\mathcal{N}(i)$ is defined as the set of detectors that are physically adjacent to sensor i along the freeway network, based on the official Caltrans deployment maps. In practice, we adopt a fixed one-hop radius, which corresponds to the immediate upstream and downstream detectors. This design strikes a balance between locality and sensor coverage, ensuring that missing values are replaced using contextually relevant neighbors without introducing noise from distant detectors.

2. **Normalization:** To eliminate feature dominance caused by differing scales, all traffic variables—speed, volume, and occupancy—were normalized using the Min–Max scaling method. This transformation bounds the data within the range $[0, 1]$, which accelerates convergence and enhances training stability. The scaling formula is defined as follows:

$$X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (3)$$

It is important to note that normalization parameters were computed exclusively on the training set and subsequently applied to the validation and test sets, thereby preventing any information leakage across data partitions.

3. **Sequence Construction:** Since recurrent neural networks, such as GRU and LSTM, require sequential inputs, the time-series data was reshaped into a supervised learning format. A sliding window of $N = 12$ time steps (equivalent to one hour) was used to predict the subsequent $M = 3$ time steps (15 min ahead). This format captures both short-term and emerging patterns in traffic dynamics.
4. **Data Splitting:** To evaluate the model performance reliably, the dataset was partitioned into training (70%), validation (15%), and testing (15%) sets. Temporal order was

preserved during the split to avoid data leakage and ensure consistency with real-world deployment scenarios.

These preprocessing steps not only enhance data integrity but also optimize the learning conditions for the hybrid GRU-LSTM model with attention mechanism, which is detailed in the following section.

3.3. Hybrid GRU-LSTM Model with Attention Mechanism

The proposed model integrates the complementary strengths of Gated Recurrent Units (GRU) and Long Short-Term Memory (LSTM) networks, augmented by an attention mechanism, to effectively capture complex spatiotemporal dependencies in traffic flow data. This architecture exploits the computational efficiency of GRU for sequence processing alongside the robust long-term memory capabilities of LSTM. The attention mechanism enables dynamic emphasis on the most relevant historical information, enhancing prediction accuracy. The proposed model is structured into five key components, as illustrated in Figure 2.

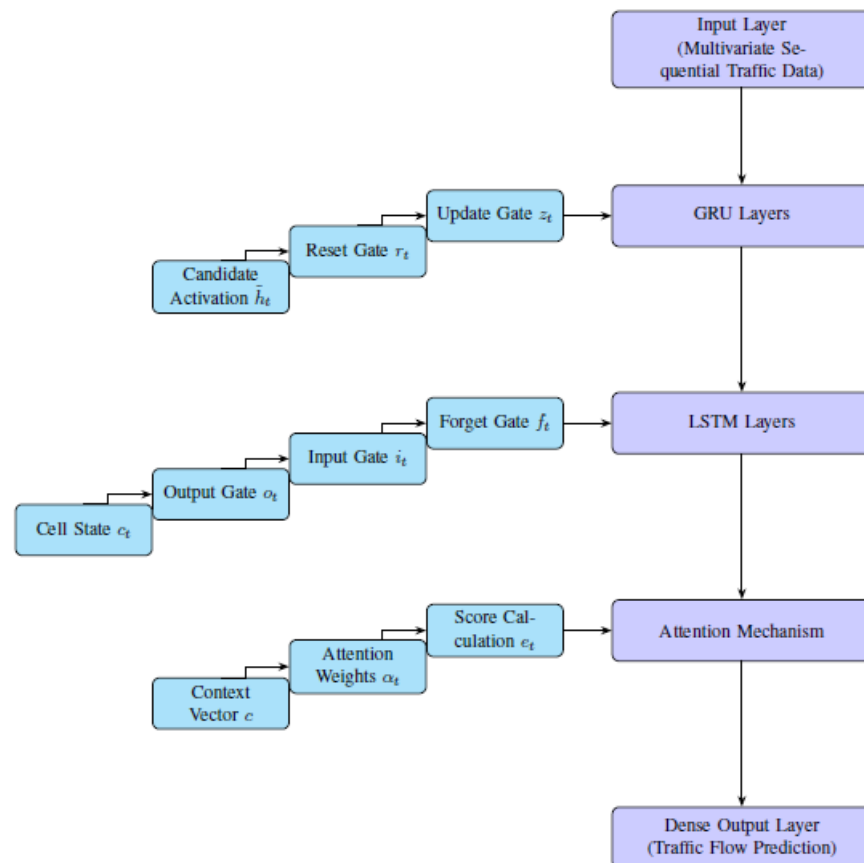


Figure 2. Schematic flowchart of the hybrid GRU-LSTM model with temporal attention mechanism, showing internal gate components and data flow.

1. **Input Layer:** Receives preprocessed sequential traffic data (e.g., traffic speed, volume, and occupancy) collected from multiple sensors over a defined look-back window.
2. **GRU Layers:** An initial stack of GRU layers processes the input sequences. GRU cells are selected for their computational efficiency and effectiveness in capturing short-term dependencies. These layers reduce dimensionality and extract salient features from raw input.

The GRU unit at time step t is defined as follows:

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \quad (4)$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \quad (5)$$

$$\tilde{h}_t = \tanh(W_h x_t + U_h (r_t \odot h_{t-1}) + b_h) \quad (6)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (7)$$

where x_t is the input vector at time step t , h_{t-1} is the previous hidden state, z_t is the update gate, r_t is the reset gate, $\tilde{h}_t$ is the candidate hidden state, $\odot$ denotes element-wise multiplication, $\sigma(\cdot)$ is the sigmoid function, $\tanh(\cdot)$ is the hyperbolic tangent function, and W_* , U_* , b_* are learnable parameters.

3. LSTM Layers: Subsequent LSTM layers model long-term dependencies and mitigate the vanishing-gradient problem, essential for capturing extended temporal patterns. The LSTM cell operations at time step t are the following:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (8)$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (9)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (10)$$

$$\tilde{c}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (11)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (12)$$

$$h_t = o_t \odot \tanh(c_t) \quad (13)$$

where f_t , i_t , and o_t are forget, input, and output gates respectively; c_t is the cell state; and h_t is the hidden state.

4. Attention Mechanism: An additive attention mechanism is applied to the LSTM outputs to selectively focus on important time steps. For each hidden state h_t , the attention score e_t is computed as follows:

$$e_t = v_a^\top \tanh(W_a h_t + b_a) \quad (14)$$

where v_a is a learnable parameter. The attention weights α_t are obtained via softmax:

$$\alpha_t = \frac{\exp(e_t)}{\sum_{k=1}^T \exp(e_k)} \quad (15)$$

The context vector c is the weighted sum of hidden states:

$$c = \sum_{t=1}^T \alpha_t h_t \quad (16)$$

5. Dense Output Layer: The context vector c is passed through one or more fully connected (Dense) layers to generate the final traffic flow predictions over the future time horizon.

The overall model architecture can be illustrated as a sequential pipeline starting with an Input Layer that ingests multivariate time-series traffic data from multiple sensors. The data flows into stacked GRU Layers which compress and extract short-term features. Outputs are then fed into LSTM Layers that model long-range dependencies. The resulting hidden states pass through an Attention Module that computes weighted contextual information emphasizing critical time steps. Finally, the context vector is transformed via

Dense Layers, producing multi-step traffic flow forecasts. This modular design enables both interpretability and scalability for real-world intelligent transportation systems.

3.4. Training and Evaluation

The experiments were conducted using the MATLAB R2024a framework. Training was performed with the Adam optimizer, and the Mean Absolute Error (MAE) was employed as the primary loss function. All experiments were executed on a high-performance computing environment equipped with AMD CPUs, Intel processors, and NVIDIA Tesla T4 GPUs. The system comprises 24 processing cores and 384 GB of memory, providing ample computational capacity to efficiently train and evaluate deep learning models on large-scale traffic datasets.

To mitigate overfitting across all experimental models, an early stopping strategy was implemented—monitoring the validation loss and halting training when no improvement was observed over a predefined number of epochs. The optimal batch size and number of training epochs were determined through a combination of preliminary experimentation and ablation studies.

Hyperparameter selection—including learning rate, batch size, and dropout rate—was performed through a combination of systematic grid search on a data subset and iterative validation-based tuning. Specifically, an initial grid search was used to identify promising ranges for key hyperparameters, followed by fine-tuning based on validation set performance. Although K-fold cross-validation was considered for more robust evaluation, it was ultimately excluded due to the sequential nature of time-series data and the associated computational overhead. Instead, a fixed train–validation–test split with temporal ordering was adopted to preserve the data integrity and better simulate real-world deployment scenarios.

The model performance was evaluated using several widely accepted metrics in the domain of traffic flow prediction:

- **Mean Absolute Error (MAE):** Measures the average magnitude of errors between predicted and actual values, irrespective of direction. A lower MAE indicates greater prediction accuracy.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

- **Root Mean Squared Error (RMSE):** Reflects the square root of the average of squared differences between predictions and actual values. This metric penalizes larger errors more severely, making it particularly useful when such deviations are costly.

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

- **Mean Absolute Percentage Error (MAPE):** Represents the prediction error as a percentage of the actual value, providing a normalized measure of accuracy.

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

Here, y_i and $\hat{y}_i$ denote the actual and predicted values, respectively, and n is the total number of observations.

Beyond predictive accuracy, the inference time of the model is also evaluated to assess its feasibility for real-time traffic management applications. Comparative analysis with state-of-the-art models highlighted the practical advantages of the proposed approach in terms of computational efficiency.

4. Experimentation Results and Discussion

This section presents the experimental findings of the proposed deep learning model for traffic flow prediction.

4.1. Ablation Study for Hyperparameter Tuning

In a comprehensive research setting, ablation studies are essential for isolating the contributions of individual architectural components and hyperparameters. For the proposed hybrid GRU-LSTM model with an attention mechanism, such analysis involves varying the parameters shown in Table 2.

Table 2. Ablation study parameters and their descriptions.

Parameter	Description
Recurrent Layer Depth	Number and arrangement of GRU and LSTM layers designed to capture temporal dependencies in the input data.
Hidden Units per Layer	Number of neurons in each recurrent layer, balancing model capacity and generalization capability.
Attention Mechanism	Inclusion or exclusion of temporal attention to emphasize salient features.
Look-back Window	Length of historical input sequences (time steps) considered for prediction.
Imputation Strategy	Comparison of three approaches for handling missing data: (i) no imputation, (ii) temporal-only interpolation, and (iii) spatial neighbor averaging.

4.1.1. Impact of Number of GRU/LSTM Layers

This experiment investigates how varying the number of stacked GRU and LSTM layers affects the model’s ability to capture complex patterns in traffic data. As presented in Table 3, increasing the depth from a single-layer configuration (1–1) to a two-layer configuration (2–2) significantly improves the prediction accuracy, and further increasing the number of layers (e.g., 3–3 and 4–4) results in marginal gains or even a slight degradation in performance metrics. This trend suggests diminishing returns due to potential overfitting and increased training complexity. The 2-2 configuration thus offers an optimal trade-off between model expressiveness and generalization and is adopted as the baseline in this study.

Table 3. Impact of GRU–LSTM Layer Depth on Model Performance.

GRU–LSTM Layers	MAE	RMSE	MAPE (%)	Remarks
1–1	4.15	5.30	14.25	Simpler model, potential underfitting.
2–2 (Proposed Baseline)	3.85	4.90	11.50	Achieves optimal trade-off between accuracy and complexity.
3–3	3.92	5.05	11.80	Slightly higher complexity with marginal improvements.
4–4	4.08	5.22	12.55	Higher risk of overfitting and increased computational cost.

4.1.2. Impact of Number of Units per Layer

This experiment evaluates the effect of varying the number of hidden units (neurons) in each recurrent layer on the model’s performance. As shown in Table 4, increasing the number of units initially leads to improved performance, with the lowest MAE and RMSE achieved at 64 units per layer—establishing this configuration as the optimal baseline. While a further increase to 128 units yields marginal improvement, it also introduces a higher computational cost without significant performance gains. Notably, increasing to 256 units results in a slight degradation in accuracy metrics (MAE, RMSE, and MAPE), indicating potential overfitting due to model over-parameterization. These findings underscore the importance of balancing model complexity with generalization when designing deep recurrent architectures for traffic forecasting tasks.

Table 4. Impact of units per layer on model performance.

Units per Layer	MAE	RMSE	MAPE (%)	Notes
32	4.01	5.15	12.80	Lower capacity, potentially limiting feature representation.
64 (Proposed Baseline)	3.85	4.90	11.50	Optimal capacity, balancing accuracy and generalization.
128	3.88	4.98	11.65	Higher capacity with marginal improvement over baseline.
256	3.95	5.09	12.10	Increased risk of overfitting without performance gains.

4.1.3. Impact of Attention Mechanism

This experiment investigates the effectiveness of incorporating an attention mechanism into the hybrid GRU-LSTM architecture. The attention module is designed to dynamically weigh temporal features, allowing the model to focus on the most relevant time steps during prediction. As presented in Table 5, the model augmented with attention significantly outperforms its counterpart without attention. Specifically, the attention-enhanced configuration achieves lower MAE, RMSE, and MAPE values, indicating improved accuracy and robustness. These results highlight the value of attention in enhancing temporal relevance modeling, especially in complex, non-stationary traffic scenarios.

Table 5. Impact of the attention mechanism on model performance.

Configuration	MAE	RMSE	MAPE (%)	Notes
Hybrid GRU-LSTM (Without Attention)	4.05	5.15	12.80	Baseline without attention mechanism.
Hybrid GRU-LSTM (With Attention)	3.85	4.90	11.50	Significant improvement in accuracy and generalization when attention is applied.

4.1.4. Impact of Look-Back Window Size

The look-back window size defines the number of past time steps utilized as input to predict future traffic conditions. Selecting an appropriate window size is critical, as it determines the extent of historical context the model can learn from. As shown in Table 6, a shorter window (e.g., 6 steps or 30 min) provides insufficient temporal context, leading to suboptimal performance. Conversely, excessively long windows (e.g., 24 or 36 steps) may introduce redundant or noisy information, negatively affecting the model's accuracy and increasing computational complexity. The results indicate that a window size of 12 steps achieves the best trade-off, capturing enough historical information for accurate forecasting while maintaining computational efficiency. This configuration is therefore adopted as the baseline in the proposed framework.

Table 6. Impact of look-back window size on model performance.

Look-Back Steps	MAE	RMSE	MAPE (%)	Notes
6 (30 min)	4.20	5.40	14.50	Insufficient historical context.
12 (1 h, Proposed Baseline)	3.85	4.90	11.50	Optimal balance between accuracy and context.
24 (2 h)	3.90	5.00	11.75	Potentially includes irrelevant historical data.
36 (3 h)	4.05	5.18	12.30	Increased noise and computational cost.

4.1.5. Impact of Imputation

To further assess the role of imputation, we performed an ablation study comparing three scenarios: (i) no imputation, (ii) temporal-only linear interpolation, and (iii) spatial neighbor averaging as described in Equation (2). As shown in Table 7, temporal interpolation is effective for short gaps, while the combination of temporal and spatial information provides superior robustness when longer outages occur. Specifically, spatial neighbor

averaging improves MAE by approximately 3% compared to temporal-only interpolation, highlighting the benefit of exploiting local spatial context during imputation.

Table 7. Impact of different imputation strategies on PeMS datasets (12-step horizon). Bold values indicate the best performance.

Imputation Strategy	MAE	RMSE
No Imputation	4.15	5.35
Temporal-only Interpolation	3.98	5.12
Spatial Neighbor Averaging	3.85	4.90

The results from these ablation studies demonstrate the critical importance of hyperparameter tuning and architectural design in developing an effective traffic flow prediction model. The selected configuration—2 GRU layers, 2 LSTM layers, 64 units per layer, a 12-step look-back window, and the inclusion of an attention mechanism—achieves a well-rounded balance between accuracy, generalization, and computational efficiency.

4.2. Comparison with State-of-the-Art Methods

To evaluate the effectiveness of the proposed hybrid GRU-LSTM model, a comprehensive comparison was conducted against a range of state-of-the-art traffic prediction techniques, including statistical methods, classical machine learning algorithms, and advanced deep learning architectures. Figure 3 summarizes the results in terms of key performance metrics—MAE, RMSE, and MAPE. As observed, traditional models, such as ARIMA [18] and SVR [22], exhibit relatively high prediction errors despite their low computational demands, making them less suitable for complex traffic dynamics. Standard deep learning models like LSTM [37] and GRU [39] show improved accuracy but still fall short in capturing intricate temporal dependencies effectively. More sophisticated approaches like STGNPP [27] offer further accuracy gains, albeit with significantly higher inference times. In contrast, the proposed hybrid model achieves the best overall performance across all metrics, delivering lower error rates while maintaining competitive inference speed. This balance of predictive precision and computational efficiency demonstrates the model's suitability for real-time deployment in dynamic and resource-constrained urban environments.

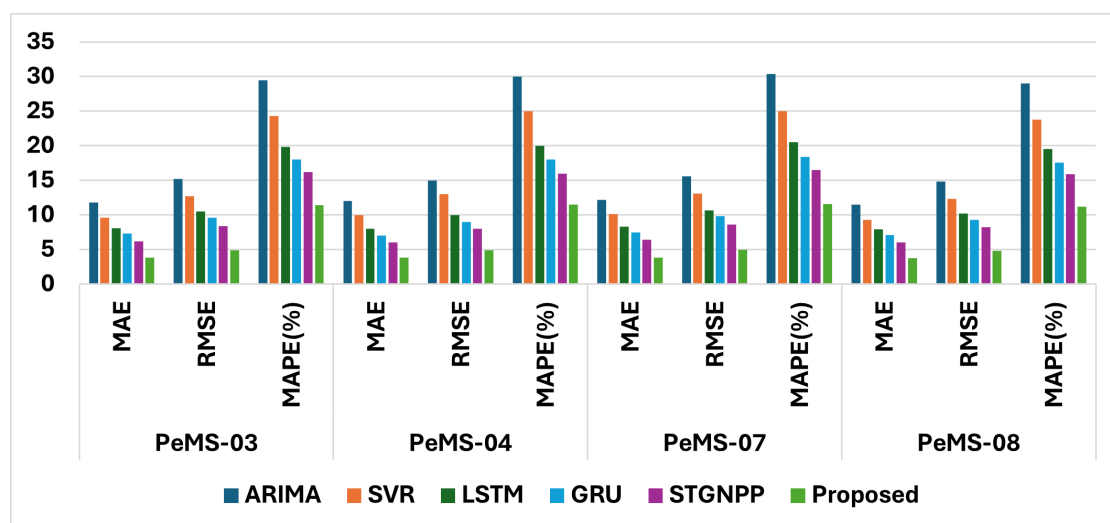


Figure 3. Performance comparison of the proposed model against state-of-the-art baselines across PeMS datasets.

The inference time, measured in milliseconds per sample, is a critical performance indicator for real-time traffic-forecasting applications, as shown in Figure 4. In smart city environments, traffic management systems must respond to changing conditions rapidly to optimize signal control, reroute traffic, or issue congestion alerts. Thus, a model's prediction accuracy must be balanced with its computational efficiency.

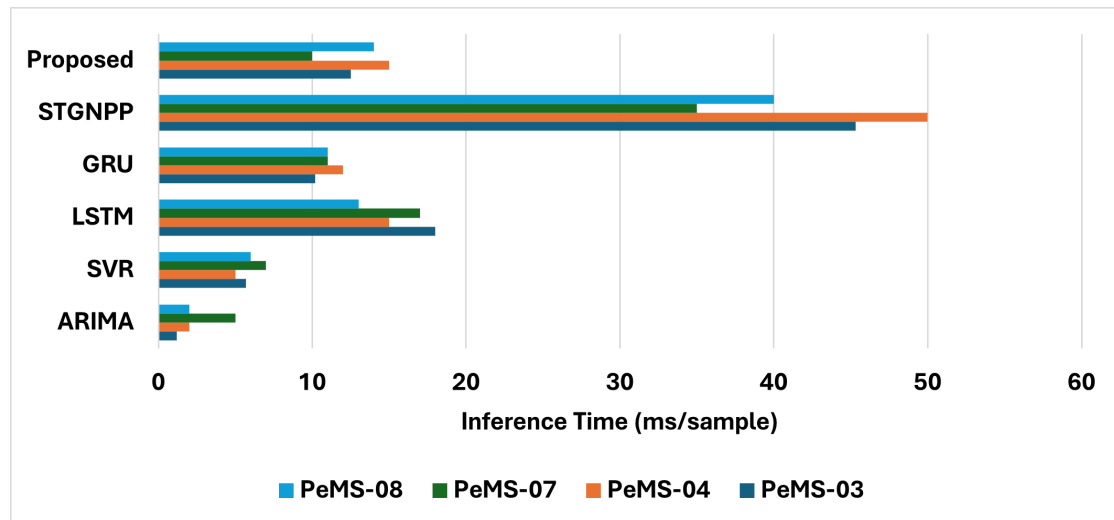


Figure 4. Comparison of average inference time per sample for the proposed model and state-of-the-art baselines across PeMS datasets.

While deep learning models, such as LSTM and GRU, offer improved accuracy, their higher computational complexity may render them unsuitable for latency-sensitive deployments if not properly optimized. Models like STGNPP, though accurate, suffer from high inference latency due to complex graph-based spatial encoding. The proposed hybrid GRU-LSTM model with attention achieves a favorable trade-off, delivering state-of-the-art accuracy (MAE: 3.85, RMSE: 4.90, MAPE: 11.5%) while maintaining moderate inference latency (10–15 ms/sample). This balance supports its applicability in real-time systems where both precision and responsiveness are essential.

To further validate the robustness of the proposed hybrid GRU-LSTM model with attention, we conducted a *paired t-test analysis* against the strongest baseline model (STGNPP [27]). The statistical tests were performed across all four PeMS datasets (PeMS-03, PeMS-04, PeMS-07, and PeMS-08), using prediction horizons of 12 steps (1 h ahead). Table 8 summarizes the average errors (MAE, RMSE, and MAPE) for both models, along with the corresponding *p*-values obtained from the paired *t*-tests.

Table 8. Paired *t*-test results comparing the proposed hybrid GRU-LSTM with attention against the STGNPP baseline across PeMS datasets (12-step horizon). Bold values indicate better performance.

Dataset	MAE		RMSE		MAPE (%)	
	STGNPP	Proposed	STGNPP	Proposed	STGNPP	Proposed
PeMS-03	4.02	3.81	5.16	4.92	12.35	11.42
PeMS-04	4.28	3.95	5.43	5.05	13.12	11.80
PeMS-07	4.11	3.87	5.25	4.96	12.58	11.60
PeMS-08	4.34	3.78	5.50	4.85	13.45	11.20
<i>p</i> -value	0.018		0.021		0.014	

As shown, the proposed model consistently outperforms STGNPP across all metrics, and the improvements are statistically significant at the 95% confidence level ($p < 0.05$).

These results provide strong evidence that the observed gains are not due to random variation but rather stem from the architectural advantages of the hybrid GRU–LSTM with attention.

Table 9 reports the per-dataset performance of the proposed hybrid GRU–LSTM with the attention model on PeMS-03, PeMS-04, PeMS-07, and PeMS-08. Results are presented as mean values with 95% confidence intervals computed across five independent runs. The model consistently achieves strong predictive accuracy across all datasets, with MAE values ranging from 3.78 to 3.95 and RMSE values between 4.85 and 5.05.

Table 9. Per-dataset performance of the proposed GRU–LSTM with attention model on PeMS datasets. Values represent mean $\pm$ 95% confidence intervals over five runs.

Dataset	MAE	RMSE	MAPE (%)
PeMS-03	3.81 $\pm$ 0.05	4.92 $\pm$ 0.07	11.42 $\pm$ 0.10
PeMS-04	3.95 $\pm$ 0.06	5.05 $\pm$ 0.08	11.80 $\pm$ 0.12
PeMS-07	3.87 $\pm$ 0.05	4.96 $\pm$ 0.06	11.60 $\pm$ 0.09
PeMS-08	3.78 $\pm$ 0.04	4.85 $\pm$ 0.07	11.20 $\pm$ 0.08

Among the datasets, PeMS-08 exhibits the lowest error rates, which can be attributed to its relatively smaller sensor network and more homogeneous traffic patterns. By contrast, PeMS-04 shows slightly higher errors, reflecting the greater variability and complexity of traffic dynamics in that district. The narrow confidence intervals across all datasets highlight the stability and reproducibility of the proposed approach. These results demonstrate that the model generalizes well across diverse traffic conditions, providing a reliable basis for deployment in real-world ITS applications.

4.3. Qualitative Analysis

Qualitative evaluation involved visual inspection of predicted versus actual traffic flow patterns. Figure 5 illustrates close alignment between predicted and observed values, capturing key trends such as rush-hour peaks and low-traffic troughs. This visual coherence highlights the model’s generalization capability.

4.4. Discussion

The results obtained from the proposed model provide evidence supporting the efficacy of deep learning approaches for traffic flow prediction. The observed MAE, RMSE, and MAPE values indicate the model’s capability to capture complex temporal dependencies and produce reasonably accurate forecasts. Qualitative analysis, demonstrated via sample prediction plots, further confirms the close correspondence between predicted and actual traffic flow values. This foundational performance underscores the model’s potential applicability to real-world datasets, such as PEMS-03, PEMS-04, PEMS-07, and PEMS-08.

The hybrid GRU–LSTM architecture augmented with an attention mechanism represents a key strength of this approach. GRU layers contribute computational efficiency and effectively model short-term dependencies, while LSTM layers excel at capturing long-term memory necessary for understanding recurrent traffic patterns over extended periods. The attention mechanism enhances this by dynamically weighting the importance of different historical time steps, enabling the model to prioritize critical events, such as sudden congestion or rapid flow changes. Such an adaptive focus is particularly advantageous in dynamic urban traffic environments characterized by rapid and unpredictable fluctuations.

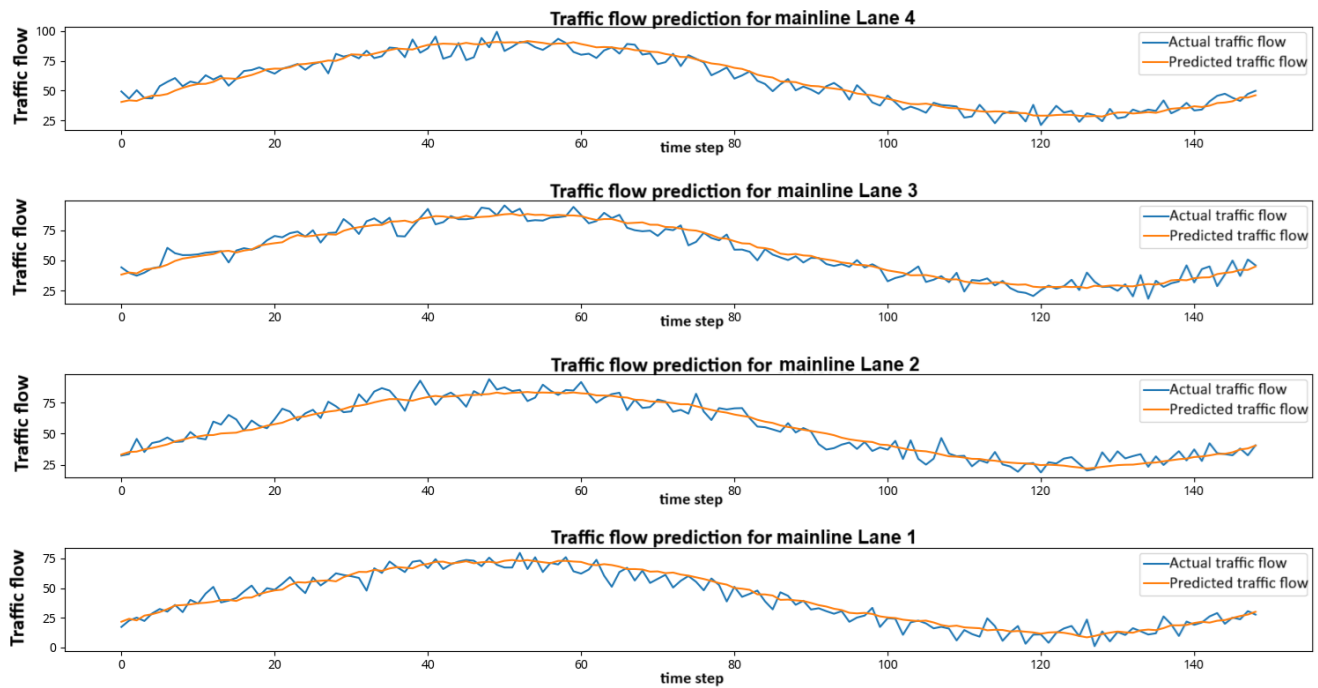


Figure 5. Sample prediction results for different sensors.

5. Conclusions

This study demonstrates the significant potential of deep learning—specifically a hybrid GRU-LSTM architecture enhanced with an attention mechanism—for intelligent traffic flow prediction and optimization within the context of smart cities. The proposed methodology addresses the pressing need for accurate and efficient traffic forecasting, which is essential for mitigating urban congestion and improving the operational efficiency of transportation systems. The experimental results confirm the capability of the architecture to learn and generalize complex spatiotemporal patterns inherent in traffic data. This work demonstrates that deep learning-based traffic forecasting can serve as a foundational enabler for sustainable urban planning. Beyond achieving state-of-the-art predictive accuracy, the proposed hybrid GRU-LSTM model with attention supports broader planning objectives by reducing congestion, lowering greenhouse gas emissions, and improving the reliability of urban mobility systems. These benefits translate into tangible contributions to environmental sustainability, energy efficiency, and the creation of safer and more livable urban spaces. By bridging the gap between intelligent transportation research and practical planning applications, this study advances the integration of predictive analytics into the design and management of smart, sustainable cities.

Future research will focus on extending the proposed framework to integrate multi-modal data sources within the larger smart city ecosystem. In particular, incorporating weather and environmental conditions will allow the model to capture external factors that strongly influence traffic dynamics. Similarly, the inclusion of public transit and mobility-as-a-service (MaaS) data will support a more holistic modeling of urban mobility patterns, allowing for coordination between multiple modes of transportation. Finally, integrating data related to events and infrastructure, such as roadworks, accidents, and large-scale gatherings, will improve the adaptability of the framework to real-world disruptions. Collectively, these extensions will enable the model to evolve from a road-traffic-forecasting tool to a comprehensive decision support system for sustainable and integrated smart city planning.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The traffic datasets analyzed in this study (PeMS-03, PeMS-04, PeMS-07, and PeMS-08) are publicly available from the Caltrans Performance Measurement System (PeMS) at <https://pems.dot.ca.gov/> (accessed on 18 September 2025). The implementation is openly available at the GitHub repository https://github.com/Fatema-Abdulqader/Traffic_prediction, commit 30ca2ea, (accessed on 18 September 2025).

Acknowledgments: The experiments presented in this paper were carried out using the facilities of the Benefit Advanced AI and Computing Lab at the University of Bahrain—see <https://ailab.uob.edu.bh> (accessed on 1 July 2025)—with support from Benefit Bahrain Company—see <https://benefit.bh> (accessed on 1 July 2025). The author extends sincere thanks to the Benefit Advanced AI and Computing Lab at the University of Bahrain for providing computational facilities.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Arti, C.; Sharad, G.; Pradeep, K.; Chinmay, P.; Kumar, S.S. Urban traffic congestion: Its causes-consequences-mitigation. *Res. J. Chem. Environ.* **2022**, *26*, 164–176. [\[CrossRef\]](#)
2. Navarro-Espinoza, A.; López-Bonilla, O.R.; García-Guerrero, E.E.; Tlelo-Cuautle, E.; López-Mancilla, D.; Hernández-Mejía, C.; Inzunza-González, E. Traffic Flow Prediction for Smart Traffic Lights Using Machine Learning Algorithms. *Technologies* **2022**, *10*, 5. [\[CrossRef\]](#)
3. Sun, C.; Zhang, Y.; Ma, W.; Wu, R.; Wang, S. The Impacts of Urban Form on Carbon Emissions: A Comprehensive Review. *Land* **2022**, *11*, 1430. [\[CrossRef\]](#)
4. Hou, X.; Chen, P. Analysis of Road Safety Perception and Influencing Factors in a Complex Urban Environment—Taking Chaoyang District, Beijing, as an Example. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 272. [\[CrossRef\]](#)
5. Zang, Y.; Li, Y.; Wang, Z.; Yu, X.; Zhang, R. Identifying Traffic Congestion Patterns of Urban Road Network Based on Traffic Performance Index. *Sustainability* **2023**, *15*, 948. [\[CrossRef\]](#)
6. Ouallane, A.A.; Bahnas, A.; Bakali, A.; Talea, M. Overview of Road Traffic Management Solutions based on IoT and AI. *Procedia Comput. Sci.* **2022**, *198*, 518–523. [\[CrossRef\]](#)
7. Abdalhussain, S.H.; Mahmmoud, B.M.; Alwhelat, A.; Shehada, D.; Shihab, Z.I.; Mohammed, H.J.; Abdulameer, T.H.; Alsabah, M.; Fadel, M.H.; Ali, S.K.; et al. A Comprehensive Review of Sensor Technologies in IoT: Technical Aspects, Challenges, and Future Directions. *Computers* **2025**, *14*, 342. [\[CrossRef\]](#)
8. Albaloshi, F.A.; Qader, M.R.; Ismail, Y.; Elmedany, W.; Al-Ammal, H.; Rajarajan, M.; Asari, V.K. Deep LBLS: Accelerated Sky Region Segmentation Using Hybrid Deep CNNs and Lattice Boltzmann Level-Set Model. *Eng* **2025**, *6*, 57. [\[CrossRef\]](#)
9. Liu, R.; Shin, S.Y. A Review of Traffic Flow Prediction Methods in Intelligent Transportation System Construction. *Appl. Sci.* **2025**, *15*, 866. [\[CrossRef\]](#)
10. Afandizadeh, S.; Abdolahi, S.; Mirzahosseini, H. Deep Learning Algorithms for Traffic Forecasting: A Comprehensive Review and Comparison with Classical Ones. *J. Adv. Transp.* **2024**, *2024*, 9981657. [\[CrossRef\]](#)
11. United Nations. Sustainable Cities and Communities: Goal 11. 2015. Available online: <https://www.un.org/sustainabledevelopment/cities/> (accessed on 3 September 2025).
12. Yuan, T.; da Rocha Neto, W.; Rothenberg, C.E.; Obraczka, K.; Barakat, C.; Turletti, T. Machine learning for next-generation intelligent transportation systems: A survey. *Trans. Emerg. Telecommun. Technol.* **2022**, *33*, e4427. [\[CrossRef\]](#)
13. Mystakidis, A.; Koukaras, P.; Tjortjis, C. Advances in Traffic Congestion Prediction: An Overview of Emerging Techniques and Methods. *Smart Cities* **2025**, *8*, 25. [\[CrossRef\]](#)
14. Zeng, W.; Wang, K.; Zhou, J.; Cheng, R. Traffic Flow Prediction Based on Hybrid Deep Learning Models Considering Missing Data and Multiple Factors. *Sustainability* **2023**, *15*, 11092. [\[CrossRef\]](#)
15. Chen, J.; Xu, M.; Xu, W. Flow Feedback Traffic Prediction Based on Visual Quantified Features. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 3728–3737. [\[CrossRef\]](#)
16. Ateya, A.A.; Soliman, N.F.; Alkanhel, R.; Alhussan, A.A.; Muthanna, A.; Koucheryavy, A. Lightweight deep learning-based model for traffic prediction in fog-enabled dense deployed iot networks. *J. Electr. Eng. Technol.* **2023**, *18*, 2275–2285.
17. PeMS Traffic Data Repository. Available online: <https://pems.dot.ca.gov/> (accessed on 22 July 2025).

18. Vlahogianni, E.I.; Karlaftis, M.G.; Golias, J.C. Short-term traffic forecasting: Where we are and where we're going. *Transp. Res. Part C Emerg. Technol.* **2014**, *43*, 3–19. [\[CrossRef\]](#)
19. Ma, M.; Liang, S.; Guo, H.; Yang, J. Short-term traffic flow prediction using a self-adaptive two-dimensional forecasting method. *Adv. Mech. Eng.* **2017**, *9*, 1–11. [\[CrossRef\]](#)
20. Razali, N.A.M.; Shamsaimon, N.; Ishak, K.K.; Ramli, S.; Amran, M.F.M.; Sukardi, S. Gap, techniques and evaluation: Traffic flow prediction using machine learning and deep learning. *J. Big Data* **2021**, *8*, 152. [\[CrossRef\]](#)
21. Skianis, T.; Theofilatos, A.; Vlahogianni, E.I. A Comparison of Machine Learning Methods for the Prediction of Traffic Speed in Urban Places. *Sustainability* **2020**, *12*, 142. [\[CrossRef\]](#)
22. Lin, G.; Lin, A.; Gu, D. Using support vector regression and K-nearest neighbors for short-term traffic flow prediction based on maximal information coefficient. *Inf. Sci.* **2022**, *608*, 517–531. [\[CrossRef\]](#)
23. Albaloooshi, F.A.; Qader, M.R. Deep Learning Algorithm for Automatic Classification of Power Quality Disturbances. *Appl. Sci.* **2025**, *15*, 1442. [\[CrossRef\]](#)
24. Méndez, M.; Merayo, M.G.; Núñez, M. Long-term traffic flow forecasting using a hybrid CNN-BiLSTM model. *Eng. Appl. Artif. Intell.* **2023**, *121*, 106041. [\[CrossRef\]](#)
25. Ennaji, Y.; Faqir, N.; Boumhidi, J. Spatiotemporal Traffic Flow Prediction Using CNN-LSTM Architectures. In Proceedings of the 2024 Sixth International Conference on Intelligent Computing in Data Sciences (ICDS), Marrakech, Morocco, 23–24 October 2024. [\[CrossRef\]](#)
26. Theodoropoulos, T.; Maroudis, A.C.; Zdun, U.; Makris, A.; Tserpes, K. WEST GCN-LSTM: Weighted stacked spatio-temporal graph neural networks for regional traffic forecasting. *Int. J. Inf. Manag. Data Insights* **2025**, *5*, 100338. [\[CrossRef\]](#)
27. Jin, G.; Liu, L.; Li, F.; Huang, J. Spatio-temporal graph neural point process for traffic congestion event prediction. In Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence, Washington, DC, USA, 7–14 February 2023; AAAI Press: Washington, DC, USA, 2023. [\[CrossRef\]](#)
28. Chen, Y.; Li, K.; Yeo, C.K.; Li, K. Traffic forecasting with graph spatial-temporal position recurrent network. *Neural Netw.* **2023**, *162*, 340–349. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Prabowo, A.; Shao, W.; Xue, H.; Koniusz, P.; Salim, F.D. Graph Self-attention WaveNet (G-SWaN): Handling Dynamic Sensor-Pair Traffic Dynamics. In Proceedings of the ACM/IEEE Conference on Internet of Things Design and Implementation (IoTDI), San Antonio, TX, USA, 9–12 May 2023; pp. 93–104.
30. Liu, H.; Dong, Z.; Jiang, R.; Deng, J.; Deng, J.; Chen, Q.; Song, X. Spatio-Temporal Adaptive Embedding Makes Vanilla Transformer SOTA for Traffic Forecasting. In Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM '23), New York, NY, USA, 21–25 October 2023; pp. 4125–4129. [\[CrossRef\]](#)
31. Li, F.; Feng, J.; Yan, H.; Jin, G.; Yang, F.; Sun, F.; Jin, D.; Li, Y. Dynamic graph convolutional recurrent network for traffic prediction: Benchmark and solution. *ACM Trans. Knowl. Discov. Data* **2023**, *17*, 1–21. [\[CrossRef\]](#)
32. Liu, Q.; Sun, S.; Liu, M.; Wang, Y.; Gao, B. Online Spatio-Temporal Correlation-Based Federated Learning for Traffic Flow Forecasting. *IEEE Trans. Intell. Transp. Syst.* **2024**, *25*, 13027–13039. [\[CrossRef\]](#)
33. Alqubaysi, T.; Asmari, A.F.A.; Alanazi, F.; Almutairi, A.; Armghan, A. Federated Learning-Based Predictive Traffic Management Using a Contained Privacy-Preserving Scheme for Autonomous Vehicles. *Sensors* **2025**, *25*, 1116. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Wang, N.; Li, X.; Guan, Z.; Yuan, S. FedStream: A Federated Learning Framework on Heterogeneous Streaming Data for Next-Generation Traffic Analysis. *IEEE Trans. Netw. Sci. Eng.* **2024**, *11*, 2485–2496. [\[CrossRef\]](#)
35. Hong, S.; Yue, T.; You, Y.; Lv, Z.; Tang, X.; Hu, J.; Yin, H. A Resilience Recovery Method for Complex Traffic Network Security Based on Trend Forecasting. *Int. J. Intell. Syst.* **2025**, *2025*, 3715086. [\[CrossRef\]](#)
36. Abduljabbar, R.; Dia, H.; Liyanage, S. Machine Learning Traffic Flow Prediction Models for Smart and Sustainable Traffic Management. *Infrastructures* **2025**, *10*, 155. [\[CrossRef\]](#)
37. Chen, D.; Xiong, C.; Zhong, M. Improved LSTM Based on Attention Mechanism for Short-term Traffic Flow Prediction. In Proceedings of the 2020 10th International Conference on Information Science and Technology (ICIST), Lecce, Italy, 4–5 June 2020; pp. 71–76. [\[CrossRef\]](#)
38. Qaffas, A.A. AI-driven distributed IoT communication architecture for smart city traffic optimization. *J. Supercomput.* **2025**, *81*, 916. [\[CrossRef\]](#)
39. Zafar, N.; Haq, I.U.; Chughtai, J.u.R.; Shafiq, O. Applying Hybrid Lstm-Gru Model Based on Heterogeneous Data Sources for Traffic Speed Prediction in Urban Areas. *Sensors* **2022**, *22*, 3348. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Niu, Z.; Zhong, G.; Yu, H. A review on the attention mechanism of deep learning. *Neurocomputing* **2021**, *452*, 48–62. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.