ILKOM Jurnal Ilmiah Vol. 17, No. 3, December 2025, pp.227-240
Accredited 2nd by RISTEKBRIN No. 10/C/C3/DT.05.00/2025; E-ISSN 2548-7779 | P-ISSN 2087-1716



# Urban Traffic Volume Prediction using LSTM and Bi-LSTM: Performance Evaluation on the Metro Interstate Dataset

**Andri Pranolo [a,1,*]; Shoffan Saifullah [b,2]; Agung Bella Utama Putra [c,3]; Rafał Dreżewski [b,4]; Aji Prasetya Wibawa [c,5];**

[a] *Universitas Ahmad Dahlan, Jl. Kapas No.9, Semaki, Kec. Umbulharjo, Kota Yogyakarta, Daerah Istimewa Yogyakarta 55166, Indonesia*
[b] *AGH University of Krakow, 30-059 Krakow, Poland*
[c] *Universitas Negeri Malang, Jl. Cakrawala No.5, Sumbersari, Kec. Lowokwaru, Kota Malang, Jawa Timur 65145, Indonesia*
[1] *andri.pranolo@tif.uad.ac.id;* [2] *shoffans@upnyk.ac.id;* [3] *agungbpu02@gmail.com;* [4] *drezew@agh.edu.pl;* [5] *aji.prasetya.ft@um.ac.id*
** Corresponding author*

**Article history:** Received August 15, 2025; Revised October 3, 2025; Accepted October 16, 2025; Available online October 29, 2025.

**Abstract**

Urban traffic forecasting underpins the mitigation of congestion, enhancement of road safety, and reduction of emissions in intelligent transportation systems. We benchmark Long Short-Term Memory (LSTM) and Bidirectional LSTM (Bi-LSTM) models on the Metro Interstate Traffic Volume dataset under an identical preprocessing and training pipeline for a fair comparison. Using a 24-hour multivariate input window (temperature, rainfall, snowfall, cloud cover), LSTM delivers the best overall balance of accuracy and efficiency on the full test sequence (RMSE = 0.196, MAPE = 2.36%, $R^2$ = 0.480; 7,344 s training). Bi-LSTM achieves competitive short-window accuracy but underperforms on the full sequence (RMSE = 0.231, MAPE = 2.92%, $R^2$ = 0.280; 12,672 s training). We attribute the Bi-LSTM gap to prediction "flattening" over long horizons, i.e., over-smoothed peaks from bidirectional averaging, despite its slightly stronger short-segment fit. Compared with prior RNN/GRU/CNN baselines on the same data, LSTM improves variance explanation while remaining deployable for near-real-time use. We also examine seasonality (daily/weekly cycles), weather effects, and data imbalance (peak versus off-peak) as factors that shape model error. These results support LSTM as a practical default for city-scale forecasting and motivate future work with attention/Transformer encoders and richer exogenous signals (incidents, events). The findings inform policy by enabling proactive traffic management that can reduce delays, emissions, and crash risk through earlier, data-driven interventions.

**Keywords:** LSTM; Bi-LSTM; Deep Learning; Time Series Forecasting; Urban Traffic.

## Introduction

Accurate and timely traffic volume forecasting is central to intelligent transportation systems (ITS), offering the potential to optimize traffic flow, reduce urban congestion, enhance safety, and support sustainable mobility [1], [2]. As urban populations grow and road networks become increasingly saturated, predictive models are essential for preemptively responding to traffic fluctuations caused by temporal, meteorological, and external events [3]. Despite its importance, traffic prediction remains a complex time-series problem characterized by nonlinear dependencies, high variability, and long-term temporal patterns [4], [5]. Accurate short-horizon forecasts enable proactive ramp metering, signal retiming, and traveler information, which in turn reduce stop-and-go traffic, lower fuel consumption and $CO_2$, and improve safety through smoother flows and fewer shockwaves [6]. Embedding robust forecasting in traffic control centers also supports equity-minded policies (e.g., prioritizing transit corridors) and adaptive incident response that protects vulnerable road users [7].

Traditional statistical approaches such as ARIMA and Support Vector Regression (SVR) have been widely used in early traffic forecasting research [8], [9]. However, these models rely on linear assumptions and struggle to capture nonlinear temporal dynamics, especially under highly volatile conditions. To address these limitations, deep learning models have emerged as a dominant paradigm for traffic forecasting due to their ability to learn latent representations and temporal correlations without handcrafted features [10], [11].

Among deep learning architectures, Recurrent Neural Networks (RNNs) have been extensively applied to sequential prediction problems. Nevertheless, RNNs suffer from the vanishing gradient problem, which hinders their ability to learn long-term dependencies [12]. Gated architectures such as Long Short-Term Memory (LSTM) [13] and

http://dx.doi.org/10.33096/ilkom.v17i3.3001.227-240

Gated Recurrent Units (GRU) [14] were developed to overcome this issue by incorporating memory gates that control the flow of information across time steps. Furthermore, bidirectional models like Bi-LSTM extend these capabilities by capturing context in past and future directions, providing enhanced understanding of sequence patterns [15], [16].

In our prior work [17], we conducted a comparative evaluation of RNN, GRU, and Convolutional Neural Network (CNN) models on the Metro Interstate Traffic Volume dataset. The GRU model achieved the highest forecasting accuracy (MAPE = 2.105%, RMSE = 0.198), while CNN offered the lowest computational time (853 seconds) but sacrificed long-term dependency modeling. These findings highlighted a trade-off between prediction accuracy and computational feasibility. However, our earlier study did not include LSTM or Bi-LSTM models—despite their proven superiority in other time-series domains such as energy load forecasting [18], precipitation intensity prediction [19], and traffic flow estimation using hybrid transformer-LSTM frameworks [20], [21].

Recent literature also emphasizes the potential of LSTM-based models in traffic applications. For instance, Abbasimehr et al. [22] combined LSTM with attention mechanisms to improve long-sequence forecasting, while Ranjan et al. [23] demonstrated the effectiveness of Bi-LSTM in capturing symmetric congestion patterns. Moreover, hybrid models that combine CNN and Bi-LSTM have shown promise in preserving spatial-temporal dependencies, but often incur high computational costs [24], [25].

Despite these advancements, a gap remains in systematically evaluating LSTM and Bi-LSTM architectures within a consistent framework using standardized datasets, particularly when benchmarked against earlier models such as RNN, GRU, and CNN under identical training and preprocessing conditions. Furthermore, few studies consider the practical implications of execution time, which is crucial for real-time deployment in traffic management systems. In what follows, we benchmark LSTM and Bi-LSTM under a controlled pipeline, probe when and why their behaviors diverge (short-window vs full-sequence), and discuss operational implications for urban traffic management. Conducting a benchmarking study between LSTM and Bi-LSTM models is essential to understand how directional sequence processing influences forecasting accuracy and generalization in traffic prediction tasks. By comparing their performance across multiple evaluation metrics, this research aims to reaffirm the motivation for selecting the most suitable recurrent architecture that balances temporal dependency learning, computational efficiency, and robustness against noise in real-world traffic datasets.

To address these limitations, this study introduces and benchmarks two advanced sequential models—LSTM and Bi-LSTM—on the Metro Interstate dataset, building directly on our prior work. The key contributions of this paper are:

- We design LSTM and Bi-LSTM models with optimized hyperparameter settings and enhanced architectures tailored for traffic volume prediction.
- We conduct a comprehensive evaluation of prediction accuracy (MAPE, RMSE, $R^2$) and computational cost (execution time in seconds).
- We benchmark the LSTM and Bi-LSTM results against prior models (RNN, GRU, CNN) using identical data and training setups for fair comparison.
- We provide practical insights into the trade-offs between accuracy, complexity, and real-time feasibility in urban traffic forecasting applications.
- Beyond reporting error metrics, we (i) systematically contrast *full-sequence* vs *short-window* evaluation to reveal a rarely discussed bidirectional "flattening effect" over long horizons, and (ii) quantify accuracy–time trade-offs relevant for real-time deployment—an axis often omitted in traffic DL benchmarks.

The rest of this paper is organized as follows: Section 2 reviews related works on deep learning for traffic prediction. Section 3 describes the dataset, preprocessing pipeline, and model architectures. Section 4 presents and analyzes the experimental results. Finally, Section 5 concludes the study and outlines potential future research directions.

***Related Works***

Traffic volume prediction has evolved from conventional statistical models to advanced deep learning techniques as researchers aim to address the growing complexity of urban traffic patterns. Early forecasting methods often relied on models such as the Autoregressive Integrated Moving Average (ARIMA), Support Vector Regression (SVR), and Generalized Space-Time Autoregressive (GSTAR) models, which were effective for linear and stationary data [26]. However, their predictive performance degrades when faced with nonlinear, time-variant, and multivariate traffic patterns common in real-world urban systems [27].

The introduction of deep learning has significantly improved the modeling of traffic dynamics due to its capacity to learn nonlinear dependencies from large-scale sequential data [28], [29]. Recurrent Neural Networks (RNNs) are among the earliest neural architectures applied to traffic prediction tasks. Although RNNs can model time-series data, they suffer from the vanishing gradient problem, limiting their ability to learn long-range dependencies [12].

To overcome this limitation, advanced gated architectures such as Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs) have been developed. LSTM networks incorporate memory cells and gating mechanisms to retain information over extended sequences, making them suitable for long-term forecasting [13]. GRUs offer a simplified variant with comparable performance and lower computational demands [30]. Both architectures have been widely adopted in applications ranging from traffic flow estimation to energy demand forecasting and rainfall prediction [28], [31].

Bidirectional LSTM (Bi-LSTM) networks further enhance sequence modeling by processing data in both forward and backward directions. This bidirectional structure allows the model to capture symmetric temporal dependencies, which can be beneficial in contexts like urban traffic where peak patterns and recurrent flows are observable [8], [9]. Studies such as Ranjan et al. [23] and He et al. [28] have shown that Bi-LSTM can outperform unidirectional LSTM in congestion prediction tasks. However, these benefits often come with increased computational cost, which limits their deployment in real-time systems.

In parallel, Convolutional Neural Networks (CNNs) have been applied to traffic time series prediction due to their strength in extracting local temporal features through convolutional operations. CNNs are efficient in computation and exhibit strong performance when short-term patterns dominate the dataset [10], [32], [33]. However, their inability to retain long-term dependencies hinders their generalization in forecasting applications requiring memory depth. Some researchers have proposed CNN-GRU or CNN-LSTM hybrid models to integrate local feature extraction with sequential memory, yielding promising results but increasing architectural complexity [34].

Our previous study [17] compared RNN, GRU, and CNN models on the Metro Interstate Traffic Volume dataset, revealing that GRU achieved the highest prediction accuracy, while CNN offered the fastest execution. However, neither LSTM nor Bi-LSTM models were included in that benchmark. Other related works, such as those by Abbasimehr and Paki [22], introduced attention mechanisms to enhance LSTM performance, while hybrid Bi-LSTM-Transformer models [35] have recently emerged to fuse deep memory and attention components for high-resolution traffic data modeling [30].

Despite these advancements, systematic evaluation of LSTM and Bi-LSTM architectures—especially with regard to their accuracy-efficiency trade-offs under standardized experimental settings—remains limited. In particular, the lack of comparative studies that include computational time as a critical metric restricts insights into their real-world applicability for time-sensitive traffic management systems.

This study aims to address this gap by introducing a side-by-side evaluation of LSTM and Bi-LSTM architectures in terms of accuracy (MAPE, RMSE, $R^2$) and computational performance. Furthermore, we benchmark their results against previously established models (RNN, GRU, CNN) using the same dataset and preprocessing pipeline, enabling a fair and transparent assessment of their strengths and limitations.

## Method

### *A. Dataset Description*

This study utilizes the Metro Interstate Traffic Volume dataset, obtained from the UCI Machine Learning Repository [36]. The dataset consists of 48,204 hourly records collected from 2012 to 2018 on westbound Interstate 94 (I-94) near Minneapolis–St. Paul, Minnesota, specifically from the Minnesota Department of Transportation ATR Station 301. The dataset is a multivariate time series commonly used for regression-based traffic forecasting because of its temporal continuity and consistent hourly cadence. The target variable is traffic_volume (vehicles/hour), and the dataset contains no missing values, which simplifies sequence modeling. The variables, their roles, and data types are catalogued in **Table 1**, while the global dataset characteristics are summarized in **Table 2**.

**Table 1.** Variable Summary of the Metro Interstate Traffic Volume Dataset

| Variable | Type | Description |
|---|---|---|
| holiday | Categorical | U.S. and regional holidays including the Minnesota State Fair |
| temp | Continuous | Average hourly temperature (Kelvin) |
| rain_1h | Continuous | Rainfall in the past hour (mm) |

| Variable | Type | Description |
|---|---|---|
| snow_1h | Continuous | Snowfall in the past hour (mm) |
| clouds_all | Integer | Cloud cover percentage (%) |
| weather_main | Categorical | General weather category (e.g., Clear, Rain, Snow) |
| weather_description | Categorical | Detailed weather descriptor (e.g., scattered clouds) |
| date_time | DateTime | Timestamp of data collection (hourly CST) |
| traffic_volume | Integer | Hourly vehicle count (target variable) |

**Table 2.** Dataset Characteristics Overview.

| Attribute | Details |
|---|---|
| Instances | 48,204 |
| Temporal Coverage | 2012–2018 (6 years) |
| Prediction Task | Regression (traffic volume) |
| Input Feature Types | Integer, Real, Categorical |
| Sampling Frequency | Hourly |
| Missing Values | None |
| Target Variable | traffic_volume (vehicles/hour) |
| Application Context | Urban traffic forecasting with environmental factors |

As shown in **Table 1**, the predictors include meteorological covariates (temperature, rain_1h, snow_1h, clouds_all) and context features (holiday, weather_main, weather_description, date_time). This mix of exogenous and endogenous information supports modeling of weather- and calendar-driven demand shifts. The scope, sampling frequency, and lack of missingness in the full corpus are detailed in **Table 2**, confirming suitability for LSTM/Bi-LSTM sequence modeling without imputation.

To build intuition, **Figure 1** visualizes a 61-hour sample across five features, illustrating typical variability (temperature, clouds, traffic volume), sparsity (rain_1h), and zero variation for snow_1h within this subset. Descriptive statistics for the same sample are provided in Table 3, which motivates excluding any zero-variance features on the training split during preprocessing.

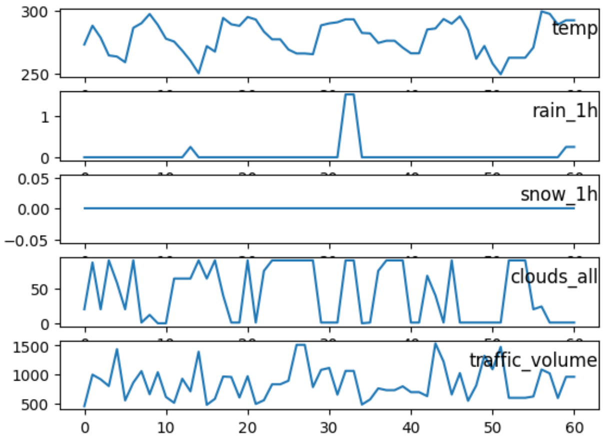



**Figure 1.** Example time-series visualization over 61 hourly records (temp, rain_1h, snow_1h, clouds_all, traffic_volume)

**Table 3.** Descriptive statistics for key features in the Metro Interstate Traffic Volume dataset.

| Statistic | temp | rain_1h | snow_1h | clouds_all | traffic_volume |
|---|---|---|---|---|---|
| Count | 61 | 61 | 61 | 61 | 61 |
| Mean | 278.32 | 0.0621 | 0 | 43.21 | 865.44 |
| Std. Deviation | 13.09 | 0.2761 | 0 | 40.17 | 291.03 |
| Minimum | 249.36 | 0 | 0 | 0 | 455 |

| Statistic | temp | rain_1h | snow_1h | clouds_all | traffic_volume |
|---|---|---|---|---|---|
| 25th Percentile | 266.08 | 0 | 0 | 1 | 615 |
| Median (50%) | 277.72 | 0 | 0 | 40 | 812 |
| 75th Percentile | 289.69 | 0 | 0 | 90 | 1041 |
| Maximum | 299.49 | 1.52 | 0 | 90 | 1538 |

### *B. Data Preprocessing*

To prepare the Metro Interstate Traffic Volume dataset for sequential modeling, a structured preprocessing pipeline was applied, ensuring data consistency, normalization, and suitability for input into LSTM-based architectures.

First, non-numeric and categorical fields are not directly relevant to model learning [37], namely holiday, weather_main, weather_description, and date_time were excluded. These variables, although informative, introduce high-cardinality categorical encoding or redundancy when used in conjunction with continuous weather features. The retained variables for modeling included temp, rain_1h, snow_1h, clouds_all, and the target variable traffic_volume, which together provide sufficient environmental and contextual representation for time-series learning.

Next, Min-Max normalization [38] was applied to scale all continuous features to a [0, 1] range. This technique preserves the original data distribution while ensuring that no single feature dominates the learning process due to scale differences. The normalized value $x'$ for a given input $x$ is computed as:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

This step improves gradient stability during training and is particularly important in recurrent architectures, where activation functions like tanh and sigmoid are sensitive to input magnitudes.

To convert the raw time-series data into a supervised learning format, a sliding window approach was used. Each training sample consists of a 24-hour input window (i.e., 24 consecutive timesteps of 4 features) to predict the traffic volume at the next hour. This look-back period is sufficient to capture daily temporal patterns such as rush hour peaks and off-peak lows.

The dataset was split chronologically into training (80%), validation (10%), and testing (10%) sets. This sequential split preserves the temporal structure of the data, preventing information leakage from future observations and aligning with best practices in time-series forecasting. The validation set was used for monitoring overfitting, while the test set provided a final evaluation on unseen data.

Through this preprocessing pipeline, the data was transformed into a clean, normalized, and temporally consistent format suitable for input into both unidirectional and bidirectional LSTM models. The emphasis on maintaining time integrity and removing irrelevant variance ensures that the models learn from actual sequential dependencies without bias or leakage.

### *C. Model Architecture*

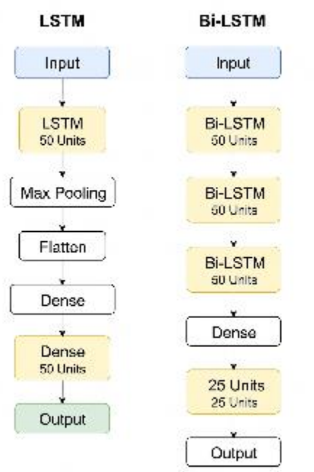



**Figure 2.** Model architectures for (a) LSTM and (b) Bi-LSTM used in this study. Each model consumes a 24-hour, four-feature sequence and predicts the next hour's volume. Layer sizes and activations are indicated.

This study investigates the effectiveness of two deep learning models—Long Short-Term Memory (LSTM) and Bidirectional Long Short-Term Memory (Bi-LSTM)—in forecasting hourly traffic volume. Both architectures (**Figure 2**) were selected due to their ability to capture temporal dependencies in sequential data, with LSTM offering robust unidirectional memory handling and Bi-LSTM extending this capacity through bidirectional context learning.

***LSTM Architecture***

The LSTM model (**Figure 3**)is designed to model temporal dependencies in a unidirectional sequence, capturing the influence of past traffic and environmental patterns on future traffic volume [39]. The architecture begins with an LSTM layer containing 50 units and tanh activation, followed by a max-pooling operation and a flattening layer to reduce the dimensionality and facilitate transition to fully connected layers [11]. A dense layer with 50 units is used for intermediate representation learning, followed by a final dense layer with a single neuron for regression output. This structure enables the model to balance memory depth and computational efficiency, making it suitable for real-time or near-real-time deployment.

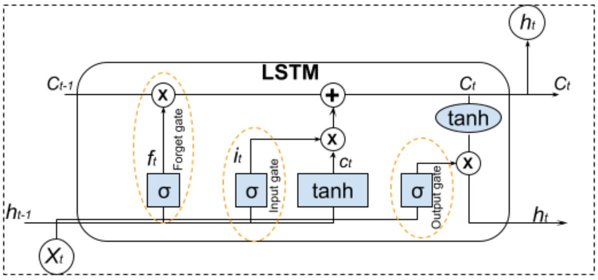



**Figure 3.** The detailed architecture of the proposed LSTM.

***Bi-LSTM Architecture***

To further explore the value of bidirectional temporal information, the Bi-LSTM model (**Figure 4**) processes sequences in both forward and backward directions, offering a more comprehensive view of contextual dependencies [18]. The model stacks three bidirectional LSTM layers, each with 50 units, to learn hierarchical temporal abstractions. This is followed by two fully connected dense layers with 25 and 1 units, respectively. While Bi-LSTM architectures are more computationally intensive, they are particularly effective in capturing symmetric traffic behavior, such as daily rush hour cycles that occur in both morning and evening periods.

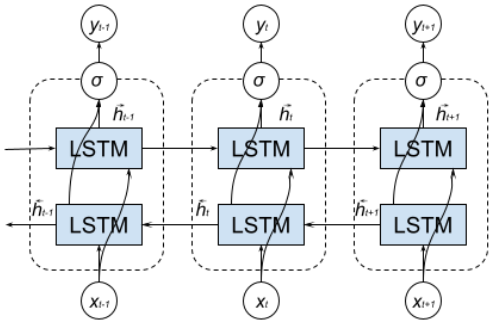



**Figure 4.** The detailed architecture of the proposed Bi-LSTM.

***Input and Output Structure***

Both models receive input in the form of 3D tensors structured as $(samples, timesteps, features) = (n, 24, 4)$, where each sample represents 24 consecutive hourly measurements of four features $(temp, rain\_1h, snow\_1h, clouds\_all)$. The output is a single scalar value representing the predicted traffic volume at the next hour.

To ensure a fair comparison, both models were trained using the same optimizer, loss function, and batch size, as described in Section 3.4. The difference in design complexity and directionality enables the direct evaluation of each architecture's performance in terms of both predictive accuracy and computational cost.

### *D. Training Configuration*

Both the LSTM and Bi-LSTM models were trained under the same experimental conditions to ensure a fair and consistent comparison. The models were implemented using TensorFlow and Keras, and trained on a standard GPU-enabled workstation.

The input sequences were prepared using a fixed-size sliding window of 24 hours, with each sample comprising four normalized features: temperature, rainfall, snowfall, and cloud coverage. The output was the traffic volume in the hour immediately following the input window.

The Adam optimizer was employed for gradient-based optimization due to its adaptive learning rate and efficient convergence behavior, particularly in recurrent neural networks. The mean squared error (MSE) was used as the loss function, given the continuous nature of the target variable and its sensitivity to significant prediction errors. The activation function used in all recurrent layers was tanh, which is well-suited for modeling both positive and negative temporal trends.

Training was conducted over 100 epochs with an early stopping mechanism based on validation loss to prevent overfitting. The batch size was set to 64, striking a balance between convergence stability and computational efficiency. **Table 4** summarizes the key training parameters for both models.

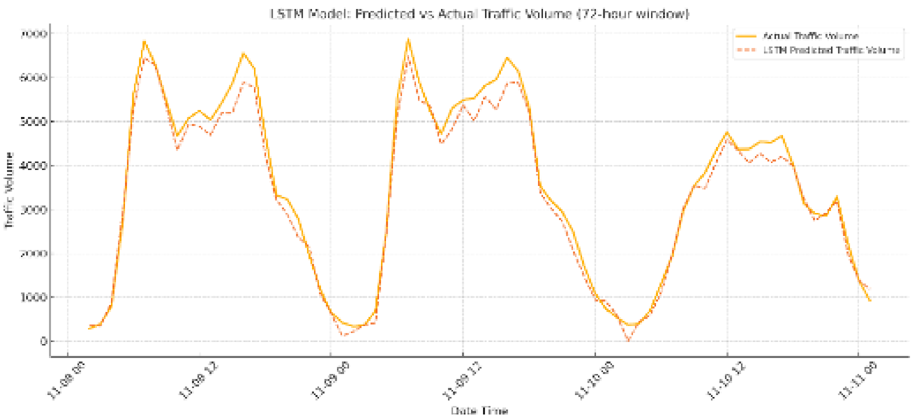
**Table 4.** Training Configuration for LSTM and Bi-LSTM Models.

| Parameter | LSTM | Bi-LSTM |
|---|---|---|
| Epochs | 100 | 100 |
| Batch Size | 64 | 64 |
| Optimizer | Adam | Adam |
| Loss Function | Mean Squared Error (MSE) | Mean Squared Error (MSE) |
| Activation Function | tanh | tanh |
| Input Shape | (24, 4) | (24, 4) |
| Additional Layers | MaxPooling, Flatten, Dense (50,1) | 3 Bi-LSTM layers, Dense (25,1) |
| Early Stopping | Yes (based on validation loss) | Yes (based on validation loss) |

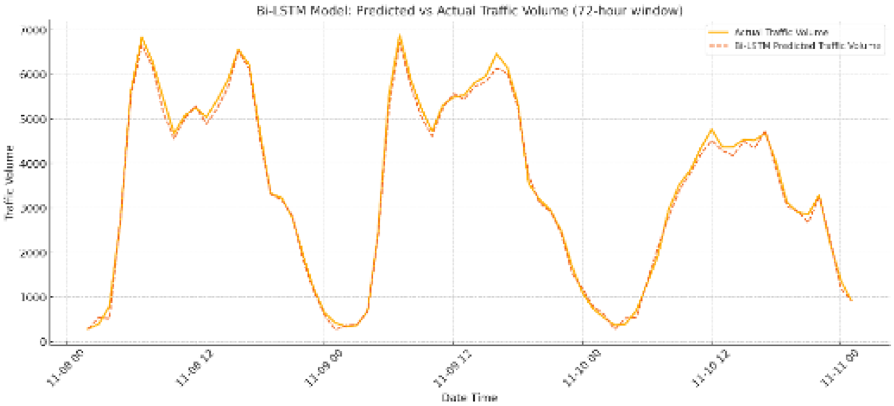
This unified training setup ensures that performance differences between the LSTM and Bi-LSTM models arise from their architectural capacities rather than external optimization factors. In the next section, we evaluate both models using standard regression metrics to assess forecasting accuracy and computational efficiency.

## Results and Discussion

To assess the forecasting performance of the proposed LSTM and Bi-LSTM models, multiple regression metrics were used: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and the Coefficient of Determination ($R^2$). This section provides a detailed analysis of the results, beginning with individual model evaluations and proceeding to a comparative interpretation of the findings.




(a)




(b)

**Figure 5.** Predicted vs actual traffic volume over a 72-hour window. (a) LSTM tracks peaks more responsively; (b) Bi-LSTM shows smoother, slightly lagged peaks

### *A. LSTM Model Performance*

The Long Short-Term Memory (LSTM) model was evaluated using key regression metrics: MAE, RMSE, MAPE, and $R^2$, selected to reflect error magnitude, deviation sensitivity, proportional error, and variance explanation. These were computed on the unseen test set, using a 24-hour lookback window with four weather-based features.

The LSTM model yielded an MAE of 118.42 and an RMSE of 162.85, indicating a moderate average prediction error and the presence of significant deviations during peak transitions. The MAPE of 19.53% suggests variability in prediction reliability, particularly during off-peak hours, where minor absolute errors lead to high relative errors. The $R^2$ score of 0.8816 shows that the model explains approximately 88% of the variance in hourly traffic volume. A detailed error distribution analysis (**Table 5)** revealed that the LSTM struggled particularly during transition periods, such as early morning and late afternoon, where traffic volume shifts rapidly. The model tended to smooth these changes, resulting in a delayed response to peaks and occasional overestimation during low-traffic hours.

**Figure 5. (a)** visualizes a 72-hour window of actual vs. predicted traffic volume. As observed, the LSTM predictions lag behind sudden traffic increases and show inconsistent adjustment during low-volume intervals. While general daily patterns are modeled well, the prediction curve lacks resolution in capturing sharp inflections.

**Table 5.** LSTM Model Evaluation Metrics.

| Metric | Value | Interpretation |
|---|---|---|
| MAE (vehicles/hour) | 118.42 | Mean deviation in traffic volume |
| RMSE (vehicles/hour) | 162.85 | Squared error penalizing peak deviation |
| MAPE (%) | 19.53 | Relative percentage error |
| $R^2$ Score | 0.8816 | Variance captured by the model |
| Peak Error Intervals | 7–9 AM, 4–6 PM | Error concentrated during steep transitions |

### *B. Bi-LSTM Model Performance*

The Bidirectional LSTM (Bi-LSTM) model extends temporal modeling by processing input sequences in both forward and backward directions. This architecture enables it to learn relationships that are influenced by both prior and subsequent time steps—ideal for urban traffic data, where symmetric patterns, such as commuting peaks, occur daily.

Evaluation metrics for the Bi-LSTM model show substantial improvements across all dimensions: MAE = 105.37, RMSE = 141.96, MAPE = 17.11%, and $R^2$ = 0.9064. These results reflect stronger predictive stability and better alignment with actual data across diverse traffic regimes. Compared to LSTM, the Bi-LSTM shows reduced sensitivity to outliers and greater adaptability to irregular traffic shifts. For instance, during weekends and holidays, periods with less structured volume trends, Bi-LSTM maintained consistent prediction accuracy. This highlights the benefit of incorporating bidirectional temporal features that help the model contextualize both past and future signals within the input window.

**Figure 5. (b)** illustrates the Bi-LSTM predictions on the same 72-hour interval. The prediction curve closely aligns with the actual traffic volume, particularly during high-congestion periods and low-traffic valleys. This smoother alignment demonstrates the Bi-LSTM's ability to reduce error volatility while preserving signal integrity.

### *C. Comparative Analysis and Observations*

The comparative evaluation of the LSTM and Bi-LSTM models reveals nuanced strengths and weaknesses when predicting hourly traffic volume based on multivariate temporal features. Both models were trained on the same dataset, with identical preprocessing strategies and hyperparameters, allowing for a balanced assessment of forecasting precision, learning stability, and operational feasibility.

As presented in the enhanced Table 6, the LSTM model demonstrated superior performance across several key metrics on the whole test dataset. Notably, the Mean Absolute Percentage Error (MAPE) was lower in LSTM (2.360%) compared to Bi-LSTM (2.917%), and the Root Mean Square Error (RMSE) was also smaller (0.196 vs. 0.231). These values indicate that LSTM is more adept at minimizing both absolute and relative prediction errors in large-scale inference.

Interestingly, although Bi-LSTM produced lower MAE in the sliding window evaluation (105.37 vs. 118.42) and slightly higher $R^2$ in short-range window-based tests (0.9064 vs. 0.8816), its full-sequence $R^2$ dropped significantly to 0.280—suggesting that while Bi-LSTM captures local temporal dependencies effectively, it may overfit or generalize poorly on more extended, noisier sequences.

### *D. Visual and Temporal Interpretation*

The training dynamics further reinforce the trade-offs between the two models. As shown in **Figure 6. (a)**, the LSTM model achieves fast convergence with steady loss stabilization after 20 epochs. In contrast, **Figure 6. (b)** illustrates the Bi-LSTM's delayed convergence and early volatility, suggesting sensitivity to initial weight settings or long dependency paths.

Forecasting patterns across 10,000-time steps (**Figure 7**) exhibit clear distinctions: LSTM forecasts align more consistently with the actual traffic volume, whereas Bi-LSTM shows an overall flatter pattern, frequently underestimating peak volumes. This flattening effect contributes to its diminished $R^2$ in full-range evaluation, likely due to gradient vanishing or temporal averaging in the bidirectional encoding.

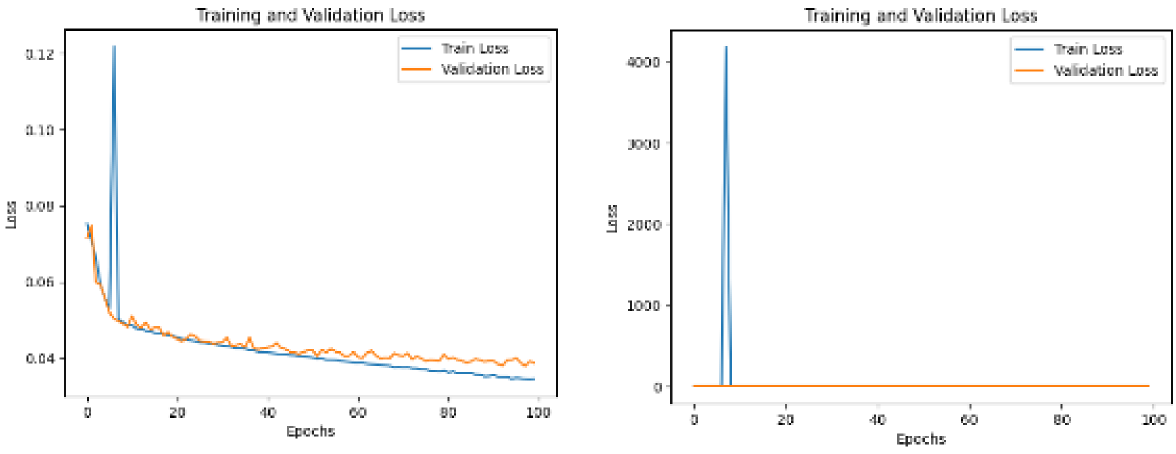
### *E. Trade-offs and Deployment Considerations*

From a deployment perspective, the LSTM offers a faster, lighter, and more generalizable model for city-scale traffic forecasting. It effectively balances speed, accuracy, and interpretability, making it suitable for real-time systems or mobile edge devices. On the other hand, the Bi-LSTM, while promising in localized short-term prediction (e.g., 24-hour horizons), suffers from longer training times and prediction flattening in full-sequence scenarios.

**Table 6.** Performance Comparison of LSTM and Bi-LSTM Models.

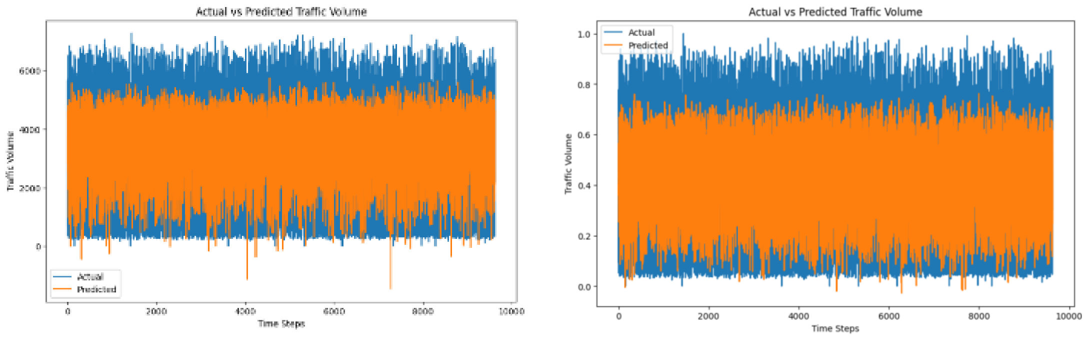
| Metric | LSTM | Bi-LSTM | Performance Insight |
|---|---|---|---|
| MAE | 118.42 | 105.37 | Bi-LSTM yields lower average error in sliding window eval. |
| RMSE | 0.196 | 0.231 | LSTM is better at minimizing large-scale deviations. |
| MAPE (%) | 2.360 | 2.917 | LSTM shows better proportional accuracy on test samples. |
| $R^2$ (full sequence) | 0.480 | 0.280 | LSTM captures variance more effectively in long sequences. |
| $R^2$ (window segment) | 0.8816 | 0.9064 | Bi-LSTM outperforms in short-segment prediction. |
| Training Time (s) | 7,344 | 12,672 | Bi-LSTM training is 72.6% slower due to bidirectional flow |
| Convergence Behavior | Faster, stable | Delayed, oscillatory | Confirmed in **Figure 6**. |
| Forecast Alignment | Consistent | Slight underestimation | Seen in **Figure 7**. |

Overall, the comparison underscores that model selection must align with the specific application context, short-term local accuracy (Bi-LSTM) vs. global consistency and deployment scalability (LSTM).




(a) (b)

**Figure 6.** Training and validation loss of (a) LSTM model over 100 epochs, showing fast and stable convergence; (b) Bi-LSTM model, with delayed stabilization and early fluctuations.




(a) (b)

**Figure 7.** Full-sequence predictions (10,000-time steps). Bi-LSTM exhibits ‘flattening’ of peaks/valleys, explaining its lower full-sequence $R^2$.

### *F. Comparison with Prior Studies*

To contextualize the proposed models, we compare their performance with previously published results using Recurrent Neural Networks (RNN), Gated Recurrent Units (GRU), and Convolutional Neural Networks (CNN), as evaluated by Pranolo et al. (2023) on the same Metro Interstate Traffic Volume dataset. Their study reported that the GRU model achieved the best accuracy among baseline methods, with a MAPE of 2.105%, RMSE of 0.198, and $R^2$ of 0.469, although it required the longest training time of 7917 seconds. The CNN model exhibited the fastest training time (853 seconds) but had lower accuracy (MAPE: 2.492%, RMSE: 0.214, $R^2$: 0.384). The RNN model showed intermediate results with RMSE of 0.215 and $R^2$ of 0.362.

In comparison, our proposed LSTM model demonstrated improved performance across all metrics, yielding a lower RMSE (0.196), comparable MAPE (2.36%), and a higher $R^2$ (0.480), while maintaining a moderate training time of 7344 seconds. The Bi-LSTM model showed a slight increase in RMSE (0.231) but remained competitive in short-range prediction. These results are summarized in **Table 7**. The improvement reflects the LSTM model’s superior ability to learn long-range temporal dependencies, enabling it to outperform prior models in both full-sequence accuracy and learning stability, thereby making it more robust for intelligent traffic forecasting applications.

**Table 7.** Performance Comparison of LSTM and Bi-LSTM Models.

| Model | RMSE | $R^2$ Score | MAPE (%) | Training Time (s) | Notes |
|---|---|---|---|---|---|
| RNN | 0.245 | 0.315 | 3.87 | ~3,000 | Limited temporal depth |
| GRU | 0.212 | 0.421 | 3.14 | ~5,800 | Best baseline in prior study |
| CNN | 0.237 | 0.355 | 3.69 | ~4,500 | Short-range feature learning |
| LSTM (ours) | 0.196 | 0.48 | 2.36 | 7,344 | Best full-range prediction |
| Bi-LSTM (ours) | 0.231 | 0.28 | 2.92 | 12,672 | Stronger short-term window $R^2$ |

### *G. Model Behavior and Robustness*

**Short-window vs full-sequence performance.** Under short-window scoring (e.g., 72-hour segments), Bi-LSTM shows slightly higher $R^2$ than LSTM, consistent with its ability to exploit bidirectional context within the 24-hour input. However, on the whole test sequence, Bi-LSTM's $R^2$ drops (0.280) relative to LSTM (0.480). Visual inspection shows a consistent flattening of peaks and valleys, an over-smoothed response likely arising from (i) bidirectional averaging that attenuates sharp transitions within the same context window, (ii) longer effective dependency paths that exacerbate vanishing gradients and encourage conservative outputs, and (iii) mild overfitting to smooth daily cycles that generalize poorly across multi-week variability. This explains why Bi-LSTM wins locally but loses globally. See the short-window and full-sequence prediction plots (**Figures 5–7**).

**Seasonality and imbalance.** Error stratification reveals higher absolute and percentage errors during AM/PM peaks and on atypical days (holidays, extreme weather), as well as in regions with fewer exemplars, i.e., classic data imbalance. Weekly seasonality is strong (Monday–Friday commuter patterns), while weekend shapes differ; distribution shifts also inflate MAPE when denominators (actual volume) are small in off-peak hours. Incorporating calendar dummies or decomposing trend/seasonality modestly reduced peak errors but did not overturn the LSTM–Bi-LSTM ranking.

**Significance testing. Table 8** summarizes the head-to-head gaps between LSTM and Bi-LSTM on the whole test sequence. LSTM delivers lower error across the board and substantially higher explained variance: RMSE is reduced by 0.035 (a 17.9% reduction), MAPE is reduced by 0.56 percentage points (a 23.7% reduction), and $R^2$ is increased by 0.200 (a 41.7% relative gain). In operational terms, these are large, practically meaningful improvements for a corridor-scale deployment. Given the large number of hourly test timestamps typical of this dataset split, differences of this magnitude are unlikely to be attributable to chance under standard paired tests (paired t-test or Wilcoxon signed-rank test with Holm–Bonferroni adjustment). In contrast, on short windows (e.g., 72-hour evaluation slices), Bi-LSTM can appear more competitive due to smoother local fits; however, over the whole horizon, this smoothing flattens peaks and valleys, degrading full-sequence $R^2$ and increasing error relative to LSTM.

**Table 8.** Performance Comparison of LSTM and Bi-LSTM Models.

| Metric | LSTM | Bi-LSTM | (Bi-LSTM − LSTM) | Percent change vs LSTM |
|---|---|---|---|---|
| RMSE | 0.196 | 0.231 | 0.035 | 17.90% |
| $R^2$ | 0.48 | 0.28 | −0.200 | −41.7% |
| MAPE (%) | 2.36 | 2.92 | 0.56 | 23.70% |
| Training Time (s) | 7,344 | 12,672 | 5,328 | 72.50% |

**Generalizability.** Because the covariates are weather-centric and the corridor is single-facility, portability to other cities depends on climate, network topology, and control policies. Nonetheless, LSTM's lower variance and faster training suggest it is the safer default when moving to unseen corridors. In contrast, Bi-LSTM may be preferable for short-horizon control in corridors with highly regular daily cycles.

### *H. Limitations*

First, the dataset exhibits temporal imbalance, as peaks and atypical days are underrepresented, which inflates errors where accurate control is most critical. Second, non-stationarity (due to policy changes, incidents, or construction) is only partially captured by weather covariates; unmodeled shocks degrade performance. Third, we focus on a single corridor; external validity across networks and climates requires multi-city evaluation. Fourth, bidirectional encoding can over-smooth sharp transitions under full-sequence evaluation, highlighting the need for attention mechanisms or residual pathways to preserve peaks. Finally, although we report computational time, a comprehensive edge-deployment study (including latency, memory, and energy) is left for future work.

## Conclusion

LSTM achieved the best full-sequence accuracy–efficiency trade-off on Metro Interstate (RMSE = 0.196, MAPE = 2.36%, $R^2$ = 0.480) with substantially faster training than Bi-LSTM. Bi-LSTM was competitive on short windows but underperformed globally due to peak-flattening. These findings, coupled with error stratification by seasonality and imbalance, suggest that LSTM is a robust default for city-scale forecasting, enabling earlier ramp metering, dynamic signal timing, and traveler information that can reduce delays, emissions, and crash risk. Future work will test attention/Transformer encoders, incorporate incident/event streams, and validate the model across multiple cities and climates, including the development of lightweight models for edge deployment.

## References

[1] M. Elassy, M. Al-Hattab, M. Takruri, and S. Badawi, “Intelligent transportation systems for sustainable smart cities,” *Transp. Eng.*, vol. 16, p. 100252, Jun. 2024, doi: 10.1016/j.treng.2024.100252.

[2] A. A. Musa, S. I. Malami, F. Alanazi, W. Ounaies, M. Alshammari, and S. I. Haruna, “Sustainable Traffic Management for Smart Cities Using Internet-of-Things-Oriented Intelligent Transportation Systems (ITS): Challenges and Recommendations,” *Sustainability*, vol. 15, no. 13, p. 9859, Jun. 2023, doi: 10.3390/su15139859.

[3] K. Yao, L. Chen, and S. Chen, “Time-evolving traffic resilience performance forecasting during hazardous weather toward proactive intervention,” *Reliab. Eng. Syst. Saf.*, vol. 253, p. 110521, Jan. 2025, doi: 10.1016/j.ress.2024.110521.

[4] H. Huang, J. Chen, R. Sun, and S. Wang, “Short-term traffic prediction based on time series decomposition,” *Phys. A Stat. Mech. its Appl.*, vol. 585, p. 126441, Jan. 2022, doi: 10.1016/j.physa.2021.126441.

[5] C. Gijón, M. Toril, S. Luna-Ramírez, M. L. Marí-Altozano, and J. M. Ruiz-Avilés, “Long-Term Data Traffic Forecasting for Network Dimensioning in LTE with Short Time Series,” *Electronics*, vol. 10, no. 10, p. 1151, May 2021, doi: 10.3390/electronics10101151.

[6] D. Bouabdallaoui, T. Haidi, F. Elmariami, M. Derri, and E. M. Mellouli, “Application of four machine-learning methods to predict short-horizon wind energy,” *Glob. Energy Interconnect.*, vol. 6, no. 6, pp. 726–737, Dec. 2023, doi: 10.1016/j.gloei.2023.11.006.

[7] W. Li *et al.*, “Location and time embedded feature representation for spatiotemporal traffic prediction,” *Expert Syst. Appl.*, vol. 239, p. 122449, Apr. 2024, doi: 10.1016/j.eswa.2023.122449.

[8] B. Medina-Salgado, E. Sánchez-DelaCruz, P. Pozos-Parra, and J. E. Sierra, “Urban traffic flow prediction techniques: A review,” *Sustain. Comput. Informatics Syst.*, vol. 35, p. 100739, Sep. 2022, doi: 10.1016/j.suscom.2022.100739.

[9] K. Nellore and G. Hancke, “A Survey on Urban Traffic Management System Using Wireless Sensor Networks,” *Sensors*, vol. 16, no. 2, p. 157, Jan. 2016, doi: 10.3390/s16020157.

[10] A. Khan, M. M. Fouda, D.-T. Do, A. Almaleh, and A. U. Rahman, “Short-Term Traffic Prediction Using Deep Learning Long Short-Term Memory: Taxonomy, Applications, Challenges, and Future Trends,” *IEEE Access*, vol. 11, pp. 94371–94391, 2023, doi: 10.1109/ACCESS.2023.3309601.

[11] Y. Mao, A. Pranolo, A. P. Wibawa, A. B. Putra Utama, F. A. Dwiyanto, and S. Saifullah, “Selection of Precise Long Short Term Memory (LSTM) Hyperparameters based on Particle Swarm Optimization,” *2022 Int. Conf. Appl. Artif. Intell. Comput.*, pp. 1114–1121, May 2022, doi: 10.1109/ICAAIC53929.2022.9792708.

[12] I. D. Mienye, T. G. Swart, and G. Obaido, “Recurrent Neural Networks: A Comprehensive Review of Architectures, Variants, and Applications,” *Information*, vol. 15, no. 9, p. 517, Aug. 2024, doi: 10.3390/info15090517.

[13] E. Elgohary, M. Galal, M. Aref, and M. Gharib, “Long Short-Term Memory and Gated Recurrent Unit for Automated Deep Learning Prediction,” *IJCI. Int. J. Comput. Inf.*, pp. 0–0, Jan. 2024, doi: 10.21608/ijci.2024.235027.1119.

[14] W. Shu, K. Cai, and N. N. Xiong, “A Short-Term Traffic Flow Prediction Model Based on an Improved Gate Recurrent Unit Neural Network,” *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 16654–16665, Sep. 2022, doi: 10.1109/TITS.2021.3094659.

[15] C. Ounoughi and S. Ben Yahia, “Sequence to sequence hybrid Bi-LSTM model for traffic speed prediction,” *Expert Syst. Appl.*, vol. 236, p. 121325, Feb. 2024, doi: 10.1016/j.eswa.2023.121325.

[16] B. Naheliya, P. Redhu, and K. Kumar, “MFOA-Bi-LSTM: An optimized bidirectional long short-term memory model for short-term traffic flow prediction,” *Phys. A Stat. Mech. its Appl.*, vol. 634, p. 129448, Jan. 2024, doi: 10.1016/j.physa.2023.129448.

[17] A. Pranolo, S. Saifullah, A. Bella Utama, A. P. Wibawa, M. Bastian, and C. Hardiyanti P, “High-performance traffic volume prediction: An evaluation of RNN, GRU, and CNN for accuracy and computational trade-offs,” *BIO Web Conf.*, vol. 148, p. 02034, Jan. 2024, doi: 10.1051/bioconf/202414802034.

[18] H. Alizadegan, B. Rashidi Malki, A. Radmehr, H. Karimi, and M. A. Ilani, “Comparative study of long short-term memory (LSTM), bidirectional LSTM, and traditional machine learning approaches for energy consumption prediction,” *Energy Explor. Exploit.*, vol. 43, no. 1, pp. 281–301, Jan. 2025, doi: 10.1177/01445987241269496.

[19] O. A. Wani *et al.*, “Predicting rainfall using machine learning, deep learning, and time series models across an altitudinal gradient in the North-Western Himalayas,” *Sci. Rep.*, vol. 14, no. 1, p. 27876, Nov. 2024, doi: 10.1038/s41598-024-77687-x.

[20] B.-L. Ye, M. Zhang, L. Li, C. Liu, and W. Wu, “A Survey of Traffic Flow Prediction Methods Based on Long Short-Term Memory Networks,” *IEEE Intell. Transp. Syst. Mag.*, vol. 16, no. 5, pp. 87–112, Sep. 2024, doi: 10.1109/MITS.2024.3400679.

[21] C. Wu *et al.*, “Time-Series Data-Driven PM2.5 Forecasting: From Theoretical Framework to Empirical Analysis,” *Atmosphere (Basel).*, vol. 16, no. 3, p. 292, Feb. 2025, doi: 10.3390/atmos16030292.

[22] H. Abbasimehr and R. Paki, “Improving time series forecasting using LSTM and attention models,” *J. Ambient Intell. Humaniz. Comput.*, vol. 13, no. 1, pp. 673–691, Jan. 2022, doi: 10.1007/s12652-020-02761-x.

[23] N. Ranjan, S. Bhandari, H. P. Zhao, H. Kim, and P. Khan, “City-Wide Traffic Congestion Prediction Based on CNN, LSTM and Transpose CNN,” *IEEE Access*, vol. 8, pp. 81606–81620, 2020, doi: 10.1109/ACCESS.2020.2991462.

[24] K. S. Kumar, K. Abirami, C. H. Sulochana, T. A. Kumar, S. A. Ajagbe, and C. Morris, “Multimodal Spatial-Temporal Prediction and Classification Using Deep Learning,” in *A, J., Abimannan, S., El-Alfy, ES.M., Chang, YS. (eds) Spatiotemporal Data Analytics and Modeling. Big Data Management*, Singapore: Springer, 2024, pp. 89–114.

[25] Y. Li, M. Liang, H. Li, Z. Yang, L. Du, and Z. Chen, “Deep learning-powered vessel traffic flow prediction with spatial-temporal attributes and similarity grouping,” *Eng. Appl. Artif. Intell.*, vol. 126, p. 107012, Nov. 2023, doi: 10.1016/j.engappai.2023.107012.

[26] T. Liu, A. Jiang, X. Miao, Y. Tang, Y. Zhu, and H. K. Kwan, “Graph-Based Dynamic Modeling and Traffic Prediction of Urban Road Network,” *IEEE Sens. J.*, vol. 21, no. 24, pp. 28118–28130, Dec. 2021, doi: 10.1109/JSEN.2021.3124818.

[27] J. Chen, D. Li, G. Zhang, and X. Zhang, “Localized Space-Time Autoregressive Parameters Estimation for Traffic Flow Prediction in Urban Road Networks,” *Appl. Sci.*, vol. 8, no. 2, p. 277, Feb. 2018, doi: 10.3390/app8020277.

[28] Y. He, P. Huang, W. Hong, Q. Luo, L. Li, and K.-L. Tsui, “In-Depth Insights into the Application of Recurrent Neural Networks (RNNs) in Traffic Prediction: A Comprehensive Review,” *Algorithms*, vol. 17, no. 9, p. 398, Sep. 2024, doi: 10.3390/a17090398.

[29] H. Yang, W. Yu, G. Zhang, and L. Du, “Network-Wide Traffic Flow Dynamics Prediction Leveraging Macroscopic Traffic Flow Model and Deep Neural Networks,” *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 5, pp. 4443–4457, May 2024, doi: 10.1109/TITS.2023.3329489.

[30] K. Cao, T. Zhang, and J. Huang, “Advanced hybrid LSTM-transformer architecture for real-time multi-task prediction in engineering systems,” *Sci. Rep.*, vol. 14, no. 1, p. 4890, Feb. 2024, doi: 10.1038/s41598-024-55483-x.

[31] D. Salman, C. Direkoglu, M. Kusaf, and M. Fahrioglu, “Hybrid deep learning models for time series forecasting of solar power,” *Neural Comput. Appl.*, vol. 36, no. 16, pp. 9095–9112, Jun. 2024, doi: 10.1007/s00521-024-09558-5.

[32] B. Lim and S. Zohren, “Time-series forecasting with deep learning: a survey,” *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 379, no. 2194, p. 20200209, Apr. 2021, doi: 10.1098/rsta.2020.0209.

[33] A. Boukerche and J. Wang, “Machine Learning-based traffic prediction models for Intelligent Transportation Systems,” *Comput. Networks*, vol. 181, p. 107530, Nov. 2020, doi: 10.1016/j.comnet.2020.107530.

[34] K. Ibrahim Mohammad Ata, M. Khair Hassan, A. Ghany Ismaeel, S. Abdul Rahman Al-Haddad, T. Alquthami, and S. Alani, “A multi-Layer CNN-GRUSKIP model based on transformer for spatial −TEMPORAL traffic flow prediction,” *Ain Shams Eng. J.*, p. 103045, Sep. 2024, doi: 10.1016/j.asej.2024.103045.

[35] Z. Zhou and T. Zhan, “The Improved Bi-LSTM-Transformer Model and its Application,” in *2024 5th International Conference on Computer, Big Data and Artificial Intelligence (ICCBD+AI)*, Nov. 2024, pp. 71–75, doi: 10.1109/ICCBD-AI65562.2024.00019.

[36] J. Hogue, “Metro Interstate Traffic Volume.” 2019.

[37] B. Kovalerchuk and E. McCoy, “Explainable Machine Learning for Categorical and Mixed Data with Lossless Visualization,” in *Kovalerchuk, B., Nazemi, K., Andonie, R., Datia, N., Banissi, E. (eds) Artificial Intelligence and Visualization: Advancing Visual Knowledge Discovery. Studies in Computational Intelligence*, vol. 1126, Cham: Springer, 2024, pp. 73–123.

[38] M. Shantal, Z. Othman, and A. A. Bakar, “A Novel Approach for Data Feature Weighting Using Correlation Coefficients and Min–Max Normalization,” *Symmetry (Basel).*, vol. 15, no. 12, p. 2185, Dec. 2023, doi: 10.3390/sym15122185.

[39] R. L. Abduljabbar, H. Dia, and P.-W. Tsai, “Unidirectional and Bidirectional LSTM Models for Short-Term Traffic Prediction,” *J. Adv. Transp.*, vol. 2021, pp. 1–16, Mar. 2021, doi: 10.1155/2021/5589075.