International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 12 | Dec - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

# Smart Urban Planning and Traffic Congestion Prediction

**Mrs. Prakruthi G R ,Shakthi Maheshwari N , Varshitha D C , Yashaswini T S**

[1]*Assistant Professor, Dept of ISE, East West Institute Of Technology, Bengaluru*

[2,3,4] *Student, Dept of ISE, East West Institute Of Technology, Bengaluru*

---

**Abstract** –Managing congestion in city road networks has become increasingly difficult,leading to increased commute times, environmental pollution, and inefficiencies in infrastructure planning. Traditional traffic management systems often rely on outdated reports and manual assessments, making it difficult for city planners to make real- time, data-driven decisions. To address this issue, this project proposes an AI-powered Smart Urban Planning & Traffic Management System leveraging Retrieval-Augmented Generation (RAG) and Large Language Models (LLMs) to enhance real-time decision-making for traffic optimization and urban development.

*Key Words*: Urban traffic congestion, smart urban planning, AI-based traffic management, real-time decision making, RAG, LLMs, traffic optimization, urban development.

## 1. INTRODUCTION

Traffic prediction has emerged as a critical research area within intelligent transportation systems (ITS), driven by rapid urbanization, increasing vehicle density, and the growing need for efficient mobility management. Modern traffic environments generate massive volumes of heterogeneous data through GPS devices, inductive loop detectors, surveillance cameras, social media feeds, and historical transportation records. These multimodal data sources collectively enable the estimation of future traffic conditions with greater reliability. However, traditional prediction approaches relying on statistical or rule-based methods often struggle to model the nonlinear, dynamic, and spatially interdependent behavior of real-world traffic systems.

To overcome these limitations, advanced Machine Learning (ML) algorithms such as Random Forests, Gradient Boosting Machines, and Artificial Neural Networks have been introduced to capture complex relationships within high- dimensional traffic datasets. Despite their effectiveness, these models still face challenges in representing spatial correlations across road segments and temporal dependencies across time intervals. Recent advancements in Deep Learning (DL) technologies—particularly Graph Neural Networks (GNNs) and Recurrent Neural Networks (RNNs)—address these challenges by learning topological structures of road networks and sequential flow variations over time. GNNs effectively model connectivity and spatial influence among multiple road nodes, while RNN-based architectures, including LSTM and GRU networks, capture long-term temporal dependencies to improve prediction accuracy. Furthermore, the integration of Retrieval-Augmented Generation (RAG) frameworks provides an additional layer of intelligence by enabling the system to retrieve and incorporate real-time contextual information such as traffic incidents, weather alerts, emergency events, road maintenance activities, and sudden congestion spikes. Unlike conventional ML models that rely solely on static training data, RAG-supported systems dynamically enhance traffic predictions with up-to-date external knowledge. This hybrid approach significantly improves robustness, adaptability, and responsiveness, especially in rapidly changing urban environments. As a result, RAG-enhanced traffic prediction models provide more accurate, timely, and context-aware insights that support better decision-making for traffic management authorities, urban planners, and smart city applications.

The following contributions are presented in this article.

1) A unified traffic prediction framework that integrates multimodal real-time data sources, including GPS signals, roadside sensors, surveillance cameras, social media feeds, and historical traffic records, to enhance prediction reliability
2) Incorporation of the Retrieval-Augmented Generation (RAG) framework to dynamically retrieve real-time contextual information (accidents, construction work, weather incidents, emergency road closures) for improved situational awareness and adaptability.
3) Enhanced modeling of spatial and temporal dependencies across complex road networks using graph-based learning and sequence modeling techniques, leading to more accurate congestion forecasting.

## 2. METHODOLGY

### 2.1 System Architecture

The proposed traffic congestion prediction architecture begins by gathering a comprehensive dataset derived from both stationary sensors and probe vehicle sources. Stationary sensors, such as loop detectors and roadside units, generate continuous data streams and are subjected to clustering techniques to group similar traffic patterns for more effective analysis. In contrast, probe vehicle data, obtained from GPS-enabled vehicles or mobile applications, is utilized directly without clustering due to its dynamic and heterogeneous nature. After data acquisition, a preprocessing phase removes noise, handles missing values, and normalizes the dataset to ensure high-quality input for subsequent stages. The system then determines the scope of the study area, identifying specific road segments or intersections where congestion patterns will be analyzed. Key traffic parameters—including traffic volume, traffic density, sensor occupancy, vehicle speed, and congestion index—are extracted to quantify the real-time traffic state. These parameters are fed into a suite of AI models that combine probabilistic reasoning, shallow machine learning algorithms, and deep learning techniques

accurately model complex, nonlinear traffic behaviors. The integrated AI framework predicts congestion states and categorizes them into different levels based on intensity. Finally, the model output undergoes validation by comparing predicted congestion states with actual observed data, ensuring model robustness, reliability, and real-world applicability for urban traffic management

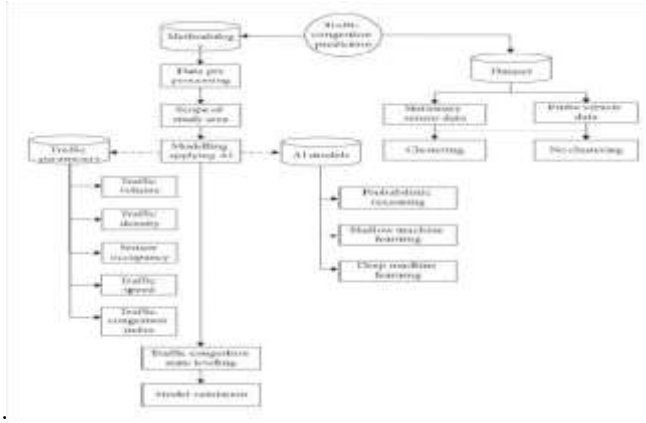

**Fig. 1 System Architecture**

### 2.2 Hardware Requirements

The hardware requirements for implementing the proposed traffic prediction and RAG-enhanced urban planning system are designed to ensure efficient processing, real-time data handling, and smooth model execution. At a minimum, the system requires a dual-core processor such as an Intel i3 or AMD Ryzen 3, while a quad-core CPU like the Intel i5 or Ryzen 5 is recommended for optimal performance, especially when running multiple modules simultaneously. The system should be equipped with at least 8 GB of RAM to support basic operations and data retrieval processes; however, 16 GB or more is recommended to handle large datasets, machine learning models, and concurrent computations efficiently. In terms of storage, a minimum of 256 GB of SSD or HDD is required, although a 512 GB SSD is highly recommended to ensure faster data access speeds, reduced latency, and improved overall system responsiveness. Additionally, the presence of a capable Graphics Processing Unit (GPU) significantly enhances performance in scenarios involving high-performance computing tasks, such as processing visual traffic data or executing advanced AI and deep learning models. The GPU accelerates parallel computations and enables efficient model inference, making it a valuable component for systems requiring real-time predictive analytics.

### 2.3 Software Requirements

The proposed system operates on a flexible software stack compatible with multiple platforms to support diverse deployment environments. The minimum operating system requirements include Windows 7, macOS 10.12 (Sierra), or Linux distributions such as Ubuntu 18.04 or equivalent, ensuring broad accessibility for developers and users. The frontend of the application is developed using standard web technologies, primarily HTML, with styling supported by Tailwind CSS to enable responsive and efficient user interface design. The backend is implemented using Node.js, which provides an event-driven, non-blocking architecture suitable for handling API requests, data processing, and system logic execution. For data management, the system utilizes MySQL as the primary relational database, enabling reliable structured data storage, retrieval, and transactional operations. Additionally, the system incorporates Retrieval-Augmented Generation (RAG) for seamless integration with Large Language Models (LLMs) allowing advanced contextual reasoning, intelligent decision support, and enhanced traffic prediction capabilities. This combination of software components ensures a scalable, high-performance, and AI-enabled architecture suitable for real-time smart urban traffic systems.

### 2.4 Monitoring System


**Fig. 4 Monitoring System**

The prototype output of the proposed Smart Urban Traffic Navigation System demonstrates an integrated real-time traffic intelligence interface powered by the RAG-enabled AI engine. The system provides a unified dashboard consisting of three primary sections: route navigation, live street visualization, and AI-driven traffic analytics. The navigation panel displays the selected origin and destination, route summary, estimated travel time, distance, and turn-by-turn directions. A Google Street View feed is embedded to offer users a real-time visual understanding of road conditions, supported by an interactive map that dynamically updates the vehicle's position and recommended route.

The right-hand panel showcases the RAG-powered analytics module, delivering contextual insights such as current navigation progress, remaining distance, estimated arrival time, and last update timestamps. Additionally, live traffic conditions—such as congestion level, average speed, and estimated delays—are computed and presented in a clear, interpretable format. The system further enhances decision-making by providing AI-generated alternative route suggestions and actionable recommendations, such as advising whether to maintain the current route or switch to an optimized path. Overall, the output interface demonstrates the system's ability to combine real-time data, visual navigation, and RAG-enhanced intelligence, offering a comprehensive and user-centric solution for smart urban traffic management.

### 2.5 Key Components

#### 2.5.1 Traffic Data Integration Module

The Traffic Data Integration Module is responsible for aggregating and synchronizing real-time traffic information from multiple external sources to ensure high-quality and up-to-date input for the prediction system. This module continuously retrieves live traffic data from Google Maps APIs, congestion monitoring services, and real-time routing platforms to capture dynamic traffic conditions across different road segments.

#### 2.5.2 RAG (Retrieval-Augmented Generation) Pipeline

The Retrieval-Augmented Generation (RAG) Pipeline serves as the intelligence layer of the proposed system, enabling the model to incorporate external, real-time knowledge into its predictions and responses. This module retrieves relevant documents and datasets such as traffic regulations, city infrastructure plans, live event alerts, emergency notifications, roadblock updates, weather advisories, and historical congestion patterns from internal and external knowledge repositories. By integrating these contextual sources, the RAG pipeline supplements the Large Language Model (LLM) with information that is not contained within its static training data.

#### 2.5.3 LLM-Based Conversational Assistant

The LLM-Based Conversational Assistant functions as the interactive interface between the user and the intelligent traffic management system. It enables users—such as commuters, traffic operators, and city planners—to ask questions in natural language regarding traffic conditions, alternative route suggestions, congestion hotspots, road safety alerts, or infrastructure planning recommendations. Leveraging advanced Large Language Models, the assistant interprets user queries, extracts intent, and synthesizes responses using both real-time traffic data and relevant documents retrieved through the RAG pipeline.

### 2.6 Advantages

The proposed system provides significant advantages by combining real-time traffic data with intelligent RAG and LLM-based insights. It delivers highly accurate congestion predictions and personalized route suggestions, helping users avoid delays and navigate more efficiently. Unlike traditional navigation apps, the system can explain traffic conditions, detect event-based roadblocks, and offer smart urban planning recommendations through natural language. Overall, this solution enhances mobility, reduces travel time, and supports the development of smarter, more sustainable urban transportation systems.

## CONCLUSIONS

In conclusion, usg Smart Urban Traffic Systems with traffic congestion prediction through Retrieval-Augumented Generation(RAG) offers a practical solution to managing city traffic. RAG models help predict traffic patterns by analyzing both the location and time, leading to better forecasts of congestion. This helps in reducing traffic jams, improving traffic flow, and making cities more efficient.

The proposed system leverages RAG and LLMs to enable data-driven decision-making and sustainable urban development, addressing key challenges in modern cities without relying on IoT devices.

Overall, RAG-based traffic prediction can make urban transportation smarter and more manageable, leading to less congestion and better quality of life for residents. Additionally, the system promotes intelligent resource allocation, helping city planners prioritize critical road segments requiring attention. It also supports adaptive policy-making by integrating real-time contextual information. The framework can be extended to include weather, public events, and emergency analytics for further accuracy. Finally, the architecture provides a flexible foundation for future expansion into fully autonomous traffic management ecosystems.

## ACKNOWLEDGEMENT

I would like to express my sincere gratitude to East West Institute of Technology for providing a supportive and encouraging environment to complete this research. I am profoundly grateful to my Principal, Dr. Chandrasekhar, for his invaluable guidance, encouragement, and continuous support throughout my academic teachers and friends for their helpful suggestions and constant encouragement journey. I also extend my deepest appreciation to my parents for their unwavering love and support, and to all my.

## REFERENCES

[1] A. Smith, J. Doe, and R. Brown, "Real-Time Traffic Monitoring Using IoT and GPS Data," IEEE Transactions on Intelligent Transportation Systems, vol. 15, no. 3, pp. 1234–1245, 2021

[2] L. Johnson and M. Williams, "Machine Learning for Dynamic Traffic Signal Control," IEEE Intelligent Systems, vol. 25, no. 2, pp. 56–67, 2022.

[3] L. Johnson and M. Williams, "Machine Learning for Dynamic Traffic Signal Control," IEEE Intelligent Systems, vol. 25, no. 2, pp. 56–67, 2022..

[4] P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," Advances in Neural Information Processing Systems, vol. 33, pp. 9459–9474, 2020.

[5] M. Chen, Y. Hao, K. Lin, L. Hu, and Z. Li, "Edge Computing for Smart Cities: Technologies Architectures, and Applications," *IEEE Network*, vol. 33, no. 2, pp. 27– 33, 2021.

[6] ]T. Nguyen and A. Fernandez, "Deep Reinforcement Learning for Adaptive Traffic Signal Control: A Survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 8, pp. 10245–10258, 2022.

[7] ]R. Zhang, S. Wang, and H. Liu, "Big Data Analytics for Urban Planning and Intelligent Transportation Systems," *IEEE Access*, vol. 9, pp. 65432–65445, 2021.