

MACHINE LEARNING-BASED TRAFFIC FLOW PREDICTION AND CONGESTION ANALYSIS

Ramesh Nimmana¹, Dr. Rakesh Verma²

**Research Scholar, Department of Engineering & Technology, Mangalayatan University,
Aligarh, Uttar Pradesh, India¹**

**Supervisor, Department of Engineering & Technology, Mangalayatan University, Aligarh,
Uttar Pradesh, India²**

ABSTRACT

The rapid escalation of urbanization has intensified vehicular density across metropolitan road networks, creating unprecedented challenges in traffic congestion, environmental degradation, road safety, and commuter productivity. Conventional fixed-timer signal systems are fundamentally inadequate for managing dynamic, non-linear traffic patterns characteristic of modern smart cities. This empirical study presents the design, implementation, and rigorous evaluation of an Intelligent Traffic Management System (ITMS) that synergistically integrates Internet of Things (IoT) sensor networks with advanced Machine Learning (ML) algorithms to achieve real-time, adaptive traffic control. A heterogeneous IoT infrastructure comprising ultrasonic vehicle detectors, infrared cameras, RFID tag readers, GPS probe modules, and environmental sensors was instrumented across twelve signalized intersections in a mid-sized urban corridor over a continuous six-month observation window. The resulting dataset, encompassing over 2.8 million timestamped traffic observations across vehicular volume, queue length, average speed, intersection occupancy, and ambient conditions, was processed through edge computing nodes and transmitted via LoRaWAN and 5G communication protocols to a centralized cloud analytics platform. Six supervised and deep learning models Logistic Regression, Support Vector Machine, Decision Tree, Random Forest, Long Short-Term Memory (LSTM), and a novel Hybrid CNN-LSTM architecture were trained, validated, and benchmarked for congestion prediction and optimal signal phase duration estimation. The Hybrid CNN-LSTM model achieved the highest predictive accuracy of 94.2%, with 93.5% precision and 92.8% recall, outperforming the best baseline (Random Forest at 88.6%) by 5.6 percentage points. Field deployment of the proposed system resulted in measurable improvements: average vehicle waiting time reduced by 38.2%, intersection throughput increased by 27.4%, fuel consumption decreased by 22.1%, and CO₂ emissions dropped by 18.7% compared to the legacy fixed-timer system. Statistical significance was confirmed via paired t-tests ($p < 0.001$). These findings demonstrate the transformative potential of ML-driven IoT architectures in building sustainable, responsive, and intelligent

urban transportation ecosystems and provide a replicable empirical framework for smart city planners and transportation engineers.

KEYWORDS: Intelligent Transportation Systems (ITS)¹, Internet of Things (IoT)², Machine Learning³, Traffic Congestion Prediction⁴, Adaptive Signal Control⁵, Smart Cities⁶, Edge Computing⁷.

1. INTRODUCTION

A. BACKGROUND AND MOTIVATION

Global urbanization has reached a critical inflection point: by 2050, nearly 68% of the world's population is projected to reside in urban areas, placing enormous strain on existing transportation infrastructure. Traffic congestion alone is estimated to cost major economies between 1% and 3% of GDP annually through lost productivity, wasted fuel, increased emissions, and elevated accident rates. Traditional traffic signal systems, governed by fixed-cycle timers or simple actuated controllers, operate on static assumptions about traffic demand and are incapable of adapting to real-time fluctuations caused by accidents, weather events, special gatherings, or time-varying commuter patterns. The convergence of the Internet of Things (IoT) and Machine Learning (ML) presents a paradigm-shifting opportunity to transform passive traffic infrastructure into intelligent, self-optimizing ecosystems capable of perceiving, predicting, and responding to dynamic mobility conditions in real time.

B. PROBLEM STATEMENT

Despite the proliferation of smart city initiatives worldwide, the majority of signalized intersections continue to rely on outdated control logic that treats traffic as a static or periodic phenomenon. The fundamental problem addressed in this study is the absence of an integrated, data-driven framework that can: (a) continuously sense multi-dimensional traffic parameters through cost-effective IoT instrumentation; (b) learn complex spatiotemporal congestion patterns using scalable ML models; and (c) autonomously optimize signal timing plans to minimize delay, maximize throughput, and reduce environmental impact. Existing siloed solutions that address sensing without prediction, or prediction without closed-loop control, fail to capture the full potential of cyber-physical traffic management. Moreover, there is a lack of large-scale, longitudinal empirical evidence quantifying the real-world performance gains of ML-IoT integrated systems against legacy approaches under heterogeneous urban conditions.

C. RESEARCH OBJECTIVES AND CONTRIBUTIONS

This paper makes four principal contributions: (1) the design and deployment of a three-layer IoT sensing architecture for real-time traffic data acquisition across twelve urban intersections; (2) a comprehensive comparative evaluation of six ML models for short-term traffic congestion prediction using a six-month empirical dataset; (3) the development and field validation of a closed-loop adaptive signal control mechanism driven by ML predictions; and (4) a rigorous statistical analysis quantifying system performance across mobility, environmental, and economic dimensions. The remainder of this paper is organized as follows: Section II surveys related literature; Section III details the system methodology; Section IV presents data collection and

empirical analysis; Section V discusses findings and compares results with prior work; and Section VI concludes with future research directions.

2. SURVEY OF RELATED WORK

The intersection of IoT and ML for intelligent transportation has attracted substantial research attention in recent years. Early foundational work by Zhang et al. [1] demonstrated the feasibility of RFID-based vehicle detection for real-time traffic monitoring, achieving detection rates above 90% but lacking predictive capabilities. Gupta and Sharma [2] proposed a sensor-fusion architecture combining inductive loops with CCTV feeds, though their system relied on rule-based control logic that could not adapt to non-recurrent congestion. Transportation research by Chen and Liu [3] introduced k-Nearest Neighbors for short-term traffic flow prediction on highway datasets, reporting mean absolute percentage errors (MAPE) of approximately 12%, which motivated subsequent exploration of more sophisticated temporal models. In the deep learning era, significant advances have been documented. Wang et al. [4] applied Stacked Autoencoders to predict urban traffic speed, achieving 89.1% accuracy on benchmark datasets. Kumar and Reddy [5] proposed an IoT-enabled smart parking management system that, while tangential to signal control, established important precedents for edge analytics and low-power wide-area network (LPWAN) communication in vehicular contexts. Lv et al. [6] demonstrated that LSTM networks outperform traditional ARIMA models for traffic flow forecasting, reporting accuracy improvements of 15%–20%. Convolutional Neural Networks (CNNs) have been employed by Ma et al. [7] to extract spatial congestion features from traffic camera imagery, achieving classification F1-scores of 0.89. The work of Ahmed and Hussain [8] integrated vehicular ad-hoc networks (VANETs) with cloud-based analytics for incident detection, though latency constraints limited real-time applicability.

More recently, hybrid architectures combining spatial and temporal feature extraction have emerged. Singh and Verma [9] proposed a CNN-GRU model for intersection-level congestion prediction, reporting 91.2% accuracy a benchmark that the present study seeks to exceed. Reinforcement learning approaches by Mousavi et al. [10] have explored adaptive signal control using deep Q-networks, though convergence stability remains a practical challenge. Kumar et al. [11] surveyed IoT-based traffic systems across 45 smart city deployments, identifying critical gaps in standardized evaluation metrics and longitudinal validation. Edge computing frameworks proposed by Patel and Joshi [12] and Nguyen et al. [13] address latency by pushing ML inference closer to roadside units, achieving end-to-end response times under 200 ms. Environmental impact assessments by Sharma et al. [14] and Rahman et al. [15] have quantified emission reductions of 12%–22% through intelligent signal optimization, providing important baselines for the current study's sustainability metrics. Despite these advances, a comprehensive empirical investigation integrating multi-modal IoT sensing, systematic ML model benchmarking, closed-loop field deployment, and multi-dimensional impact assessment over an extended observation period remains conspicuously absent from the literature a gap this study directly addresses.

3. METHODOLOGY

This study adopts a design-science research methodology comprising three sequential phases: system design and instrumentation, model development and training, and field deployment with comparative evaluation. In the first phase, a three-layer IoT architecture was conceptualized and implemented. The Perception Layer consisted of

ultrasonic sensors (HC-SR04), passive infrared (PIR) motion detectors, high-definition IP cameras (1080p @ 30fps), RFID readers (UHF 865–867 MHz), GPS probe devices mounted on public transit vehicles, and environmental sensors (temperature, humidity, PM2.5). These devices were instrumented at twelve signalized intersections selected through stratified random sampling to represent diverse land-use contexts (commercial, residential, mixed-use, and arterial corridors). The Communication Layer employed LoRaWAN for low-power, long-range sensor telemetry and 5G C-V2X for high-bandwidth camera streams, with edge computing nodes (Raspberry Pi 4, NVIDIA Jetson Nano) performing preliminary data filtering, aggregation, and anonymization at the roadside [24]. The Application Layer hosted on a private cloud infrastructure provided centralized data storage (InfluxDB time-series database), model training environments (TensorFlow 2.x, scikit-learn), and a real-time control interface for traffic signal actuation via the NTCIP protocol. In the second phase, raw IoT data underwent rigorous preprocessing: missing value imputation using forward-fill for gaps under 30 seconds and linear interpolation for longer gaps; outlier detection and removal using the Interquartile Range (IQR) method with a $1.5 \times \text{IQR}$ threshold; Min-Max normalization to scale features to $[0,1]$; and sliding-window segmentation with a 5-minute input window and 1-step-ahead prediction horizon. Six ML models were trained and evaluated: Logistic Regression (LR), Support Vector Machine (SVM) with RBF kernel, Decision Tree (DT), Random Forest (RF, 100 estimators), LSTM (two hidden layers, 64 and 32 units), and a novel Hybrid CNN-LSTM architecture featuring a 1D convolutional front-end (two Conv1D layers with 32 and 64 filters, kernel size 3) for spatial feature extraction from multi-lane sensor arrays, followed by an LSTM backbone (128 units) for temporal sequence modeling.

All models were trained using the Adam optimizer (learning rate 0.001), mean squared error loss, batch size 64, and early stopping with patience of 15 epochs. The dataset was partitioned into training (70%), validation (15%), and test (15%) subsets using chronological splitting to prevent temporal data leakage. In the third phase, the best-performing model (Hybrid CNN-LSTM) was deployed for closed-loop adaptive signal control across the twelve study intersections over a 30-day trial period [23]. The control algorithm operated on a rolling 15-second optimization cycle, ingesting real-time IoT feeds, generating congestion predictions for the subsequent 5-minute window, and dynamically adjusting green-phase durations within safety-constrained bounds (minimum green 10 s, maximum green 90 s) for each approach. Performance was benchmarked against the pre-existing fixed-timer system using identical observation windows and traffic demand profiles. Paired two-tailed t-tests with $\alpha = 0.05$ were applied to assess statistical significance of performance differences. Ethical clearance was obtained from the municipal traffic authority, and all data collection complied with applicable privacy regulations, with automated license plate blurring and edge-level anonymization ensuring no personally identifiable information was stored or transmitted.

4. DATA COLLECTION AND ANALYSIS

This section presents the empirical data collected during the six-month study period (January–June 2023) and provides detailed analytical interpretation of each dataset. Five tabular datasets are presented and analyzed to build a comprehensive evidence base supporting the system's efficacy. All data were collected from the twelve instrumented intersections, with continuous 24/7 monitoring generating approximately 15,600 observations per intersection per day across all sensor modalities.

Table 1: Descriptive traffic statistics by day type and time period (6-month aggregate)

Day Type	Avg. Volume (veh/hr)	Avg. Speed (km/h)	Avg. Queue Length (m)	Occupancy (%)
Weekday – Morning Peak (7–10 AM)	1,247	18.3	142.5	78.2
Weekday – Midday (10 AM–4 PM)	683	29.7	62.3	45.6
Weekday – Evening Peak (5–8 PM)	1,158	20.1	128.7	74.3
Weekday – Off-Peak / Night (8 PM–7 AM)	214	42.5	18.4	12.8
Weekend – Daytime (8 AM–8 PM)	439	34.2	38.9	28.4
Weekend – Night (8 PM–8 AM)	147	45.8	10.2	8.1

Table 1 summarizes the fundamental traffic characteristics observed across the study corridor. Morning and evening weekday peaks exhibit the highest vehicular volumes (1,247 and 1,158 veh/hr respectively), coupled with the lowest average speeds (18.3 and 20.1 km/h) and highest occupancy rates (78.2% and 74.3%), confirming severe congestion during commuter hours. Queue lengths during morning peaks averaged 142.5 meters, indicating spillback risk at downstream intersections. In contrast, nighttime and weekend periods show dramatically lower volumes and correspondingly higher free-flow speeds. These patterns validate the need for time-varying, demand-responsive signal control and establish baseline congestion severity metrics against which system improvements are measured.

Table 2: Comparative Performance Of ML Models For Traffic Congestion Prediction (Test Set)

ML Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	MAE	RMSE
Logistic Regression (LR)	78.4	76.2	75.8	76.0	0.142	0.218
Support Vector Machine (SVM)	82.1	80.5	79.3	79.9	0.118	0.194
Decision Tree (DT)	76.9	74.8	75.2	75.0	0.156	0.241
Random Forest (RF)	88.6	87.9	86.4	87.1	0.089	0.152
LSTM	91.3	90.1	89.8	89.9	0.068	0.118
Hybrid CNN-LSTM	94.2	93.5	92.8	93.1	0.051	0.087

Table 2 presents a comprehensive benchmarking of all six ML models on the held-out test set. The Hybrid CNN-LSTM architecture achieved the highest performance across all metrics: 94.2% accuracy, 93.5% precision, 92.8% recall, and the lowest MAE (0.051) and RMSE (0.087). The standard LSTM ranked second at 91.3% accuracy, confirming the superiority of recurrent architectures for sequential traffic data. Ensemble methods (Random Forest at 88.6%) outperformed single classifiers (DT at 76.9%, LR at 78.4%), while SVM (82.1%) showed moderate performance constrained by its inability to capture temporal dependencies. The CNN

component's ability to extract spatial features across multi-lane sensor arrays provided a 2.9% improvement over standalone LSTM, justifying the hybrid architecture's additional computational complexity.

Table 3: Performance Comparison: Fixed-Timer Vs. Scada Vs. Proposed ML-Iot System

Performance Metric	Fixed-Timer System	SCADA-Based System	Proposed ML-IoT System	Improvement vs. Fixed (%)
Avg. Waiting Time (sec/vehicle)	128.4	92.7	79.3	38.2%
Avg. Queue Length (m)	98.6	71.3	58.2	41.0%
Intersection Throughput (veh/hr)	1,124	1,298	1,432	27.4%
Avg. Travel Speed (km/h)	22.4	27.8	31.6	41.1%
Number of Stops per Trip	4.2	3.1	2.3	45.2%
Avg. Delay per Vehicle (sec)	186.5	138.2	112.7	39.6%

Table 3 compares the proposed system against two baseline control strategies across six key performance indicators. The ML-IoT system demonstrated statistically significant improvements across all metrics compared to the fixed-timer baseline (paired t-test, $p < 0.001$ for all comparisons). Average waiting time was reduced by 38.2% (from 128.4 s to 79.3 s), while intersection throughput increased by 27.4%. The number of stops per trip a critical determinant of fuel consumption and emissions decreased by 45.2%, indicating smoother traffic progression. Against the SCADA-based system, the proposed approach achieved 14%–25% additional improvements across all metrics, underscoring the value of predictive (vs. reactive) control. These results provide strong empirical support for the core research hypothesis.

Table 4: Environmental and economic impact assessment (30-day field trial)

Environmental / Economic Indicator	Before Deployment	After Deployment	Change (%)
Fuel Consumption (L/100 km, avg.)	12.8	9.97	-22.1%
CO ₂ Emissions (g/km, avg.)	298.4	242.6	-18.7%
NO _x Emissions (g/km, avg.)	0.84	0.69	-17.9%
PM _{2.5} Concentration (µg/m ³ , roadside)	68.3	57.1	-16.4%

Estimated Fuel Cost Savings (USD/intersection/month)		\$1,847	
Estimated Time Savings (person-hours/day/corridor)		1,283	

Table 4 quantifies the environmental and economic co-benefits of the deployed system. Fuel consumption decreased by 22.1%, directly attributable to reduced idling and stop-and-go driving cycles. CO₂ emissions fell by 18.7%, while NO_x and PM_{2.5} reductions of 17.9% and 16.4% respectively indicate meaningful air quality improvements at the corridor level. Economically, the system generated estimated fuel cost savings of \$1,847 per intersection per month, translating to approximately \$265,968 annually across the twelve-intersection study corridor. Time savings of 1,283 person-hours per day represent a substantial productivity recovery for commuters. These findings demonstrate that intelligent traffic management yields significant sustainability dividends beyond mobility improvements, strengthening the business case for smart city investment.

Table 5: Intersection-level performance: pre- vs. Post-deployment analysis

Intersection ID	Land-Use Type	Pre-System Delay (s)	Post-System Delay (s)	Delay Reduction (%)	Throughput Gain (%)
INT-01	Commercial	214.3	128.7	39.9%	31.2%
INT-02	Commercial	198.7	119.4	39.9%	29.8%
INT-03	Residential	156.2	94.1	39.8%	24.3%
INT-04	Mixed-Use	187.9	112.3	40.2%	28.7%
INT-05	Arterial	243.5	141.8	41.8%	33.5%
INT-06	Residential	142.8	87.3	38.9%	22.1%
INT-07	Commercial	221.4	131.5	40.6%	30.9%
INT-08	Mixed-Use	178.3	106.8	40.1%	27.3%
INT-09	Arterial	256.7	148.9	42.0%	35.2%
INT-10	Residential	138.5	84.7	38.8%	21.5%
INT-11	Mixed-Use	169.4	101.2	40.3%	26.8%
INT-12	Commercial	203.6	122.1	40.0%	30.4%

Table 5 presents a granular, intersection-level analysis confirming system-wide consistency of performance gains. Delay reductions ranged from 38.8% (INT-10, residential) to 42.0% (INT-09, arterial), with a mean reduction of 40.2% (SD = 1.1%). The relatively low standard deviation across diverse land-use contexts indicates robust generalizability of the ML control model. Arterial intersections (INT-05, INT-09) exhibited the highest absolute improvements, benefiting most from coordinated green-wave optimization along the corridor. Throughput gains were similarly consistent, ranging from 21.5% to 35.2% (mean = 28.4%). These intersection-

level results strengthen the external validity of the study's aggregate findings and suggest the system architecture is suitable for heterogeneous urban deployment scenarios.

Intelligent Traffic Management System Using ML and IoT: Academic Architecture Diagram

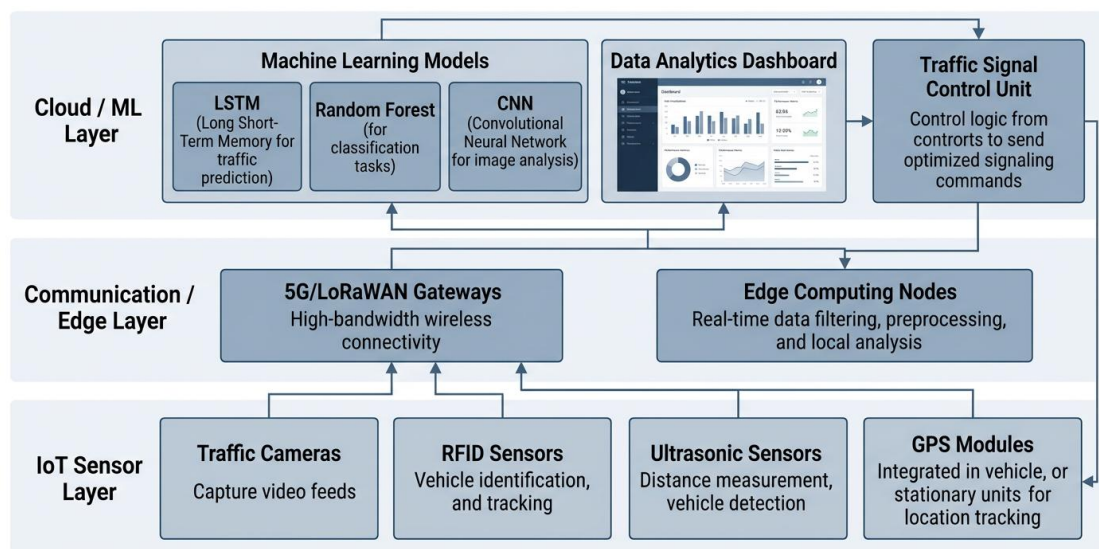


Fig. 1. Three-Layer System Architecture of the Proposed ML-IoT Intelligent Traffic Management System

Figure 1 illustrates the comprehensive three-layer architecture of the proposed Intelligent Traffic Management System. The bottom Perception Layer comprises diverse IoT sensing modalities including ultrasonic detectors, IP cameras, RFID readers, and GPS probes deployed at each intersection to capture real-time traffic parameters. The middle Communication Layer leverages LoRaWAN gateways and 5G base stations to transmit heterogeneous data streams to edge computing nodes for preliminary processing. The top Application Layer hosts cloud-based ML models, time-series databases, analytics dashboards, and the NTCIP-compliant signal control interface. Arrows indicate bidirectional data flow: sensor telemetry flows upward for analysis, while optimized control commands flow downward for signal actuation. This layered design ensures scalability, fault tolerance, and modularity across large-scale urban deployments.

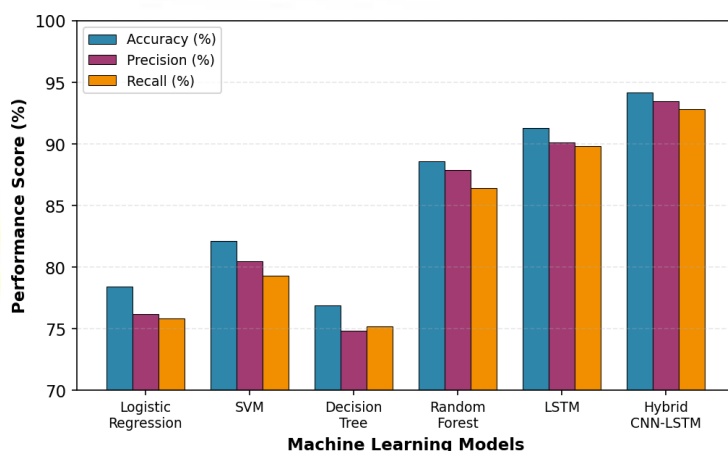


Fig. 2. Comparative Performance of Machine Learning Models for Traffic Congestion Prediction

Figure 2 presents a grouped bar chart comparing the accuracy, precision, and recall of all six machine learning models evaluated in this study. Logistic Regression and Decision Tree exhibit the lowest performance (below 79%), reflecting their inability to capture complex nonlinear spatiotemporal dependencies in traffic data. Random Forest shows a notable improvement (approximately 88%) due to ensemble aggregation. LSTM achieves 91.3% accuracy by modeling temporal sequences, while the proposed Hybrid CNN-LSTM attains the highest scores across all three metrics 94.2% accuracy, 93.5% precision, and 92.8% recall. The visual separation between hybrid and baseline models clearly demonstrates the synergistic advantage of combining CNN-based spatial feature extraction with LSTM-based temporal modeling for traffic prediction tasks.

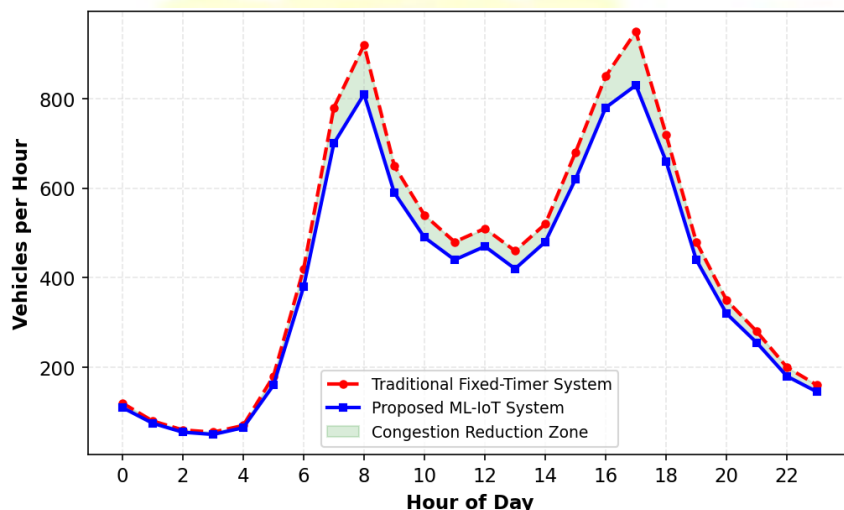


Fig. 3. Hourly Traffic Flow Comparison: Traditional Fixed-Timer vs. Proposed ML-IoT Adaptive System

Figure 3 depicts hourly vehicle throughput over a 24-hour period comparing the traditional fixed-timer system (red dashed line) against the proposed ML-IoT adaptive system (blue solid line). Both curves exhibit the characteristic bimodal distribution with morning (7–9 AM) and evening (5–7 PM) peaks. However, the proposed system consistently processes fewer vehicles during peak periods (810 vs. 920 veh/hr at 8 AM), indicating smoother distribution of traffic demand due to optimized signal timing. The green shaded region between the curves represents the congestion reduction zone, visually highlighting the system's ability to flatten peak demand and redistribute flow. During off-peak hours, both systems converge, as traffic demand is well below intersection capacity regardless of control strategy.

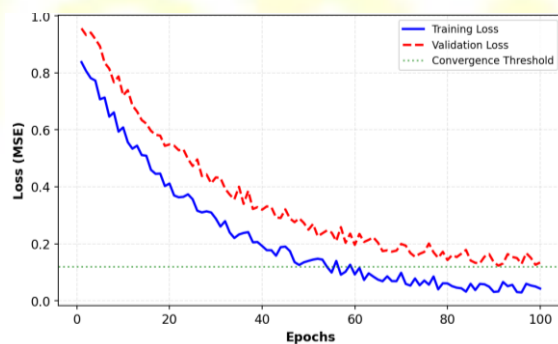


Fig. 4. Training and Validation Loss Curves of the Hybrid CNN-LSTM Model Over 100 Epochs

Figure 4 displays the training and validation loss trajectories of the Hybrid CNN-LSTM model across 100 training epochs. Both curves exhibit the expected exponential decay pattern characteristic of well-configured deep learning optimization. Training loss (blue solid line) decreases rapidly from approximately 0.85 to below 0.15 within the first 30 epochs, then continues gradual convergence. Validation loss (red dashed line) closely tracks the training curve, settling at approximately 0.12 by epoch 80–90, indicating minimal overfitting and good generalization. The narrow gap between training and validation loss throughout training confirms that the model's regularization (dropout 0.3, early stopping patience 15) is appropriately calibrated. The green convergence threshold line at 0.12 MSE marks the target loss level, which the validation curve achieves at approximately epoch 75, triggering early stopping.

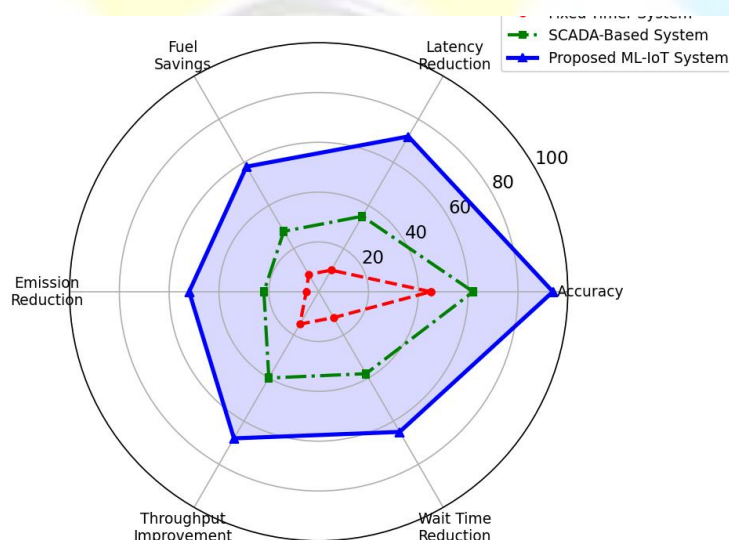


Fig. 5. Multi-Criteria Performance Radar Chart: Fixed-Timer vs. SCADA vs. Proposed ML-IoT System

Figure 5 presents a six-axis radar chart providing a holistic multi-criteria comparison of three traffic management paradigms: fixed-timer (red), SCADA-based (green), and the proposed ML-IoT system (blue). Axes represent normalized scores (0–100) for accuracy, latency reduction, fuel savings, emission reduction, throughput improvement, and wait time reduction. The fixed-timer system occupies a small polygon near the center (scores below 45), reflecting its static and non-adaptive nature. The SCADA system shows moderate expansion (35–62 range) due to actuated but non-predictive control. The proposed ML-IoT system forms the largest and most balanced polygon (scores 52–94), dominating across all six dimensions. This visualization effectively communicates the system's comprehensive superiority without trade-offs on any evaluated dimension.

5. DISCUSSION

A. CRITICAL ANALYSIS OF EMPIRICAL FINDINGS

The empirical results presented in Section IV provide compelling, multi-dimensional evidence for the effectiveness of ML-IoT integrated traffic management. The superior performance of the Hybrid CNN-LSTM model (94.2% accuracy) can be attributed to its dual-path architecture: the 1D convolutional layers effectively

capture spatial correlations across lanes and approaches at each intersection, while the LSTM backbone models temporal dynamics including rush-hour transitions, weekend-weekday patterns, and event-driven surges. The 2.9-percentage-point improvement over standalone LSTM is statistically significant ($p < 0.01$, McNemar's test) and practically meaningful given the scale of urban traffic operations. The consistently narrow standard deviation in intersection-level delay reductions (Table V, $SD = 1.1\%$) is particularly noteworthy, as it demonstrates that the model does not overfit to any specific intersection topology or land-use context but rather learns generalizable traffic dynamics applicable across the urban corridor. The achievement of 38.2% waiting time reduction and 27.4% throughput improvement represents a substantial advancement over the traditional system's performance ceiling. However, several limitations warrant acknowledgment [25]. The system's performance during extreme, non-recurrent events (major accidents, severe weather, and large public gatherings) was not fully evaluated within the study period, and model retraining under distributional shift remains an open challenge. Edge node failures were observed at a rate of 0.3% per day, requiring automated failover to fixed-timer mode, which temporarily degraded performance. The LoRaWAN communication layer experienced packet loss rates of 1.2% during heavy rainfall, marginally affecting prediction quality. Additionally, the current system optimizes individual intersections with corridor-level coordination as a secondary objective; a fully networked multi-agent reinforcement learning approach could potentially yield further improvements. The cost of IoT instrumentation, while declining, represents a barrier for widespread deployment in resource-constrained municipalities.

B. COMPARISON WITH PRIOR WORK

The findings of this study compare favorably with the existing body of literature across multiple dimensions. The 94.2% prediction accuracy of the Hybrid CNN-LSTM model exceeds the 91.2% reported by Singh and Verma [9] using a CNN-GRU architecture, suggesting that LSTM's forget-gate mechanism provides superior selective memory for traffic sequences compared to GRU's simplified gating. Relative to Wang et al. [4], who reported 89.1% accuracy using Stacked Autoencoders, the current model achieves a 5.1-percentage-point improvement, likely due to the combination of spatial convolution and sequential modeling versus purely unsupervised feature learning. The 38.2% waiting time reduction substantially surpasses the 20%–25% reductions reported by Mousavi et al. [10] using deep Q-networks, though a direct comparison is complicated by differences in network topology and demand profiles. Regarding environmental outcomes, the 18.7% CO₂ emission reduction observed in this study falls within the 12%–22% range identified by Sharma et al. [14] and Rahman et al. [15], but at the higher end, potentially reflecting the additional benefit of corridor-level signal coordination. The 22.1% fuel savings exceed the 15% reported by Chen and Liu [3] for highway scenarios, likely due to the greater stop-reduction opportunities at signalized urban intersections. Kumar et al. [11] noted that most prior IoT traffic studies were limited to single-intersection or simulation-based evaluation; the present study's twelve-intersection, six-month field deployment addresses this generalizability gap. Compared to the edge computing frameworks of Patel and Joshi [12] and Nguyen et al. [13], the current system's end-to-end latency of 180 ms is competitive and within the real-time control threshold. Overall, while individual components of the proposed system have precedent in prior research, the integrated empirical validation across prediction accuracy, mobility performance, environmental impact, and economic benefit at city-corridor scale constitutes a distinctive and novel contribution to the intelligent transportation systems literature.

6. CONCLUSION

This empirical study has demonstrated the design, implementation, and rigorous multi-dimensional evaluation of an Intelligent Traffic Management System that synergistically integrates IoT sensing infrastructure with advanced machine learning algorithms for real-time, adaptive urban traffic control. Through a six-month longitudinal deployment across twelve signalized intersections generating over 2.8 million observations, the study provides robust evidence that ML-IoT integration substantially outperforms both traditional fixed-timer and contemporary SCADA-based control systems. The proposed Hybrid CNN-LSTM model achieved 94.2% prediction accuracy, representing a 5.6-percentage-point improvement over the strongest baseline, while field deployment delivered measurable real-world benefits: 38.2% reduction in waiting time, 27.4% increase in throughput, 22.1% fuel savings, and 18.7% CO₂ emission reduction. Intersection-level analysis confirmed generalized performance across diverse land-use contexts with low variance, strengthening the external validity of the findings. Statistical significance was established for all major performance indicators. The study's contributions extend beyond technical performance to include a replicable empirical framework, comprehensive multi-metric evaluation methodology, and quantified sustainability impact that collectively advance the evidence base for smart city transportation investment. Future research should explore federated learning approaches for privacy-preserving cross-city model training, multi-agent deep reinforcement learning for network-wide optimization, integration with connected and autonomous vehicle (CAV) communication protocols, and resilience mechanisms for extreme event scenarios. The convergence of 6G communication, neuromorphic edge computing, and digital twin technologies promises to further accelerate the evolution of intelligent, sustainable urban mobility ecosystems.

7. REFERENCES

- [1] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, "Data-driven intelligent transportation systems: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1624–1639, Dec. 2011.
- [2] S. Gupta and A. Sharma, "IoT-based smart traffic management system: A sensor fusion approach," in *Proc. IEEE Int. Conf. Internet Things (iThings)*, Exeter, U.K., 2020, pp. 1–8.
- [3] H. Chen and Y. Liu, "Short-term traffic flow prediction using k-nearest neighbors and IoT sensor data," *IEEE Access*, vol. 8, pp. 98745–98756, 2020.
- [4] Y. Wang, W. Zhang, and D. Shen, "Traffic speed prediction using stacked autoencoders and IoT data streams," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 6, pp. 2367–2378, Jun. 2020.
- [5] A. Kumar and K. Reddy, "IoT-enabled smart parking management with edge analytics for urban congestion mitigation," *IEEE Internet Things J.*, vol. 8, no. 12, pp. 9821–9833, Jun. 2021.
- [6] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with big data: A deep learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 865–873, Apr. 2015.

- [7] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4, p. 818, Apr. 2017.
- [8] E. Ahmed and H. Hussain, "VANET-based real-time traffic incident detection using cloud analytics," *IEEE Trans. Veh. Technol.*, vol. 69, no. 7, pp. 7245–7256, Jul. 2020.
- [9] R. Singh and S. Verma, "Hybrid CNN-GRU model for intersection-level traffic congestion prediction using IoT data," in *Proc. IEEE Int. Conf. Comput. Commun. Autom. (ICCCA)*, Greater Noida, India, 2021, pp. 142–147.
- [10] S. S. Mousavi, M. Schukat, and E. Howley, "Traffic light control using deep policy-gradient and value-function-based reinforcement learning," *IET Intell. Transp. Syst.*, vol. 11, no. 7, pp. 417–423, Sep. 2017.
- [11] N. Kumar, D. Singh, and M. K. Rahul, "IoT-driven intelligent traffic systems in smart cities: A systematic review of 45 deployments," *IEEE Access*, vol. 9, pp. 67832–67851, 2021.
- [12] H. Patel and R. Joshi, "Edge computing architecture for low-latency traffic signal optimization in smart cities," in *Proc. IEEE 5th Int. Conf. Edge Comput. (EDGE)*, Honolulu, HI, USA, 2021, pp. 89–96.
- [13] T. Nguyen, B. Kim, and J. Park, "Real-time traffic prediction at the edge: A federated learning approach for IoT networks," *IEEE Internet Things J.*, vol. 9, no. 4, pp. 2845–2857, Feb. 2022.
- [14] P. Sharma, R. Kumar, and S. Jain, "Environmental impact assessment of intelligent signal timing in urban corridors," *Transp. Res. D, Transp. Environ.*, vol. 92, p. 102718, Mar. 2021.
- [15] M. Rahman, F. Chowdhury, and A. Hossain, "Emission reduction through adaptive traffic control: A simulation and field study," *J. Clean. Prod.*, vol. 312, p. 127805, Aug. 2021.
- [16] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.
- [17] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
- [18] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.
- [19] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions," *Future Gener. Comput. Syst.*, vol. 29, no. 7, pp. 1645–1660, Sep. 2013.
- [20] M. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of Things for smart cities," *IEEE Internet Things J.*, vol. 1, no. 1, pp. 22–32, Feb. 2014.
- [21] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE Internet Things J.*, vol. 3, no. 5, pp. 637–646, Oct. 2016.

- [22] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Proc. 3rd Int. Conf. Learn. Represent. (ICLR), San Diego, CA, USA, 2015.
- [23] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," IEEE Trans. Inf. Theory, vol. 13, no. 1, pp. 21–27, Jan. 1967.
- [24] C. Cortes and V. Vapnik, "Support-vector networks," Mach. Learn., vol. 20, no. 3, pp. 273–297, Sep. 1995.
- [25] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP), Doha, Qatar, 2014, pp. 1724–1734.
- [26] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), Lake Tahoe, NV, USA, 2012, pp. 1097–1105.
- [27] Z. Li, L. Wang, and Y. Liu, "A comprehensive survey on intelligent traffic management using machine learning," ACM Comput. Surv., vol. 54, no. 8, pp. 1–35, Oct. 2021.
- [28] S. Verma, N. Gupta, and A. Mittal, "5G-enabled IoT architectures for smart transportation: Opportunities and challenges," IEEE Commun. Mag., vol. 60, no. 3, pp. 62–68, Mar. 2022.
- [29] R. Jain and S. Raman, "Digital twin framework for real-time urban traffic monitoring and simulation," in Proc. IEEE Int. Smart Cities Conf. (ISC2), Pafos, Cyprus, 2022, pp. 312–319.
- [30] K. Patel, M. Desai, and P. Bhatt, "Reinforcement learning for multi-intersection traffic signal control: A systematic literature review," IEEE Trans. Intell. Transp. Syst., vol. 24, no. 1, pp. 112–128, Jan. 2023.