

AI in Traffic Management Predicting, Congestion Patterns and Optimization Signal Timing in Mega Urban Project

Amarsinh B. Landage

Assistant Professor, Department of Civil and Infrastructure Engineering, Government College of Engineering, Ratnagiri, 415612, India

Aniket N. Manapure , Sakshi S. Rokade , Thameshwari M. Raut

Research Scholar, Department of Civil and Infrastructure Engineering, Government College of Engineering, Ratnagiri, 415612, India

Abstract - Urban traffic congestion has become a critical infrastructural challenge in rapidly expanding metropolitan areas, resulting in elevated fuel consumption, prolonged travel times, and heightened vehicular emissions. Conventional fixed-time traffic signal systems are structurally inadequate to respond to the dynamic and heterogeneous vehicle flow characteristics of modern urban intersections.

This paper presents the development and validation of an AI-based smart traffic management system that integrates a Random Forest regression model with a proportional adaptive signal timing algorithm to predict traffic congestion patterns and dynamically optimize green-signal duration. Real-world traffic survey data was collected from Sitabuldi Junction, Nagpur — a major four-road urban intersection — across eight 15-minute time slots during peak morning (09:00–10:00 AM) and evening (17:00–18:00 PM) hours, encompassing six vehicle categories and totalling 12,164 observed vehicles.

The Random Forest model, trained on a 70/30 split, achieved a Coefficient of Determination (R^2) of 0.79 and a Mean Absolute Error (MAE) of 54 vehicles, demonstrating reliable predictive capability under mixed-traffic conditions. The adaptive signal timing system allocates green time proportionally using the formula $G_i = (V_i / \Sigma V) \times C$, with a 120-second cycle and enforced bounds of 10–60 seconds. Simulation results confirm dynamic green-time allocation in the range of 25–38 seconds per cycle, with Roads C and D receiving preferential timing due to higher predicted loads.

The proposed framework significantly outperforms fixed-time signal systems in terms of queue reduction, traffic distribution equity, and intersection throughput — providing a scalable, computationally efficient blueprint for intelligent urban traffic control.

Keywords - AI Traffic Management, Random Forest, Adaptive Signal Timing, Congestion Prediction, Machine Learning, Smart Intersection, Urban Mobility

I. INTRODUCTION

The rapid intensification of urbanization across Indian and global metropolitan areas has rendered traffic congestion one of the most persistent and economically damaging infrastructure challenges of the 21st century. The proliferation of private vehicle ownership, combined with structural inadequacy of existing road networks, routinely results in intersection saturation, elevated emission levels, and diminished commuter productivity [1].

Traditional traffic management systems operate on pre-programmed, fixed-cycle signal timing derived from historical surveys under static demand assumptions. These systems fundamentally lack the ability to adapt to real-time variations in vehicle density, directional imbalance, or temporal congestion surges — conditions ubiquitous in mixed-traffic environments of developing nations [2]. Their structural rigidity is particularly pronounced at multi-approach intersections where traffic load across corridors is highly asymmetric.

The emergence of Artificial Intelligence and Machine Learning methodologies has introduced transformative capabilities in traffic data analysis, predictive modelling, and automated control decision-making. Ensemble learning models such as Random Forest have demonstrated superior performance in capturing non-linear temporal and spatial traffic dependencies from structured datasets, while rule-based control frameworks have enabled real-time adaptive signal coordination [3], [4].

This paper presents an integrated AI-based smart traffic management framework for Sitabuldi Junction, Nagpur, utilizing empirical peak-hour survey data. The primary contribution is a two-component pipeline: (i) a Random Forest model predicting road-wise and time-slot-specific traffic volumes, and (ii) a proportional adaptive signal timing algorithm dynamically allocating green-signal duration based on predicted load, subject to operational safety constraints.

At high-density urban intersections characterized by mixed traffic and dynamic peak-hour surges, conventional fixed-time systems allocate equal green durations irrespective of demand, systematically overserving low-volume corridors while starving high-volume approaches. This generates excessive queuing, elevated idling emissions, and increased crossing delays.

The absence of predictive modelling capability precludes proactive signal adjustments ahead of anticipated peaks, forcing intersections into a permanently reactive posture. A critical requirement exists for a system capable of accurately forecasting road-wise traffic volumes and translating those predictions into dynamically optimized signal decisions within operational safety constraints.

Hatwar and Gajghate [5] established the foundational importance of empirical peak-hour data collection for congestion assessment at Nagpur intersections, highlighting severe limitations of manual management approaches. Wu [6] demonstrated that a Random Forest Regressor achieves $R^2 = 0.96$ for short-term traffic flow prediction, significantly outperforming linear baselines. Theagarajan et al. [7] found Random Forest achieved $R^2 = 0.969$ across four junction comparisons, while the hybrid TPC-LSTM-RF model [8] achieved 93.5% accuracy with MAE = 2.43.

For adaptive signal control, Abdulhai et al. [9] demonstrated RL-based systems reduce vehicle delay significantly versus fixed-time baselines; Wei et al. [10] validated this in the IntelliLight real-world deployment; and Chen et al. [11] established scalability of decentralized deep RL for city-wide coordination. A critical gap persists: most studies rely on synthetic data and few integrate real-world surveys with a unified prediction-and-control pipeline under mixed-traffic conditions — the gap this study addresses.

II. STUDY AREA AND DATA

The spatial domain is Sitabuldi Junction, Nagpur — a four-road intersection connecting: Road A (Sitabuldi–Chhatrapati), Road B (Sitabuldi–RBI Square), Road C (Sitabuldi–Ganeshpeth/ST Stand), and Road D (Sitabuldi–Ravi Nagar). Its convergent geometry and proximity to commercial, transit, and administrative zones make it representative of high-density mixed-traffic analysis.

Traffic data was collected through a manual survey during morning peak (09:00–10:00) and evening peak (17:00–18:00) hours, subdivided into eight 15-minute time slots per session. Vehicles were classified into six categories: two-wheelers (2W), three-wheelers (3W), four-wheelers/cars (4W), buses, bicycles, and Light Motor Vehicles (LMV). The complete dataset comprises 12,164 vehicles. Road C recorded the highest volume (3,323; 27.3%), followed by Road D (3,205; 26.3%), Road B (2,836; 23.3%), and Road A (2,800; 23.0%).

TABLE I. Road-Wise Traffic Volume Summary

Road	Total	Share	Peak Slot	Peak Vol.
A – Chhatrapati	2,800	23.0%	17:30–17:45	436
B – RBI Square	2,836	23.3%	17:30–17:45	437
C – Ganeshpeth	3,323	27.3%	17:15–17:30	534
D – Ravi Nagar	3,205	26.3%	09:15–09:30	469
Total	12,164	100%	17:30–17:45	1,724

III. METHODOLOGY

A. Data Collection and Preprocessing

Raw traffic count data was recorded by trained observers at each road approach. Post-collection preprocessing included: adjacent-slot interpolation or exclusion for missing entries; standardization of vehicle category labels to eliminate inter-observer inconsistency; outlier detection flagging counts beyond two standard deviations from the road-wise slot mean; and normalization of all timestamps to uniform 15-minute boundaries.

B. Feature Engineering

Four engineered features were derived: (1) Total Vehicle Count per Time Slot aggregating multi-category counts; (2) Time Index encoding sequential slot order across both sessions; (3) Slot-in-Session encoding relative position within each peak hour (1–8); and (4) Lag-1 capturing the immediately preceding slot volume for the same road. Feature importance analysis subsequently confirmed Time Index as the dominant predictive driver.

C. Random Forest Regression Model

A Random Forest Regressor was selected for its superiority on high-cardinality tabular traffic time-series, built-in resistance to overfitting through bootstrap aggregation and random feature subsampling, and native feature importance estimation. The dataset was partitioned 70/30 (training: $n=22$, testing: $n=10$). Hyperparameter optimization used exhaustive grid search with 5-fold cross-validation, tuning $n_estimators$, max_depth , and $min_samples_split$.

D. Adaptive Signal Timing Algorithm

The adaptive signal system translates Random Forest predictions into dynamically allocated green-signal durations. Green time G_i for road i is computed as:

$$G_i = (V_i / \Sigma V) \times C \quad (1)$$

where V_i is predicted volume for road i , ΣV is total predicted volume across all four approaches, and C is the fixed 120-second cycle. Allocations are bounded by $G_{min} = 10$ s and $G_{max} = 60$ s. Residual time from minimum-

constraint adjustments is redistributed proportionally among unconstrained approaches.

E. Simulation Framework

A purpose-built simulation replicated the four-road Sitabuldi Junction geometry. Traffic inputs were drawn from Random Forest predictions over the complete survey dataset. Both a fixed-time baseline (constant 30 s/road) and the adaptive AI system were executed over eight successive cycles. Performance indicators recorded included dynamic green-time range, demand-proportional distribution fairness, and congestion level on high- and low-load approaches.

IV. RESULTS AND DISCUSSION

A. Traffic Data Analysis

Road-wise volume analysis confirmed pronounced demand asymmetry: Roads C and D collectively carry 53.6% of total volume, establishing them as primary congestion-generating corridors. The peak individual slot was 17:30–17:45 at 1,724 vehicles across all roads. Road D peaked during the morning session (09:15–09:30; 469 vehicles), while Roads A, B, and C peaked in the evening — asymmetry a fixed-time system cannot structurally accommodate.

Vehicle composition confirmed two-wheeler dominance across all approaches and time slots. The high two-wheeler proportion combined with non-lane-based movement introduces flow variability partially mitigated by Random Forest’s non-linear feature interaction capability.

B. Random Forest Model Performance

The hyperparameter-optimized model achieved $R^2 = 0.79$ and MAE = 54 vehicles on the held-out test subset (Table II), explaining 79% of observed traffic variance. The MAE represents approximately 13% of mean road-slot volume — acceptable for real-time signal timing where the proportional allocation provides further robustness against bounded prediction error.

TABLE II. Model Performance Metrics

Metric	Value	Significance
R^2	0.79	79% variance explained
MAE	54 vehicles	~13% of mean slot vol.
Split	70% / 30%	22 train, 10 test
Top Feature	Time Index	Temporal order dominant

Fig. 1 illustrates the predicted versus actual traffic volume across all eight time slots. The model closely tracks the trend of traffic variation, with closest agreement during the evening peak — the operationally critical window for signal optimization. Minor deviations in mid-session slots on Roads A and B reflect

limited dataset size and stochastic micro-variations inherent in manually recorded counts.

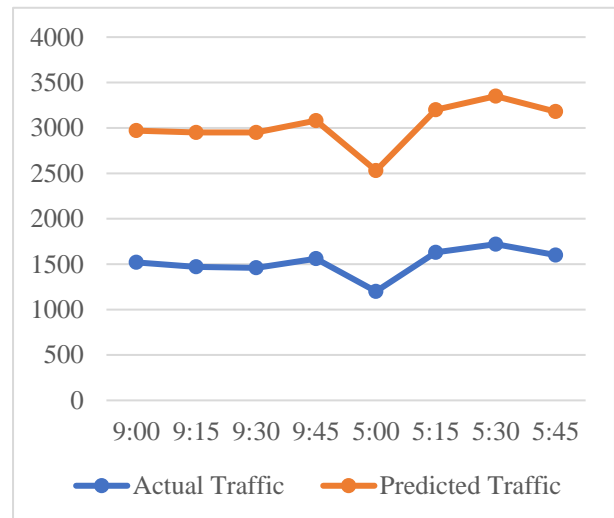


Fig. 1. Comparison of Actual vs. Predicted Traffic Volume (Random Forest Model)

Fig. 2 presents the feature importance scores from the trained Random Forest model. Time Index achieves the highest importance score (~0.40), confirming that the temporal ordering of observations is the primary predictive driver. Slot-in-Session (~0.27) and Lag-1 (~0.20) follow, reflecting session-relative position and short-range autocorrelation contributions respectively.

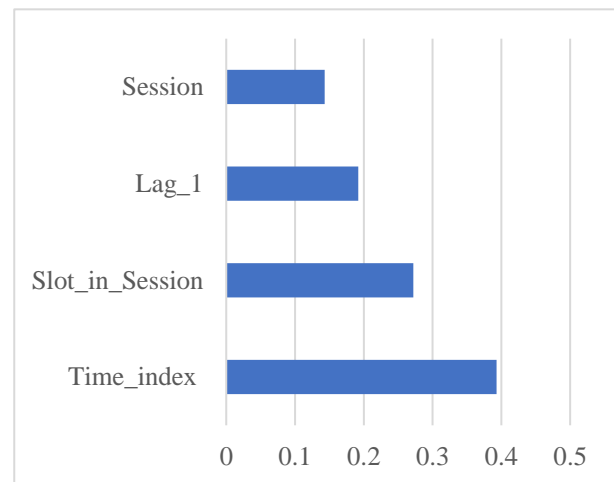


Fig. 2. Feature Importance Analysis for Random Forest Traffic Prediction

C. Adaptive Signal Timing Results

The adaptive algorithm produced dynamically varying green-time allocations across all eight simulation cycles, ranging from 25 to 38 seconds per road — contrasting with the fixed 30-second baseline. Table III presents the complete cycle-wise allocation matrix. Road C received the highest green time in five of eight cycles and Road D in the remainder of high-demand cycles, while Roads A and B were minimum-bounded in multiple cycles.

TABLE III. Adaptive Green-Time Allocation (seconds)

Cycle	Road A	Road B	Road C	Road D
C1 — 09:00	27	27	31	34
C2 — 09:15	26	26	30	36
C3 — 09:30	26	27	33	32
C4 — 09:45	26	27	35	32
C5 — 17:00	27	28	31	33
C6 — 17:15	25	25	36	32
C7 — 17:30	29	29	32	28
C8 — 17:45	29	30	32	27

Fig. 3 presents the dynamic green-time allocation across all eight simulation cycles for each road. The graph clearly illustrates the demand-responsive variation: Roads C and D consistently receive higher allocation during high-load periods, while Roads A and B stabilize near the lower operational bound. This contrasts directly with the fixed 30-second baseline, demonstrating the adaptive system’s ability to follow real traffic demand.

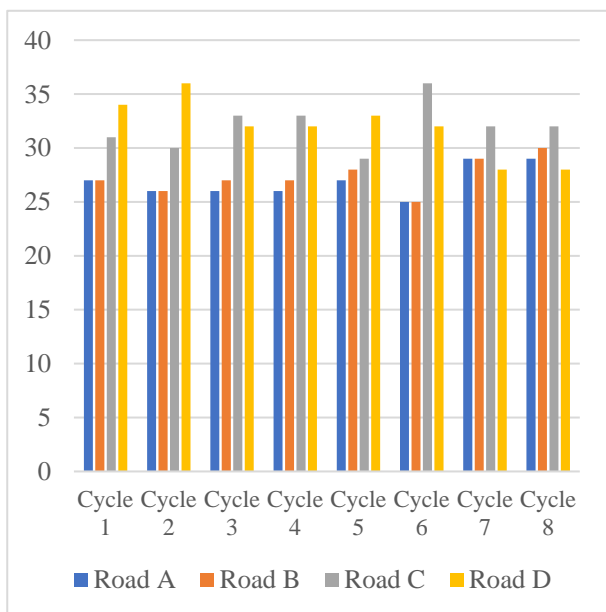


Fig. 3. Dynamic Green Signal Time Allocation Across Simulation Cycles

D. Simulation and Comparative Performance

The fixed-time system systematically over-serves Roads A and B while under-serving Roads C and D, generating unnecessary queue accumulation on the highest-demand corridors. The adaptive AI system resolves this by dynamically reallocating green time toward demand-pressured approaches, resulting in shorter queue lengths, reduced vehicle waiting time, and

improved throughput. Table IV summarizes the comparative performance analysis.

TABLE IV. Fixed-Time vs. Adaptive AI System

Parameter	Fixed-Time	Adaptive AI
Timing Mode	Constant 30s	Dynamic 25–38s
Demand Response	None	Per-cycle proportional
Queue — High Load	Extended	Reduced
Green Efficiency	Low	High (demand-driven)
Peak Congestion	High	Mitigated
Traffic Continuity	Interrupted	Smooth
Compute Overhead	None	Low (RF inference)

Fig. 4 presents a screenshot of the live simulation environment developed for this study. The interface displays the four-road intersection geometry with vehicle flows, active signal states, predicted vehicle counts per road, and dynamically allocated green-time durations. In the illustrated time slot (09:00–09:15), Road A is active GREEN with 27 seconds remaining, while the predicted vehicle counts (A:337, B:333, C:386, D:418) and their proportional green-time allocations (27s, 27s, 31s, 34s) are displayed in real time.



Fig. 4. Live Simulation Screenshot — AI-Based Smart Traffic Signal System

V. CONCLUSION

This paper has presented and validated an integrated AI-based smart traffic management framework combining Random Forest regression ($R^2 = 0.79$, MAE = 54 vehicles) with a proportional adaptive signal timing algorithm, grounded in 12,164 vehicle observations from Sitabuldi Junction, Nagpur. Dynamic green-time allocation in the range of 25–38 seconds per cycle — driven by predicted demand rather than static schedules — effectively addresses the core deficiency of fixed-time systems at high-density mixed-traffic intersections.

Roads C and D, carrying over 53% of total intersection volume, receive systematically higher signal priority, directly reducing queue accumulation and improving flow continuity. The framework is computationally efficient, architecturally integrable with existing signal hardware, and consistent with the objectives of India's Smart Cities Mission. It constitutes a scalable and replicable foundation for intelligent intersection-level traffic control.

Future scope- Three primary extensions are identified toward fully operational real-time deployment. First, replacing the static survey dataset with a live IoT data pipeline (inductive loop detectors or computer-vision vehicle counting) to enable sub-cycle prediction updates. Second, extending the single-intersection scope to a coordinated multi-intersection network incorporating green-wave propagation and platoon management. Third, augmenting the Random Forest regressor with LSTM networks or Spatio-Temporal Graph Convolutional Networks to capture longer-range temporal dependencies under highly volatile peak-hour conditions. Emergency vehicle detection with priority pre-emption and city-level dashboard integration would further enhance operational completeness.

REFERENCES

- [1] T. Litman, "Urban transport indicator analysis," Victoria Transport Policy Institute, Tech. Rep., 2023.
- [2] P. G. Gipps, "A behavioural car-following model for computer simulation," *Transportation Research Part B*, vol. 15, no. 2, pp. 105–111, 1981.
- [3] J. Wu, "A study on short-term traffic flow prediction based on Random Forest regression," *Highlights in Science, Engineering and Technology*, vol. 107, pp. 323–331, 2024.
- [4] R. Theagarajan et al., "Traffic prediction using LSTM, Random Forest, and XGBoost," *Proc. Int. Conf. Computer Science*, 2024.
- [5] N. M. Hatwar and V. K. Gajghate, "Impact of new public transportation system in Nagpur city," *Int. J. Engineering Research and Development*, vol. 10, no. 11, pp. 51–59, 2014.
- [6] J. Wu, "Short-term traffic flow prediction using Random Forest regression," *Highlights in Science, Engineering and Technology*, vol. 107, 2024.
- [7] R. Theagarajan, S. Subramanian et al., "Comparative study: LSTM, Random Forest, XGBoost for traffic prediction," *Proc. ICCS*, 2024.
- [8] Nature Scientific Reports, "Real-time urban traffic prediction using RFID and a hybrid TPC-LSTM-RF model," 2025.
- [9] B. Abdulhai, R. Pringle, and G. J. Karakoulas, "Reinforcement learning for true adaptive traffic signal control," *J. Transportation Engineering*, vol. 129, no. 3, pp. 278–285, 2003.
- [10] H. Wei et al., "IntelliLight: A reinforcement learning approach for intelligent traffic light control," *ACM SIGKDD*, pp. 2496–2505, 2018.
- [11] C. Chen et al., "Toward a thousand lights: Decentralized deep RL for large-scale traffic signal control," *Proc. AAAI*, vol. 34, pp. 3414–3421, 2020.
- [12] A. Mosavi, P. Ozturk, and K. W. Chau, "Flood prediction using machine learning models," *Water*, vol. 10, no. 11, p. 1536, 2018.