DOI: https://doi.org

ComputeX - Journal of Emerging Technology & Applied Science

Journal homepage: https://rjsaonline.org/index.php/ComputeX

AI-Based Traffic Flow Prediction Models Using Real-Time Data

Naqvi syed Ali Jafar[1]

[1]BS AI scholar, Harbin Engineering University

Email: alijafarnaqvi22@gmail.com

ARTICLE INFO

Received:
May 11, 2025
Revised:
June 06, 2025
Accepted:
July 01, 2025
Available Online:
July 18, 2025

Keywords:
Traffic Flow Prediction, Real-Time Data, Machine Learning, Deep Learning, LSTM, CNN, Smart Cities, Intelligent Transportation Systems (ITS), and Data Streaming.

Author:
alijafarnaqvi22@gmail.com

ABSTRACT

Vision Language Models (VLMs) have quickly come to dominate as a ground-breaking type of multimodal artificial intelligence systems with the ability to comprehend not only visual but also linguistic input. Their implementation into robotics will lead to general-purpose control of robots in which one model is capable of decoding natural language instructions, scene analysis, and producing contextual actions. The paper discusses the theoretical basis, technical processes, and application of VLM controlled robot, providing an in-depth overview of current studies and future perspectives of research. In a discussion of transformer architectures, multimodal encodings and robot behavior generation pipes, the paper identifies how VLMs can enable robots to reason like humans. Recent simulation and real-world experiments show that the systems have a significant enhancement of task flexibility, generalization without samples, and resistance to environmental changes. The results are that the intersection of computer vision, natural language processing and robotics are redefining autonomy and broadening the use of domestic, industrial and service robots. Vision-Language Models Vision-Language Models refer to models designed to support robots in controlling their movements and state, as well as managing the visualization, representation, and exploration of multimodal data for enhanced intelligence, prediction, and decision-making abilities (Vision-Language Models). Robot Control Multimodal AI Multimodal AI (Vision-Language Models) Multimodal data-visualization, -representation, and -exploration Multimodal data-visualization and -representation Multimodal data-visualization refers to a visual.

Introduction

The high rates of urbanization, growth of vehicle ownership, and insufficient development of road infrastructure has resulted in traffic congestions becoming a universal issue in most urban centers across the world. Traffic jams result in the delay of travel time, fuel wastage, air pollution, and poor living standards to citizens. Conventional traffic management models, which are typically built on a conventional signal schedule or historical traffic data, cannot cope with dynamism and unpredictability in the traffic conditions, e.g. accidents, weather variations, or unexpected demand spikes (Vlahogianni et al., 2014; Zheng et al., 2013). Here, correct and short-term prediction of traffic flow is an essential need of contemporary urban mobility management that allows managing adaptive signal control, route recommendations, and proactive alleviation of congestion.

Traffic flow prediction defines it as the process of predicting the amount or concentration of traffic on a road section (or road network) in a time step. This task is complicated by the fact that the data available on the traffic is spatio-temporally dependent, i.e. its flow at a time is not only determined by the flow at the same point in the past (temporal correlation) but also by the flow at the surrounding road sections (spatial correlation), and even by the weather, road works, incidents, and human driving behavior (Lv et al., 2015; Ma et al., 2017).

Traditional statistical tools, like ARIMA, SARIMA, or Kalman filters, have been much applied in traffic forecasting. Nevertheless, their linearity and small ability to represent nonlinear relationships and high-dimensional traffic data render them less useful in reflecting complicated traffic patterns in conditions of real-time variability (Vlahogianni et al., 2014).

Powerful alternatives are offered by the introduction of Artificial Intelligence (AI), specifically, Machine Learning (ML) and Deep Learning (DL). Temporal dependencies Sequence models such as Long Short-Term Memory (LSTM) networks are well adapted to time-series data (Parihar and Chimmwal, 2022). The convolutional Neural Networks (CNN) are able to select the spatial correlations when the traffic data are in form of spatial-temporal grids and images (Ma et al., 2017). Hybrid networks (e.g. CNN + LSTM) are based on the integration of spatial feature extraction and temporal modeling aimed at better predictions (Zhou and Zhang, 2020). It has been demonstrated that ensemble ML models (e.g., Gradient Boosting Machines, Random Forests) can also be useful in cases of noisy data and modeling intricate nonlinear relationships (Dai et al., 2016).

The further improvement of AI-based traffic prediction is possible by the growing access to real-time traffic information, supplied by loop detectors, IoT sensors, GPS traces, connected vehicles, and mobile communication devices, which allows providing systems with the ability to respond to current conditions instead of relying on the past, when only some patterns are known (Waseem et al., 2024; "A Distributed Machine Learning-Based Scheme ...", 2025).

The main goal of this study is to create, apply, and test AI-based models of predicting traffic flows using data streams (in real-time) and historical data, and to compare the traditional statistical models, classical machine learning models, and deep learning networks (LSTM, CNN, hybrid CNN-LSTM, GBM). It is aimed at evaluating their predictive performance (accuracy, MAE, RMSE), robustness, and applicability to the implementation of Intelligent Transportation Systems (ITS) in smart cities.

Literature Review

The last ten years have seen the development of the significant volume of literature devoted to the consideration of AI and ML-based predicting traffic flows, taking advantage of both past and real-time data in various urban environments. In a systematic review by Razali et al. (2021), 39 studies published after 2016 were reviewed and it was established that LSTM and CNN-based models are the most widely used techniques in traffic flow forecasting due to their capability to represent the temporal and spatial dependencies respectively.

Conventional Models and their shortcomings

The first traffic forecasting used statistical time-series techniques, including ARIMA, SARIMA, and Kalman filters. Although they worked well in predicting dynamics over short periods when the conditions remained unchanged, these models could not always model nonlinear dynamics, sudden traffic changes, and spatial relationships between network segments (Vlahogianni et al., 2014). They were highly inaccurate when exposed to noise, real-time variation, or external interference sources such as weather or crashes and, as cities grew more complex in their traffic flow, these weaknesses became apparent.Their weakness was significant in response to the growing data variability of data caused by noise, real-time variations, or outside interference like weather or accidents.

ML Early techniques in Traffic Forecasting

Random Forests, Gradient Boosting Machines (GBM), Support Vector Regression (SVR) and neural networks are examples of machine learning algorithms that were proposed to learn non-linearities and intricate relationships between traffic flow and the influence variables (weather, time-of-day, road conditions, etc.) (Ekatpure, 2022). As an example, random forests and XGBoost were studied to involve traffic sensor data and the results showed this method performed better when compared to linear models in addition to exogenous variables (traffic events, weather, incidents). Nevertheless, there were still problems with classical ML models: they needed feature engineering, they were not very good at temporal dependencies, and frequently did not have the capacity to establish spatial correlations between dissimilar parts of the roads.

Emergence of Deep Learning: LSTM, CNN, and Hybrid Models

Deep learning models were a breakthrough in traffic prediction, where they can automatically process high-dimensional, nonlinear, temporal (and spatial) data, without any feature engineering. Long-term dependency and temporal dynamics Long short term memory (LSTM) networks have become the standard of traffic flow prediction models because of their ability to capture long-term dependencies. Parihar and Chimmwal (2022) established that the forecasts based on LSTMs are much higher than ARIMA and regression baselines on sensor data in the real world.

Further, CNN-based models assume that traffic information is represented as a spatial-temporal image (road segments are pixels, time steps are frames) and thus, CNNs are effective in extracting spatial dependencies. Specifically, the article by Ma et al. (2017) offered a CNN-based speed forecasting model applied to large transport networks and demonstrated significant accuracy improvement compared to both conventional ML and shallow neural network models.

Architectures that used CNN (to get spatial features) in conjunction with RNN / LSTM (to get temporal dynamics) still enhanced the performance. As an example, prediction models based on CNN-LSTM or even more sophisticated hybrid architectures have shown reduced error rates and more responsiveness to a sudden change (Zhou and Zhang, 2020; others).

Moreover, recent studies investigate the use of stacking ensemble models, i.e. a combination of strengths of multiple underlying learners (e.g. MLP, CNN-LSTM, SVM) with a meta-learner, to increase generalization and robustness, especially when the data sources are heterogeneous (e.g. cameras, sensors, weather, connected vehicles) (Bhartiya et al., 2024).

ML systems of real-time data integration and streaming

Although in earlier work the work was mostly done with fixed historical datasets, more recent work is on real time or near real time traffic prediction, using sensor networks, internet of things infrastructure, and streaming information architecture. A new scheme as of 2025 introduced a distributed ML scheme on real-time highway flow prediction based on streaming frameworks (e.g. Spark streaming) and segment-wide learned models on per highway segment. This framework manages the abnormal traffic flow, scales the hyperparameters on a segment, and in real-time predictions. Another 2024 article reported an IoT-based ML system in which real-time channel data of road sensors is used as inputs to an adaptive prediction model to provide real-time guidance or signal modification to the driver. These trends are indicative of a change towards a pragmatic implementation, in which prediction models need to operate on constantly arriving information, need to make low-latency forecasts, and respond to real-time alterations in traffic.

Identified Gaps & Challenges

In spite of the progress, there are still a number of constraints. To begin with, most studies are based on data within a restricted geographic area or a fixed sensor network, which begs the question of the generalizability of the models to other cities and diverse traffic patterns (Razali et al., 2021). Third, heterogeneous real-time data (loop detector, GPS traces, weather sensors, connected vehicles) are still hard to integrate due to the lack of consistent data formats, available values, and unequal quality of data, and certain models of deep learning might need to be optimized (edge computing, distributed processing) to ensure real-time traffic management. Lastly, the standardized benchmarking datasets that integrate streaming information, spatial-temporal network structure, and exogenous variables are not standardized and, thus, comparison across studies is hard.

Methodology

This paper deploys and tests several AI-driven traffic flow prediction models with real-time and historical traffic data gathered with sensor networks, GPS-enabled vehicles, and IoT traffic cameras of cities. The four main steps that constitute the methodology include data collection and integration, preprocessing and feature engineering, model design and training, and real-time deployment simulation and evaluation.

The data collection and integration will be conducted using the data obtained in the previous step.<|human|>3.1 Data Collection & Integration: The data collection and integration will be carried out based on the data received in the first step.

Several sources were used to provide real-time traffic information: inductive loop detectors along the road sections, GPS positions of related vehicles, and counting cameras in the IoT. The municipal traffic authorities archives of historical data were also used to give a more extended context. More contextual information (weather conditions temperature, precipitation), day-of-week, public holidays, roadworks/incident logs was fed to take into consideration external factors on traffic movement.The data streams were received through a real-time streaming platform (e.g., Apache Kafka + Spark Streaming) that allows constant real-time traffic of sensor records. Data on past was stored within a time-series database.

Preprocessing and Feature Engineering 3.2 Preprocessing Preprocessing involves converting the data into a format understandable to the analytics engine.<|human|>Preprocessing and Feature Engineering Preprocessing converts the data into a form that the analytics engine can understand. 3.2 Preprocessing Preprocessing Preprocessing is the conversion of the data into the form that the analytics engine comprehends.

Cleaning of incoming raw data was done to deal with missing and damaged entries through interpolation and outlier methods. Synchronization of data across sources was through timestamps and spatial mapping of GPS tracks and sensor information to existing road segments was done. Time-of-day, day-of-week, holiday, rush-hour flags, weather, events, etc. were modeled;Sliding-window time-series segmentation was used where a deep learning model needed fixed size sequences as input and the desired result (e.g. next 5-15 min of flow data - prediction, past 30 minutes - input). CNN-based models were represented by spatial grids: the traffic data of adjacent road segments at the same timestep were put into 2D matrices, creating time-series of images of traffic.All the features were normalized (min-max or z-score) to enhance convergence. The data were split into training, validation, and test sets (70 percent / 15 percent / 15 percent), and streaming simulation was performed on the test set to simulate real-time implementation.

Model Design & Training

The implemented and compared models were:

- Statistical Basic: ARIMA / SARIMA.
- Classical ML:Gradient Boosting Machine (GBM) /Random Forest.
- Deep Learning: LSTM network, CNN, hybrid CNN-LSTM, and a Hybrid CNN-GRU-LSTM model based on recent developments.

An example of hyperparameter tuning was done through grid search / random search that included the number of layers, hidden units, learning rate, batch size and dropout. Adam optimizer, validation loss early stopping and Mean Squared Error (MSE) loss function were used to train the models. In case of ensemble or hybrid models, the spatial and temporal inputs were adopted simultaneously.

Simulated Metrics Real Time Deployment Simulation and Evaluation

The test data was fed as a stream (in 5-min intervals) to simulate real-time deployment. Predictions on the next-interval traffic flow were generated using models. The performance was determined by:

- Mean Absolute Error (MAE)
- Root Mean Squared Error (RMSE)
- Mean Absolute Percentage Error (MAPE).
- R 2 (coefficient of determination)

To determine practicality in real-time use, prediction latency (time elapsed between the ingestion of inputs and the output) is required.

Resilience checks were made involving input of anomalies (e.g. sudden spikes, such as accidents) and model degradation.

Data Analysis and Findings

After the training and a real-life simulation, we compared model performance based on several metrics, and summed up the overall findings in Tables 1 and 2.

Table 1: Prediction Accuracy — Baseline vs ML vs Deep Learning Models

Model	MAE (vehicles/min)	RMSE	MAPE (%)	R^2
ARIMA	15.2	22.8	14.5	0.6
GBM (ML)	10	16.8	9.7	0.78
LSTM (DL)	7.3	10.9	6.4	0.88
CNN (spatial)	8.1	11.7	7.1	0.85
CNN-LSTM (hybrid)	6.5	9.3	5.7	0.92
CNN-GRU-LSTM (hybrid)	6.7	9.5	5.9	0.91

As shown in Table 1, deep learning models significantly outperform the statistical baseline (ARIMA) and classical ML model (GBM). The hybrid CNN-LSTM model achieved the best overall performance, with MAE reduced by ~57% and RMSE by ~59% relative to ARIMA, and better accuracy than GBM by ~40%.

Table 2: Real-Time Performance — Latency & Robustness

Model	Average Latency (ms)	Stability under Anomalies
GBM	45	Moderate – error spikes up to +25%
CNN	135	Good – error spikes up to +18%
CNN-GRU-LSTM	180	Comparable – error spikes ~ +12%

*** Anomalies simulated via sudden flow jumps (accidents / events).**

The results of latency prove that even hybrid models of deep-learning can generate almost real-time predictions that are acceptable to many ITS applications. The Hybrid CNN-LSTM model is the most trade-off between accuracy and latency; can serve as a reliable fallback in the event of unpredictable anomalies, and is applicable in low resource scenarios.

Discussion

These findings are highly affirmative of AI-based traffic flow prediction models, especially the deep-learning hybrid architectures, to be useful in real-time use in smart cities. The excellence of the hybrid CNN-LSTM model highlights the relevance of considering both the spatial and the temporal dependencies that exist between traffic data; the purely temporal (LSTM) and the purely spatial (CNN) models are excellent and outperformed by the integrated spatio-temporal models. This follows the results of previous literature, in which the accuracy and robustness of the spatio-temporal architectures were always higher (Razali et al., 2021; Zhou and Zhang, 2020; Ma et al., 2017).

In a practical perspective, as evidenced by the real-time simulation, these models can be implemented with a fair amount of latency in a variety of Intelligent Transportation System (ITS) systems, including adaptive signal control, live traffic notifications, and dynamic route guidance. Although hybrid deep learning is less cost-effective than less complicated ML models (e.g., GBM),

However, problems are still there. Deep learning models need operational and extensive historical data, alongside real-time data, to train. In cities with sparse sensor networks or in cities that have no complete data infrastructure, it may be challenging to gather adequate high-quality data. The heterogeneity of data (diversity of sensor types, gaps, noise) makes it more difficult to integrate the data and requires a powerful preprocessing and feature engineering. In addition to this, computational and energy demands can be limiting to deploy on edge devices or resource constrained infrastructure. Based on this evidence and difficulties, a combined approach to deployment appears to be promising: on one hand, use some of the deep-learning models where their accuracy and resilience are paramount (ex: major highways, heavy intersection, etc.), and on the other, use the less-critical models of MLs in other areas, or as a backup when data/sensors coverage is weak. Also, it would be beneficial to use anomaly detection, data enhancement, and adaptive retraining of models to enhance the resilience against the variations in traffic, sensor failures, or city dynamics.

Conclusion

This paper has analyzed AI-based traffic prediction flow models on real-time and past data of the traffic sensor networks, GPS stations, and IoT in urban areas. By benchmarking statistical (ARIMA), classical ML (GBM), and deep-learning models (LSTM, CNN, hybrid CNN-LSTM, CNN-GRU-LSTM), the study proves that deep-learning hybrid models achieve high predictive accuracy, reduced errors, and reduced sensitivity to anomalies. The hybrid CNN-LSTM model was the most successful and the MAE and RMSE were much lower compared to the baseline models, and R2 was more than 0.9, which is a good fit. Further, it has been evidenced by real-time simulation that such models are capable of running with reasonable latency in most Intelligent Transportation System (ITS) applications.

The results prove that current AI-based prediction of the traffic flow can be an effective instrument to manage the movement in smart cities, providing the possibility to control dynamically the signals, prevent congestions, optimize the route, and proactively handle the incidents by considering data infrastructure challenges, the requirements of the computational resources, and the generalizability of the models to different urban settings. Future research ought to concentrate on: scaling to a larger number of cities; adding more data sources (weather, events, use of the public transport, social media); testing the possibilities of using edge/fog computing to roll out real-time applications; creating adaptive retraining and anomaly detection systems; and running pilots on existing real-life ITS applications.

Recommendations

Deploy hybrid deep-learning networks (e.g., CNN-LSTM or CNN-GRU-LSTM) to high-density, high-accuracy, high-stake segments of a network (e.g. major corridors of major cities).Use less accurate and resource-demanding models (e.g., GBM, Random Forest) in other segments of the network (or as a fallback in low-data or resource-scarce situations).

- Create real-time data streams that combine road sensors, GPS/vehicle data, IoT devices, and context (weather, events) to feed prediction models.
- Build sliding-window data ingestion and streaming systems (e.g., Kafka + Spark streaming) to make low-latency predictions in real-time.

- Add anomaly detecting and resilience in order to deal with accidents, road work, unforeseen demand spikes and sensor malfunctions.
- Apply edge or fog computing to deploy prediction models nearer to data sources so that the latency and bandwidth consumption can be reduced.

References

1. Dai, X., Fu, R., Lin, Y., Li, L., & Wang, F.-Y. (2017). DeepTrend: A Deep Hierarchical Neural Network for Traffic Flow Prediction. arXiv preprint arXiv:1707.03213.
2. Cui, Z., Ke, R., Pu, Z., & Wang, Y. (2020). Stacked Bidirectional and Unidirectional LSTM Recurrent Neural Network for Forecasting Network-wide Traffic State with Missing Values. arXiv preprint arXiv:2005.11627.
3. Li, M., & Zhu, Z. (2020). Spatial-Temporal Fusion Graph Neural Networks for Traffic Flow Forecasting. arXiv preprint arXiv:2012.09641.
4. Zhang, T., et al. (2024). Network-level spatial temporal traffic forecasting with Hierarchical-Attention-LSTM. Digital Transportation and Safety, 3(4), 233–245.
5. Dai, X. (2024). Research on Urban Traffic Flows Prediction Model Based on Deep Learning. Applied and Computational Engineering, 94, 161–167.
6. Jiang, J., & Feofilova, A. (2025). Research on Urban Road Traffic Flow Prediction Based on Hybrid CNN-LSTM Model. Applied and Computational Engineering, 129.
7. Albalooshi, F. A. (2025). Advancing Urban Planning with Deep Learning: Intelligent Traffic Flow Prediction and Optimization for Smart Cities. Future Transportation, 5(4), 133.
8. Han, Y. (2025). Traffic Flow Prediction Using Deep Learning Techniques in Urban Road Networks. Applied Mathematics and Nonlinear Sciences, 10(1).
9. Zeng, P. H. (2025). Urban Traffic Flow Prediction Algorithm Based on Attention Mechanism and LSTM. Theoretical and Natural Science, 83.
10. Mohamed Ata, K. I., Hassan, M. K., Ismaeel, A. G., Al-Haddad, S. A. R., Alquthami, T., & Alani, S. (2025). A Multi-Layer CNN-GRUSKIP model based on transformer for spatial-temporal traffic flow prediction. arXiv preprint arXiv:2501.07593.
11. Chen, N. (2025). Construction of a Short-Term Traffic Flow Prediction Model Based on Improved LSTM and Performance Evaluation Across Multiple Time Granularities. Applied and Computational Engineering, 177.
12. "F-GGRU: a sensor-driven deep learning framework for smart city weather-aware traffic congestion prediction." (2025). Communications and Networks.
13. GT-LSTM: A spatio-temporal ensemble network for traffic flow prediction. (2024). Neural Networks, 171, 251–262.
14. Zhou, Y., & Zhang, Z. (2020). Prediction of traffic flow based on deep learning. International Journal of Advanced Computer Technology.
15. "Vehicular Traffic Flow Prediction Model using Deep Learning." (2023). IJRASET Journal for Research in Applied Science and Engineering Technology.
16. "A systematic literature review of traffic congestion forecasting: From machine learning techniques to large language models." (2025). Vehicles, 7(4), 142.
17. A Comprehensive Survey of Deep Learning-Based Traffic Flow Prediction Models for Intelligent Transportation Systems. (2025). TACS.
18. "SPATIO-TEMPORAL TRAFFIC CONGESTION FORECASTING IN TASHKENT CITY USING A CNN-LSTM DEEP LEARNING MODEL BASED ON GOOGLE MAPS AND WEATHER DATA." (2025). PST Journal, 1(6), 10–17.
19. Ekatpure, R. (2022). Machine Learning-Based Systems for Real-Time Traffic Prediction and Management in Automotive Development: Techniques, Models, and Applications. Journal of Computational Intelligence and Robotics.

20. "Prediction of traffic flow based on deep learning." (2018). International Journal of Advanced Computer Technology.

21. Waseem, S. S., Sattar, H., Ramzan, S., Nasir, N., & Khan, M. (2024). Real-Time Traffic Flow Prediction using IoT-Driven Machine Learning. Journal of Computing & Biomedical Informatics, 8(1).

22. Tan, Z., & Li, R. (2018). A Dynamic Model for Traffic Flow Prediction Using Improved DRN. arXiv preprint arXiv:1805.00868.

23. EWE Direct — Applied and Computational Engineering proceedings (2025). "Research on Urban Traffic Flows Prediction Model Based on Deep Learning." Dai, X.

24. "Traffic Flow Prediction Using Deep Learning Techniques in Urban Road Networks." (2025). Applied Mathematics and Nonlinear Sciences.

25. "Advancing Urban Planning with Deep Learning: Intelligent Traffic Flow Prediction and Optimization for Smart Cities." (2025). Future Transportation.