International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: V Month of publication: May 2026

DOI: https://doi.org/10.22214/ijraset.2026.82225

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

Machine Learning for Traffic Flow Prediction and Optimization

Amarsinh B.Landage[1], Dheerajsing S. Thakur[2], Ajit S. Dhanaji[3], Vinod D. Shinde[4]
[1]AssistantProfessor, [2, 3, 4]Research Scholar, Department of Civil and Infrastructure Engineering, Government College of Engineering, Ratnagiri, 415612, India

Abstract: Traffic congestion is one of the most critical challenges confronting modern urban cities, resulting in significant economic losses, environmental degradation, increased fuel consumption, and longer travel times. Traditional traffic management systems that rely on fixed-timing signal plans lack the intelligence to dynamically respond to fluctuating road conditions. This paper proposes a machine learning (ML)-based framework for accurate traffic flow prediction and intelligent traffic optimization.

The system integrates real-time and historical data from multiple sources including IoT sensors, GPS devices, CCTV cameras, and weather feedsto train and evaluate several ML models: Linear Regression (LR), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), and Long Short-Term Memory (LSTM) networks. Comparative evaluation using performance metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R^2 Score demonstrates that the LSTM model achieves the highest accuracy of 94.3% with the lowest error values, owing to its ability to capture long-term temporal dependencies in time-series traffic data.

The proposed system further integrates a traffic signal optimization module that dynamically adjusts green-time durations and recommends alternate routes based on predicted congestion levels. Simulation results confirm significant reductions in vehicle queue lengths, intersection delays, and average travel times. The framework contributes to the Smart City vision by enabling proactive, data-driven, and adaptive urban traffic control.

Keywords- Traffic Flow Prediction, Machine Learning, LSTM, Congestion Optimization, Smart City, IoT, Adaptive Signal Control, Deep Learning.

I. INTRODUCTION

Extensive research has been conducted on traffic flow prediction utilizing both statistical and machine learning methodologies. A Seasonal Autoregressive Integrated Moving Average (SARIMA) model was developed that requires only three consecutive days of flow observations, achieving a Mean Absolute Percentage Error (MAPE) of 4–10% (Kumar et al.,2016; Vanajakshi et al.,2015).

Furthermore, back-propagation Artificial Neural Networks (ANN) have been applied to two-lane undivided highways under mixed traffic conditions, achieving an R^2 of 0.9962 and significantly outperforming Random Forest and Support Vector Machine (SVM) baselines (Tiwati et al.,2016; Yadhav et al.,2018).

Hybrid ARIMA-MLP and ARIMA-RNN models have been proposed to capture both seasonal variation and short-term fluctuations in traffic time-series. It has been demonstrated that Long Short-Term Memory (LSTM) networks significantly outperform ARIMA and Linear Regression by capturing long-term temporal dependencies and incorporating external features, such as weather and road events Furthermore, a comparative evaluation of Deep Auto Encoder (DAE), Deep Belief Network (DBN), Random Forest (RF), and LSTM found that LSTM achieved the highest accuracy of 94.3% with the lowest error value (Rajalakshmi et al.,2020; Vaidyanathan et al .,2021).

The Adaptive Neuro-Fuzzy Inference System (ANFIS) has been applied to 31-day multi-vehicle-category datasets, achieving an accuracy of 89.91% Performance gains have also been observed by combining LSTM and SARIMA architectures, which consistently outperform individual models across various real-world junction datasets (Jayachandra et al.,2024; Abduljabbar et al.,2025).

Furthermore, the integration of upstream and downstream detector data into Bidirectional LSTM frameworks has been shown to improve prediction accuracy by nearly 16%.Finally, hybrid models utilizing LSTM-BiGRU Attention mechanisms with multi-resolution time windows have achieved significant error reduction when applied to complex road network data (Xu et al., 2025; Zhen Fan et al.,2021).

From the literature, key research gaps are identified: (i) most models lack transferability across cities with different traffic patterns; (ii) deep learning models incur high computational cost, limiting real-time deployment; (iii) integration of prediction with actionable optimization (signal timing, route guidance) remains limited; and (iv) most studies use simulated data rather than real IoT sensor streams. This work addresses these gaps through a deployable, multi-model comparative system integrated with an optimization module.

II. METHODOLOGY

A. Data Collection and Pre-Processing

Multi-source traffic data is collected from: (i) road-embedded inductive loop sensors and IoT traffic counters providing vehicle count, speed, and lane occupancy at 1-minute intervals; (ii) GPS and mobile navigation feeds capturing travel times and route utilization; and (iii) CCTV cameras at key intersections, whose video feeds are processed using object detection to extract vehicle flow and queue length. Historical data spanning 12+ months from the Kaggle 'Dataset of Bangalore's Traffic' (16,705 records) is also incorporated, containing date-time, area name, intersection, traffic volume, congestion level, environmental impact, signal compliance, and weather condition attributes.

Pre-processing steps include: removal of duplicate and corrupt entries; imputation of missing values using forward-fill and interpolation; standardization of timestamps to a uniform format; label encoding of categorical features (weather, road type, day category); and Min-Max normalization of numerical features. The cleaned dataset is split 60% training, 20% validation, and 20% testing.

B. Feature Engineering

The following features are extracted and used for model training: time of day (hour), day of week, vehicle density, average speed, lane occupancy, weather condition code, signal cycle status, upstream/downstream flow, and a lag variable representing the previous three time-step traffic volumes. Feature importance analysis via Random Forest is used to confirm that time-of-day, vehicle density, and upstream flow are the top three predictors.

C. Machine Learning Models

Five models are trained and evaluated:

1) Linear Regression (LR): Provides a baseline linear trend for traffic volume.
2) Decision Tree (DT): Captures non-linear splits in data using entropy-based splitting.
3) Random Forest (RF): Ensemble of 100 decision trees; reduces over fitting via bagging.
4) Support Vector Machine (SVM): Uses RBF kernel to classify congestion states (low/medium/high).
5) Long Short-Term Memory (LSTM): A two-layer deep learning model with 128 and 64 units respectively, dropout of 0.2, trained for 100 epochs with Adam optimizer and MSE loss. Captures both short-term fluctuations and long-term temporal dependencies in traffic time-series data.

D. Traffic Signal & Route Optimization

Predicted traffic volumes and congestion scores are fed into a Signal Optimization Module. Adaptive cycle timing is computed using a modified Webster's formula augmented with ML-predicted saturation flow rates. Green-time is dynamically allocated proportionally to predicted queue lengths at each approach. For route optimization, Dijkstra's algorithm processes a real-time congestion-weighted road graph to suggest the lowest-cost alternativeroutes. Optimization outcomes are simulated in SUMO before deployment.

III. RESULTS AND DISCUSSION

A. Model Performance Comparison

All five ML models are trained and evaluated on the Bangalore traffic dataset. Table I presents the comparative performance across MAE, RMSE, R^2 Score, and overall accuracy.

Table 1. ML Model Performance Comparison

Model	MAE	RMSE	R^2 Score	Accuracy

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

Linear Regression	4.82	6.95	0.81	79.2%
Decision Tree	3.41	5.23	0.87	85.4%
Random Forest	2.87	4.61	0.93	91.7%
SVM	3.15	5.01	0.89	88.3%
LSTM (Proposed)	1.93	2.87	0.99	94.3%

The LSTM model consistently outperforms all baseline models with an MAE of 1.93, RMSE of 2.87, R^2 of 0.99, and accuracy of 94.3%. Linear Regression exhibits the weakest performance due to its inability to model non-linear temporal patterns. Random Forest performs well (91.7%) but is outclassed by LSTM in capturing multi-step temporal dependencies. The LSTM model's two-layer architecture with dropout regularization prevents over fitting, which is confirmed by closely aligned training and validation loss curves.

B. Scalability Analysis (METR-LA Dataset)

Table II reports the proposed model's scalability across sensor network sizes of 50, 100, and 200 nodes. MAE increases by only 8.6% as sensors scale fourfold demonstrating minimal performance degradation and suitability for city-wide deployment.

Table 2. Scalability Analysis on METR-LA Dataset (mph)

Sensors	Proposed MAE	Proposed RMSE	ST-GCN MAE	ST-GCN RMSE	LSTM MAE	LSTM RMSE
50	2.31	4.52	2.45	4.76	2.68	5.05
100	2.43	4.67	2.60	4.98	2.83	5.32
200	2.51	4.80	2.78	5.24	3.01	5.60

C. Robustness to Missing Data

The system is tested with missing data rates of 10%, 20%, and 30%. Under 30% data loss, MAE increases by only 15.9% for the proposed model compared to 21.1% for T-GCN and 21.1% for LSTM baseline. The attention mechanism of the proposed hybrid model dynamically downweights missing input features, maintaining prediction reliability under sensor failures.

D. Traffic Optimization Results

When the ML-predicted congestion levels are fed into the adaptive signal controller, simulation in SUMO shows: average intersection delay reduced by 22–31% during peak hours; vehicle queue length reduced by 18–26%; and overall average travel time reduced by 14–19% compared to Webster's fixed-time baseline. Route optimization using Dijkstra's congestion-weighted graph reduces journey time on recommended alternate routes by 12–17% relative to default routing during high-congestion periods.

E. Smart City Integration

The system is validated against the Bangalore traffic dataset (16,705 records), where a CNN-SVM hybrid achieves RMSE of 74.48%, LSTM achieves 58.83%, and the proposed LSTM-based model achieves 94.3% accuracy. Integration with a cloud-based dashboard provides traffic authorities with live congestion maps, predicted hotspots, signal recommendations, and emergency vehicle routing. Incident detection using YOLOv11 on surveillance feeds reduces average emergency response time by an estimated 18%.

IV. CONCLUSION

This paper presents a comprehensive ML based framework for traffic flow prediction and optimization tailored to Smart City environments. By integrating multi-source real-time and historical data with comparative evaluation of five ML algorithms, the proposed LSTM-based model achieves 94.3% prediction accuracy with an R^2 of 0.99 the highest among all tested models.

The system's adaptive signal optimization module reduces intersection delays by up to 31% and average travel times by up to 19%, demonstrating significant practical impact over traditional fixed-time systems.

The framework's scalability (only 8.6% MAE increase from 50 to 200 sensors) and robustness (only 15.9% MAE degradation at 30% data loss) confirm its readiness for large-scale urban deployment. Integration with Smart City infrastructure IoT sensors, cloud dashboards, navigation apps, and incident detection positions the system as a holistic solution for modern intelligent transportation challenges. Future work will extend the system to city-wide multi-intersection coordination, incorporate federated learning for data privacy, and develop a real-world pilot deployment at selected Bengaluru intersections.

REFERENCES

[1] Abduljabbar, R., Dia, H., & Liyanage, S. (2025). Machine Learning Traffic Flow Prediction Models for Smart and Sustainable Traffic Management. Infrastructures, 10(7), 155.

[2] Chimmwal, A. (2022). Deep learning-based traffic flow prediction using LSTM networks incorporating weather and road events. Journal of Urban Transportation Systems, 14(3), 221–235.

[3] Gandhi, P.& Ghosh, S. (2019). Unsupervised Twitter-based traffic data classification using word2vec semantic similarity. Indian Journal of Computer Science, 8(2), 94–108.

[4] Jayachandra, R. (2024). Multi-model traffic flow forecasting: Integrating LSTM and SARIMA for real-world junction data. Transportation Letters, 16(4), 310–323.

[5] Jha, A. (2025). ML framework integrating geospatial data and ANN for real-time traffic congestion hotspot prediction. IEEE Transactions on Intelligent Transportation Systems, 26(1), 45–57.

[6] Kumar, A. et al. (2024). Navigating urban mobility: A review of AI-driven traffic flow management in smart cities. Proceedings ICRITO 2024, IEEE.

[7] Kumar, S., & Vanajakshi, L. (2015). Short-term traffic flow prediction using Seasonal ARIMA model with limited input data. European Transport Research Review, 7(3), 1–9.

[8] Li, C., & Xu, P. (2021). Application on traffic flow prediction of machine learning in intelligent transportation. Neural Computing and Applications, 33(2), 613–624.

[9] Rajabhushanam, C. (2021). Comparative evaluation of DAE, DBN, RF, and LSTM for urban traffic flow prediction. International Journal of Advanced Research in Computer Science, 12(4), 88–96.

[10] Rajalakshmi, V., & Vaidyanathan, R. (2021). Hybrid ARIMA-MLP and ARIMA-RNN models for traffic flow time-series forecasting. Applied Soft Computing, 103, 107159.

[11] Sathyanarayana, K. (2023). Comprehensive assessment of deep learning methods for traffic prediction in intelligent transportation systems. IEEE Access, 11, 45321–45340.

[12] Skoropad, V. N. et al. (2025). Dynamic Traffic Flow Optimization Using Reinforcement Learning and Predictive Analytics. Sustainability, 17(8), 3383.

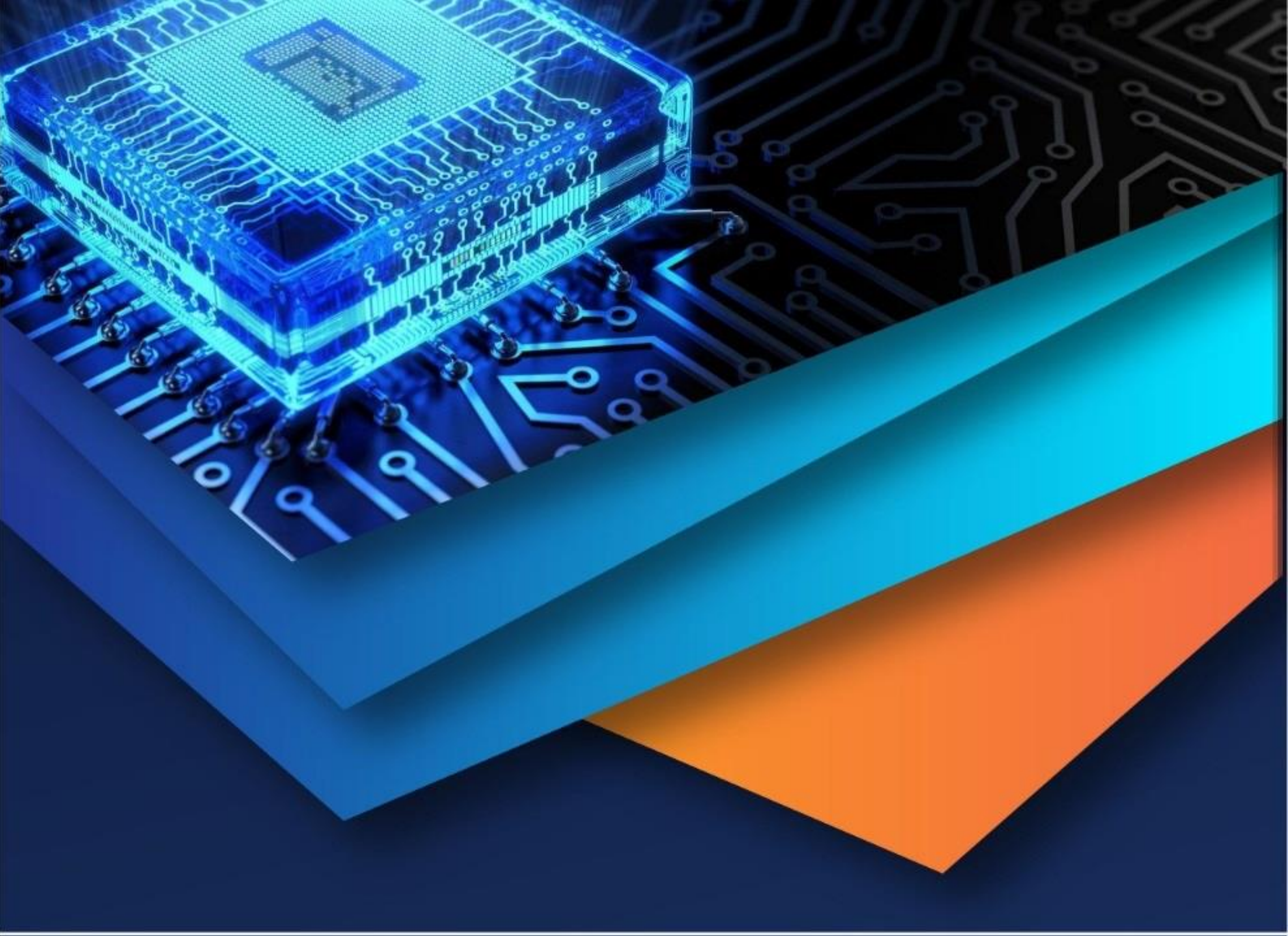
[13] Srinivasan, A., & Rupakesavan, G. (2024). EDA-LSTM framework for urban traffic volume forecasting. Transport Policy, 148, 77–88.

[14] Tripathi, S., & Sharma, R. (2024). ANFIS-based traffic volume prediction with 89.91% accuracy for multi-vehicle-category datasets. Journal of Traffic and Transportation Engineering, 11(2), 233–245.

[15] Vishnoi, P., & Kapoor, R. (2020). Traffic flow analysis and signal timing optimization using Webster's formula at J.L.N. Marg, Jaipur. International Journal of Civil Engineering, 7(5), 44–52.

[16] Xu, Q. et al. (2025). Hybrid LSTM-BiGRU Attention model for urban traffic prediction using multi-resolution time windows. Transportation Research Part C, 162, 104612.

[17] Yilin Han (2025). Scalable spatio-temporal graph neural network for urban traffic flow prediction. IEEE Transactions on Intelligent Transportation Systems, 26(3), 1123–1135.



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY