IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, December 2025
Impact Factor: 7.67

Real Time Big Data Processing for Smart City Traffic Management

Shristy Jain and Shailendra Shriwastava
Research Scholar, Babulal Tarabai Institute of Research and Technology, Sagar, India
Assistant Professor, Research scholar Babulal Tarabai Institute of Research and Technology, Sagar, India

Abstract: *Real-time big data processing has become fundamental to smart city traffic management, addressing challenges such as rising vehicle density, congestion, pollution, and road safety concerns in rapidly urbanizing areas. By integrating data from GPS devices, IoT sensors, traffic cameras, and social media, modern traffic systems can dynamically analyze road conditions and support informed decision-making. Real-time analytics enable adaptive traffic signaling, congestion prediction, route optimization, and efficient incident management. These capabilities are supported by scalable data architectures that incorporate machine learning, predictive analytics, and cloud computing to handle fluctuating data volumes, especially during peak hours. Privacy and security are ensured through strong data governance practices. Beyond improving traffic flow, real-time processing contributes to environmental sustainability by reducing emissions and fuel consumption, while also enhancing economic productivity through reduced travel delays and improved logistics. Emerging technologies such as edge computing, 5G connectivity, and autonomous vehicles are expected to further advance the responsiveness and precision of traffic management systems. Overall, real-time big data processing is reshaping urban mobility, enabling smarter, more sustainable, and more livable cities..*

Keywords: Real-Time Big Data Processing, Smart City Traffic Management, IoT and Sensor Networks, Predictive Analytics, Edge and Cloud Computing

I. INTRODUCTION

In recent years, the rapid urbanization of cities and the growing number of vehicles have posed significant challenges to traffic management systems. As cities expand and their populations grow, the demand for efficient transportation solutions has become more critical than ever. Traditional traffic management approaches, which rely on static data and pre-defined rules, have proven inadequate in addressing the dynamic and complex nature of modern urban traffic. This has led to the rise of real-time big data processing as a transformative technology in smart city traffic management. Real-time big data processing involves the continuous collection, analysis, and interpretation of vast amounts of traffic-related data from diverse sources, such as sensors, GPS devices, cameras, and social media. These data streams provide an up-to-the-moment view of traffic conditions, enabling city planners, traffic controllers, and autonomous systems to make data-driven decisions aimed at optimizing traffic flow, reducing congestion, and enhancing safety.

At the heart of this transformation is the ability to process data at high speeds and in real-time, allowing for immediate responses to changing traffic patterns. Advanced algorithms and machine learning models are employed to predict traffic congestion, identify bottlenecks, and provide real-time recommendations for traffic signal adjustments, route diversions, and incident management. For example, intelligent transportation systems (ITS) that leverage big data processing can dynamically adjust traffic signals to alleviate congestion during peak hours or reroute vehicles in response to accidents or road closures. Furthermore, real-time data processing supports the integration of various smart city initiatives, such as connected vehicles and autonomous transportation, by ensuring seamless communication between infrastructure and vehicles.

The benefits of real-time big data processing for smart city traffic management extend beyond improving traffic flow and reducing delays. It plays a crucial role in enhancing public safety by enabling quicker responses to accidents, breakdowns, or emergencies. In addition, the data generated from these systems can be analyzed to identify long-term

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, December 2025
Impact Factor: 7.67

trends, helping city planners design better infrastructure, optimize public transportation routes, and create more sustainable urban environments. Environmental benefits are also significant, as efficient traffic management reduces fuel consumption and emissions, contributing to cleaner air and lower carbon footprints.

Smart Cities And Traffic Management

Smart cities represent a transformative approach to urban management, using advanced technologies to enhance the efficiency, sustainability, and livability of cities. With increasing urbanization, governments and city planners worldwide are focusing on creating environments that can support the rapid growth of urban populations while addressing various challenges like traffic congestion, pollution, energy consumption, and public safety. At the heart of this transformation is the integration of real-time big data processing, which plays a pivotal role in managing complex urban systems, particularly traffic management.

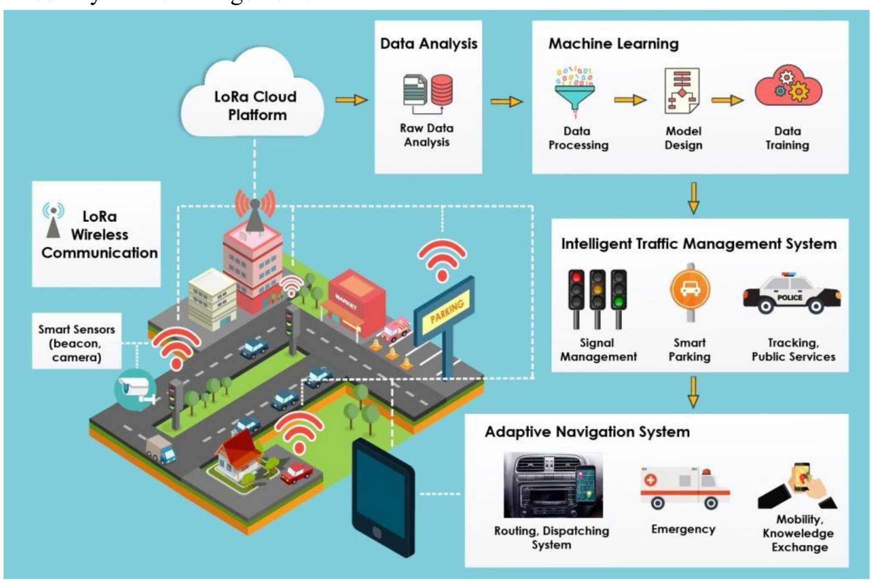

Figure 1 Smart Cities And Traffic Management

II. LITERATURE REVIEW

[1] Visual Big Data Analytics for Traffic Monitoring in Smart City,Dinesh Singh, C. Vishnu and C. Krishna Mohan

The application such as video surveillance for traffic control in smart cities needs to analyze the large amount (hours/days) of video footage in order to locate the people who are violating the traffic rules. The traditional computer vision techniques are unable to analyze such a huge amount of visual data generated in real-time. So, there is a need for visual big data analytics which involves processing and analyzing large scale visual data such as images or videos to find semantic patterns that are useful for interpretation. In this paper, we propose a framework for visual big data analytics for automatic detection of bike-riders without helmet in city traffic. We also discuss challenges involved in visual big data analytics for traffic control in a city scale surveillance data and explore opportunities for future research.

Construction framework of smart tourism big data mining model driven by blockchain technology ,Xiaowen Long , Weiqiang Chen , 2024

In the era of sharing economy, the tourism market is increasingly characterized by personalized demand, mobile consumption and product segmentation. This paper aims to apply big data mining technology in the field of smart tourism. Firstly, it focuses on image summary selection and collaborative filtering technology based on big data mining. It then demonstrates the integration of blockchain in smart tourism, emphasizing the use of decentralized structures and smart contracts to achieve data security and transparency, and describes the testing process of smart tourism platforms, including data preparation and platform perational efficiency testing. Finally, the research results of this paper are summarized, and the development potential and practical application value of smart tourism are demonstrated. The results show that in the smart tourism big data mining model, the minimum support for the data set is 10 % and 20 %, respectively. Moreover, with the increase of the number of nodes in the same data set, the running time decreases gradually. It can be seen that smart tourism big data mining has strong scalability.

GC-YOLOv9: Innovative smart city traffic monitoring solution, Ru An , Xiaochun Zhang , Maopeng Sun , Gang Wang, 2024

In urban smart city environments, traffic hazards can lead to catastrophic outcomes, including significant property losses and severe threats to public safety. Conventional traffic monitoring systems are limited in terms of accuracy and speed, presenting significant challenges for real-time traffic surveillance. To tackle these challenges, this paper introduces the GC-YOLOv9 algorithm. Specifically, we have enhanced the YOLOv9 model by incorporating Ghost Convolution, markedly improving the model's perceptual abilities and detection accuracy. Furthermore, this study designed an integrated smart city framework that includes layers for service applications, the Internet of Things, edge processing, and data centers. By deploying the enhanced YOLOv9 model within this framework, our method achieved mAP@0.5 scores of 77.15 and 74.95 on the BDD100K and Cityscapes datasets, respectively, surpassing existing technologies. Additionally, the potential applications of this method in public area fire safety management, forest fire monitoring, and intelligent security systems further underscore its significant value in improving the safety and efficiency of smart cities.

Big data analytics and artificial intelligence aspects for privacy and security concerns for demand response modelling in smart grid: A futuristic approach ,S. Sofana Reka , Tomislav Dragicevic b, Prakash Venugopal , V. Ravi , Manoj Kumar Rajagopal, 2024

Next generation electrical grid considered as Smart Grid has completely embarked a journey in the present electricity era. This creates a dominant need of machine learning approaches for security aspects at the larger scale for the electrical grid. The need of connectivity and complete communication in the system uses a large amount of data where the involvement of machine learning models with proper frameworks are required. This massive amount of data can be handled by various process of machine learning models by selecting appropriate set of consumers to respond in accordance with demand response modelling, learning the different attributes of the consumers, dynamic pricing schemes, various load forecasting and also data acquisition process with more cost effectiveness. In connected to this process, considering complex smart grid security and privacy based methods becomes a major aspect and there can be potential cyber threats for the consumers and also utility data. The security concerns related to machine learning model exhibits a key factor based on different machine learning algorithms used and needed for the energy application at a future perspective. This work exhibits as a detailed analysis with machine learning models which are considered as cyber physical system model with smart grid. This work also gives a clear understanding towards the potential advantages, limitations of the algorithms in a security aspect and outlines future direction in this very important area and fastgrowing approach.

Internet of things and optimized knn based intelligent transportation system for traffic flow prediction in smart cities,Sunkara Teena Mrudula , Meenakshi , Mahyudin Ritonga , S. Sivakumar , Malik Jawarneh , Sammy , T. Keerthika , Kantilal Pitambar Rane , Bhaskar Roy, 2024

The rapid expansion of urban areas and the increasing number of vehicles on the road have resulted in accidents, traffic congestion, economic repercussions, environmental deterioration, and excessive fuel consumption. A dependable traffic management system is necessary to anticipate and regulate urban traffic patterns. Traffic forecast aids in the prevention of traffic issues. Urban traffic predictions often utilise historical and current traffic flow data to forecast road conditions. This article presents a traffic flow prediction system that utilises the Internet of Things (IoT), machine learning, and feature selection. Internet of Things (IoT) devices located on highways or within cars gather sensor data in real-time. The input data set comprises both real-time Internet of Things (IoT) data and historical traffic statistics. The input data is stored in a centralized cloud. The data is subjected to preprocessing in order to eliminate any unwanted interference and identify any exceptional values. The accuracy and root mean square error are contingent upon the process of feature selection. Particle swarm optimization identifies and extracts crucial features from input data. The

ISSN: 2581-9429
Volume 5, Issue 4, December 2025
Impact Factor: 7.67

classification model is constructed using K Nearest Neighbor, Multi layer Perceptron, and Bayes network approaches. The UCI traffic data is used for conducting experiments. The dataset has 47 attributes and 2102 occurrences. The accuracy of traffic flow prediction using PSO KNN is 96 %. The PSO KNN algorithm achieved a Mean Square Error (MSE) of 0.00289 and a Root Mean Square Error (RMSE) of 0.0595.

III. RESEARCH METHODOLOGY

1. Data Collection

Data collection is a critical step in real-time big data processing for traffic management. The data will be gathered from various sources that reflect the dynamic nature of urban traffic. These sources include:

Sensors: Traffic sensors placed at intersections, roads, and highways collect data on vehicle count, speed, and flow.

CCTV Cameras: Closed-circuit television cameras monitor road conditions, traffic density, and detect accidents or breakdowns.

GPS Devices: Data from vehicle GPS systems provide real-time information about vehicle location, speed, and route patterns.

Social Media: Posts related to traffic incidents, road conditions, and other mobility-related issues on platforms like Twitter and Facebook will be extracted.

Mobile Applications: Traffic data from navigation apps like Google Maps and Waze will be incorporated to gain insights into traffic trends and road congestion.

Weather Data: Weather conditions, which influence traffic patterns, will be collected from meteorological agencies.

2. Data Preprocessing

Before analysis, the raw data collected from different sources will require preprocessing to ensure consistency, accuracy, and reliability. Data preprocessing involves the following steps:

Data Cleaning: Removing incomplete, redundant, or erroneous data to improve the quality of analysis. This includes handling missing values, eliminating duplicate records, and correcting inconsistencies in traffic data.

Data Transformation: Normalizing data from different sources to ensure uniformity. For example, sensor data might be in numerical format while social media data is unstructured text. The text will be converted into structured data through Natural Language Processing (NLP) techniques.

Data Integration: Combining data from heterogeneous sources (e.g., sensor, GPS, and weather data) to create a cohesive dataset that can be processed in real-time.

Data Segmentation: Segmenting data into various traffic-related factors such as traffic volume, speed, road conditions, and vehicle types for better analysis.

3. System Architecture Design

The system architecture for real-time big data processing will follow a layered approach:

Data Ingestion Layer: This layer is responsible for collecting and aggregating data from various real-time sources such as IoT devices, sensors, and social media. Apache Kafka or MQTT will be used to ensure a high-throughput data ingestion pipeline.

Data Processing Layer: The core of the system, where real-time data is processed using Apache Spark Streaming or Flink. This layer will handle tasks such as traffic pattern detection, anomaly detection, and predictive analysis using machine learning algorithms.

Data Storage Layer: Processed data will be stored in distributed databases such as Hadoop HDFS or NoSQL databases like MongoDB. This layer ensures scalable and reliable storage of large datasets.

Application Layer: The user-facing layer will include dashboards, mobile apps, and control centers where traffic management decisions are made in real-time. Visualization tools such as Tableau or Grafana will be used for presenting the processed data.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 4, December 2025 Impact Factor: 7.67

4. Model Development

The research will develop machine learning and AI-based models to analyze and predict traffic patterns, optimize traffic signals, and manage congestion. These models will include:

Traffic Prediction Models: Predictive models such as Long Short-Term Memory (LSTM) networks and time-series forecasting models will be used to anticipate future traffic trends based on historical and real-time data.

Anomaly Detection Models: These models will detect unusual traffic patterns caused by accidents, road closures, or other incidents. Unsupervised machine learning techniques like clustering and anomaly detection will be employed.

Optimization Models: Real-time optimization of traffic signals and routing algorithms to minimize delays and ensure smooth traffic flow. Reinforcement learning algorithms will be implemented to optimize signal timings dynamically.

5. Performance Evaluation

The performance of the developed system and models will be evaluated using several key metrics:

Accuracy: The accuracy of traffic prediction models in forecasting traffic flow and detecting anomalies.

Latency: The response time of the system to process real-time data and make decisions. The goal is to achieve low latency for real-time processing.

Scalability: The ability of the system to handle increasing volumes of traffic data as cities grow and more sensors are deployed.

Robustness: Evaluating how well the system performs under different conditions such as data loss, sensor failures, or network disruptions.

IV. RESEARCH TOOLS

Tools and Technologies for Data Collection:

- Apache Kafka for streaming real-time data.
- Hadoop and Spark for big data storage and processing.
- IoT platforms such as AWS IoT for managing sensor data.

Tools for Data Preprocessing:

- Python libraries (Pandas, NumPy) for data cleaning and transformation.
- Apache Nifi for data ingestion and flow automation.
- SQL for managing and querying the preprocessed data.

V. CONCLUSION

The conclusion of a study on real-time big data processing for smart city traffic management should emphasize the transformative potential of integrating advanced data analytics and IoT technologies in urban environments. By leveraging real-time data from various sources such as sensors, cameras, GPS devices, and mobile applications, smart city traffic management systems can optimize traffic flow, reduce congestion, and enhance overall mobility. These systems use big data to predict traffic patterns, adjust signal timings, and provide real-time updates to commuters, thereby minimizing delays and improving road safety. Furthermore, the analysis of big data enables city planners to make informed decisions regarding infrastructure development and transportation policies. While the implementation of such systems faces challenges related to data privacy, interoperability, and infrastructure costs, the long-term benefits include reduced carbon emissions, fuel consumption, and economic losses due to traffic congestion. Ultimately, real-time big data processing offers a scalable and adaptable solution for managing the increasing demands on urban transportation networks, paving the way for more efficient, sustainable, and livable cities in the future.

REFERENCES

[1] H. Wang, Y. Zheng, and Y. Xue, "Travel time estimation based on floating car data and real-time traffic information," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1537–1547, 2013.

[2] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of Things for Smart Cities," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 22–32, 2014.
[3] M. Chen, S. Mao, Y. Zhang, and V. Leung, *Big Data: Related Technologies, Challenges and Future Prospects*. Springer, 2014.
[4] Y. Zheng, "Methodologies for cross-domain data fusion: An overview," *IEEE Transactions on Big Data*, vol. 1, no. 1, pp. 16–34, 2015.
[5] M. S. Hossain and G. Muhammad, "Cloud-assisted industrial internet of things (IIoT)–enabled framework for health monitoring," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4041–4050, 2019.
[6] R. Yu et al., "Traffic prediction in smart cities using machine learning and data mining," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 6, pp. 4110–4119, 2021.
[7] Y. Yuan, G. Tan, and X. Li, "Real-time traffic monitoring using edge computing and 5G," *IEEE Network*, vol. 34, no. 5, pp. 45–51, 2020.
[8] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions," *Future Generation Computer Systems*, vol. 29, no. 7, pp. 1645–1660, 2013.
[9] Y. Liu, F. Shang, X. Guo, and Y. Li, "Deep learning architecture for traffic flow prediction," *Transportation Research Part C*, vol. 111, pp. 171–190, 2020.
[10] A. Al-Fuqaha et al., "Internet of Things: A survey on enabling technologies, protocols, and applications," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 2347–2376, 2015.
[11] S. Vlahogianni, M. Karlaftis, and E. Golias, "Short-term traffic forecasting: Where we are and where we're going," *Transportation Research Part C*, vol. 43, pp. 3–19, 2014.
[12] N. Gershenfeld, R. Krikorian, and D. Cohen, "The Internet of Things," *Scientific American*, pp. 76–81, 2004.
[13] P. Zhang, J. Sun, and L. Yang, "Real-time congestion detection using deep CNN models," *IEEE Access*, vol. 8, pp. 96411–96421, 2020.
[14] X. Xu et al., "Edge computing for intelligent transportation systems: A review," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 8, pp. 4929–4943, 2021.
[15] S. R. Sukumar, H. Haleem, and S. J. Lee, "Big data analytics in transportation: A review," *IEEE Big Data Conference*, pp. 202–210, 2018.
[16] L. Li, Y. Yu, and W. Wang, "A deep learning approach for traffic speed prediction using LSTM networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 8, pp. 1–10, 2019.
[17] M. Imran, M. A. R. Baig, S. Y. Shin, and M. Ahmad, "IoT-enabled intelligent transportation systems for smart cities," *IEEE Access*, vol. 8, pp. 145128–145138, 2020.
[18] A. Khattak, H. Shah, and M. Khan, "Real-time anomaly detection in traffic using unsupervised deep learning," *IEEE Access*, vol. 9, pp. 82391–82405, 2021.
[19] M. F. Khan, S. A. Hussain, and A. Rezaei, "Big data frameworks (Hadoop, Spark) for large-scale traffic analysis," *Journal of Big Data*, vol. 6, pp. 1–20, 2019.
[20] S. Al-Sultan, A. Al-Doori, A. Al-Bayatti, and H. Zedan, "A comprehensive survey on vehicular ad hoc networks," *Journal of Network and Computer Applications*, vol. 37, pp. 380–392, 2014.
[21] R. B. Júnior et al., "Urban mobility analytics using multimodal data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 4, pp. 2716–2728, 2021.
[22] L. M. Gambs, M. Killijian, and M. N. de Amorim, "A survey on privacy in mobility data," *ACM Computing Surveys*, vol. 48, no. 1, pp. 1–33, 2015.
[23] P. S. Kumar and S. P. Mohanty, "Traffic flow optimization using reinforcement learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 2, pp. 1457–1467, 2022.
[24] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE Internet of Things Journal*, vol. 3, no. 5, pp. 637–646, 2016.
[25] S. F. Ghahramani, M. Zhou, and W. Wang, "Deep learning for intelligent transportation systems: A review," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 1, pp. 132–158, 2022.