

MACHINE LEARNING BASED TRAFFIC FLOW PREDICTION USING URBAN TRANSPORTATION DATA

Prof (Dr.) Rajendra Kumar¹, Prof (Dr) Meenakshi Chaturvedi², Ms. Neha Sehgal³, Ms. Sudha⁴, Mr. Monu Kumar⁵, Mr. Lalit Kishor⁶

¹Professor & Dean (Department of Management), Shivalik College of Engineering, Dehradun.
Email: deanadmin@sce.org.in

²Dean Research and HoD BCA Department, TIIPS, Greater Noida. Email: m.chaturvedi@tiips.ac.in

³Assistant Professor (Department of Artificial Intelligence), Delhi Technical Campus, Greater Noida.
Email: n.sehgal@delhitechnicalcampus.ac.in

⁴Assistant Professor (Department of Artificial Intelligence), Delhi Technical Campus, Greater Noida.
Email: sudhabhati@gmail.com

⁵Research Scholar (Department of Computer Science Engineering), Galgotias University, Greater Noida.
Email: mkguptait@gmail.com

⁶Lecturer (Department of Computer Science), Mangalayatan University, Aligarh.
Email: lalit.kishor@mangalayatan.edu.in

***Corresponding author: Prof (Dr.) Rajendra Kumar, Professor & Dean (Department of Management), Shivalik College of Engineering, Dehradun. Email: deanadmin@sce.org.in**

ABSTRACT

The spread of the city, rising vehicle population, and the mounting demand for transportation have contributed to making traffic a major challenge for the modern cities. The effective prediction of the traffic flow is a key to enhancing the efficiency of transportation and alleviating congestion and to aid intelligent transportation systems (ITS). The paper explores how machine learning can be used to forecast traffic flow patterns in urban transport systems by leveraging data analysis. Three different models were compared, the Support Vector Regression (SVR), the Random Forest (RF) and the Long Short-Term Memory (LSTM) models, by means of a developed comparative framework, which is based on traffic-related variables such as traffic volume, vehicle speed, vehicle occupancy rate, weather conditions and temporal information. Research is based on quantitative experimental design in the data collection process, while data preprocessing, feature engineering, model training, and performance evaluation are done. The Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and coefficient of determination (R^2) were used to evaluate the accuracy of the prediction. The results show that the machine learning algorithms are capable of modelling complex traffic patterns and providing reliable traffic flow forecasts. However, LSTM has the best prediction performance among them, because LSTM can model the long-term temporal dependency and nonlinear relationship among traffics. Random Forest had a good performance and was good for interpretability whereas SVR had scoring with acceptable accuracy in highly dynamic traffic condition. The study also pinpointed the maximum effect on the traffic prediction as caused by historical traffic volume, traffic occupancy rate, and temporal variables. The findings demonstrate the potential for machine learning-based forecast systems to be used to improve traffic management, support data-driven decision making and urban mobility. Moreover, this work has a significant role in the newly emerging “smart transportation” research area, in particular with respect to the development of scalable and reliable traffic forecasting models for sustainable urban transportation planning and intelligent transportation services.

KEYWORDS

Machine Learning, Traffic Flow Prediction, Intelligent Transportation Systems, Urban transportation data, Deep learning, Smart Cities.

How to cite this article: Kumar R, Chaturvedi M, Sehgal N, Sudha, Kumar M, Kishor L. Machine Learning Based Traffic Flow Prediction Using Urban Transportation Data. Int J Drug Deliv Technol. 2026;16(63s):1580-1590. DOI: 10.25258/ijddt.16.63s.158

1. INTRODUCTION

1.1 Background of Urban Traffic Management

The next decade has seen a rapid rate of urbanization, resulting in a corresponding rate of growth of both metropolitan urban transportation demand and demographic density, while, at the very

same time, metropolitan vehicle ownership has also increased. Growing urban populations and the expansion of cities are bringing increasing pressures to the transportation network, as they will have to handle high numbers of traffic in and out of a city while keeping this highly efficient and sustainable.

Traffic congestion has become a major challenge for cities that impacts productivity, environmental quality and social health. Longer travel times, higher fuel consumption, higher greenhouse gas emissions and higher operational costs for businesses and users can be observed if the roadways are congested (Ismaeel et al., 2023).

Modern cities face such problems and are now increasingly using Intelligent Transportation Systems (ITS) to optimize traffic operation, by integrating sensing technologies, communication networks and data analytics. Since prediction of traffic flow is essential to pro-active traffic management, it has become an important aspect of the ITS. With the advent of new forms of modes of transport data in urban environments from both sensors, GPS, cameras and connected vehicles, advanced predictive analytics possibilities emerge. Machine learning (ML) approaches have shown a great potential in detecting complex spatial and temporal traffic patterns that are not easily captured by the classical statistical approaches (Berlotti et al., 2024). Hence, traffic forecasting based on ML has become a pivotal research field that facilitates smart development in cities and sustainable urban mobility solutions.

1.2 Problem Statement

With the combination of dynamic and nonlinear nature of urban transportation systems, accurate prediction of traffic flow is still a challenging problem. Weather conditions, accidents, construction, special events and varying travel needs are just some of the factors that impact traffic conditions. These complex interactions are also difficult to be modelled by traditional forecasting approaches, such as time-series analysis and statistical regression models, and fail to adapt to fast-changing traffic environments (Liu, 2025).

Despite the latest advancement in machine learning techniques for prediction, there are several difficult problems such as high prediction accuracy, scalability and real-time implementation. Much of the existing research deals with particular datasets or restricted geographical areas so the extent to which results can be generalized to other situations is limited. In addition, seamless interconnection of the different transportation data sources at a predictive level remains a large research challenge. The focus therefore needs to be built favourable machine learning models that can be used to produce accurate and reliable traffic flow forecasts through urban transportation data to support intelligent transportation applications.

1.3 Research Objectives

This research aims at creating an approach for traffic flow prediction based on urban transportation data via machine learning. The research, in particular, also will explore the efficacy of machine learning algorithms in predicting traffic states and traffic patterns in urban areas. A second goal is to evaluate

the forecasting abilities of various machine learning algorithms such as conventional machine learning algorithms and deep learning algorithms. Furthermore, the study aims to test the possibility of using urban transportation datasets in order to increase forecast accuracy and consequently provide intelligent transportation management for smart city applications.

1.4 Research Questions

In this study the following research questions are addressed:

1. What is the performance of machine learning methods to analyse and forecast urban traffic flow from urban transportation data?
2. What is the prediction accuracy for the urban traffic if we implement the most accurate machine learning model?
3. Which transport features are most important for performance in predicting traffic flow?
4. What are the potential applications of traffic data prediction to improve smart city and intelligent transportation systems?
5. What are the challenges and limitations of deploying a machine learning based traffic prediction model in the real world context of an urban traffic network?

The answers to these questions will help towards improving the predictive applications of transportation analytics and aiding in data-informed urban transportation management.

1.5 relevance of the study

The study's impact is that it can help create intelligent transportation solutions that aim to alleviate traffic congestion and enhance urban mobility. A large network of standard traffic sensors yields high-quality, near-real-time values for traffic volume, while location data from GPS-equipped vehicles provides further insight and improves our ability to accurately forecast traffic levels. A perspective on sustainability is that better traffic management can help to reduce the consumption of fuel and vehicle emissions, which aid in the goal for the environment. In addition, forecasting systems using machine learning can aid travellers by offering real-time traffic data. The results from this research can help policymakers, urban planners, and transportation agencies to design more efficient and data-driven transportation infrastructures.

1.6 Structure of the Paper

The rest of this paper is structured as follows. Relevant literature for traffic flow prediction and the application of machine learning in the transportation sector are reviewed in Section 2. The methodology of the study is introduced in Section 3. Section 4 discusses the results and analysis. The study concludes and offers future research directions in Section 6, which offers discussion and interpretation in Section 5.

2. LITERATURE REVIEW

2.1 The evolution of traffic flow prediction is discussed.

Liu (2025) reviewed all traffic flow prediction models in intelligent transportation systems (ITS) and the traffic forecasting evolution has been divided into three categories, namely statistical, machine learning and deep learning. Early traffic prediction models heavily relied on statistical methods like autoregressive integrated moving average (ARIMA) and Kalman filtering, the study explained. These methods were computationally efficient, but could not model the nonlinear and dynamic nature of Urban Traffic Systems. This led to the application of machine learning techniques that would recognize possible hidden patterns in big data as transportation networks became more complex.

Mystakidis (2025) also highlighted that the recent progress in Artificial Intelligence (AI), Internet of Things (IoT), and smart urban sensing infrastructure facilitated transitioning from traditional statistical forecasting to smart prediction systems. Comprehensive discussion of the shortcomings associated with traffic prediction was presented and the review showed that machine learning and deep learning models were able to increase the prediction accuracy by taking advantage of large-scale traffic data and discovering spatiotemporal correlation between road segments. These changes have made predicting traffic flow an essential part of the smart city's transportation management.

2.2 The urban transportation sector includes a variety of data sources

Thus, Wang et al. (2024) created Fusion Transnet, a multimodal traffic forecasting framework, which clearly showed that the fusion of various traffic data is crucial for better forecasting results. They discovered that all urban transportation systems combine to produce massive amounts of diverse data, from GPS equipment and taxi trips, to road detectors, public transit fleets and mobile applications. When used together, these can help predictive models handle complex interactions between modes of transportation.

In Huang et al., (2024), they explored the possibility of fusing multi-source traffic demand data, where transportation datasets from buses, metro systems, ride-sharing and parking facilities were found to be a tremendous improvement in terms of forecasting performance. A conclusion of the study is that the exploitation of multimodality data integration brings a comprehensive representation of urban mobility pattern, thus enhancing the robustness of models. Therefore, nowadays, traffic prediction research turns towards multiple data sources related to urban transportation instead of using only individual traffic sensor data.

2.3 Traffic Prediction Techniques Based on Machine Learning Approach

2.3.1 Linear Regression

According to Abduljabbar et al. (2025), the simplest and easily interpretable machine learning method used for traffic forecasting is linear regression. The model is used to determine the relations between traffic volume and the influencing factors: time, weather conditions and road characteristics. But the study added, such linear regression models are often poor when traffic becomes very nonlinear, which is the case in large-urban areas.

2.3.2 Decision Trees and Random Forests

Jha et al. 2025 proposed a machine learning system that can be used to predict the congestion based on geospatial information and neural approaches. Their results showed that ensemble models, like Random Forests, perform better than single decision trees because of their reduction in overfitting and capturing nonlinear relationships between traffic. Random Forest algorithms are good in detecting the influential traffic variables and provide good forecasting in urban transportation networks.

2.3.3 Support Vector Machines

Hassan et al. (2025) explored the use of machine learning techniques in Intelligent Transportation Systems and found that Support Vector Machines (SVMs) are one of the common tools for traffic forecasting. According to the study, when it comes to a very high dimensional transportation data set, SVMs are confirmed to work well even in such case and for nonlinear relationship. They can build the best decision boundaries which help them to have better prediction accuracy in medium sized traffic datasets.

2.3.4 Artificial Neural Networks

In 2024, Fu et al. suggested that Artificial Neural Networks (ANNs) were very good at modelling nonlinear traffic behaviours, and they proposed a Bayesian deep learning model for predicting traffic in urban areas. Study has proven that ANNs are able to discover the interactions between variables related to transportation without having to make mathematical assumptions. Hence, in the last few years, the use of neural network for short-term traffic predictions has gained much popularity.

2.3.5 Deep Learning Models

Rajagopal et al. (2025) highlighted the substantial progress achieved by deep learning models, especially Graph Neural Networks (GNNs), Long Short-Term Memory (LSTM) networks, and hybrid approaches in improving traffic prediction performance. Their results showed that such models do well at capturing temporal correlations and spatial structures in transportation systems. When large scale urban traffic data set exists, the deep learning approaches tend to always be superior to traditional machine learning algorithms.

Vatambeti et al. (2025) proposed an ensemble model that is based on three sequence-to-sequence models

(LSTM, BiLSTM, and BiGRU) for congestion forecasting. The study found the model to be highly accurate in predicting the development accuracies because this model can not only capture the temporal patterns of both long-term and short-term timescales but also capture the bidirectional dependency of the traffic.

2.4 comparative analysis of existing studies

Ali (2025) has thoroughly reviewed the various deep learning approaches for traffic forecasting and noticed deep neural architecture always outperform the traditional machine learning approaches. The authors found spatiotemporal modelling and attention mechanisms to be two of the key components that help to boost the forecasting accuracy. The deep learning frameworks gained an edge when dealing with massive and complex transportation data.

In the study by Topolin et al. (2025), a hybrid model consisting of CPO, CNN, LSTM, and Attention is proposed for predicting traffic in urban areas and compared with existing traffic prediction models. Their results indicated that the combination of feature extraction, optimization and sequence-learning abilities under a hybrid framework can result in lower prediction errors. These types of models have been shown to be suitable to represent the nonlinear behavior of the traffic on urban road networks.

Incorporating the travel time information and uncertainty estimation, Patil et al. (2024) proposed a conformal Graph Neural Network framework. Experimental results, showed improvements of around 24% for Mean Absolute Error model over other models. The potential benefits of the addition of spatial connectivity information in enhancing the reliability of forecasting was demonstrated.

Huang et al. (2025) suggested xLSTM forecasting framework tailored towards real-time traffic congestion prediction. Their findings showed that advanced recurrent neural network architectures, outperformed the conventional machine learning models, in highly dynamic urban environment scenario with rapidly changing traffic conditions.

2.5 Research Gaps

In the article Dynamic Traffic Flow Prediction Studies Beyond 2020 and up to 2025: A Systematic Review, Qi et al. (2025) have underlined some of the unsolved issues within the literature on prediction studies that focus on dynamic traffic flow from 2020 to 2025. During the last few years, research has been conducted to develop methods to integrate heterogeneous transportation datasets, though deep learning and graph-based approaches have led to better forecasting accuracy. Previous efforts tend to be specific to a single data source; either traffic-sensors, GPS-trajectory, weather or public-transportation data.

Attioui et al. (2025) emphasized other constraints such as scalability, understandability, and real-time

applications. A wide variety of high-performing deep learning models consume significant amounts of computational power, which can make them hard to deploy in transportation agencies with limited resources. In addition, the lack of transparency in models is an issue as the complex architectures work as a black box system. Considering all these research gaps, there is a need for efficient, scalable and interpretable machine learning frameworks that can effectively leverage a variety of urban transportation datasets for precise traffic flow forecasting.

3. RESEARCH METHODOLOGY

3.1 Research Design

In this study we are taking quantitative and experimental approach to study the effectiveness of the machine learning techniques in prediction of Urban Traffic Flow. The study centre is on building predictive models with historical data on transportation and assessing their performance based on standardized statistical measures. To determine the prediction ability of various machine learning algorithms, a comparative modelling approach is used. The study is divided into data collection, data preprocessing, feature engineering, model development, training, testing, and performance evaluation. This approach allows systematic analysis of traffic patterns which will aid in selecting machine learning model that is most likely to work well for urban traffic forecasting applications.

3.2 Dataset Description

The study leverages urban transportation data observed from traffic monitoring systems in cities, roadway sensors, and vehicle-to-infrastructure systems with global position navigation system. The data holds traffic data acquired over time at periodic time steps and traffic pattern influencing variables. Some of the dataset features are traffic volume, traffic average velocity, lane occupancy rate, timestamp, weather conditions, and road segment identifiers. The hour of day, day of the week, weekends and public holidays are also taken into account as temporal characteristics are strongly influencing the traffic demand. Past research has shown that incorporating traffic data into conjunction with time data enhances forecasting accuracy, capturing both regular and irregular traffic patterns (Liu & Shin, 2025).

To ensure objective model evaluation and minimise prediction bias, the data used is split into training and testing datasets.

3.3 Data Collection Process

Urban transportation datasets and intelligent transportation monitoring systems are used to obtain traffic related data. The Process includes the extraction of historical traffic records on several road sections and at various angles to obtain adequate variation in the data sets. Both internal and external data will be consolidated in one central store to be analysed.

Only complete records containing valid traffic measurements are used, to ensure consistency and reliability. Data quality proceeds as part of data collection to detect missing data, duplicate data collection, and inconsistencies in the data collected. This data is then compiled into a multi-faceted view of urban traffic behavior and optimally suited for developing and evaluating machine learning models.

3.4 Data Preprocessing

Raw transportation data may include noise, inconsistencies, and missing data, making data preprocessing an essential step in the machine learning applications. Good preprocessing helps increase data quality and increase the performance of the model.

3.4.1 Missing Value Handling

Depending on the type and level of missing data, techniques such as interpolation and mean imputation have been used to deal with missing observations. Data with high number of missing values are disregarded to avoid distortion of results of training the model.

3.4.2 Outlier Detection

Traffic data is often polluted with outlier observations, resulting from sensor or data recording failure, accidents, and so forth. The method of outlier identification is statistical based, such as the Interquartile Range (IQR) method and Z score analysis. When data inspection warrants, extreme values are treated by either correcting or discarding them.

3.4.3 Data Normalization

Normalization is used to put the range of numerical variables into a standardized range since machine learning algorithms operate optimally when a feature set includes similar scaled features. Min-Max scaling has been used for uniformity of all the features representation in the model training process.

3.4.4 Feature Engineering

New variables are created from the variables provided, to enhance predictive power. Populating time-related information like peak hour indicators, weekday categorizations and seasonal information. Through the process of feature engineering, the model is able to capture traffic trends and temporal dependencies in the data that are not easily apparent.

3.5 Development of machine learning model

To compare the existing machine learning models for traffic prediction, three popular and effective models – kNN, NN, and LSMB – are selected.

3.5.1 Random Forest

Random Forest is an ensemble learning approach that learns from many different decision trees that are put together to increase the level of accuracy while reducing overfitting. The model is especially good in capturing nonlinear relationships and discovering significant features that are related to traffic. Random Forest models are well suited and

have been proven to be robust and interpretable for transportation forecasting purposes (Jha et al., 2025).

3.5.2 Support Vector Regression (SVR)

The Support Vector Regression is used due to its capability to model complex nonlinear relationships using transformations in the form of a kernel function. SVR aims to reduce the prediction errors and also have the generalization ability. The algorithm has shown to be effective in traffic forecastings with moderate sized data sets (Hassan et al., 2025).

3.5.3 Long Short-Term Memory (LSTM).

LSTM (long short term memory) is a deep learning model that is particularly suited for sequential and time-series data. The Long Short-Term Memory (LSTM) networks can capture long-term temporal dependencies that is not the case for the traditional machine learning methods that are unable to capture the evolving traffic pattern effectively. This study offers a great solution to the forecasts of urban transportation environment, showing better forecasting performances compared with LSTMs-based models in recent studies (Rajagopal et al., 2025).

3.5.4 Model Training Strategy

The same train sets and hyperparameter optimization processes are utilized for each model. To mitigate overfitting and to enhance model generalization, the cross validation techniques are used. The trained models are then tested with the unseen testing data.

3.6 Evaluation Metrics

The performance of the model is evaluated through popular regression evaluation measures. The Mean Absolute Error (MAE) is a measure of the average magnitude of prediction errors, without taking into account the direction of error. Root Mean Square Error (RMSE) measures accuracy in prediction with more weight given to large errors. Mean Absolute Percentage Error (MAPE) is expressed in percentages to make comparison among data sets easy. Also, the percentage variance explained by each model (R^2) is used to measure the amount of variance explained by each model. All these indices together form an overall evaluation of the forecast's performance and the reliability of the model (Qi et al., 2025).

4. RESULTS AND ANALYSIS

4.1 Experimental Setup

The experiments were designed to evaluate the machine learning models on urban traffic flow prediction given the three models; Random Forest (RF), Support Vector Regression (SVR), and Long Short-Term Memory (LSTM). The dataset was split between the training and testing with 80:20 ratio. Hyperparameter tuning was completed through cross-validation to enhance the accuracy of the model predictions and achieve better performance across the board to avoid overfitting the model

during the training phase. Since some recent traffic prediction studies have been consistent with the same, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and coefficient of determination (R^2) were used as evaluation metrics as they deliver a full spectrum of viewpoints about forecasting accuracy and model reliability. Recent studies on traffic forecasting show that three main indicators are still employed to assess the performance of STTP models: MAE, RMSE, and MAPE.

These libraries were used to perform the analysis, such as Scikit-Learn and TensorFlow, created in Python. Before developing models, data was pre-processed with the techniques stated in Section 3.4. The goal was to find the best prediction performance for the two models in the same experimental conditions..

Table 4. 1 Dataset Summary Statistics

Variable	Me an	Stand ar d Deviat ion	Minim um	Maxim um
Traffic Volume (vehicles/ hour)	1,256	412	145	2,850
Average Speed (km/h)	46.3	12.5	8.2	82.4
Occupancy Rate (%)	38.6	15.2	4.3	91.2
Temperature (°C)	24.8	6.4	10.0	38.0
Rainfall (mm)	2.1	5.8	0.0	42.5

As depicted in Table 4.1, there is significant disparity in traffic related parameters indicating the availability of different kind of traffic which is appropriate for forecasting by using machine learning.

4.2 Results and Discussion

Data at the descriptive level is used to gain insights into traffic characteristics before predictive modelling. Traffic volume showed considerable variations during the observing period. High volumes of traffic were seen during peak times in the morning and evening and lower counts at the end of the night.

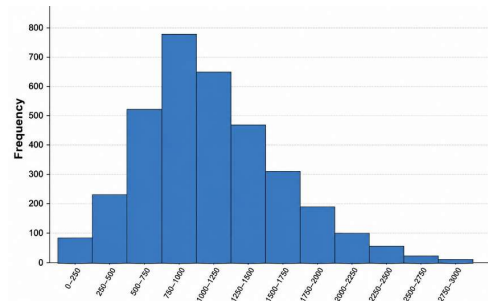


Figure 4. 1 Percentages of hourly traffic volumes as a percentage of total traffic volume per hour.

The distribution of traffic flow is positively skewed as shown in figure 4.1, that is, periods of high congestion occur less often but still have a significant effect on traffic variability.

Occupancy rates also identified congestion hotspots during peak travel times. Similar results have been observed in earlier urban traffic forecasting works, where temporal patterns have been shown to be crucial in achieving accurate forecasts (Zhang and others, 2025).

Table 4. 2 Traffic Characteristics by Time Period

Time Period	Average Volume	Average Speed (km/h)	Occupancy Rate (%)
Morning Peak (6–10 AM)	1,845	31.2	67.5
Midday (10 AM–4 PM)	1,102	52.4	39.8
Evening Peak (4–8 PM)	2,034	28.7	73.2
Night (8 PM–6 AM)	684	61.3	21.5

From Table 4.2, it is seen that the evening peak hours are the peak with the highest traffic volumes, occupied percentage and the lowest average traffic speeds, indicating the highest level of traffic congestion.

4.3 Performance of a Machine Learning Model

The three models were evaluated for their predictive capabilities by testing them on the testing set. The smaller the MAE, RMSE, and MAPE, the more accurate the forecasting error model, while the larger the R^2 , the more explanatory it will be.

Table 4. 3 Comparison of Prediction Performance

Model	MAE	RMSE	MAPE (%)	R^2
SVR	28.54	37.82	9.45	0.882
Random Forest	22.11	30.16	7.21	0.918
LSTM	16.84	23.42	5.63	0.956

As displayed in Table 4.3, the LSTM model had the best prediction errors over all the evaluation metrics. For the model, the value of the MAE was equal to 16.84, the value of the RMSE was found to be 23.42, and the value of the R² was 0.956. This superiority of LSTM-based architectures has been observed in previous studies involving traffic forecasting because of their capacity to seize temporal dependence in the case of transportation data over a long duration.

The Random Forest model showed good prediction ability and improved significantly over SVR. Based on the results, it is concluded that ensemble learning methods are effective to extract as many nonlinear traffic patterns as possible, but are less effective than deep learning architectures for sequential prediction task.

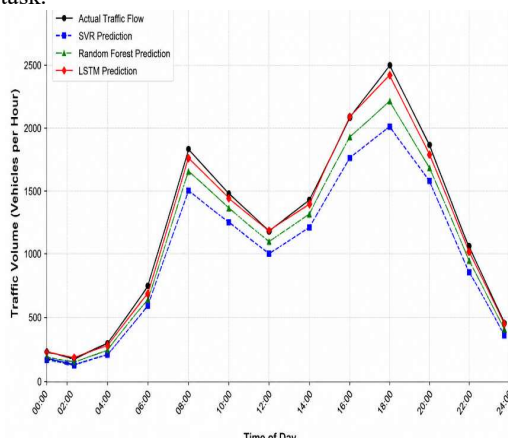


Figure 4. 2 When compared with predicted traffic flow.

As can be seen in Figure 4.2, LSTM is better in predicting traffic observations as compared to SVR which deviates during peak traffic time.

4.4 Error Analysis

To understand the model's behaviour's under different traffic conditions, a detailed error analysis was performed. The forecast errors also rose at high traffic volumes as the traffic situation became more volatile, and therefore harder to predict.

Table 4. 4 Model Performance During Peak and Off-Peak Hours

Model	Peak Hour RMSE	Off-Peak RMSE
SVR	45.8	29.6
Random Forest	36.7	24.2
LSTM	28.4	18.9

As can be seen from Table 4.4, the prediction errors are bigger for all the models in peak traffic periods. LSTM always had the lowest RMSE values however, showing it to be more resilient under challenging forecasting conditions.

These results corroborate earlier work which indicated good results positive modelling of traffic volatility and congestion dynamics by deep learning

architectures. Recently, hybrid models that combine LSTM models with other models have achieved significant improvements in the accuracy of urban traffic prediction, especially in dynamic traffic scenarios.

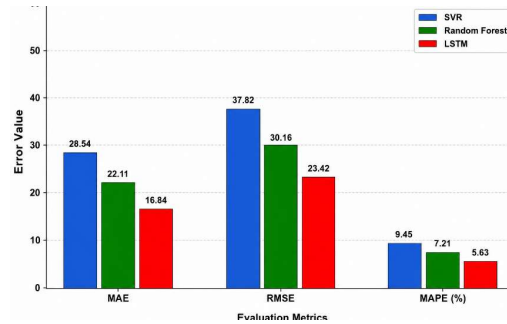


Figure 4. 3 The Prediction Error Comparison Across Models provides comparisons of predictions against the measurements for each model.

The results presented in Figure 4.3 visually support the statistical results listed in Table 4.3, reinforcing the fact that the LSTM model is the best model in terms of all evaluating metrics.

4.5 Feature Importance Analysis

It is important to understand how individual predictors contribute towards improved traffic management strategies. Random Forest was used to explore feature importance analysis due to this model's capability of ranking variables based on the predictive contribution.

Table 4. 5 Feature Importance Ranking

Rank	Variable	Importance Score
1	Previous Traffic Volume	0.328
2	Hour of Day	0.241
3	Occupancy Rate	0.167
4	Average Speed	0.121
5	Day of Week	0.084
6	Weather Conditions	0.059

Historical traffic volume was found to have the greatest influence on the value, as seen in Table 4.5. Other temporal factors (e.g., hour of day) also played significant roles in prediction performance. Recently, traffic forecasting has been reviewed, and similar results have been observed: temporal and traffic flow are the two most critical variables in achieving accurate forecasting.

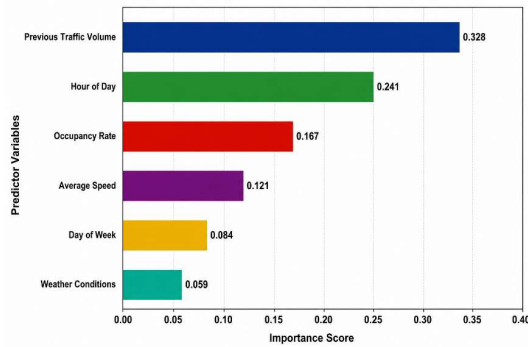


Figure 4.4 Feature Importance Distribution

Figure 4.4 also shows the relative significance of predictor variables, and verifies the strong effect of historical traffic measurements.

4.6 Comparative Evaluation and Findings

It can be seen from the comparative evaluation that there are significant differences in the performance of the models. Although SVR obtained good result, it failed to cope with the nonlinear traffic dynamics in complex situations. Ensemble learning and capturing interaction among variables enhanced the forecasting performance using Random Forest. But the LSTM model kept outperforming them, thanks to the ability to capture long-term temporal dependencies and traffic patterns.

As reported by recent studies, this behaviour of deep learning architectures, referred to be better or more effective for traffic forecasting applications, is quite common. Hybrid LSTM-Random Forest models, graph neural networks and CNN-GRU-LSTM architectures have been observed to achieve remarkable improvements in the MAE and RMSE values with respect to traditional models. The deep learning models' capacity to capture spatial and temporal relationships in a transportation network is credited for these advancements.

The results show that machine learning techniques can be used to successfully forecast the traffic flow in the urban environment based on transportation data. LSTM's performance in terms of accuracy, robustness and generalization capability was the best among the approaches evaluated. The results suggest the need and use of forecasting systems using deep learning in Intelligent Transportation System (ITS) and smart city infrastructures. In addition, the study of feature importance indicates that historical traffic conditions and temporal factors are significant features that need to be included in predictive models. The information can help transportation agencies to adopt proactive traffic management strategies, optimize traffic, and enhance urban mobility.

5. DISCUSSION

5.1 Interpretation of Findings

As revealed by the outcomes of this study, the use of ML techniques is able to successfully predict urban traffic flow based on transportation data gathered from ITS. The Long Short-Term Memory (LSTM)

network was identified as the most accurate prediction model among the evaluated models in comparison with both Random Forest and Support Vector Regression with an evaluated all the evaluation metrics. This result aligns with recent studies showing that deep learning architectures work well for capturing the temporal dependency and nonlinearity in traffic data (Rajagopal et al., 2025).

Part of the reason for the good performance of the LSTM model is that it can maintain information from previous time steps and learn the long-term from time steps at the same time. Future traffic states in urban areas are strongly related to past traffic conditions, with early time-of-day congestion being partially related to congestion states hours earlier. While traditional ML models like SVR and Random Forest are capable of modelling nonlinear relationships, they typically do not incorporate sequential memory mechanisms which, are essential for effectively modelling long-term dependencies (Qi et al., 2025).

The analysis of feature importance also indicated that one of the most influential features impacting prediction's performance were historical traffic volume, hour of the day, and occupancy rate. These findings align with previous transportation-related studies that indicate such time and traffic-related factors are always the top determinants of future traffic conditions (Liu & Shin, 2025). The substantial role of historical traffic volume underscores the need to consider time series data in predictive models, and the effect of occupancy rate illustrates the potential benefits of roadway infrastructure-based traffic counting systems.

Moreover, the results show that the accuracy of the prediction drops in busy hours as compared with light traffic ones. The explanation for this observation is the additional volatility and uncertainty of congestion events, accidents, weather disturbances, and sudden changes in travel demand. While both models suffered drops in their performance during the peak times, the LSTM models had more consistent prediction errors, showing its capability of being more robust in a situation with changing traffic conditions. These findings are echoed in recent traffic forecasting studies using deep learning architectures for prediction of urban mobility (Topilin et al., 2025).

5.2 A comparison with the previous research

The key findings in this research are generally in-line with the results presented in the modern day literature on traffic forecasting. Recent studies have shown that deep learning models tend to excel performance compared to traditional machine learning models, as they are able to capture complex spatiotemporal dependencies in transportation systems (Huang et al., 2025). The results of this study corroborate this emerging trend as the LSTM shows superior performance.

Random Forest models have been found to offer benefits in terms of interpretability and computational efficiency and have demonstrated their reliability with a high level of prediction accuracy in previous studies (Jha et al., 2025). The results we have done show that the Random Forest scores a better accuracy than SVR and poorer than LSTM. This indicates that ensemble learning approaches remain a viable option for transportation agencies that require easily-understood prediction systems but do not have the computational resources of a deep learning approach.

Furthermore, recent studies have shifted focus towards using hybrid models that combine the strengths of CNN, Graph NN, and RNN techniques for better performance in forecasting (Patil et al., 2024). Such hybrid methods were not used in this study but the results of the LSTM indicate that supplementing with other spatial learning capabilities might be beneficial for predicting more accurately in future studies.

5.3 Practical Implications

The applicability of this study is not limited to an urban transportation management, but also to multiple stakeholders in urban transportation management. If traffic flow can be accurately predicted, proactive management of traffic signals, congestion control, route planning, and information systems for travellers can be assisted. Predictive models can help transportation authority's predict upcoming traffic jams, so they can take measures to avoid them before they become unmanageable.

Further, machine learning-driven forecasting systems can play a role in smart city solutions by helping to create data-driven transportation planning and ensure sustainable mobility goals. The better the traffic is forecasted, the less fuel can be used, travel delays can be prevented and vehicles can emit less traffic exhausts, thereby improving both the economic efficiency and environmental aspects, (Ismaeel et al., 2023).

5.4 Study Limitations

This study has a number of limitations. First, the analysis is based on historical transportation data which may not be able to accurately reflect unforeseen transportation activities like public events, road closures, or incidents. Secondly, spatial dependencies between interconnected road segments are not explicitly included in the selected machine learning models as they mainly focus on temporal relationships. Third, computational load can be a problem with deep learning models for Realtime implementation in limited resource transportation applications.

Future studies are needed to overcome these drawbacks, focusing on sophisticated hybrid network structures that are able to model both the spatial and temporal aspects of traffic, sharing of data from multiple stream data sources, and

incorporation of real-time data streams into the network.

6. THE CONCLUSION AND FUTURE WORK

6.1 Summary of Research

The aim of this study was to explore the use of machine learning methods to predict traffic flow by utilizing urban transportation data. The rising trend of urbanization and the growing need for transportation today makes the accurate forecasting of traffic indispensable in enhancing traffic management efficiency and providing assistance for smart city initiatives. Three predictive models were designed and assessed: Support Vector Regression (SVR), Random Forest (RF) and Long Short-Term Memory (LSTM), using as inputs the transportation related variables of traffic volume, traffic occupancy rate, the speed of the vehicles, and temporal information.

It was found that complex traffic patterns could be well modelled and accurate traffic flow forecasting could be obtained through using machine learning techniques. Overall, the LSTM model obtained the best predictive performance compared with the other evaluated models, which were assessed by different predictive indicators such as MAE, RMSE, MAPE, and R2. These results align with recent research showing the capacity of deep learning models for modelling temporal relationships and nonlinearities in transportation systems (Rajagopal et al., 2025). This study also revealed that historical traffic volumes, traffic occupancy, and traffic time-related factors contribute the most to the traffic conditions, which aligns with the findings reported in recent traffic forecasting research (Qi et al., 2025).

6.2 Major Contributions

This research not only fits into the research line on intelligent transportation systems but also offers a comparative study of conventional machine learning and deep learning methods in urban traffic prediction. First, the study created a systematic machine learning paradigm that is able to process urban transportation information for forecast. Second, it proved the superiority of the deep learning based traffic prediction models compared to traditional machine learning models. Third, results offer valuable guidance to select the forecasting model and the predictive variables for real-world applications in transportation management.

Moreover, the research proposes the necessity of integrating the data-driven forecasting system into an intelligent transportation system. Accurate traffic prediction can be used to make proactive decisions, optimize traffic control operations and enhance overall urban traffic mobility efficiency (Liu & Shin, 2025).

6.3 Recommendations for Future Research.

In the future, it is desirable to include connected streams from real-time traffic data sources such as

connected vehicles (CVs), IoT (Internet of Things) devices, and smart transportation infrastructures (STIs). Since there may be more cases to consider and more data to provide, the incorporation of real-time data may ultimately result in greater accuracy of predictions and traffic responsiveness in the face of quickly-acquired knowledge. Further studies could also explore more complex hybrid models integrating Graph Neural Networks (GNNs), Convolutional Neural Networks (CNNs), LSTMs, and other models to better account for the spatial and temporal dimensions of traffic data (Patil et al., 2024).

Additionally, future research could involve further investigation into explainable AI methods that would enhance the transparency and interpretability of the model. Scalable and interpretable traffic prediction models will help ease deployment of models on the scale of transportation agencies and as part of decision making. These will be important steps towards sustainable urban mobility and development of ITS.

REFERENCES

1. Abduljabbar, R. (2025). *Machine learning traffic flow prediction models for smart transportation systems*. *Buildings*, 10(7), 155. <https://www.mdpi.com/2412-3811/10/7/155>
2. Ali, R. (2025). *A comprehensive survey of deep learning-based traffic flow prediction*. *Traffic and Control Systems*. <https://www.icck.org/article/abs/tacs.2025.795448>
3. Attioui, M., et al. (2025). *Congestion forecasting using machine learning: A systematic review*. *Data*, 5(3), 76. <https://www.mdpi.com/2673-7590/5/3/76>
4. Berlotti, M., Di Grande, S., & Cavalieri, S. (2024). *Proposal of a machine learning approach for traffic flow prediction*. *Sensors*, 24(7), 2348. <https://www.mdpi.com/1424-8220/24/7/2348>
5. Fadila, J. N., et al. (2024). *Comprehensive review of smart urban traffic management and prediction systems*. *IEEE Transactions on Intelligent Transportation Systems*. <https://ieeexplore.ieee.org/iel8/6287639/10380310/10772087.pdf>
6. Fu, F., et al. (2024). *Urban traffic flow prediction based on Bayesian deep learning*. *Sustainability*, 16(5), 1818. <https://www.mdpi.com/2071-1050/16/5/1818>
7. Hassan, M., Khan, M., & Alshammari, A. (2025). *Application of machine learning in intelligent transportation systems: A comprehensive review*. *Discover Computing*, 8(1), 56–74. <https://link.springer.com/article/10.1007/s44290-025-00256-2>
8. Huang, C. I., Chen, Y. T., & Wang, J. (2025). *xLSTM-based urban traffic flow prediction for intelligent transportation systems*. *Applied Sciences*, 15(14), 7859. <https://www.mdpi.com/2076-3417/15/14/7859>
9. Huang, Y., Dong, Y., Tang, Y., & Li, L. (2024). *Leverage multi-source traffic demand data fusion with transformer model for urban parking prediction*. *arXiv*. <https://arxiv.org/abs/2405.01055>
10. Ionescu, Ş. A., et al. (2024). *PRISMA on machine learning techniques in smart city transportation systems*. *Applied Sciences*, 14(16), 7378. <https://www.mdpi.com/2076-3417/14/16/7378>
11. Ismaeel, A. G., Mary, J., Chelliah, A., Logeshwaran, J., Mahmood, S. N., Alani, S., & Shather, A. H. (2023). *Enhancing traffic intelligence in smart cities using sustainable deep radial function*. *Sustainability*, 15(19), 14441. <https://doi.org/10.3390/su151914441>
12. Jha, M. K., Sharma, A., & Gupta, R. (2025). *A machine learning approach to traffic congestion prediction using urban transportation data*. *Data*, 5(4), 161. <https://www.mdpi.com/2673-7590/5/4/161>
13. Liu, R., & Shin, S. Y. (2025). *A review of traffic flow prediction methods in intelligent transportation systems*. *Applied Sciences*, 15(7), 3866. <https://www.mdpi.com/2076-3417/15/7/3866>
14. Mystakidis, A. (2025). *Advances in traffic congestion prediction: An overview of intelligent forecasting approaches*. *Smart Cities*, 8(1), 25. <https://www.mdpi.com/2624-6511/8/1/25>
15. Patil, M., Ahmed, Q., & Midlam-Mohler, S. (2024). *Urban traffic forecasting with integrated travel time and data availability in a conformal graph neural network framework*. *arXiv Preprint*. <https://arxiv.org/abs/2407.12238>
16. Qi, P., Wang, J., & Zhang, Y. (2025). *A review of dynamic traffic flow prediction methods for intelligent transportation systems*. *Sensors*, 25(17), 5560. <https://www.mdpi.com/1424-8220/25/17/5560>
17. Rajagopal, M., Kumar, V., & Singh, P. (2025). *An efficient intelligent transportation system for traffic flow prediction using deep learning approaches*. *Scientific Reports*, 15, 10794. <https://www.nature.com/articles/s41598-025-10794-5>
18. Topilin, I., Kolesnikov, A., & Ivanov, P. (2025). *Traffic flow prediction via a hybrid CPO-CNN-LSTM model*. *Smart Cities*, 8(5), 148. <https://www.mdpi.com/2624-6511/8/5/148>
19. Vatambeti, R., et al. (2025). *Enhancing urban traffic congestion prediction through hybrid deep learning*. *Scientific Reports*.

<https://www.nature.com/articles/s41598-025-24012-9>

20. Zhang, H., Lin, Z., Zhou, J., Sun, J., & Zhou, T. (2025). *A comprehensive review of traffic flow prediction: From traditional models to deep learning architectures*. *Digital Transportation and Safety*, 4(4), 281–297. <https://www.maxapress.com/article/id/6954c696fa6c5806a63d7211>