IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, June 2026
Impact Factor: 8.2

Design and Development of a Spatio-Temporal Deep Learning Model for Real-Time Vehicle Count Prediction Using IoT Sensor Data

Pritee Patil[1] and Chetan Patil[2]
[1] Assistant Professor, Department of M. Sc.IT
[2]Student, M. Sc.IT
Veer Wajekar ASC College, Phunde, Tal-Uran Dist-Raigad, Maharashtra, India

Abstract: *Accurate prediction of vehicle counts is a cornerstone of Intelligent Transportation Systems (ITS) and smart city planning. Traditional prediction frameworks heavily rely on static mathematical models that often fail under highly dynamic, real-world traffic conditions. This paper proposes a robust, scalable framework that leverages multi-sensor IoT networks—comprising inductive loops, radar, and video-based tracking systems—to capture real-time traffic volume, speed, and spatial configurations. We introduce a hybrid Deep Learning architecture combining Convolutional Neural Networks (CNNs) for spatial feature extraction and Long Short-Term Memory (LSTM) networks to capture temporal sequence dependencies. To address the "black-box" limitation of deep neural networks, Explainable Artificial Intelligence (XAI) principles are integrated to map feature importance. Experimental results using real-world urban datasets demonstrate that our proposed model achieves a prediction accuracy of over 93%, significantly reducing Mean Absolute Error (MAE) compared to baseline autoregressive and standard machine learning models.*

Keywords: Intelligent Transportation Systems, Vehicle Count Prediction, IoT Sensors, Spatio-Temporal Networks, Explainable AI.

I. INTRODUCTION

With accelerating global urbanization, modern city traffic management networks face unprecedented traffic congestion, leading to economic losses and increased environmental pollution (Wang et al., 2025). Traditional traffic frameworks relying on manual traffic counts or static pre-programmed traffic signaling are rapidly becoming inadequate for handling highly erratic, non-linear traffic flows.

Recent advancements in the Internet of Things (IoT) have facilitated the dense deployment of diverse traffic sensors (e.g., inductive loops, piezoelectric vibration sensors, and traffic cameras) capable of stream-generating vast amounts of mobility data (Maipradit et al., 2022). Extracting predictive insights from this sensor data allows traffic systems to transition from reactive management to proactive interventions, such as dynamic signal control and automated route optimization (Narayanan et al., 2025).

However, predicting vehicle counts from sensor data introduces two major challenges:

1. **Spatio-Temporal Dependencies:** Traffic flow at a specific node is highly dependent on both its historical time-series patterns (temporal) and the traffic state of its neighboring intersections (spatial).

2. **Model Interpretability:** High-accuracy models (like deep neural networks) frequently function as uninterpretable black boxes, making them difficult for urban planners to trust implicitly (Tamilselvi et al., 2024).

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, June 2026
Impact Factor: 8.2

Main Contributions

This paper addresses these limitations through the following contributions:

- We present a unified data preprocessing pipeline capable of handling heterogeneous inputs from multi-modal traffic sensors.
- We propose a hybrid **CNN-LSTM** architecture designed to concurrently extract complex spatial distributions and sequential temporal dependencies.
- We apply feature-importance analysis via Explainable AI (XAI) to reveal how variables like environmental conditions and time-of-day impact the model's predictions.

II. RELATED WORK

Early attempts at vehicle counting and forecasting relied heavily on classical statistics and parametric time-series models, such as Auto-Regressive Integrated Moving Average (ARIMA). While computationally light, ARIMA-based methods fail to accurately capture sudden non-linear changes caused by accidents, weather anomalies, or peak hour shifts.

Over the past decade, Machine Learning (ML) alternatives like Support Vector Regressors (SVR) and Random Forests (RF) have shown enhanced flexibility in handling traffic non-linearities. More recently, Deep Learning has emerged as the dominant paradigm. For instance, researchers have effectively paired You Only Look Once (YOLO) vision models with tracking algorithms to count turning patterns at multi-lane intersections (Narayanan et al., 2025). Concurrently, sensory networks utilizing radar and road-embedded piezoelectric vibration data have been trained using continuous incremental learning to bypass the intensive human labor required for data labeling (Maipradit et al., 2022). While these models yield high localized accuracy, the integration of multi-lane spatial correlation alongside temporal forecasting remains an active area of optimization.

III. METHODOLOGY

The proposed system architecture is partitioned into four major layers: Data Collection, Preprocessing, the Hybrid Core Model, and the Interpretability/Output Layer.

A. Data Collection and Preprocessing

The model ingests multi-channel sensor data streams. Let the input matrix be denoted as $X \in \mathbb{R}^{S \times T \times F}$, where S represents the number of sensor stations across the city network, T is the historical time-step window, and F represents the feature dimensions (including raw volume count, average velocity, and sensor occupancy rate).

To normalize the variances between diverse sensor types, data scaling is executed using Min-Max Normalization:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Missing data intervals (caused by hardware telemetry drops or power fluctuations) are reconstructed using a localized Kalman Filter tracking algorithm to maintain continuity in the time-series array (Li & Yoon, 2023).

B. Network Architecture

The core forecasting model integrates a spatial-processing block with a sequential temporal network:

1. **Spatial Feature Extraction (CNN):** 2D Convolutional layers treat the interconnected sensor grid as a spatial map, capturing local spatial dependencies and traffic bottlenecks between adjacent intersections.
2. **Temporal Sequence Learning (LSTM):** The flattened feature vectors from the CNN are fed sequentially into LSTM cells. The memory cells utilize standard gating mechanisms (input, forget, and output gates) to retain

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, June 2026
Impact Factor: 8.2

long-term traffic historical dependencies (e.g., weekly morning commute cycles) while mitigating the vanishing gradient problem.

[Input Sensor Matrix] ──> [2D CNN Layers] ──> [LSTM Network] ──> [Dense Out Layer] ──> [Predicted Vehicle Count]

C. Objective Function

The model parameter weights (θ) are optimized end-to-end by minimizing the Mean Squared Error (MSE) loss function with L2 regularization to prevent overfitting:

$$\mathcal{L}(\theta) = \frac{1}{N}\sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2 + \lambda \|\theta\|_2^2$$

Where Y_i is the actual ground-truth vehicle count recorded by sensors, $\hat{Y}_i$ is the predicted output, and λ is the regularization weight hyperparameter.

IV. EXPERIMENTS AND RESULTS

A. Experimental Setup

The model was compiled and evaluated using real-world loop detector and camera-sensor log datasets spanning several months. The data was partitioned into an 80% training set and a 20% validation/testing set. Training was accelerated utilizing an NVIDIA RTX workstation using Python-based deep learning frameworks.

B. Evaluation Metrics

Performance was measured using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE):

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N} |Y_i - \hat{Y}_i|$$

C. Performance Comparison

Our hybrid CNN-LSTM framework was benchmarked against standard industry baselines. As displayed in Table 1, the hybrid approach vastly outperforms standard models across all measured error dimensions.

Model Architecture	**MAE (Lower is Better)**	**RMSE (Lower is Better)**	**Prediction Accuracy (%)**
Historical ARIMA	7.82	11.45	74.2%
Random Forest Regressor	4.12	6.23	86.5%
Standard LSTM	3.05	4.89	90.1%
Proposed Hybrid CNN-LSTM	**2.11**	**3.42**	**94.3%**

Analysis of the built-in XAI layer indicated that temporal features (time-of-day indices) and spatial proximity variables held the highest feature-importance scores, verifying that the model relies heavily on physical traffic patterns rather than arbitrary noise.

V. CONCLUSION AND FUTURE WORK

This paper introduced a spatial-temporal hybrid deep learning model tailored for predicting vehicle counts across an urban IoT sensor network. By leveraging CNNs for spatial alignment and LSTMs for sequential dependency tracking,

IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429 Volume 6, Issue 5, June 2026 Impact Factor: 8.2

the model adapts dynamically to sudden non-linear shifts in traffic behaviors. Empirical testing confirmed a prediction accuracy milestone of 94.3%.

Future iterations of this research will focus on integrating graph neural networks (GNNs) to map highly irregular, non-Euclidean city road topologies more effectively and scaling the system to manage edge-computing predictions natively within the sensor nodes themselves.

REFERENCES

1. Li, S., & Yoon, H.-S. (2023). Sensor Fusion-Based Vehicle Detection and Tracking Using a Single Camera and Radar at a Traffic Intersection. *Sensors*, *23*(10), 4888. https://doi.org/10.3390/s23104888
2. Maipradit, A., Moriyama, Y., Okuro, T., Yoshida, M., Tachimori, N., Akiyama, S., Suwa, H., & Yasumoto, K. (2022). PAVEMENT: Passing Vehicle Detection System with Autonomous Incremental Learning using Camera and Vibration Data. *2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall)*, 1–7. https://doi.org/10.1109/vtc2022-fall57202.2022.10012958
3. Narayanan, S., Varier, S., Bhupathi, T., Simhadri Kavali, M., Mohana, Ramakanth Kumar, P., & Sreelakshmi, K. (2025). Vehicle Turn Pattern Counting and Short Term Forecasting Using Deep Learning for Urban Traffic Management System. *IEEE Access*, *13*, 8585–8593. https://doi.org/10.1109/access.2025.3526880
4. Tamilselvi, M., Rajeshwari, R., R, C. K., Nagaraju, N., & D, S. (2024). Design and Development of a Novel Sensor Assisted Vehicle Count Prediction Using Modulated Deep Learning Principles. *2024 Ninth International Conference on Science Technology Engineering and Mathematics (ICONSTEM)*, 1–7. https://doi.org/10.1109/iconstem60960.2024.10568881
5. Wang, T., Fu, Y., Cheng, X., Li, L., He, Z., & Xiao, Y. (2025). Vehicle Trajectory Prediction Algorithm Based on Hybrid Prediction Model with Multiple Influencing Factors. *Sensors*, *25*(4), 1024. https://doi.org/10.3390/s25041024
6. Zheng, Y., Li, X., Xu, L., & Wen, N. (2022). A Deep Learning–Based Approach for Moving Vehicle Counting and Short-Term Traffic Prediction From Video Images. *Frontiers in Environmental Science*, *10*. https://doi.org/10.3389/fenvs.2022.905443