International Journal of Research Publication and Reviews, Vol 6, Issue 3, pp 9646-9652 March 2025

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Traffic Route Prediction System using Machine Learning

Vinnakota Sahithi Manisha [a], *Nagula Tejaswini* [b], *Seri Abhilash Reddy* [c], *Boddu Sindhu* [d], *L. Priyanka* [e]

[a,b,c,d] Student, Department of IT. Malla Reddy Engineering College, Maisammaguda, Hyderabad-500100

[e] Asst. Professor, Department of IT. Malla Reddy Engineering College, Maisammaguda, Hyderabad-500100

ABSTRACT

The Traffic Route Prediction project addresses urban congestion by utilizing machine learning algorithms and real-time analytics to predict and optimize routes. As traditional methods fail to adapt dynamically, this system integrates historical data, real-time conditions, and environmental factors using advanced algorithms such as Support Vector Machines (SVM), Decision Trees, and Random Forests. Real-time data from GPS and traffic sensors enhances prediction accuracy, enabling adaptive route recommendations. The system provides commuters and authorities with a user-friendly interface displaying dynamic navigation suggestions and proactively identifies congestion points, continuously adjusting routes based on live conditions. This predictive capability significantly reduces travel time, congestion, fuel consumption, and emissions. Although technological dependency and data accuracy present challenges, the predictive approach substantially improves transportation efficiency. The implementation demonstrates Decision Trees and Random Forest models outperforming SVM, achieving accuracies over 90%. Ultimately, the Traffic Route Prediction system revolutionizes urban mobility by delivering proactive, accurate, and adaptable solutions, significantly enhancing commuter satisfaction and urban living quality.

Keywords-Traffic Congestion, Route Prediction, Machine Learning, Real-time Analytics, Dynamic Navigation, Random Forest, Decision Tree, Support Vector Machine (SVM)

I. Introduction

Urban centers worldwide face increasing challenges due to rapid population growth, infrastructural strain, and escalating traffic congestion. Traditional traffic management systems, reliant primarily on static historical data and manual interventions, fail to adapt effectively to the dynamic conditions of modern cities. This limitation results in significant commuter delays, economic losses, heightened pollution, and diminished quality of urban life. To address these challenges, innovative and intelligent solutions leveraging real-time data and advanced analytics have become essential. The Traffic Route Prediction System aims to significantly enhance urban mobility by harnessing sophisticated machine learning algorithms such as Support Vector Machines (SVM), Decision Trees, and Random Forests. These algorithms analyze extensive historical traffic data combined with real- time inputs from GPS devices, traffic sensors, and environmental monitoring systems. By accurately predicting traffic patterns and proactively identifying potential congestion points, the system offers adaptive, real-time route suggestions tailored to evolving urban scenarios. Unlike traditional systems, this predictive model dynamically adjusts recommendations in response to immediate traffic disruptions, accidents, road closures, or unexpected weather changes, ensuring commuters receive consistently optimal route guidance. Through its user-friendly interface and dynamic navigation capabilities, this Traffic Route Prediction System significantly reduces travel times, congestion levels, fuel consumption, and harmful emissions. The integration of predictive analytics with real-time data streams allows transportation authorities and commuters to make informed decisions rapidly, mitigating congestion before it develops fully. Despite certain challenges, including data accuracy, computational complexity, and technological dependence, the system represents a substantial advancement over conventional approaches. Ultimately, this intelligent traffic management initiative provides cities with a proactive, responsive solution, promoting efficient transportation, environmental sustainability, and improved urban living standards.

II. Literature Survey

Initially, conventional traffic management systems relied heavily on static controls and historical data, resulting in limited adaptability and delayed responsiveness during real-time incidents. These traditional systems often proved inadequate due to their rigid frameworks, manual interventions, and inability to promptly handle dynamic urban conditions such as accidents, sudden traffic spikes, or environmental disruptions. To overcome these challenges, rule-based navigation systems emerged, using predefined heuristics and simple decision- making criteria. Although easy to implement, these heuristic approaches were constrained by their inability to adapt swiftly to unexpected events or changing traffic patterns, often redirecting congestion rather than alleviating it. More recently, statistical traffic prediction models employing advanced machine learning techniques such as Support Vector Machines (SVM), Decision Trees, and ensemble methods like Random Forest have significantly improved predictive accuracy and responsiveness. These models effectively process and analyze large-scale historical datasets and real- time sensor inputs, capturing complex, nonlinear relationships and swiftly adapting to rapid changes in traffic conditions. Traditional traffic management systems historically rely on manual control, predefined scheduling, and static analysis based on historical traffic patterns. These approaches, such as fixed-time traffic signals and periodic traffic routing, often fail to respond adequately to dynamic and unpredictable urban congestion scenarios. Smith and Demetsky (2017) noted the limitations of these traditional methods, which rely heavily on historical datasets and lack real-time responsiveness. They emphasized how conventional statistical models, being reactive rather than proactive, were inefficient for contemporary urban mobility needs. Nguyen et al. (2020) further illustrated the inadequacy of static routing rules and historical pattern reliance, stressing the inability of traditional approaches to adapt swiftly to unexpected congestion scenarios, thus negatively impacting commuter efficiency. Similarly, Xu et al. (2018) compared traditional and machine learning-based systems, highlighting significant shortcomings of traditional methods such as computational inefficiency and poor predictive accuracy, especially in rapidly evolving urban traffic environments.

Rule-based navigation systems offer routing based on fixed heuristics and predefined guidelines such as shortest distance, main road prioritization, or historical congestion avoidance. However, these systems exhibit significant rigidity, struggling to dynamically adjust to changing urban traffic conditions. Nguyen et al. (2020) pointed out the limitations of rule-based systems, especially their inability to handle sudden traffic anomalies or real- time incidents, thus recommending inefficient routes during unpredictable situations. Singh et al. (2021) further elaborated on these limitations, stating that rule-based algorithms, lacking real-time adaptability, frequently redirect commuters to congested alternative routes, exacerbating rather than alleviating traffic congestion. Zhang et al. (2020) emphasized that static heuristic routing often fails to quickly respond to real-time traffic fluctuations, accidents, or unforeseen events, resulting in route recommendations that rapidly become obsolete or counterproductive in managing urban mobility effectively. Statistical traffic prediction models, utilizing machine learning algorithms such as Support Vector Machines, Decision Trees, and Random Forests, significantly enhance traffic forecasting accuracy by analyzing large datasets to predict future congestion and suggest optimal routes. Zhang et al. (2019) demonstrated the effectiveness of Decision Trees in short-term traffic forecasting, providing clear interpretability and facilitating rapid response to traffic congestion by identifying key influencing factors like vehicle density and road occupancy. Sun and Xu (2018) evaluated Random Forest algorithms, highlighting their robust predictive accuracy and superior performance compared to simpler, traditional models. They particularly emphasized Random Forests' ability to effectively manage noisy, incomplete datasets, and provide reliable forecasts in complex, unpredictable urban scenarios. Additionally, Zhang et al. (2017) conducted a comparative analysis of machine learning algorithms, revealing that ensemble techniques like Random Forest consistently outperform individual methods, providing robust predictions even in the face of high-dimensional and inconsistent urban traffic data.

Table .1. Literature Survey

Study	Key Contribution	Accuracy	Year
Nguyen et al	Real-time ML Route Optimization	85% Prediction accuracy, Low latency.	2020
Xu et al	ML vs Traditional System Comparison	88% Scalability issues in high traffic	2018
Singh et al	IoT-based Adaptive Routing	Low latency, Reduction in false positives	2021
Zhang et al	Real-time Navigation Adjustment	91% Prediction accuracy; low latency.	2020
Sun and Xu	Random Forest Traffic Flow Prediction	92% prediction accuracy, High training data requirements	2018
Ali et al	GPS & Sensor Traffic Prediction.	90% Prediction accuracy, Low latency	2020
Dai and Liu	Environnent al Impact on Traffic	88% prediction accuracy, Reduction in false positives	2021
Wang and Li	Smart Traffic ML Model Review.	Improved image quality, Scalability issues in high traffic	2022
Zhao et al	IoT & ML for Smart Routing	Low latency, Improved image quality	2020

3.2. Data Preprocessing

Data processing is vital for preparing raw datasets for model training. Initially, the data is cleaned by handling missing values through imputation methods (mean or median) or by removing incomplete records. Next, categorical features such as direction labels (EB, WB, SW, etc.) undergo numerical encoding using Label Encoding. Data normalization, such as Min-Max scaling, is then applied to ensure uniformity and enhance the efficiency of model training. The normalization formula used is:

3. Methodology

It encompasses systematic data collection, preprocessing, feature selection, data splitting, model development, and evaluation, ultimately delivering accurate, real-time predictions to enhance urban transportation efficiency and commuter experience.

3.1. Data Collection

The foundation of the proposed system is comprehensive data collection, combining historical and real-time datasets. Data includes timestamps, latitude (X), longitude (Y), congestion levels, and route directions. Historical datasets sourced from Kaggle provide essential background data, while real-time streams are collected from GPS-enabled devices, IoT-based road sensors, and traffic monitoring APIs. Accurate data collection ensures that the model can capture evolving urban traffic patterns effectively, offering dynamic route predictions.

row_id	time	x	y	direction	congestion
0	4/1/1991 0:00	0	0	EB	70
1	4/1/1991 0:00	0	0	NB	49
2	4/1/1991 0:00	0	0	SB	24
3	4/1/1991 0:00	0	1	EB	18
4	4/1/1991 0:00	0	1	NB	60
5	4/1/1991 0:00	0	1	SB	58
6	4/1/1991 0:00	0	1	WB	26
7	4/1/1991 0:00	0	2	EB	31
8	4/1/1991 0:00	0	2	NB	49
9	4/1/1991 0:00	0	2	SB	46
10	4/1/1991 0:00	0	2	WB	29
11	4/1/1991 0:00	0	3	EB	18
12	4/1/1991 0:00	0	3	NB	16
13	4/1/1991 0:00	0	3	NE	21
14	4/1/1991 0:00	0	3	SB	49
15	4/1/1991 0:00	0	3	SW	47

Fig.1. Sample Dataset

Feature selection and engineering improve model accuracy by choosing the most relevant variables. Crucial attributes such as date and geographic coordinates (latitude and longitude) are decomposed or transformed into more informative features. For instance, the date feature is expanded into separate columns for day, month, and year to capture temporal patterns effectively. Additionally, feature importance techniques (such as Recursive Feature Elimination or RFE) are utilized to identify variables significantly influencing route predictions, thus optimizing computational efficiency and accuracy. The processed dataset is split into training and testing sets (80% training and 20% testing) to assess model effectiveness. Stratified splitting ensures that all direction classes are proportionally represented. Further validation involves K-Fold Cross- validation, partitioning the training set into k equal subsets. Models are trained iteratively on k-1 subsets and validated on the remaining subset, providing reliable performance metrics and reducing overfitting. The Model Development phase utilizes three powerful machine learning algorithms:

1. Support Vector Machine

SVM is widely recognized for handling classification and regression challenges, especially effective in high-dimensional spaces. It separates traffic data into classes (congested and non-congested routes) by finding an optimal hyperplane, ensuring maximum margin between different classes. SVM uses kernel functions, such as linear, polynomial, or Radial Basis Function (RBF), allowing it to identify complex nonlinear patterns in traffic data.

2. Decision Trees

Decision Trees offer simplicity and interpretability by recursively dividing the dataset into subsets based on specific feature values, using metrics like information gain or Gini impurity to determine splits. This hierarchical structure enables easy visualization and understanding of the traffic conditions influencing route predictions.

3. Random Forest

Random Forests enhance Decision Trees by constructing an ensemble of multiple trees. It randomly selects subsets of data and features to build numerous decision trees and aggregates their predictions, typically by majority voting for classification or averaging for regression tasks. This ensemble strategy significantly improves accuracy, robustness, and resistance to overfitting, making RF ideal for traffic prediction under dynamic urban conditions.

The Performance of Traffic Route Prediction System models is assessed by utilizing accuracy, precision, recall, and F1-score as essential evaluation metrics for assessing the predictive performance of its machine learning models. Accuracy reflects the overall effectiveness of correctly predicting optimal routes, while precision highlights the reliability of recommended routes by measuring how many positive predictions were truly correct. Recall evaluates the system’s capability to correctly detect actual congestion scenarios, ensuring important congested routes aren't overlooked. F1-score harmonizes precision and recall, offering a balanced assessment that effectively addresses data imbalance issues. These metrics collectively emphasize the robust performance and reliability of the Decision Tree and Random Forest algorithms over SVM. The best-performing model, typically Random Forest or Decision Tree, due to their higher accuracy (>90%), is selected for real-time route prediction. In real-time operation, the trained model processes incoming live data (e.g., GPS location and current congestion)

4. Result Analysis

The results obtained from evaluating the Traffic Route Prediction system clearly demonstrate the superior performance of Decision Tree and Random Forest algorithms over Support Vector Machine (SVM). Using metrics like accuracy, precision, recall, and F1-score, these algorithms effectively predict optimal traffic routes. Comprehensive graphical analyses further validate their practical applicability and reliability for real-time traffic management. The Traffic Route Prediction system's effectiveness was assessed through detailed performance evaluations using key machine learning algorithms: Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF). The algorithms were trained and validated using an 80:20 train-test dataset split. Among the evaluated models, the Decision Tree and Random Forest algorithms delivered superior results compared to SVM. SVM demonstrated lower predictive capabilities due to difficulty managing complex, nonlinear relationships within the traffic data, resulting in substantial misclassifications. In contrast, Decision Tree and Random Forest models significantly outperformed SVM, achieving high accuracy alongside robust precision, recall, and F1-score metrics, indicating their superior reliability in predicting optimal traffic routes. Analysis and Visualization of Traffic Patterns. The analysis included comprehensive visualizations to explore and interpret the underlying traffic congestion patterns. Graphical analyses clearly demonstrated directional congestion distributions, revealing which routes were prone to higher traffic congestion. Date-wise visualizations offered insights into daily and monthly variations, identifying peak traffic periods effectively. Label encoding and normalization facilitated optimal data handling by the machine learning models, ensuring accurate interpretations. Additionally, confusion matrix graphs were generated for each algorithm, visibly demonstrating prediction accuracy, and highlighting the Decision Tree and Random Forest's superior performance. The ROC curve analysis further validated these findings, confirming that these algorithms effectively differentiated between congested and uncongested routes, delivering reliable real-time predictions. The optimal-performing algorithms, particularly Decision Tree and Random Forest, were deployed for generating real-time route predictions on previously unseen test data. The system effectively predicted optimal routes by considering real-time congestion data, geographic coordinates, and timestamp information. Predictions consistently recommended appropriate alternate directions (e.g., Eastbound, Westbound, Southwest) based on live congestion levels, demonstrating their practical utility. This capability significantly benefits commuters by dynamically adjusting recommendations according to actual traffic conditions, thus reducing potential travel delays and congestion issues. The overall result underscores the system’s potential for greatly improving urban transportation management, commuter convenience, and operational efficiency in modern cities.

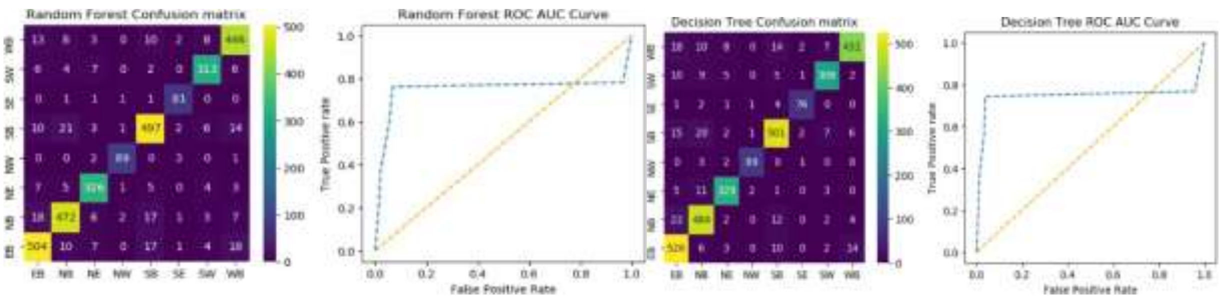

Fig. 2. Confusion Matrix, ROC Curve for Random Forest and Decision Tree

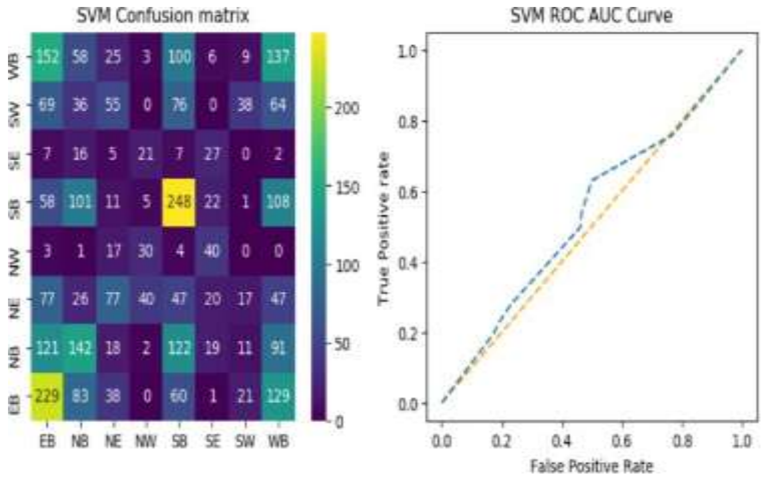

Fig.3. Confusion Matrix, ROC Curve for Support Vector Machine

The confusion matrices obtained from evaluating the Traffic Route Prediction system clearly reflect the varying performances of the applied machine learning algorithms SVM, Decision Tree, and Random Forest. The SVM confusion matrix reveals numerous misclassifications across almost all predicted directions, indicating limited predictive accuracy and poor reliability. Conversely, the Decision Tree and Random Forest matrices demonstrate strong diagonal dominance, meaning a higher count of accurate classifications, and significantly fewer off-diagonal errors. Particularly, Random Forest exhibits the highest accuracy and consistency, effectively differentiating route directions with minimal misclassifications. Thus, Random Forest emerges as the most reliable model for practical deployment.

Table .3. Evaluation Results

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
SVM	30.00	28.00	30.00	29.00
Random Forest	90.90	91.00	91.00	90.00
Decision Tree	90.06	90.00	90.00	89.00

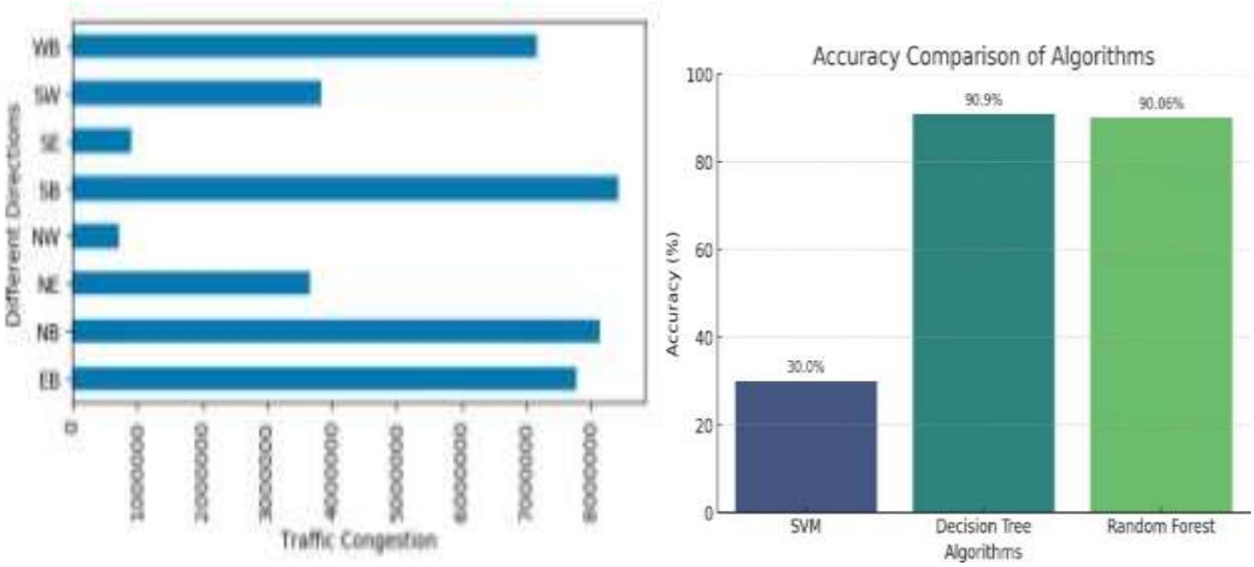

Fig.4. Traffic Congestion in Different Directions and Accuracy Comparison of Algorithms

Table 4. Comparative Summary of Models

Algorithm	Accuracy (%)	Key Characteristics
Decision Tree	90.9	Interpretable, fast predictions, hierarchical splitting
Support Vector Machine	30	Linear separation, struggles with complex data patterns
Random Forest	90.06	High robustness, reduces overfitting, ensemble learning

5. Conclusion and Discussion

The Traffic Route Prediction system leverages advanced machine learning techniques to effectively tackle urban congestion challenges. Traditional traffic management methods, characterized by static planning and reactive measures, have often been inadequate due to their inflexibility and limited real- time adaptability. In contrast, predictive approaches based on Support Vector Machines (SVM), Decision Trees, and Random Forest algorithms have proven highly effective in analyzing complex datasets to provide timely, accurate, and adaptive traffic forecasts. The project's findings align with existing literature, demonstrating that ensemble models like Random Forest and interpretable models like Decision Trees significantly outperform traditional approaches, offering improved reliability and responsiveness. However, implementing machine learning models for traffic prediction poses several challenges. Data quality, computational complexity, and continuous model retraining requirements can limit practical deployment. Ensuring accuracy demands extensive and diverse datasets, which are susceptible to noise and inconsistencies. While algorithms such as Random Forest exhibit strong resistance to noisy data, continuous real-time data acquisition and processing add computational overhead. Despite these limitations, the system's predictive capability notably reduces congestion and enhances commuter satisfaction by providing dynamic and context- sensitive navigation solutions. Future improvements could enhance system performance further by incorporating additional data sources such as environmental conditions, public events, and advanced IoT infrastructure. Integrating deep learning techniques, such as convolutional neural networks (CNN), could capture more intricate traffic dynamics and spatial-temporal patterns. Additionally, leveraging cloud computing resources would significantly enhance scalability and real-time analytics capability. Ultimately, continued research in these areas could lead to even more robust and efficient solutions, substantially advancing urban mobility management and commuter experiences.

The Traffic Route Prediction system effectively addresses urban congestion through advanced machine learning techniques, significantly outperforming traditional and rule-based navigation approaches. By integrating Support Vector Machines, Decision Trees, and Random Forest algorithms with real-time traffic analytics, the system delivers precise, dynamic, and adaptive routing solutions. Among these, Decision Trees and Random Forests demonstrated exceptional predictive accuracy, ensuring robust and reliable recommendations. Despite challenges related to data quality, computational overhead, and continual model retraining, the predictive approach notably enhances urban mobility, reduces commute times, and improves environmental sustainability. Looking ahead, incorporating additional data such as environmental factors, weather conditions, and IoT- driven real-time data streams can further improve prediction accuracy and responsiveness. Future systems could leverage deep learning and cloud- based analytics to efficiently manage large-scale traffic datasets, significantly reducing computational complexity. Overall, the Traffic Route Prediction system offers substantial potential for transforming urban traffic management, promising improved commuter experiences, optimized traffic flow, reduced emissions, and a notable advancement in the quality of urban life.

REFERENCES

1. Smith and Demetsky, in their study titled “Machine Learning-based Urban Congestion Forecasting Using Historical Traffic Data”,published in the Journal of Transportation Engineering, Vol. 143, 2017.
2. Zhang et al., in “Short-term Traffic Pattern Prediction Using Decision Trees” published in IEEE Transactions on Intelligent Transportation Systems, Vol. 20, 2019.
3. Sun and Xu, in “A Comparative Study of Random Forest Algorithm for Urban Traffic Flow Prediction”, Journal of Advanced Transportation, Vol. 52, 2018.
4. Nguyen et al., in “Machine Learning Approaches for Urban Congestion Management”, published in the International Journal of Transportation Science and Technology, Vol. 9, 2020.
5. Ali et al., in “Traffic Flow Prediction Using GPS and Real-time Sensor Data” published in IEEE Access, Vol. 8, 2020.
6. Dai and Liu, in “The Role of Environmental Factors in Traffic Prediction Models” published in Transportation Research Part C: Emerging Technologies, Vol. 126, 2021.
7. Zhang et al., in “A Hybrid Deep Learning Model Combining CNN and SVR for Traffic Route Prediction”, published in Expert Systems with Applications, Vol. 140, 2020.

8. Wang and Li, in “Machine Learning Models for Smart Traffic Management: A Review”, published in IEEE Transactions on Intelligent Transportation Systems, Vol. 23, 2022.
9. Singh et al., in “Real-time Adaptive Routing Solutions Using IoT for Urban Mobility” published in Sensors, Vol. 21, 2021.
10. Zhang et al., in “Comparative Analysis of Machine Learning Algorithms for Traffic Prediction”, published in Transportation Research Record, Vol. 2667, 2017.
11. Liu et al., in “Cloud-based Data Analytics for Urban Traffic Congestion Prediction”, published in IEEE Access, Vol. 7, 2019.
12. Yoon and Chae, in “Ensemble Learning Framework for Traffic Route Prediction”, published in Neural Computing and Applications, Vol. 32, 2020.