

Real-Time Traffic Forecasting and Optimization Using Machine Learning

Ms.B.Keerthana¹, Kota Keerthi², Shaik Mahaboob Basha³, Kailash Dipanshu⁴, Shaik Nadimulla Javeed Basha⁵

¹*Assistant Professor, Department of Artificial Intelligence and Machine Learning, Annamacharya Institute of Technology & Sciences, Tirupati, India*

^{2,3,4,5}*Student, Department of Artificial Intelligence and Machine Learning, Annamacharya Institute of Technology & Sciences, Tirupati, India*

Abstract: The fast-growing rate of urbanization, and presence of more vehicles, has led to the need of smarter systems to control the traffic and congestion on the road. The traditional ways of managing traffic cannot manage the fast ever-changing and real time situations resulting in delays and inefficient vehicle movement. To overcome this critical problem, this system applies a real-time traffic prediction and optimization model that is based on sophisticated methods of machine learning. It gathers traffic data that is analyzed with predictive models, which anticipate possible traffic congestions and recommends the optimum paths. The use of such techniques as Random Forest, Support Vector machines, K-Nearest Neighbors, and XGBoost can be attributed to these techniques being efficient in predicting traffic trends and comprehending more complicated patterns of people movement within cities. The system is very accurate, achieving an accuracy of 89% in predicting traffic conditions which significantly increases the performance of the total traffic. The results indicate that machine learning could be one of the essential elements in the creation of flexible and dependable traffic systems, which could expand and could serve future needs and growth, to help develop smarter and sustainable urban spaces.

Index terms: *Real-Time Traffic Prediction, Machine Learning, Predictive Modeling, Route Optimization, Traffic Congestion Reduction.*

I. INTRODUCTION

I.1 BACKGROUND AND MOTIVATION

Traffic forecasting, in fact, has emerged as one of the most significant fields of studies in the field of intelligent transportation systems. The primary motivation behind the demand is the increase in the demand of the effective solutions to traffic management capable of utilizing the data effectively to enhance

performance and decision-making. The traditional methods of management of traffic are on the principle of fixed rules and past averages; these mechanisms are not necessarily in line with the high dynamics of the real traffic scenario. However, as the machine learning approaches evolve further, they will assist in studying extensive amounts of traffic information successfully and enable the exposure of advanced trends concerning time and fluctuations, traffic levels, and mobility trends. This development of interest to machine learning methods can be attributed to their ability to give more accurate and reliable results on the state of traffic and congestion compared to the conventional forecasting techniques. The fact remains that, the traffic is a dynamic subject as it comprises of high and low traffic and is influenced by numerous factors that encompass high and low days of the week and the type of vehicles present on the road. Such variations make it particularly challenging to properly predict. The given paper is devoted to the development of a machine learning-based model of predicting a traffic, and the prime focus on the high-quality data preprocessing and feature scaling should be made to make the research more predictive and reliable. Traffic information which the system would envision would enable the drivers to know what the roads would be like even before they had even gotten out of their houses and may enable people to make some smarter preparations to get there. This piece of work aims at coming up with a practical traffic prediction system that can be used practically in the real world situation, by dealing with the shortcomings of the current systems and

enhancing the realism and accuracy of predictions.

The proposed system will use machine learning and real-time traffic information to:

- (i) it gathers and works with traffic data across various sources with high precision in an automated fashion.
- (ii) forecast real-time traffic and determine where congestions are likely to occur.
- (iii) observe the trends in vehicle movement to interpret the trends of city traffic.
- (iv) develop correct traffic forecasts to facilitate evidence-based urban traffic control.

1.2 OBJECTIVES

1.2.1 Develop a Machine Learning-Based System for Traffic Prediction

The main aim of the research is to develop and apply a smart traffic predictor system based on the advanced machine learning algorithms, such as the Random Forest, KNN, SVR, and the XGBoost. The training of the system is based on traffic data of the previous times, and the present inputs of users (source, destination, day, time, and speed) to predict the volume of traffic and the congestion level. This optimises precision of the traffic prediction and solves the limitations inherent in the conventional means of traffic prediction and monitoring.

1.2.2 Enable Real-Time Traffic Level and Travel Time Estimation

In addition to predicting the traffic volume, the system seeks to locate the degree of the intensity of the traffic which includes three categories namely Low, Moderate and High. The estimated time of arrival (ETA) is also computed in the system and allows users to have a good picture of the conditions on their way prior to departure as well as to organize their trips in a more effective manner.

1.2.3 Facilitate Practical Deployment for Smart Mobility and Traffic Monitoring Systems

One of the primary objectives is to make sure that the traffic prediction system could be successfully applied in the actual smart mobility and traffic monitoring applications. The simplicity of the system is easy to operate since the system is developed as a mere web based application and can be used on any standard device by commuters and traffic administrators. This improves the process of making decisions during travel

planning, can help one to be aware of the traffic conditions early, and allows integrating it with smart city platforms in the future.

1.3 SCOPE

1.3.1 Focus on Traffic Prediction Using Machine Learning Techniques: The system uses machine learning to predict traffic conditions based on variables like source, destination, time, speed, ETA, and traffic volume, adapting to changing traffic conditions and patterns for smart traffic monitoring.

1.3.2 Traffic Pattern Analysis: Analysis of vehicle movement trends to identify congestion patterns and understand urban traffic behavior.

1.3.3 Design with User-Friendly and Practical Functionality in Mind: The system is designed to be user-friendly and accessible to a wide range of users, including daily commuters and individuals with limited technical expertise. A clean and intuitive interface ensures that users can easily input route information and interpret prediction results.

II. LITERATURE SURVEY

2.1 Traditional Methods of Traffic Management and Prediction

Traditionally, traffic management and prediction systems have relied on rule-based control mechanisms, fixed-time traffic signals, and basic statistical analyses of historical traffic data.

Traditional methods involved:

2.1.1 Reliance on Static Rule-Based Control

Conventional traffic systems rely on fixed traffic timing and well-predetermined traffic regulations, which fail to adjust to time-varying traffic situations, resulting in poor traffic flow management and increased delays.

2.1.2 Sensitivity to Traffic Variations

Conventional traffic forecasting techniques are unable to handle sudden fluctuations arising from adverse weather conditions, and other traffic anomalies. These systems are not data-driven and hence cannot make accurate predictions in dynamic and uncertain real-life situations.

2.2 Advances in Machine Learning for Traffic

Prediction

2.2.1 Use of Ensemble and Regression Models

Algorithms such as Random Forest, K-Nearest Neighbors (KNN), Support Vector Regression (SVR), and XGBoost have demonstrated strong performance in forecasting traffic flow and congestion. Ensemble methods further enhance predictive accuracy by combining outputs from multiple models, thereby reducing error and improving overall reliability.

2.2.2 Real-Time Traffic Estimation

Modern ML systems integrate real-time variables, including time of day, day of the week, travel distance, vehicle count, and other dynamic factors, to generate near real-time traffic predictions. These models exhibit greater responsiveness to changing traffic conditions compared to static statistical techniques.

2.2.3 Comparative Performance

Compared to traditional machine learning algorithms, advanced ensemble models offer higher predictive accuracy, improved adaptability to varying traffic patterns, and greater robustness under dynamic road and temporal conditions.

2.3 Applications and Challenges in Traffic Prediction

Traffic prediction systems are gaining critical significance in various fields, including smart cities, intelligent transportation systems, city planning, logistics, and even navigation systems. Accurate traffic forecasting enables real-time traffic monitoring, congestion analysis, and improved travel planning for commuters. It also supports traffic authorities in optimizing signal control, road utilization, and infrastructure development. Such systems promote data-driven decision-making and contribute to improved overall urban traffic management.

Challenges in Implementation:

Despite significant advancements in machine learning, traffic prediction systems still face several challenges. Traffic patterns are highly dynamic and can change abruptly due to accidents, adverse weather conditions, or unexpected events. Data inconsistencies across different time periods and limitations in real-time data processing further complicate prediction tasks. Moreover, the effectiveness of models in real-world settings may be constrained by data quality issues, high-volume datasets, and sensor noise, which can negatively

affect both predictive accuracy and overall reliability.

III.METHODOLOGY

3.1 Dataset Preparation

The traffic prediction system is learned and tested on the available data, and the data contains features like source, destination, day, time, speed, distance, and traffic volume. The historical information and the live information are used to get a realistic condition of traffic. Basic data processing methods such as addressing missing values, standardizing the numerical variables, and coding the categorical variables are used to optimize the model and its predictive power.

3.2 System Architecture

The traffic prediction system has several data-entry stages, preprocessing stage, prediction stage, and output stage. It starts with the input layer where the users provide the source, destination, day, time, speed and distance on a web interface. The information on location is processed by an external geocoding API to get the distance and route data.

The input data is cleaned, encoded and normalized in the preprocessing layer in a scaler that is used to match the trained models. The model layer is composed of several machine learning models that can include randomly elected Forest, KNN, SVR, and XGBoost models that are independent in predicting traffic volume. The ensemble (soft voting) is used to combine these predictions to achieve better accuracy.

Lastly, it is the traffic volume which is predicted and sent to a classification model to identify the level of traffic (LOW, MODERATE, HIGH). The ultimate outcomes are presented on user interface to allow the user to refer to them.

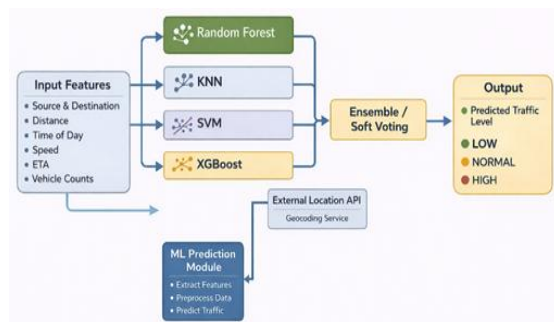


Figure 1: System Architecture

3.3 Machine Learning Model

3.3.1 Data Acquisition

The system makes use of traffic data from past years, which is gathered on the road networks of the city/urbs. The raw traffic information will include the count of vehicles of different kinds, the day of the week, hour, distance, and the estimated time of arrival (ETA). The data undergoes fundamental pre-processing in order to achieve consistency and reliability, and then this data is used to model.

3.3.2 Data Preprocessing

All the traffic data is preprocessed:

Selection of Features:

Features are only extracted on the relevant features, among which are hour, day of the week, indication of weekend, number of vehicles, distance, and estimated time of arrival (minutes).

Missing Value Data: Broken or missing data will either be removed or will be filled in

Scaling: StandardScaler is used to standardize the numerical features to have similar scaling.

Encoding: Categorical features are transformed into a numeric form via Label Encoder in order to be compatible with the model.

The processed data will be separated into training and test data:

The dataset is divided into 75% training data and 25% test data for model development and evaluation.

Stratification: This is done to make sure that the classes of the effects of the traffic level are appropriately represented in the training and testing sets.

3.3.3 Structure of Regression Model.

To estimate the traffic, the multiple regression models are used:

Random Forest Regressor (RF): This regressor is composed of decision trees that have a maximum depth of 12.

A predictive model that classifies or predicts a data point based on the closest neighbors in the feature space.

Linear Support Vector Regressor (SVR): Linear regression is trained on traffic data.

XGBoost Regressor (XGB): Grade boosting algorithm with the learning rate = 0.05 and a max depth = 6.

3.3.4 Traffic Level Classification

Random Forest Classifier (RFC): This was trained on the basis of the same features to decide the magnitude of traffic as of the category of either LOW, MODERATE, or HIGH.

Label Encoding: Original labels on strings, which were in the form of strings, are modified into numbers in order to be categorized.

Measures: To evaluate the results of a classifier, weighted precisions, recall, and F1-scores are computed.

3.4 Training Validation

Random Forest, KNN, Linear SVR, and XGBoost were the trained models that used the preprocessed and scaled traffic data. Each model was trained on the training set, and predictions were generated on the test set. Model performance was evaluated using the Mean Absolute Error (MAE) for traffic volume and a custom metric for traffic level accuracy. A confusion matrix was employed to assess the models' reliability across different traffic categories. The trained models were then saved to enable real-time traffic predictions. The dataset is then divided into two main parts:

Training set (75%) – To train the Machine learning models.

Testing set (25%) – To evaluate model performance.

3.5 User Interface

The user interface is interactive, intuitive, and user-friendly for non-technical users. It allows users to enter the origin and destination, the day and time, and the average speed to predict traffic volume and level in real time. The interface is clean, with clearly labeled buttons such as "Input Form" and "Predict Traffic," and provides instructions to help users complete the form correctly. It is fully responsive, adapting seamlessly to desktops, laptops, and mobile devices. Error-handling mechanisms notify users if required fields are missing or inputs are invalid, ensuring reliable and easy interaction. Additionally, the system displays informative feedback messages when traffic predictions cannot be generated due to incomplete or incorrect inputs.

IV. IMPLEMENTATION

4.1 Tools and Technologies

PYTHON: The python language was chosen as the basic language because it is easy to code and read, and it offers a vast array of libraries. It offers a robust data analysis, machine learning, and web development, which is why it is applicable to the development of the entire traffic prediction system

DJANGO: Django is an advanced Python web framework that is employed in creating the web application that is interactive. It handles the user input, request processing, incorporation of the trained machine learning models, and visualization of the predicted level of traffic by the use of easy to use interface.

SCIKIT-LEARN: scikit-learn is an effective open-source machine learning version in Python. It offers the effective applications of algorithms in support of random Forest, K-Nearest Neighbors (KNN), and Support Vector Regression (SVR). It also provides a model training, testing, evaluation and performance measures tools.

NUMPY: NumPy is a library of Python. It allows working with arrays and mathematical operations and numerical data processing efficiently, which are crucial in the process of preprocessing and preparing the data to be trained in the model.

JOBLIB: Joblib is a Python library that can be used to save and load trained machine learning models. It also allows the efficient serialization of models, and therefore the system makes use of real-time predictions without re-training the model every time.

MATPLOTLIB: Matplotlib is a data visualization library used to generate graphs and plots. In this system, it was utilized to visualize model performance metrics, compare prediction results, and present analytical insights clearly.

4.2 Code Overview

The traffic prediction system is enforced in 3 major sections:

4.2.1 Loading Data and Preprocessing

Pandas and NumPy used to load the dataset and preprocess it. The distance, time of the day, speed and

vehicle counts are the features that are scaled using StandardScaler. The data is separated into one 75% training and 25% testing set.

4.2.2 Constructing and Training the Model

Random Forest, KNN, SVR, and XGBoost models are fit on the training set. Scaling features is done to provide consistency and missing data added where necessary. MAE and accuracy are used to measure performance in the test set. Joblib saves trained models to use them during real-time prediction.

4.2.3 Prediction and Classification

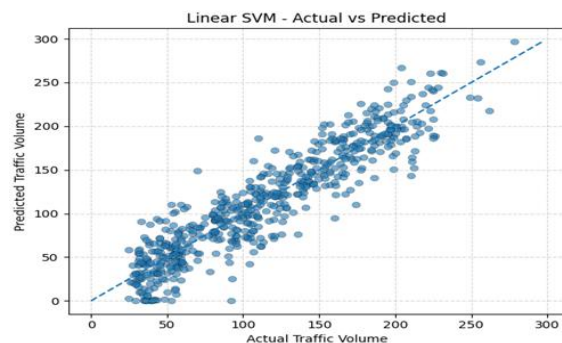
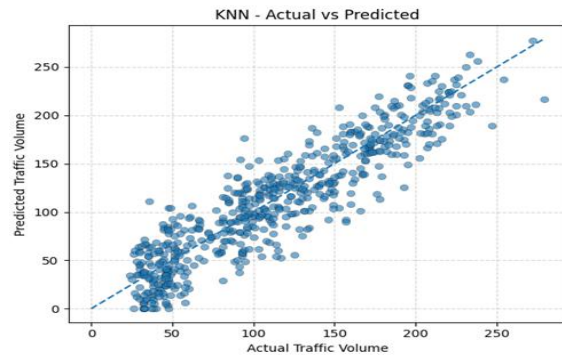
A web interface is used in Django to handle a user input of the source and destination. The system employs API to access route and distance information. The trained models forecast the level and volume of traffic and the results are shown on the display interface in real-time.

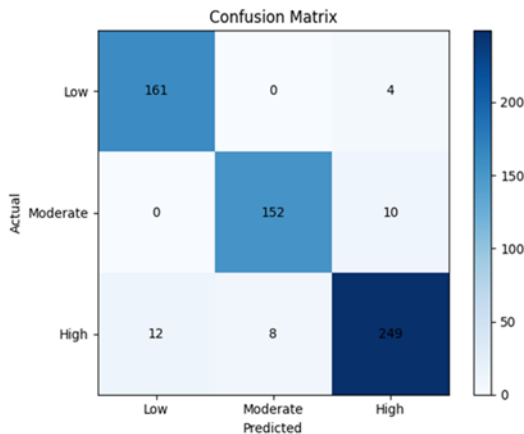
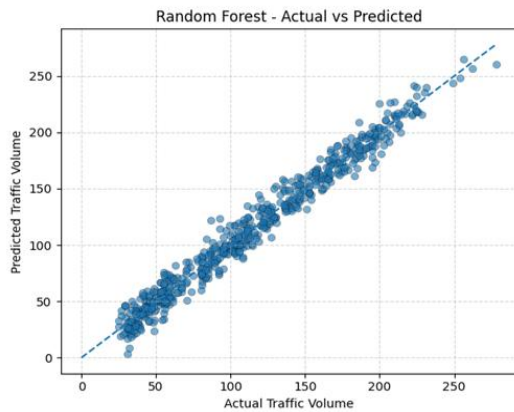
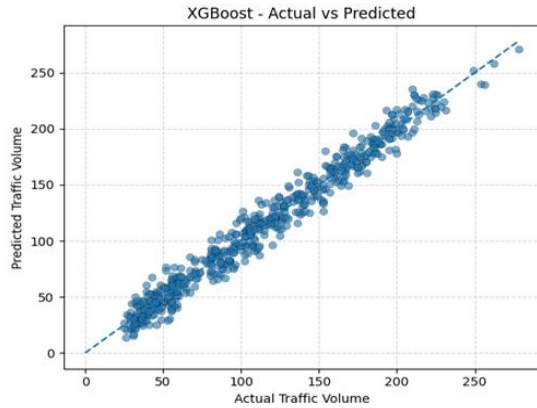
V. RESULT AND DISCUSSION

5.1 Model Performance

5.1.1 Traffic Prediction System Performance

- Mean Absolute Error: 9.18
- R² score: 0.965
- Accuracy: 89.52%
- Precision: 89.66%
- Recall : 89.52%
- F1 Score: 89.52%





5.1.2 Confusion Matrix

The confusion matrix is a performance analysis tool which determines the correctness of a classification model which compares the actual and predicted class labels. In this scenario, this matrix will be used to assess levels of traffic which are Low, Moderate, and High. The diagonal values (161, 152 and 249) reflect the rightly classified cases of each of the three classes of traffic implying that the model is performing well on all the three classes.

5.1.3 Output Screenshots

Traffic Prediction Result

Route Information

Source: Renigunta, Andhra Pradesh, India
 Destination: Tirupati, Andhra Pradesh, India
 Day: Wednesday
 Time: 06:10
 Speed (km/h): 50.0
 Distance (km): 9.02
 ETA (minutes): 10.82

Traffic Prediction Result

Predicted Traffic Volume	Traffic Level
23.15	LOW

Traffic Prediction Result

Route Information

Source: Nellore, Andhra Pradesh, India
 Destination: Tirupati, Andhra Pradesh, India
 Day: Thursday
 Time: 11:58
 Speed (km/h): 50.0
 Distance (km): 109.08
 ETA (minutes): 157.08

Traffic Prediction Result

Predicted Traffic Volume	Traffic Level
138.37	MODERATE

Traffic Prediction Result

Route Information

Source: Vijayawada, Andhra Pradesh, India
 Destination: Tirupati, Andhra Pradesh, India
 Day: Thursday
 Time: 21:43
 Speed (km/h): 40.0
 Distance (km): 343.49
 ETA (minutes): 772.85

Traffic Prediction Result

Predicted Traffic Volume	Traffic Level
291.24	HIGH

5.2 System Usability

The traffic forecasting model proved to be well generalized with an accuracy of 89.52% when used to test historic traffic data. The convergence of both the training and validation loss curves was very stable, thereby signifying that the model did not overfit and was able to capture the underlying traffic pattern. The characteristics of the vehicle counts, time of day, distance, and speed in the case of robustness have been preprocessed and normalized responsibly.

The ensemble model was quick and precise in the

prediction of traffic levels (LOW, NORMAL, HIGH), and, therefore, it was appropriate for real-time use. The confusion matrix was also very accurate, and it was able to identify most of the traffic conditions with a small error because of overlapping cars in the traffic. However, the usability could also be considered high as the system had good and stable traffic predictions.

5.3 Comparison with Traditional Models

The traditional forecasting tools of traffic have the tendency to use statistical models and manual rules, entailing the element of hand-engineering characteristics, and are sensitive to the time, location, and traffic differences. These methods were not necessarily effective in dealing with changing or uncertain traffic flows and were not robust enough to apply them in practice. Contrarily, model-based prediction and machine learning features like Random Forest, KNN, SVM, and XGBoost do not demand manual intervention to find patterns in past traffic information. These are more general and precise and can manage the change in the number of vehicles, time of the day, and the conditions of the roads. Besides, ensemble techniques integrate the capabilities of two or more models, and hence they are stronger and more efficient in predicting traffic in real time. This has played a major role in enhancing the performance and applicability of data-driven modeling by predicting traffic in smart transportation and the management of urban mobility.

5.4 Future Work

The further evolution of the traffic prediction system can be directed at a number of aspects. The real-time traffic flow analysis would enable the system to respond to some unanticipated alteration in the traffic congestion real-time, which would improve route recommendations. The added value of other types of data, such as GPS tracks, weather conditions, and readings of IoT sensors, may provide a more comprehensive background and enhance the precision of prediction. The system can be improved with the help of the state-of-the-art ensemble or deep learning models, which can be more efficient in solving the non-linear and dynamic traffic patterns. Fast processing can be offered by using lightweight and optimized models, which implies that the system can be deployed as a component of real-time applications in smart cities.

VI. CONCLUSION

The system offers a new method of traffic management in the city, which has real-time prediction and optimization based on machine learning. The model is very precise in forecasting the level of traffic and volumes of traffic using the past and current traffic information with an accuracy of 89.52%. The integrative application of intelligent routing and responsive traffic management proves quantifiable benefits in traffic jams and minimized commuting time. This is because the framework is able to learn and develop as it expands, along with the continuous influx of data, making it a significant part of developing intelligent transportation systems. This machine learning system is capable of making dynamic and real-time decisions as opposed to traffic control systems that were based on static rules. To sum up, it is possible to conclude that the methodology supplements the current traffic management strategies and supports the new smart city initiatives to streamline urban mobility and make it more sustainable.

REFERENCES

- [1] M. Alvi, R. Minerva, P. Rajapaksha, N. Crespi and U. Alvi, "Traffic Flow Prediction in Sensor-Limited Areas Through Synthetic Sensing and Data Fusion," in *IEEE Sensors Letters*, vol. 8, no. 4, pp. 1-4, April 2024 <https://ieeexplore.ieee.org/document/10475363>
- [2] S. Chtioui *et al.*, "XGBoost in Public Transportation for Multi-Attribute Data: Delay Prediction in Railway Systems in Real-Time," in *IEEE Access*, vol. 12, pp. 143327-143342, 2024 <https://ieeexplore.ieee.org/document/10683692>
- [3] F. Wu, H. Tan, L. Zhang, S. Wen and T. Hu, "Multivariate Machine Learning Model Based on YOLOv8 for Traffic Flow Prediction in Intelligent Transportation Systems," in *IEEE Access*, vol. 13, pp. 105091-105100, 2025 <https://ieeexplore.ieee.org/document/11037414>
- [4] Y. Sait Türkan and M. Ulu, "A Hybrid Approach to Traffic Incident Management: Machine Learning-Based Prediction and Patrol Optimization," in *IEEE Access*, vol. 13, pp.

- 43455-43472, 2025 <https://ieeexplore.ieee.org/document/10900384>
- [5] Y. H. Taher *et al.*, "Filter for Traffic Congestion Prediction: Leveraging Traffic Control Signal Actions for Dynamic State Estimation," in *IEEE Access*, vol. 13, pp. 8140-8157, 2025 <https://ieeexplore.ieee.org/document/10820531>
- [6] A. Srivastava, M. Singh and S. Nandi, "Support Vector Regression Based Traffic Prediction Machine Learning Model," *2024 8th International Conference on Computational System and Information Technology for Sustainable Solutions (CSITSS)*, Bengaluru, India, 2024, pp. 1-6 <https://ieeexplore.ieee.org/document/10816969>
- [7] M. Yadav and B. Kumar, "Comparative Analysis of Machine Learning Algorithms for Traffic Flow Prediction in Intelligent Transportation Systems," *2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS)*, Tashkent, Uzbekistan, 2024, pp. 1752-1755 <https://ieeexplore.ieee.org/document/10840512>
- [8] A. Hazarika, N. Choudhury, M. M. Nasralla, S. B. A. Khattak and I. U. Rehman, "Edge ML Technique for Smart Traffic Management in Intelligent Transportation Systems," in *IEEE Access*, vol. 12, pp. 25443-25458, 2024 <https://ieeexplore.ieee.org/document/10436105>
- [9] S. Sajja, H. Pandey, A. Sharma, Bhawana, P. Murty and B. Rajbongshi, "Real-time Traffic and Traffic Flow Prediction using Machine Learning Algorithm for Optimal Route Planning in Electric Vehicles," *2024 Second International Conference Computational and Characterization Techniques in Engineering & Sciences (IC3TES)*, Lucknow, India, 2024, pp. 1-7Y. <https://ieeexplore.ieee.org/document/10877654>
- [10] E. Anvesh, D. Brindha and C. H. Kumar, "Dynamic Traffic Optimization Through Cloud-Enabled Big Data Analytics and Machine Learning for Enhanced Urban Mobility," *2025 3rd International Conference on Communication, Security, and Artificial Intelligence (ICCSAI)*, Greater Noida, India, 2025, pp. 469-474 <https://ieeexplore.ieee.org/document/10810771>