International Journal of Engineering and Techniques - Volume 12 Issue 2, March-April - 2026

RESEARCH ARTICLE OPEN ACCESS

Time Series Traffic Congestion Prediction Using Hybrid Machine Learning Models

Mr.S.Amaresan[1], R.Vishwanath[2], S.A.Infant Vishal[3], A.Sanjay[4]

[1] Professor, Department of Computer Science and Engineering, Parisutham Institute of Technology and Science, Thanjavur, Tamil Nadu – 613006, India
Emai: amaresancse1974@gmail.com

[2]UG Student, Department of Computer Science and Engineering, Parisutham Institute of Technology and Science, Thanjavur, Tamil Nadu – 613006, India
Email: avishwanath198@gmail.com

[3] UG Student, Department of Computer Science and Engineering, Parisutham Institute of Technology and Science, Thanjavur, Tamil Nadu 613006, India
Email: infantvishal1234@gmail.com

[4]UG Student, Department of Computer Science and Engineering, Parisutham Institute of Technology and Science, Thanjavur, Tamil Nadu – 613006, India
Email: sanjaydevops.in@gmail.com

Abstract:

Traffic congestion has become a critical issue in modern urban environments due to rapid population growth and the increasing number of vehicles. Traditional traffic management systems rely heavily on real-time monitoring techniques such as surveillance cameras and sensors, which are expensive and reactive in nature. These systems fail to provide predictive insights that can help in proactive traffic management. This paper presents a machine learning-based approach for predicting traffic congestion using historical traffic data.

The proposed system, TrafiVista, utilizes supervised learning algorithms such as Random Forest and XGBoost to analyze traffic patterns and classify congestion levels into Low, Medium, and High categories. The system is implemented as a web-based platform using FastAPI, React, and SQLite, allowing users to input parameters such as time and location to obtain predictions efficiently.

The approach offers a cost-effective, scalable, and intelligent solution for traffic prediction without relying on expensive infrastructure. By leveraging historical data, the system enables better traffic planning and decision-making, contributing to improved transportation efficiency.

I. INTRODUCTION:

The rapid growth of urbanization and population has led to a significant increase in the number of vehicles on roads, resulting in severe traffic congestion in cities and semi-urban areas. Traffic congestion not only increases travel time but also contributes to higher fuel consumption, environmental pollution, and stress among commuters. Efficient traffic management has therefore become a major challenge for transportation systems worldwide.

Traditional traffic management systems primarily depend on real-time monitoring using surveillance cameras, sensors, and manual control by traffic authorities. Although these systems provide real-time insights, they are reactive in nature and respond only after congestion has already occurred. Moreover, the installation and maintenance of such infrastructure involve high costs and complexity, making them less feasible for large-scale deployment.

With advancements in technology, machine learning has emerged as a powerful tool for analyzing large volumes of data and identifying patterns. In the context of traffic management, machine learning techniques can be used to analyze historical traffic data and predict future congestion levels. This predictive capability enables proactive decision-making, allowing users and authorities to take preventive measures before congestion occurs.

The proposed system, TrafiVista, focuses on developing a machine learning-based traffic congestion prediction model using historical data. The system analyzes key features such as time, date, and location to identify traffic patterns and forecast congestion levels. Unlike traditional systems, it does not rely on expensive real-time infrastructure, making it a cost-effective and scalable solution.

Additionally, the system is implemented as a web-based application that allows users to easily access predictions by entering required parameters. This improves usability and ensures that the system can be widely adopted for real-world applications such as travel planning, traffic analysis, and urban development.Overall, this work aims to bridge the gap between traditional reactive systems and modern predictive approaches by leveraging machine learning for intelligent traffic congestion prediction.

II. RELATED WORK:

Hazarika [1] proposed an edge-based machine learning approach for smart traffic management in intelligent transportation systems. The study focuses on processing traffic data such as vehicle count and flow at the edge level to reduce latency and improve system efficiency. By performing computations closer to the data source, the system minimizes delays and enhances responsiveness. The research highlights how machine learning models can effectively analyze traffic patterns and support congestion prediction. It also emphasizes the importance of decentralized processing in modern traffic systems. The approach reduces dependency on centralized servers and improves scalability. However, it still requires infrastructure support for data collection. The study demonstrates that machine learning can significantly enhance traffic management efficiency. This work is relevant as it shows the effectiveness of ML in predicting traffic conditions. It supports the idea of using data-driven approaches for congestion analysis.

Saraff [2] presented a machine learning-based traffic surveillance system that analyzes video data to identify congestion patterns. The system extracts features such as vehicle density and movement from surveillance footage. These features are then used to analyze traffic behavior and detect congestion levels. The study shows that machine learning models can effectively process visual data for traffic analysis. It also highlights the importance of feature extraction in improving prediction accuracy. The system provides better insights compared to traditional monitoring methods. However, it relies heavily on camera-based infrastructure, making it costly and complex. The research demonstrates the role of machine learning in traffic surveillance applications. It also emphasizes the need for efficient data processing techniques. This work is relevant as it supports the use of ML for congestion prediction but highlights infrastructure limitations.

Lin and Jhang [3] proposed an intelligent traffic monitoring system using machine learning models. The system analyzes historical traffic data to identify congestion patterns and predict future conditions. The study shows that data-driven approaches provide more accurate insights compared to traditional rule-based systems. Machine learning algorithms are used to model complex relationships between traffic parameters. The research emphasizes the importance of historical data in traffic prediction tasks. It also demonstrates improved prediction performance using ML techniques. The system is capable of handling large datasets efficiently. However, the study does not focus on user accessibility or deployment aspects. The findings highlight the effectiveness of machine learning in traffic analysis. This work strongly supports the use of ML for congestion prediction systems.

Nashaat [4] introduced a machine learning-based traffic prediction system using data reduction techniques. The study focuses on handling large-scale traffic datasets efficiently by reducing data complexity. Various preprocessing methods are applied to improve model performance and reduce computational cost. The research highlights the importance of data cleaning and feature selection in machine learning models. It also demonstrates how reduced datasets can improve prediction accuracy. The system uses machine learning algorithms to forecast traffic conditions. The study shows that proper preprocessing significantly enhances model efficiency. However, it requires careful tuning of reduction techniques. The work emphasizes the role of data preprocessing in traffic prediction systems. It is relevant as it supports efficient model development using optimized datasets.

Shin [5] proposed a machine learning-based approach for traffic congestion prediction using historical data. The system classifies traffic conditions into different congestion levels using supervised learning algorithms. The study demonstrates that machine learning models can effectively learn traffic patterns and provide accurate predictions. It highlights the importance of feature selection and model training in achieving high performance. The system shows improved accuracy compared to traditional methods. It also emphasizes the role of historical data in predictive analysis. However, the model is limited to offline analysis and lacks real-time usability. The research provides a strong foundation for ML-based traffic prediction. It validates the effectiveness of classification models in congestion prediction. This work is directly relevant to the proposed system.

Lv et al. [6] proposed a deep learning-based approach for traffic flow prediction using large-scale traffic data. The study utilizes deep neural networks to capture complex patterns in traffic flow. It focuses on learning spatial and temporal relationships in traffic data. The model demonstrates improved accuracy compared to traditional machine learning methods. The research highlights the importance of big data in traffic prediction systems. It also shows that deep learning models can effectively handle large datasets. The system is capable of adapting to dynamic traffic conditions. However, the computational cost of deep learning models is relatively high. The study provides a strong foundation for modern traffic prediction systems. It is relevant as it introduces advanced techniques for handling complex traffic data.

Zhang et al. [7] introduced a deep spatio-temporal residual network for predicting citywide traffic flow. The model captures both spatial and temporal dependencies in traffic data. It uses residual learning to improve prediction accuracy. The study demonstrates that combining spatial and temporal features enhances performance. The system is capable of predicting traffic patterns across multiple regions simultaneously. It also handles large-scale urban traffic data efficiently. The research highlights the importance of deep learning in traffic forecasting. However, the model requires high computational resources. It may not be suitable for lightweight systems. This work is significant as it improves prediction accuracy using advanced neural networks.

Ma et al. [8] proposed a traffic speed prediction model using Long Short-Term Memory (LSTM) networks. The study focuses on capturing temporal dependencies in traffic data. LSTM models are effective in handling time-series data such as traffic patterns. The system learns long-term dependencies and improves prediction accuracy. The research demonstrates that LSTM outperforms traditional models in time-based prediction tasks. It is particularly useful for forecasting traffic speed variations. However, training LSTM models requires significant computational resources. The model also requires large datasets for better performance. The study highlights the importance of time-series analysis in traffic prediction. It is highly relevant for understanding traffic trends over time.

Min and Wynter [9] proposed a real-time traffic prediction system using spatio-temporal correlations. The study focuses on analyzing traffic data across both space and time dimensions. The model uses statistical techniques combined with machine learning methods. It improves prediction accuracy by considering correlations between different locations. The system is capable of providing real-time traffic predictions. It highlights the importance of spatial relationships in traffic analysis. However, the model may struggle with sudden traffic changes. The study provides insights into early traffic prediction techniques. It also demonstrates the benefits of combining spatial and temporal data. This work is relevant for understanding multi-dimensional traffic analysis.

Chen et al. [10] proposed a hybrid machine learning model for traffic flow prediction. The study combines multiple algorithms to improve prediction accuracy. Hybrid models leverage the strengths of different techniques. The system integrates traditional machine learning and advanced models. It shows improved performance compared to single-model approaches. The research highlights the importance of combining models for better results. The system is capable of handling complex traffic patterns effectively. However, hybrid models may increase system complexity. Proper tuning is required for optimal performance. This work is highly relevant as it supports the use of hybrid approaches in traffic prediction systems.

III. PROPOSED METHODOLOGY

The proposed system, TrafiVista, is a machine learning-based platform designed to predict traffic congestion using historical traffic data. The system focuses on analyzing traffic patterns and providing accurate congestion predictions without relying on expensive real-time infrastructure. It integrates machine learning models with a web-based application to ensure accessibility and ease of use. The methodology includes multiple stages such as data collection, preprocessing, model training, and prediction. Each stage is carefully designed to improve prediction accuracy and system performance. The system uses structured datasets containing traffic-related attributes for analysis. By leveraging machine learning techniques, the system can identify patterns and trends in traffic data. The overall architecture is modular, scalable, and efficient. The methodology ensures that the system can be extended for future enhancements. This approach provides a reliable solution for traffic congestion prediction.

a. System Overview

The proposed system consists of several interconnected modules that work together to perform traffic congestion prediction efficiently. The main modules include the User Authentication Module, Backend API and Data Preprocessing Module, Machine Learning Inference Module, and Database Management Module. Each module is responsible for a specific function within the system architecture. The user interacts with the system through a web interface where input parameters such as time and location are provided. The backend processes the input data and prepares it for prediction. The machine learning module analyzes the processed data and generates congestion predictions. The database module stores user data and prediction results for future reference. The system follows a structured workflow that ensures smooth data flow between modules. The modular design improves maintainability and scalability. Overall, the system provides an efficient framework for traffic congestion prediction.

b. Data Processing and Feature Extraction

Data processing and feature extraction play a crucial role in improving the accuracy of the machine learning model. The system uses historical traffic data containing attributes such as time, location, and vehicle count. During preprocessing, the raw data is cleaned by handling missing values and removing inconsistencies. Feature extraction is performed to identify relevant attributes that influence traffic congestion. Important features include date, time (hour, day, month), day type (weekday or weekend), location, and traffic volume. These features help the model understand traffic patterns effectively. Categorical data is converted into numerical form for model compatibility. Data normalization is applied to maintain consistency across features. Proper preprocessing ensures that the dataset is structured and reliable. This step significantly enhances model performance. Overall, data processing ensures that the input to the model is accurate and meaningful.

c. Machine Learning Model

The machine learning model is the core component of the proposed system, responsible for predicting traffic congestion levels. The system primarily uses the Random Forest algorithm due to its high accuracy and ability to handle complex data relationships. Additionally, XGBoost is used as an enhancement model to improve prediction performance. These algorithms are trained using historical traffic datasets containing labeled congestion levels. The model learns patterns based on input features such as time, location, and traffic volume. During training, the model adjusts its parameters to minimize prediction errors. The trained model is capable of classifying traffic conditions into Low, Medium, and High congestion levels. Ensemble techniques improve prediction accuracy by combining multiple decision trees. The model is evaluated using performance metrics such as accuracy. This ensures reliable and consistent predictions. Overall, the use of hybrid machine learning models enhances system effectiveness.

d. System Architecture

The system follows a web-based architecture that integrates frontend, backend, database, and machine learning components. The frontend is developed using React and Vite, providing a user-friendly interface for interaction. The backend is implemented using FastAPI, which handles API requests and processes input data efficiently. The machine learning model is integrated into the backend using Scikit-learn. SQLite is used as the database for storing user information and prediction results. SQLAlchemy is used for managing database operations in a structured manner. The architecture ensures smooth communication between components through REST APIs. Security is maintained using authentication mechanisms for user access. The system is designed to handle multiple users and requests efficiently. The modular structure allows easy maintenance and future upgrades. Overall, the architecture ensures scalability, efficiency, and reliability.

e. Prediction Workflow

The prediction workflow defines the step-by-step process followed by the system to generate traffic congestion predictions. Initially, the user inputs required details such as time and location through the web interface. The backend receives the input and performs validation to ensure correctness. The validated data is then preprocessed and converted into a feature vector. This feature vector is passed to the trained machine learning model for prediction. The model analyzes the input data and determines the congestion level based on learned patterns. The output is classified into categories such as Low, Medium, or High congestion. The predicted result is then sent back to the frontend and displayed to the user. The system ensures quick and efficient prediction without delay. The workflow is designed to be simple and user-friendly. Overall, it enables real-time prediction using historical data.

f. Data Collection

Data collection is a fundamental step in the proposed system, as it provides the raw input required for training the machine learning models. The system uses historical traffic datasets that include parameters such as vehicle count, speed, time, and location. These datasets may be obtained from publicly available sources or generated synthetically to simulate real-world traffic conditions. The collected data represents traffic patterns at different times and under varying conditions. It helps in identifying peak hours and congestion trends across different locations. The quality of the dataset directly affects the performance of the prediction model. Therefore, care is taken to ensure that the data is diverse and representative. The data is stored in a structured format for efficient processing. Proper data collection enables the system to learn meaningful patterns. Overall, it forms the backbone of the traffic prediction system.

g. Model Training and Evaluation

Model training and evaluation are essential processes in building an accurate traffic prediction system. The collected dataset is divided into training and testing sets to ensure proper validation of the model. During training, the machine learning algorithms learn relationships between input features and congestion levels. The system uses Random Forest and XGBoost algorithms for classification tasks. These models are trained using labeled data to predict congestion categories effectively. After training, the model is evaluated using performance metrics such as accuracy, precision, and recall. A confusion matrix is also used to analyze classification results. Hyperparameter tuning is performed to improve model performance. The evaluation process ensures that the model generalizes well to new data.

This step helps in selecting the best-performing model. Overall, it ensures reliability and efficiency in predictions.

h. Web Application Integration

Web application integration enables users to interact with the traffic prediction system in an efficient and user-friendly manner. The system is implemented using a frontend developed with React and a backend built using FastAPI. The trained machine learning model is deployed within the backend for real-time predictions. Users can input parameters such as date, time, and location through the web interface. The backend processes the input data and communicates with the machine learning model to generate predictions. The results are then displayed on the frontend in a clear and understandable format. REST APIs are used for seamless communication between the frontend and backend. Authentication mechanisms ensure secure access to the system. The integration allows the system to be accessed from any device with internet connectivity. Overall, it enhances usability and enables real-world application of the system.

IV. CONCLUSION

This paper presents a machine learning-based traffic congestion prediction system that utilizes historical traffic data to forecast congestion levels effectively. The proposed system addresses the limitations of traditional traffic management approaches by introducing a predictive model that enables proactive decision-making. By using machine learning algorithms such as Random Forest and XGBoost, the system can accurately classify traffic conditions into different congestion levels based on input parameters like time and location.

Overall, the proposed system demonstrates the effectiveness of machine learning in solving real-world traffic problems. It provides a cost-effective and scalable alternative to traditional traffic monitoring systems. By leveraging historical data and predictive analytics, the system contributes to improved traffic planning, reduced congestion, and better transportation management. This work highlights the potential of intelligent systems in transforming modern traffic management solutions.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to their project guide and faculty members of the Department of Computer Science and Engineering for their valuable guidance and continuous support throughout this work. The authors also thank their institution for providing the necessary resources and environment to complete this project successfully. Finally, they extend their thanks to their friends and family for their encouragement and support.

REFERENCES

[1] A. Hazarika, “Edge ML Technique for Smart Traffic Management in Intelligent Transportation Systems,” 2024.

[2] A. Saraff, “Indian Traffic Surveillance Video Summarization Using Machine Learning,” 2025.

[3] C.-J. Lin and J.-Y. Jhang, “Intelligent Traffic Monitoring System Using Machine Learning Models,” 2022.

[4] H. Nashaat, “Machine Learning-Based Traffic Prediction Using Data Reduction Techniques,” 2024.

[5] D.-H. Shin, “Prediction of Traffic Congestion Using Machine Learning Approaches,” 2020.

[6] Y. Lv, Y. Duan, W. Kang, Z. Li, and F. Wang, “Traffic Flow Prediction With Big Data: A Deep Learning Approach,” 2015.

[7] J. Zhang, Y. Zheng, and D. Qi, “Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction,” 2017.

[8] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, “Long Short-Term Memory Neural Network for Traffic Speed Prediction,” 2015.

[9] W. Min and L. Wynter, “Real-Time Road Traffic Prediction With Spatio-Temporal Correlations,” 2011.

[10] Z. Chen, X. He, and J. Wang, “Hybrid Machine Learning Model for Traffic Flow Prediction,” 2018.