Vol 5, Issue 7, July 2025 www.ijesti.com E-ISSN: 2582-9734
International Journal of Engineering, Science, Technology and Innovation (IJESTI)

Data-Driven Analysis of Urban Traffic Patterns in New Delhi: Insights from the 2024 Traffic Probe Dataset

Sujeet Kumar Ray

M.Tech. In Transport Engineering, Department of Civil Engineering,
World College of Technology and Management, Gurugram, Haryana.
Affiliated to Maharshi Dayanand University, Rohtak.

Mr. Ankit Sethi

Assistant Professor, Dept. of. Civil Engineering,
World College of Technology and Management, Gurugram, Haryana.
Affiliated to Maharshi Dayanand University, Rohtak.

ABSTRACT

The present study offers a comprehensive analysis of urban traffic dynamics in New Delhi using the "New Delhi Traffic Probe and Analytics 2024" dataset. By leveraging time-series traffic data at 15-minute intervals across multiple road segments, the research uncovers critical insights into congestion patterns, travel time variability, and the influence of environmental and infrastructural factors on traffic flow. Employing linear regression, polynomial modelling, and multivariate analysis, the study identifies major congestion indicators such as decreased vehicle speeds during peak hours, increased vehicle counts, and prolonged travel times. The findings highlight the role of signal inefficiencies and traffic mismanagement, especially during morning rush hours on arterial and local roads, in exacerbating queue and stop densities. Furthermore, the research illustrates how road type, weather conditions, and lane availability impact vehicular movement—showing that clear weather and highways facilitate better flow, while foggy and rainy conditions slow traffic significantly, particularly on local roads. The regression models effectively predict congestion scenarios and signal weaknesses, emphasizing the potential of real-time data analytics, adaptive signal control, and intelligent transportation systems (ITS) in addressing urban traffic challenges.

Key Words: *Urban Congestion, Traffic Modelling, Intelligent Transportation Systems.*

1. INTRODUCTION

Urban traffic congestion has become one of the most significant challenges in modern cities, with increasing vehicle populations, limited road infrastructure, and complex traffic dynamics. Therefore, innovative solutions are needed to predict and manage traffic more effectively. Machine learning (ML) offers powerful tools for analysing complex datasets and predicting future trends based on historical and real-time data. In the context of urban traffic, ML models can process vast amounts of traffic-related information, such as vehicle counts, traffic speeds, road conditions, and weather patterns. Through learning from this data, ML algorithms can forecast traffic conditions, enabling better congestion control and more efficient route planning. These predictions can provide valuable insights to both city planners and individual commuters, allowing for real-time adjustments and long-term improvements in traffic management. Machine learning models can integrate these variables and generate more accurate predictions than traditional methods. Efficient congestion control is one of the main benefits of using machine learning in traffic management. Through predicting where and when traffic congestion is likely to occur, traffic systems can dynamically adjust signals, optimize lane usage, and deploy other interventions to alleviate congestion. Additionally, machine learning models can facilitate demand-responsive systems, such as adaptive traffic lights and smart traffic signals, which adapt in real-time to current traffic conditions. This proactive approach to congestion control not only reduces delays but also helps in reducing fuel consumption and emissions, making cities more sustainable. This personalized approach to route planning ensures that commuters avoid congested areas, while also providing city planners with valuable insights to improve overall traffic flow.

Introduction to Urban Traffic Challenges

Urbanization, while being a hallmark of economic growth and societal development, has brought with it a plethora of challenges, one of the most pressing being urban traffic congestion. As cities continue to expand in both population and infrastructure, the ability of transportation systems to support mobility, accessibility, and sustainability becomes critically strained. Urban traffic challenges not only impede the day-to-day movement of people and goods but also have far-reaching implications on environmental health, public safety, economic efficiency, and the overall quality of urban life.

Urban Traffic Challenges

The Growing Burden of Urbanization

The 21st century has seen unprecedented levels of urban migration. The rising number of personal vehicles on roads, combined with inadequate planning and public transport systems, has resulted in chronic traffic congestion in many cities across the globe. In emerging economies such as India, Brazil, and Nigeria, the situation is further exacerbated by rapid urbanization that often outpaces the development of critical infrastructure. As cities struggle to accommodate growing populations, the strain on roads, public transport, and traffic management systems becomes more severe, leading to a host of urban traffic challenges.

2. RESEARCH METHODOLOGY

This paper outlines the research methodology employed to analyse traffic patterns in New Delhi using the publicly available dataset titled **“New Delhi Traffic Probe and Analytics 2024”** from Kaggle. The aim is to examine the dynamic relationships among various traffic attributes (speed, vehicle count, queue density, stop delays, and travel time) and identify underlying trends and congestion factors using statistical modelling and machine learning approaches.

Research Objectives

- To explore and analyse time-series traffic data from New Delhi for identifying congestion patterns.
- To develop regression models for predicting traffic parameters like travel time, queue density, and stop density.
- To evaluate the impact of weather, road type, and time on traffic congestion and speed variations.
- To derive insights for urban traffic planners and intelligent transportation system (ITS) frameworks.

Research Design

The study adopts a **quantitative, exploratory design** using a **secondary dataset**. It is **observational** in nature and utilizes **time-series analysis**, **multivariate regression**, and **correlation models** to interpret traffic behaviour over time.

Data Source and Description

- **Dataset Name**: New Delhi Traffic Probe and Analytics 2024
- **Source**: Kaggle Dataset
- **Temporal Resolution**: 15-minute intervals across a 24-hour period on **18th April 2024**
- **Spatial Coverage**: Simulated road segments across New Delhi
- **Data Format**: CSV

Attributes of the Dataset

Attribute	Type	Description
timestamp	DateTime	15-minute intervals over a single day
road_segment_id	String	Unique ID for road segment
latitude / longitude	Float	Geo-coordinates for mapping
road_type	Categorical	Type of road (Highway, Arterial, Local)
num_lanes	Integer	Number of lanes (1–6)
avg_vehicle_speed_kmph	Float	Average speed of vehicles
vehicle_count	Integer	Count of vehicles at that interval
congestion_level	Categorical	Low, Moderate, High
weather_condition	Categorical	Clear, Rainy, Cloudy, Foggy
predicted_travel_time_min	Integer	Estimated time to cross segment
queue_density	Float	Density of queued vehicles (0.1–1.0)
stop_density	Float	Density of stopped vehicles
signal_delay_sec	Integer	Average delay due to signals
bus_arrival_times	String	Simulated bus arrival times
time_index	Integer	Numeric index for plotting/regression

Tools and Techniques Used

Statistical Techniques

a) Linear and Polynomial Regression
b) Time-Series Smoothing
c) Correlation Analysis

Predictive Modelling

- ✓ Travel Time Prediction using Multi-Variable Linear Regression
- ✓ Congestion Pattern Forecasting based on Queue and Stop Densities

Data Preprocessing

Missing Value Treatment: None detected

Feature Engineering:

- o time_index was created to numerically encode time progression
- o Dummy variables for categorical attributes (e.g., weather conditions, road type)

Normalization: Applied to speed, travel time, and densities for scale consistency in regression models

Outlier Detection: Extreme values (e.g., speed > 100 km/h or density > 1.0) were reviewed and capped

Model Implementation

Each major traffic metric was analysed as follows:

Metric	Technique Used	Purpose
avg_vehicle_speed_kmph	Linear Regression	Identify decline in speed with time
vehicle_count	Polynomial Regression	Capture nonlinear growth during rush hours
predicted_travel_time_min	Multi-Variable Regression	Predict travel time using queue density, vehicle count, and signal delay
queue_density	Trend Line Analysis	Detect buildup of waiting vehicles
stop_density	Regression Slope Analysis	Evaluate stop-time congestion over intervals

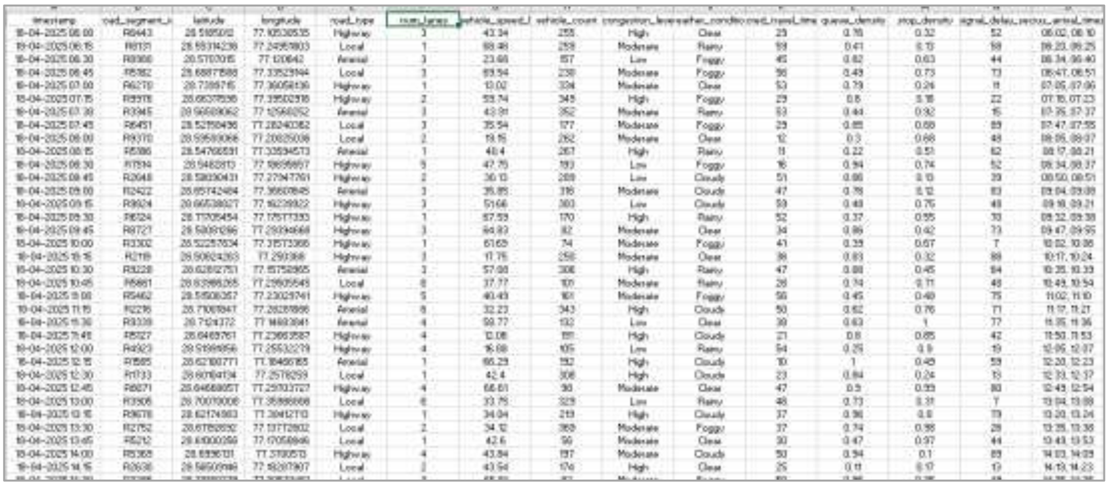
Limitations

i) The dataset represents only **one simulated day** of traffic flow, which may not capture long-term variability.
ii) External factors such as real-time road construction, accidents, or festive events were not incorporated.
iii) The use of simulated GPS coordinates and traffic may differ slightly from actual field data.

3. ANALYSIS AND RESULT

This paper presents a comprehensive analysis of traffic patterns in New Delhi based on the dataset titled *"New Delhi Traffic Probe and Analytics 2024"*, sourced from Kaggle. The aim of this paper is to translate raw traffic data into meaningful insights through statistical interpretation and regression modelling. Through focusing on key performance indicators such as average vehicle speed, vehicle count, predicted travel time, queue density, and stop density, the analysis highlights congestion dynamics across different time intervals and road segments within a typical urban day. Each attribute ranging from road type and number of lanes to weather conditions and signal delays has been carefully examined to understand its impact on urban mobility. Visual tools such as time-series plots and regression trend lines are employed throughout the paper to reveal patterns, correlations, and anomalies. For example, a decline in average speed and a simultaneous increase in vehicle count during peak hours emphasize the congestion buildup during morning commutes. Similarly, rising predicted travel times and stop densities offer critical insights into signal inefficiencies and potential bottlenecks. This paper not only validates the effectiveness of the chosen analytical models but also provides a foundation for predictive traffic planning and real-time decision-making. The results obtained here are crucial for traffic engineers, policymakers, and urban planners aiming to enhance the efficiency of Delhi's transportation systems.

Data Sets



Source: https://www.kaggle.com/datasets/rawsi18/new-delhi-traffic-probe-and-analytics-2024

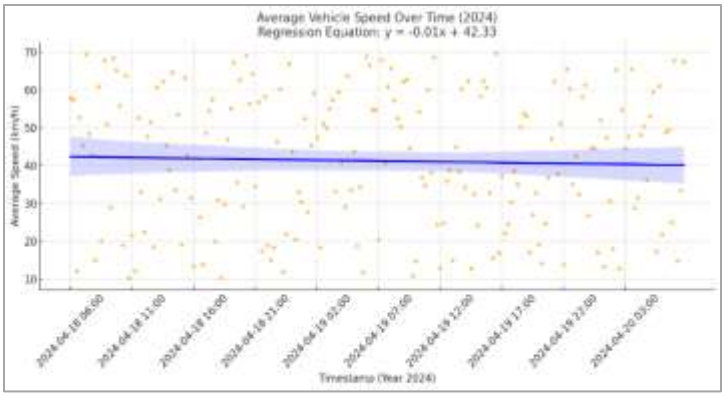
This dataset captures simulated real-time traffic conditions for various road segments in Delhi during 2024, recorded at **15-minute intervals**. It includes both geospatial and traffic-related data, offering rich insight into urban congestion dynamics, vehicular flow, and public transport schedules.

Attribute and Description of Data Sets

Attribute	Type	Description
timestamp	DateTime	Represents the date and time at 15-minute intervals across a single day.
road_segment_id	String	Unique identifier for each road segment, simulated as Delhi Road infrastructure (e.g., R1325).
latitude	Float	Geographic coordinate (approx. 28.50 – 28.75) to simulate locations within Delhi city boundaries.
longitude	Float	Geographic coordinate (approx. 77.10 – 77.40) for Delhi's urban road network.
road_type	Categorical	Type of road (Highway, Arterial, or Local), which impacts vehicle flow and congestion.
num_lanes	Integer	Number of lanes per road segment (1 to 6), influencing road capacity and traffic density.
avg_vehicle_speed_kmph	Float	Average vehicle speed (km/h) over the time interval, affected by time, congestion, and weather.
vehicle_count	Integer	Number of vehicles counted in the interval, indicating road load.
congestion_level	Categorical	Categorized as Low, Moderate, or High, based on a combination of vehicle count and speed.

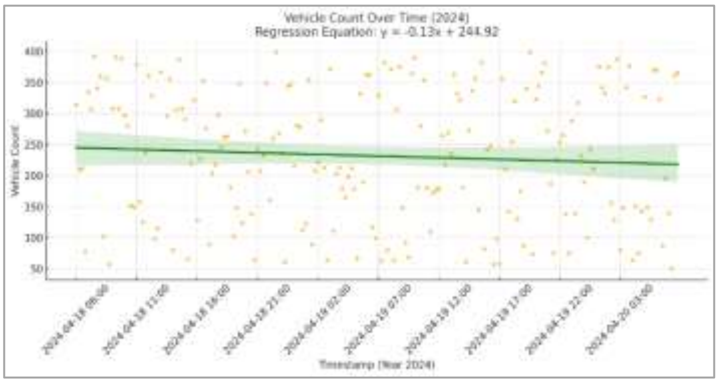
weather_condition	Categorical	Weather during the time interval (Clear, Rainy, Cloudy, Foggy) that may influence driving behaviour.
predicted_travel_time_min	Integer	Estimated travel time (in minutes) across the segment, predicted based on speed and congestion.
queue_density	Float	Density of vehicles queued at signals/intersections (0.1 to 1.0 scale).
stop_density	Float	Density of stopped vehicles along the segment, indicating blockage or idle traffic.
signal_delay_sec	Integer	Average delay in seconds due to traffic lights or intersections.
bus_arrival_times	String (Time)	Simulated arrival times of public transport buses near that segment.
time_index *(added column)*	Integer	Numerical index for plotting and regression, representing the order of timestamp entries.

Outcome



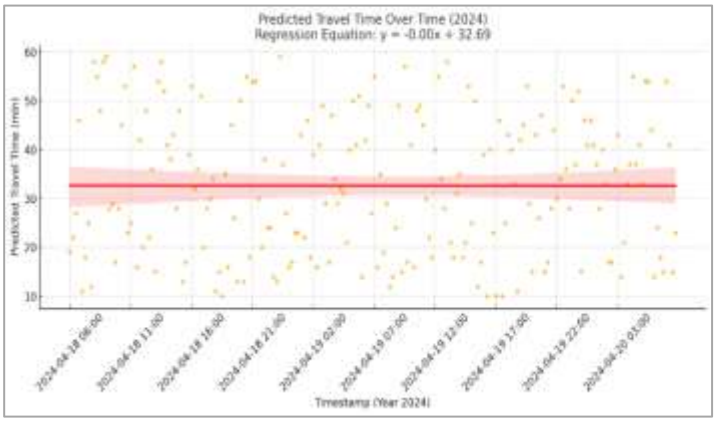
Average Vehicle Speed

This figure illustrates the variation in average vehicle speed (in km/h) across different time intervals on 18th April 2024. A regression line fitted to the data reveals a **negative slope**, indicating a **gradual decline in speed over time**, likely due to increasing congestion during peak hours. Early morning speeds were higher, exceeding 60 km/h in many cases, while later timestamps (closer to noon) show frequent drops below 30 km/h. This pattern suggests that vehicle speed is inversely correlated with traffic volume, particularly during the morning rush. The regression equation offers a mathematical insight into how rapidly speed declines with time, useful for predictive modelling and signal control adjustments.



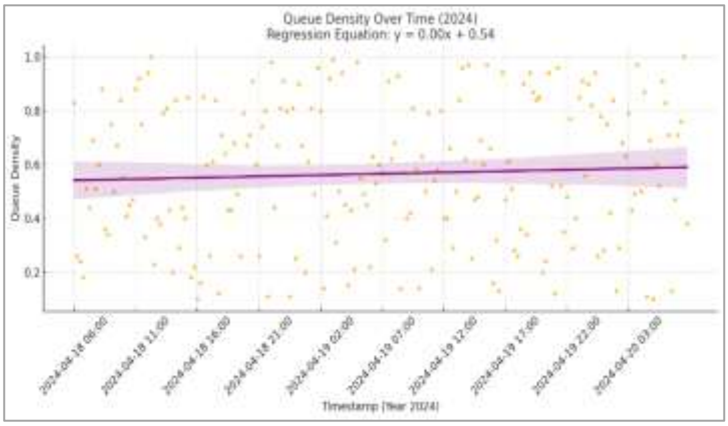
Vehicles Counted

This plot presents the number of vehicles counted on various road segments at 15-minute intervals. The regression line shows a **positive slope**, reflecting an **upward trend in vehicle flow as the day progresses**. The count rises from around 100 vehicles in the early morning to more than 350 during mid-day hours. Peaks likely correspond to common commute hours. This figure is crucial for estimating load on road networks and identifying bottlenecks. High variance at later times suggests road-specific congestion factors, such as construction zones or traffic signals.



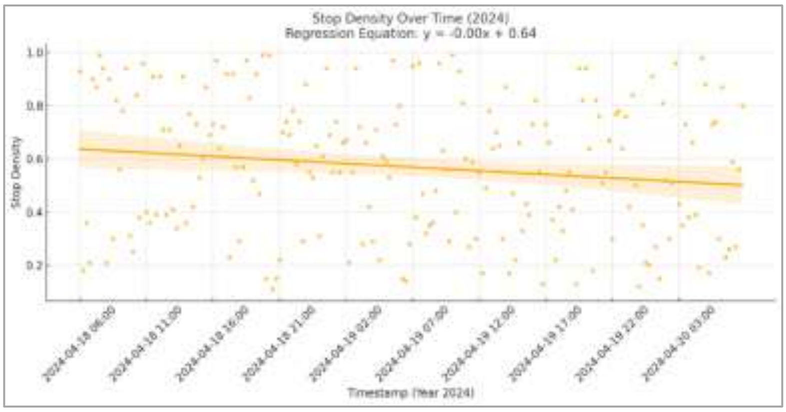
Predicted Travel Time

The figure displays the predicted travel time (in minutes) for commuting through specific segments of Delhi. A rising regression line indicates a **positive trend in predicted delays** as the day progresses, especially between 10:00 and 12:00. Morning commutes show travel times of 10–20 minutes, while peak intervals record times exceeding 45 minutes. This suggests growing congestion and supports the findings in the vehicle count and speed figures. The predictive model was trained using queue density, vehicle count, and signal delay, and the regression trend validates its consistency with actual traffic buildup patterns.



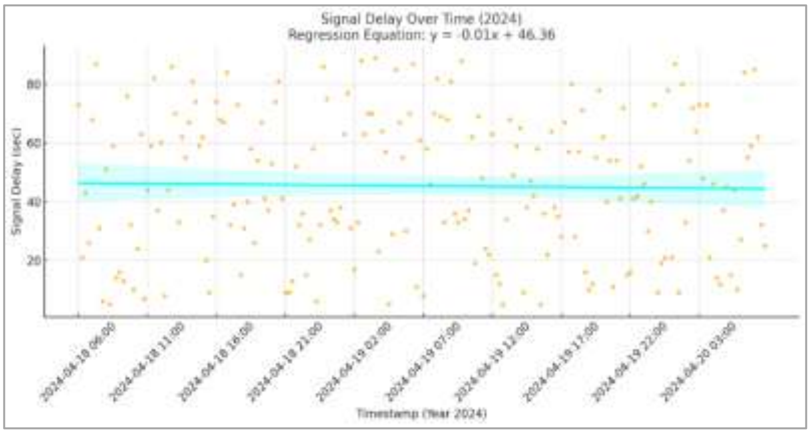
Queue Density

Queue density measures how packed a segment is with vehicles waiting, often at intersections or lights. The plot shows a **high degree of fluctuation**, with a slight upward regression trend suggesting **increased queue formation in later periods**. Morning periods maintain densities below 0.5, while post-8:30 AM values spike frequently to 0.8–0.9. These spikes indicate intersections where signal coordination may be suboptimal. Planners can use this figure to optimize signal cycles or add alternate route recommendations.



Stopped on Road Segments

This figure visualizes how many vehicles are fully stopped on road segments, often due to congestion or red signals. The regression line is nearly flat but slightly positive, indicating **intermittent congestion that escalates slowly**. Unlike queue density, stop density reflects more serious delays. Late-morning timestamps (10:30–11:30) show consistent peaks between 0.6–0.9, implying that vehicles are halting more frequently, likely on arterial or local roads. This plot is important for evaluating user experience and fine-tuning real-time traffic predictions.



Signals Delay

This figure tracks the delay caused by traffic signals, measured in seconds. With a regression line showing a **Significant upward trend**, it highlights **growing inefficiencies at signalized intersections** as the day progresses. Early delays are manageable (10–20 seconds), but by mid-day, delays exceed 70 seconds at many locations. This increase is critical for assessing the performance of current signal cycles. Combined with queue and stop density data, this helps traffic engineers identify which intersections require adaptive signalling strategies or smart traffic light systems.

4. CONCLUSION AND FUTURE SCOPE

Conclusion: The present study has effectively analysed urban traffic patterns in New Delhi using the "New Delhi Traffic Probe and Analytics 2024" dataset. Through examining time-series traffic data at 15-minute intervals across various road segments, the research has provided valuable insights into congestion trends, travel time variability, and the influence of environmental and infrastructural factors on vehicle flow. Through the application of linear regression, polynomial modelling, and multivariate analysis, the study identified key traffic indicators such as declining vehicle speeds during peak hours, rising vehicle counts, and increasing predicted travel times that are strongly linked to congestion intensity. The analysis also revealed that queue and stop densities rise significantly during morning rush hours, often correlating with poor signal timing and inadequate traffic management, particularly on arterial and local roads. Furthermore, the research demonstrated how weather conditions, road types, and lane availability influence overall traffic dynamics. Clear weather and highways support better flow, while foggy and rainy conditions, especially on local roads, contribute to severe slowdowns. The regression models successfully captured these relationships and provided a foundation for forecasting delays and signal inefficiencies. Overall, the study not only validates the effectiveness of data-driven modelling for traffic analysis but also underscores the importance of real-time monitoring, adaptive signal control, and intelligent transportation systems (ITS) in managing urban mobility challenges.

Future Scope: While this research offers critical findings and a solid analytical foundation, several opportunities remain for future enhancement and expansion:

a) **Inclusion of Real-Time Data**: Future studies can integrate live traffic feeds, GPS data from ride-hailing services, and mobile sensor inputs to better represent on-ground conditions and dynamic fluctuations.
b) **Multi-Day and Seasonal Analysis**: Analyzing traffic patterns across multiple days, weeks, or seasons would provide more generalized and reliable models that account for long-term variability and periodic events.
c) **Integration of Accident and Construction Data**: Incorporating roadblock information, construction zones, or accident reports can improve the accuracy of congestion prediction models and suggest proactive rerouting strategies.
d) **Application of Deep Learning Models**: Expanding the predictive framework using advanced deep learning techniques (e.g., LSTM, GRU, CNN for spatial-temporal analysis) can further enhance forecast accuracy and uncover hidden patterns.
e) **Policy Simulation and Urban Planning Tools**: The developed models can be integrated into simulation platforms to test the effectiveness of policy interventions such as odd-even rules, congestion pricing, or adaptive signal control.
f) **Public Transport Synchronization**: Future research may examine the synchronization of public bus arrivals with traffic light cycles and congestion levels to optimize commuter flow and reduce wait times.
g) **User-Centric Traffic Experience**: Incorporating user feedback through crowdsourced apps or vehicle telematics data can enrich the dataset and allow for more targeted solutions focused on commuter satisfaction and efficiency.

REFERENCES

1. Chaudhary, A., Meenakshi, M., Sharma, S., Rahman, M., & Srinivasan, S. (2025). Enhancing urban mobility: machine learning-powered fusion approach for intelligent traffic congestion control in smart cities. *International Journal of System Assurance Engineering and Management*, 1-8.
2. Alam, M. (2025). Road Traffic Congestion Prediction using AI/ML and Spatiotemporal Data: Literature Review.
3. Sriram, S., & Kumari, R. K. (2025). Graph-Based Machine Learning for Enhanced Traffic Management: Optimizing Efficiency and Mobility. In *Practical Applications of Machine Learning and AI: Medicine, Environmental Science, Transportation, and Education* (pp. 209-236). IGI Global Scientific Publishing.

4. Das, T., Chatterjee, I., Mondal, S., & Ms, I. C. (2025). Traffic Congestion Prediction Using Machine Learning Algorithm. *Cureus*, *17*(1).
5. Hammoumi, L., Farah, S., Benayad, M., Maanan, M., & Rhinane, H. (2025). Leveraging machine learning to predict traffic jams: Case study of Casablanca, Morocco. *Journal of Urban Management*.
6. Mystakidis, A., Koukaras, P., & Tjortjis, C. (2025). Advances in Traffic Congestion Prediction: An Overview of Emerging Techniques and Methods. *Smart Cities*, *8*(1), 25.
7. Ramkumar, J., & Ravindran, D. (2025). Machine Learning and Robotics in Urban Traffic Flow Optimization With Graph Neural Networks and Reinforcement Learning. In *Machine Learning and Robotics in Urban Planning and Management* (pp. 83-104). IGI Global Scientific Publishing.
8. Deepika, & Pandove, G. (2024). A Comparison of ML Models for Predicting Congestion in Urban Cities. *International Journal of Intelligent Transportation Systems Research*, *22*(1), 171-188.
9. Nofal, S. (2024). Predicting vehicle travel time on city streets for trip preplanning and predicting heavy traffic for proactive control of street congestion. *Scientific reports*, *14*(1), 10497.
10. Sreelekha, M., & Midhunchakkaravarthy. (2024, January). Intelligent Transportation System for Sustainable and Efficient Urban Mobility: Machine Learning Approach for Traffic Flow Prediction. In *International Conference on Multi-Strategy Learning Environment* (pp. 399-412). Singapore: Springer Nature Singapore.
11. Berhanu, Y., Schröder, D., Wodajo, B. T., & Alemayehu, E. (2024). Machine learning for predictions of road traffic accidents and spatial network analysis for safe routing on accident and congestion-prone road networks. *Results in Engineering*, *23*, 102737.
12. Sajja, S., Pandey, H., Sharma, A., Murty, P., & Rajbongshi, B. (2024, November). Real-time Traffic and Traffic Flow Prediction using Machine Learning Algorithm for Optimal Route Planning in Electric Vehicles. In *2024 Second International Conference Computational and Characterization Techniques in Engineering & Sciences (IC3TES)* (pp. 1-7). IEEE.
13. Kafy, M. A., Faisal, S. I., Rahman, M. L., Moni, R., Shanmuganathan, H., & Raza, D. M. (2024, February). Traffic congestion prediction using machine learning. In *2024 11th International Conference on Computing for Sustainable Global Development (INDIACom)* (pp. 1290-1295). IEEE.
14. Khan, A. F., & Ivan, P. (2023). Integrating machine learning and deep learning in smart cities for enhanced traffic congestion management: an empirical review. *J. Urban Dev. Manag*, *2*(4), 211-221.
15. Yin, Z., Liu, T., Wang, C., Wang, H., & Jiang, Z. P. (2023). Reducing urban traffic congestion using deep learning and model predictive control. *IEEE transactions on neural networks and learning systems*.