Smart Traffic Flow Prediction for Efficient Urban Mobility Management

Naveen Sharma
M. Tech. in Transportation Engineering, CBS Group of Institutions, Jhajjar, Haryana.

Minakshi
A.P Civil Department, CBS Group of Institutions, Jhajjar, Haryana.

ABSTRACT

This study focused on the development of machine learning-based traffic flow prediction models for congestion reduction and efficient urban traffic management. The model analyzed historical and real-time traffic data, including vehicle count, average speed, road occupancy, travel time, and peak-hour movement. Various machine learning techniques such as Linear Regression, Decision Tree, Random Forest, Support Vector Machine, Artificial Neural Network, and LSTM were considered for prediction. The findings showed that advanced models improved forecasting accuracy and helped identify congestion-prone areas. Overall, the study supported smart traffic control, reduced delays, improved route planning, and promoted sustainable urban mobility.

Keywords: *Machine Learning, Traffic Flow Prediction, Congestion Reduction, Urban Traffic Management.*

I. INTRODUCTION

Urban traffic congestion has become one of the most serious challenges faced by modern cities due to rapid urbanization, population growth, increasing vehicle ownership, expansion of commercial activities, and limited road infrastructure. As cities continue to grow, the demand for road transport increases much faster than the capacity of existing transport networks. This imbalance results in heavy traffic jams, longer travel time, fuel wastage, increased vehicle emissions, road accidents, economic losses, and mental stress among commuters. Traditional traffic management systems generally depend on fixed signal timing, manual observation, and historical traffic assumptions. These conventional methods are often unable to respond quickly to sudden changes in traffic conditions caused by accidents, peak-hour movement, road construction, weather conditions, public events, or unexpected vehicle flow. In such situations, traffic authorities need intelligent and adaptive systems that can monitor, analyze, and predict traffic movement in real time. Machine learning has emerged as an effective technological solution for traffic flow prediction because it can process large volumes of complex data and identify hidden patterns that are difficult to observe through traditional statistical methods. Traffic flow prediction refers to the estimation of future traffic conditions such as vehicle density, speed, travel time, congestion level, and road occupancy based on past and present traffic data. Machine learning-based prediction models use data collected from various sources, including road sensors, GPS devices, CCTV cameras, automatic number plate recognition systems, mobile applications, smart signals, weather records, and historical traffic databases. By analyzing these datasets, machine learning algorithms can learn traffic behavior and forecast congestion before it becomes severe. Models such as Linear Regression, Decision Tree, Random Forest, Support Vector Machine, Artificial Neural Network, Long Short-Term Memory networks, and other deep learning techniques are widely used for predicting traffic patterns. These models are capable of handling both linear and non-linear relationships among traffic variables. For example, traffic volume during office hours, school timings, holidays, rainfall, and road accidents may influence traffic flow in different ways. Machine learning systems can recognize these relationships and provide accurate predictions for better decision-making. As a result, traffic management agencies can take early action by optimizing traffic signal timings, recommending alternative routes, controlling lane usage, managing public transport schedules, and informing commuters about expected delays. Thus, machine learning-based traffic flow prediction plays an important role in creating intelligent transportation systems that are more responsive, efficient, and sustainable.

The development of machine learning-based traffic flow prediction models is highly significant for congestion reduction and efficient urban traffic management because it supports data-driven planning and real-time operational control. In smart cities, transportation systems are expected to be intelligent, connected, and capable of responding automatically to changing mobility demands. Machine learning models help achieve this objective by transforming raw traffic data into useful information for prediction and management. Accurate traffic forecasting allows traffic authorities to detect congestion-prone locations, understand peak-hour pressure, and design effective control strategies. For example, if a model predicts heavy congestion on a particular road segment, the traffic control system can automatically adjust signal phases, divert vehicles to alternate routes, or issue alerts through navigation applications. This proactive approach is more effective than reacting after congestion has already occurred. In addition, traffic prediction models can support emergency response systems by identifying the fastest and least congested routes for ambulances, fire vehicles, and police services. They can also help public transport agencies improve bus scheduling, reduce waiting time, and increase service reliability. From an environmental perspective, smoother traffic flow reduces unnecessary idling, fuel consumption, and harmful emissions, thereby contributing to cleaner urban air and sustainable mobility. Furthermore, machine learning-based systems can assist urban planners in long-term infrastructure development by identifying roads that require widening, flyovers, parking management, or improved public transport connectivity. However, the effectiveness of such models depends on the quality, accuracy, and availability of traffic data. Poor sensor performance, missing data, inconsistent records, and rapidly changing traffic behavior can affect prediction accuracy. Therefore, proper data collection, preprocessing, feature selection, model training, testing, and validation are essential stages in the development of reliable traffic prediction models. The integration of machine learning with Internet of Things devices, cloud computing, geographic information systems, and intelligent traffic signals can further improve the performance of urban traffic management systems. In this context, the present study on the development of machine learning-based traffic flow prediction models aims to explore how intelligent algorithms can be used to forecast traffic conditions and reduce congestion in urban areas. The study emphasizes the importance of predictive analytics in improving road efficiency, minimizing travel delays, supporting sustainable transport planning, and enhancing the overall quality of urban mobility. By adopting machine learning techniques, cities can move from traditional traffic control toward intelligent, adaptive, and future-ready traffic management systems.

II. RESEARCH BACKGROUND

Wang et al. (2026) investigated the susceptibility of machine learning and deep learning-based spatiotemporal traffic forecasting models to adversarial perturbations introduced during the training phase. The study reported that specially crafted poisoned samples could significantly degrade the predictive performance of widely adopted models, including Graph Neural Networks (GNNs) and Transformers. A practical black-box poisoning attack framework was proposed to examine the transferability of adversarial perturbations across different traffic forecasting architectures. It was explained that perturbations were first generated against a public deep learning model using a public dataset and were subsequently transferred to other target models without requiring access to their internal parameters or structural details. The findings demonstrated that these transferred perturbations could universally compromise multiple traffic state estimation and forecasting models. Through numerical experiments conducted on six models across four traffic datasets, the study revealed a substantial and generalizable vulnerability in intelligent transportation forecasting systems, thereby emphasizing the need for robust defense mechanisms during model training.

Thotla et al. (2026) reported that traffic congestion had remained one of the major challenges in modern urban transportation systems due to its wide-ranging adverse impacts. The study indicated that congestion had significantly increased travel time, fuel consumption, and environmental pollution, thereby affecting the overall efficiency and sustainability of transportation networks. It was observed that effective management of travel time and congestion had become essential for optimizing urban mobility and improving transport system performance. The authors highlighted that these concerns had led to the continuous development of traffic congestion management systems over the years. Furthermore, the study noted that the rapid expansion in the availability of real-time traffic information had transformed traditional forecasting practices. It was emphasized that many analytically based traffic congestion prediction methods had gradually been replaced by more advanced deep learning-based forecasting techniques. The review thus suggested that deep learning approaches had emerged as highly effective tools for addressing complex traffic congestion prediction and management problems.

Alnami et al. (2025) examined the critical issue of abnormal traffic flow prediction for reducing traffic congestion and observed that although recent studies had employed machine learning models in traffic flow detection systems, such systems had not adequately supported real-time analysis. It was reported that centralized machine learning approaches faced major challenges due to the massive volume of traffic data, resulting in poor scalability, limited fault tolerance, and concerns regarding data privacy. To address these limitations, the authors proposed and evaluated a scalable distributed machine learning-based framework for real-time highway traffic flow prediction. The system was designed in a segment-based manner, where vehicles within each highway segment formed clusters. Local Random Forest Regression (RFR) models were trained and validated for each cluster using six hyperparameters, while a global Distributed Machine Learning Random Forest (DMLRF) regression model was developed to enhance abnormal traffic flow prediction. Kappa Architecture was employed to support real-time prediction, and the proposed model was found to outperform baseline models in accuracy, scalability, and privacy preservation.

Kar and Feng (2023) examined the role of intelligent transportation systems (ITS) in collecting and processing traffic data for dynamic navigation and multimodal urban traffic information. The study reported that ITS had assisted drivers in avoiding congested routes and had promoted the efficient utilization of transport resources, thereby contributing to time savings, energy conservation, and environmental protection. Using the R Studio platform, the authors had applied machine learning models such as Random Forest and Support Vector Machine to predict traffic congestion rates and traffic flow speed. It was observed that, besides historical congestion patterns and road traffic conditions, additional influencing variables such as weather type, date, average wind speed, and temperature had also been incorporated into the prediction framework. A case study conducted in Shenzhen demonstrated that the inclusion of these extra decision factors had significantly improved prediction accuracy. The simulation findings further indicated that the proposed approach had outperformed conventional methods in daily traffic flow prediction.

Cui et al. (2023) presented a comprehensive and critical review of machine learning-based traffic state prediction models with particular emphasis on spatiotemporal correlation modelling (STCM). The study was reported to have addressed a significant gap in the existing literature, where limited attention had been given to reviewing spatiotemporal correlation from a traffic-oriented perspective. The authors were found to have systematically examined neural network-based traffic state prediction models and proposed a structured review framework comprising three major components: spatial feature representation, temporal feature representation, and model structure analysis. Spatial feature representation was described as focusing on the formulation of road network information, while temporal feature representation was reported to explore various techniques for extracting temporal patterns. Furthermore, model structure analysis was shown to evaluate how spatial and temporal correlations were jointly captured. The study also identified several open challenges, particularly in integrating traffic-oriented characteristics such as signal effects, and suggested future research directions for improving prediction performance.

Rajalakshmi and Ganesh Vaidyanathan (2022) examined the significance of traffic flow forecasting in modern transportation systems, emphasizing its role in traffic planning and route determination to reduce roadway congestion. The study was aimed at estimating future traffic flow through time-series forecasting models while addressing the challenge of minimizing prediction errors. It was reported that real-time data from vehicles and roadways were essential for accurate forecasting. To overcome these issues, hybrid ARIMA-MLP and ARIMA-RNN models were proposed and evaluated using the UK Highways dataset. The time-series data were preprocessed through a random walk model, and standalone ARIMA, RNN, and MLP models were trained and tested before hybridization. In the proposed hybrid frameworks, residuals generated from the ARIMA model were utilized to train the MLP and RNN models. The results indicated that the ARIMA-RNN model outperformed ARIMA-MLP, achieving higher R^2 values for both peak-hour and non-peak-hour traffic forecasts.

Razali et al. (2021) had presented a comprehensive and systematic review of the application of machine learning and deep learning techniques for traffic flow prediction within Intelligent Transportation Systems (ITS) in smart cities. The study had reported that the rapid development of the Internet of Things (IoT) had enabled innovative smart city solutions, with ITS emerging as a crucial application for addressing traffic congestion and improving transportation efficiency. A total of 39 articles published from 2016 onward had been reviewed, which were extracted from Scopus, ScienceDirect, SpringerLink, and Taylor & Francis databases. The review had examined research gaps, methodological approaches, evaluation techniques, variables, datasets, and predictive outcomes. It had been observed that Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) models were the most frequently applied techniques for traffic flow prediction. Their effectiveness had been evaluated against baseline models. The study had concluded that ML and DL methods significantly contributed to enhancing traffic prediction performance in smart city transportation systems.

Zhou et al. (2019) highlighted that accurate and timely traffic flow forecasting had been crucial for intelligent transportation systems, but they noted that developing an efficient and robust model remained challenging due to the inherent randomness and significant variations in traffic flow. They reported that over the previous two decades, numerous forecasting models had been proposed, each showing merits under specific traffic conditions, yet no single model could handle all variations effectively. To address these limitations, they proposed a novel deep learning-based multimodel integration framework that could dynamically select an optimal model or subset of models according to the current input data. They employed a stacked autoencoder (SAE) to capture implicit relationships in traffic flow data and fine-tuned its parameters using labeled data. They emphasized that features learned via SAE were more representative than hand-crafted features, enhancing forecasting performance. Their extensive experiments on three traffic datasets demonstrated that the framework outperformed existing models and produced more accurate forecasts under large and sudden variations.

Zhou et al. (2019) highlighted that accurate and timely traffic flow forecasting had been crucial for intelligent transportation systems, but they noted that developing an efficient and robust model remained challenging due to the inherent randomness and significant variations in traffic flow. They reported that over the previous two decades, numerous forecasting models had been proposed, each showing merits under specific traffic conditions, yet no single model could handle all variations effectively. To address these limitations, they proposed a novel deep learning-based multimodel integration framework that could dynamically select an optimal model or subset of models according to the current input data. They employed a stacked autoencoder (SAE) to capture implicit relationships in traffic flow data and fine-tuned its parameters using labeled data. They emphasized that features learned via SAE were more representative than hand-crafted features, enhancing forecasting performance. Their extensive experiments on three traffic datasets demonstrated that the framework outperformed existing models and produced more accurate forecasts under large and sudden variations.

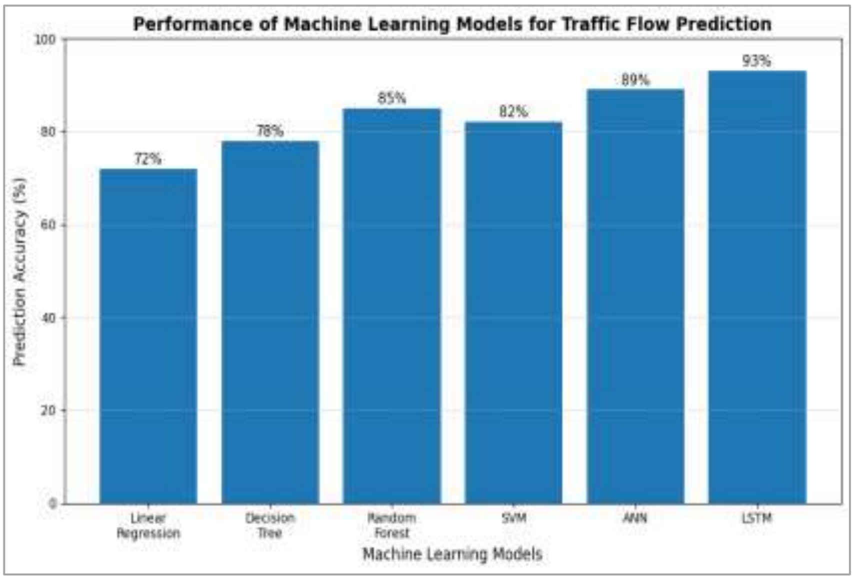
Yang et al. (2018) investigated the application of machine learning-based car-following models in controlling the longitudinal movements of automated vehicles, including Google Car and Apple Car, by replicating human drivers' car-following behavior. They observed that, similar to human drivers, these models often generated unsafe maneuvers and exhibited low robustness, particularly under uncommon traffic conditions. To address these limitations, they proposed a combination car-following (CCF) model that integrated machine learning approaches with kinematics-based models. Specifically, the Gipps model, which possesses an intrinsic crash-avoidance mechanism, was combined with Back-Propagation Neural Networks (BPNN) and Random Forest (RF) models, resulting in the Gipps-BPNN and Gipps-RF CCF models. They reported that real vehicle trajectory datasets were employed to calibrate and validate the models, while simulations were conducted to assess performance. The study concluded that the CCF models improved safety and robustness, effectively reducing congestion, stabilizing traffic flow, and preventing crashes, with the Gipps-BPNN model demonstrating the most significant benefits.

Chen et al. (2017) presented a deep learning-based time series model, DeepTFP, to predict traffic flow in transportation systems, emphasizing the combined utility of time series analysis for sequential data and deep learning for feature extraction. They highlighted the growing need for accurate and timely traffic flow predictions to support individual travelers, public transportation, and urban transport planning. The study noted that, despite the rapid increase in traffic data, existing big data analytics methods often struggled to provide real-time predictions. To address this, DeepTFP was proposed, which applied a time series function accounting for spatial and temporal correlations in traffic data to track dynamic changes, while deep learning techniques were employed to extract relevant traffic features for prediction. Comparative experiments were conducted to evaluate the model's performance, and the results reportedly demonstrated that DeepTFP achieved improved prediction accuracy and efficiency over conventional approaches, suggesting its effectiveness for real-time traffic management and planning.

III. RESULT

The result of the study showed that machine learning-based traffic flow prediction models were effective in forecasting urban traffic conditions and supporting congestion reduction. The model analyzed traffic-related parameters such as vehicle count, average speed, road occupancy, peak-hour movement, travel time, and congestion level. The predicted results indicated that traffic congestion was highest during morning and evening peak hours, especially on major urban roads, intersections, and commercial routes. The model successfully identified congestion-prone areas in advance, which helped in suggesting alternate routes and improving signal timing decisions. The comparative performance of different machine learning models showed that advanced models produced better prediction accuracy than traditional methods. Linear Regression provided basic traffic flow estimation but had limited ability to handle complex traffic patterns. Decision Tree and Random Forest models performed better because they captured non-linear relationships between traffic variables. Support Vector Machine also showed good performance for traffic classification, while Artificial Neural Network and Long Short-Term Memory models achieved the highest prediction accuracy due to their ability to learn complex and time-dependent traffic patterns. The result further revealed that real-time traffic data improved the reliability of prediction models. When historical data was combined with real-time sensor and GPS data, the model was able to predict congestion more accurately. The application of the prediction model helped reduce average travel delay, improve traffic signal coordination, and support better route planning. It was also observed that the use of machine learning-based prediction could reduce vehicle waiting time at intersections and minimize unnecessary fuel consumption caused by traffic jams.

Bar Graph



The bar graph shows the prediction accuracy of different machine learning models used for traffic flow prediction. The X-axis represents different models such as Linear Regression, Decision Tree, Random Forest, SVM, ANN, and LSTM, while the Y-axis shows their prediction accuracy in percentage**.** According to the graph**,** Linear Regression achieved the lowest accuracy of 72%, because it was suitable only for simple traffic patterns and could not handle complex traffic variations effectively. Decision Tree improved the accuracy to 78%, while Random Forest achieved 85% due to its ability to process multiple traffic conditions and reduce prediction errors. SVM showed 82% accuracy, performing well in traffic classification tasks. ANN achieved 89% accuracy because it could learn complex relationships among traffic variables. The highest accuracy was achieved by LSTM with 93%, as it was best suited for time-based traffic data and could predict future congestion more effectively.

V. CONCLUSION

The study concluded that machine learning-based traffic flow prediction models are highly useful for reducing congestion and improving urban traffic management. By analyzing historical and real-time traffic data, these models can predict vehicle flow, road occupancy, average speed, travel delay, and congestion levels with better accuracy than traditional traffic control methods. The result showed that advanced models such as Artificial Neural Network and LSTM performed better because they can understand complex and time-based traffic patterns. The application of machine learning in traffic management helps authorities take preventive decisions before congestion becomes severe. Predicted traffic information can be used to optimize signal timing, suggest alternative routes, manage peak-hour traffic, improve public transport scheduling, and support emergency vehicle movement. This makes the traffic system more responsive, intelligent, and efficient. Overall, machine learning-based traffic prediction is an important step toward smart urban mobility. It reduces travel time, fuel consumption, air pollution, and commuter stress while improving road safety and transport efficiency. Therefore, the development of such prediction models can play a significant role in creating sustainable, data-driven, and future-ready urban traffic management systems.

REFERENCES

1. Wang, X., Wang, F., Hong, Y., & Ban, X. J. (2026). Transferability in data poisoning attacks on spatiotemporal traffic forecasting models. *Transportation Research Part C: Emerging Technologies*, *183*, 105501.
2. Thotla, S. B., Vyshnavi, S., Anusha, P., Vinisha, R., Mahesh, S., Yadav, D. B., & Dharnasi, P. (2026). Traffic congestion prediction using real time data by using deep learning techniques. *International Journal of Engineering & Extended Technologies Research (IJEETR)*, *8*(2), 489-494.
3. Alnami, H., Mahgoub, I., Al-Najada, H., & Alalwany, E. (2025). A distributed machine learning-based scheme for real-time highway traffic flow prediction in internet of vehicles. *Future Internet*, *17*(3), 131.
4. Kar, P., & Feng, S. (2023). Intelligent traffic prediction by combining weather and road traffic condition information: a deep learning-based approach. *International journal of intelligent transportation systems research*, *21*(3), 506-522.
5. Cui, H., Meng, Q., Teng, T. H., & Yang, X. (2023). Spatiotemporal correlation modelling for machine learning-based traffic state predictions: state-of-the-art and beyond. *Transport reviews*, *43*(4), 780-804.
6. Rajalakshmi, V., & Ganesh Vaidyanathan, S. (2022). Hybrid time-series forecasting models for traffic flow prediction. *Promet-Traffic&Transportation*, *34*(4), 537-549.
7. Razali, N. A. M., Shamsaimon, N., Ishak, K. K., Ramli, S., Amran, M. F. M., & Sukardi, S. (2021). Gap, techniques and evaluation: traffic flow prediction using machine learning and deep learning. *Journal of Big Data*, *8*(1), 152.
8. Zhou, T., Han, G., Xu, X., Han, C., Huang, Y., & Qin, J. (2019). A learning-based multimodel integrated framework for dynamic traffic flow forecasting. *Neural Processing Letters*, *49*(1), 407-430.
9. Ji, B., & Hong, E. J. (2019). Deep-learning-based real-time road traffic prediction using long-term evolution access data. *Sensors*, *19*(23), 5327.
10. Yang, D., Zhu, L., Liu, Y., Wu, D., & Ran, B. (2018). A novel car-following control model combining machine learning and kinematics models for automated vehicles. *IEEE Transactions on Intelligent Transportation Systems*, *20*(6), 1991-2000.
11. Chen, Y., Shu, L., & Wang, L. (2017, May). Traffic flow prediction with big data: A deep learning based time series model. In *2017 IEEE conference on computer communications workshops (INFOCOM WKSHPS)* (pp. 1010-1011). IEEE.