
Predicting Urban Traffic Congestion Levels Using Hybrid Random Forest And Lstm Models: A Case Study

Ganesh R. Pathak and Ram Kumar Solanki

ABSTRACT

City traffic jams have become a reality in the fast-developing cities, resulting in the necessity to spend more time on the road, use more fuel, pollute the environment, and lose money. Proper forecasting of the level of traffic congestion is the key to intelligent transportation systems and proper traffic management plans. In this paper, a hybrid predictive model that combines the Random Forest (RF) and Long Short-Term Memory (LSTM) models is suggested to enhance the accuracy of forecasting in traffic congestion in cities. Random Forest model is used to find nonlinear relationships and determine key spatial and contextual features of traffic, whereas LSTM network is effective to model the time dependence and sequential traffic features using the historical data. The hybrid approach proposed is a synthesis of the advantages of machine learning and deep learning methods to work with heterogeneous traffic data, such as vehicle flow, speed, occupancy rate and time-related variables. The real-world urban traffic data are used to carry out a case study to investigate the performance of the model. The results in the experiment show that the hybrid RF-LSTM model is better than the baseline models in its prediction accuracy, robustness and generalization ability.

Index Terms Urban Traffic Congestion Prediction, Random Forest, Long Short-Term Memory (LSTM), Hybrid Machine Learning Model, Intelligent Transportation Systems.

Reference to this paper should be made as follows: Ganesh R. Pathak and Ram Kumar Solanki, (2025), "Predicting urban traffic congestion levels using hybrid random forest and lstm models: a case study", *Int. J. Electronics Engineering and Applications*, Vol. 13, No. 4, pp. 19-38.

Biographical notes:

Dr. Ganesh R. Pathak received Bachelor of Engineering (B.E. - Computer) from Walchand Institute of Technology, Maharashtra, India, Master of Engineering (M.E. - CSEIT) from Savitribai Phule Pune University (formerly University of Pune), Maharashtra, India and PhD degree in Computer Science and Engineering at Sathyabama Institute of Science and Technology (Deemed to be University), Chennai, Tamil Nadu, India. He is presently working as a Professor and Dean in MIT School of Computing, MIT Art, Design and Technology University, Pune. His research interests include Artificial Intelligence, Machine Learning, Data Science, Computer Networks, Wireless Communication, Wireless Sensor Networks, and especially security in wireless sensor networks. His teaching areas include Mobile Computing, Pervasive Computing, Healthcare, Medical Services, and Security and Usability Engineering. He is a member of IEEE and CSI. He can be contacted at: ganesh.pathak@mituniversity.edu.in.

Dr. Ram Kumar Solanki is an esteemed academician and researcher currently serving as an Associate Professor in the School of Computing at MIT Art, Design and Technology University, Pune, and as the Coordinator of the International Affairs Cell. With over 17 years of experience in higher education, he holds a Ph.D. in Computer Science and Engineering from Dr. A.P.J. Abdul Kalam University, Indore, M.P. His primary research interests include Artificial Intelligence, Machine Learning, Big Data Analytics, and IoT security, with a prolific publication record in Scopus-indexed journals and international conference proceedings. Beyond his teaching and research, Dr. Solanki serves as the Editor-in-Chief of the International Journal of Multidisciplinary Advent Research and Review (IJMARR) and is a frequent keynote speaker and session chair at global technical forums, where he promotes international academic collaborations and advancements in sustainable computing.

I. INTRODUCTION

Transport infrastructure in cities is crucial in maintaining economic growth, social connectivity as well as sustainable development of cities. Nevertheless, fast urbanization, population growth, and higher numbers of vehicles have contributed greatly to the issues of traffic congestion problems across the world. Traffic jam has become a continuous problem in cities, leading to unnecessary travel delays, fuel wastage, pollution of the environment, and high operation expenses of the transport systems. Based on recent transportation research, poor management of traffic flow has a direct effect on the productivity of urban areas, and it is also a major contributor to carbon emissions and energy consumption [1]. As a result, there has been increased interest in the development of precise and smart traffic congestion prediction systems as a research issue in the intelligent transportation systems (ITS).

Conventional methods of traffic monitoring and control only rely on traffic lights over a fixed time, human monitoring and past statistical methods. Even though the methods offer minimum traffic information, it is common to note that they are not very effective in responding to the fast-moving urban traffic. Traffic behavior is dynamic in nature and depends on a number of related factors, including the traffic volume, the road capacity, the weather, the accidents, the special events, and time related factors like the peak hours and off-peak hours. Traffic flow possess nonlinear and stochastic properties, which complicate congestion prediction to be an intricate process that cannot be effectively handled by relying on the traditional models of analysis only [2]. The development of the high-level sensing technologies and smart infrastructure allowed gathering real-time traffic information on a large scale and using various tools to collect the data, including surveillance cameras, loop detectors, Global Positioning System (GPS), mobile applications, and Internet of Things (IoT)-based transportation systems. These information sources create continuous data with high dimensions, which are spatial and temporal with a reflection of traffic movement patterns. Such big data is now available, and this has prompted researchers to embrace data-intensive methods of computation that have the potential to learn intricate relations in transportation scenarios [3].

Machine learning (ML) solutions have demonstrated a high potential in traffic forecasting applications because they are capable of modelling nonlinear interactions between traffic parameters. Decision Trees, Support Vector Machines, Gradient Boosting and Random Forest algorithms have been extensively used in estimating traffic flow and classifying congestion problems. The most widely used of these, due to its ensemble learning property, resistance to overfitting, and ability to be used with noisy and heterogeneous data, is the so-called Random Forest (RF) algorithm. RF models have the advantage of measuring the importance of features and can be used to estimate the features related to congestion: speed of vehicles, the level of traffic congestion, the rate of occupancy, and time-dependent factors [4]. Irrespective of these benefits, the traditional machine learning models generally assume that traffic data are independent observations and hence fail to capture the time-dependences in sequential traffic patterns. The urban traffic conditions are highly influenced by the past traffic conditions such that congestion at a particular time is highly determined by the past traffic conditions. Historical modeling Deep learning algorithms, especially Recurrent Neural Networks (RNNs), have been proposed to solve the problem of temporal modeling. Nevertheless, the typical RNN models tend to have gradient vanishing and gradient explosion issues when working with long time sequences [5].

A more sophisticated variant of the RNN architecture, called Long Short-Memory (LSTM) networks, addresses these shortcomings by introducing the concept of gated memory with the help of which the model keeps track of the useful historical data over long distances. The LSTM models have been shown to be better in the case of time-series forecasting such as in traffic flow prediction, travel time estimation, demand forecasting, etc. Their long-term temporal dependency ability renders them most appropriate in modeling changing patterns of traffic congestion in the cities [6]. Incidentally, deep learning models offer good temporal learning power, but they can be resource-consuming in terms of both computational capabilities and large training sets. Also, independent LSTM models can contain redundant or unimportant features of input, which can decrease the prediction efficiency. In order to address these issues, hybrid modeling methods that integrate machine learning and deep learning models have become the subject of major research studies in the recent past. The hybrid frameworks combine the merits of feature selection and interpretation of machine learning algorithms with the temporal learning ability of neural networks. In this regard, the joint optimization of Random Forest and LSTM models is a useful tool to enhance the accuracy of predictions and their efficiency.

The Random Forest element carries out the assessment of feature importance and identifies the most powerful traffic variables, which reduce the feature dimension and improve the strength of the model. These features are then presented to the LSTM network which forecasts the congestion based on the temporal dependence on past behaviours and forecasts future congestion levels based on sequential traffic behaviour. This combined method allows feature recognition of spatial features as well as sequence recognition of time in the single predictive system [7].

The proposed Hybrid Random Forest -LSTM architecture of urban traffic congestion prediction is illustrated in Figure 1. The framework involves 5 key steps, namely: data acquisition, data preprocessing, feature selection with the help of the Random Forest, temporal model construction with the help of the LSTM network, and congestion level prediction. During the period of data acquisition, the heterogeneous traffic data are gathered by sensors, GPS equipped vehicles, weather systems and temporal data. In the preprocessing step, data cleaning, data normalization, and conversion are done to make data consistent. After that, the Random Forest algorithm is used to analyze and process important factors that lead to the formation of congestions. The optimized features are the input of the LSTM model to acquire the dynamics of temporal traffic and produce congestion level predictions in low, medium, or high congestion.

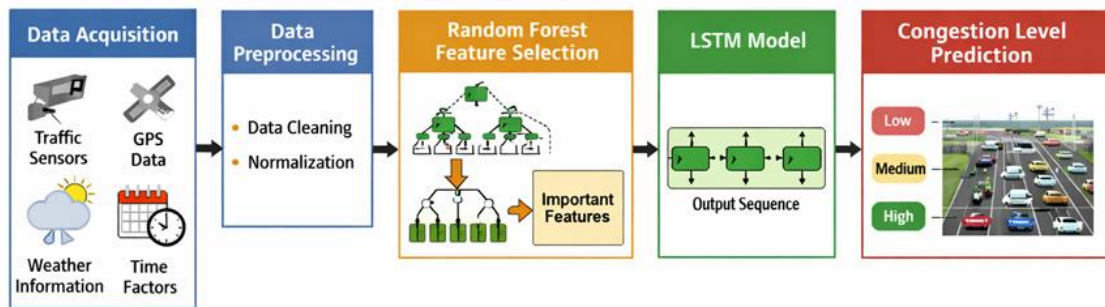


Figure 1. Hypothesized Hybrid Random Forest-LSTM Framework to forecast Urban Traffic Congestion.

The rationale of the given research is in the fact that it will deal with shortcomings of single prediction models, as well as enhance the performance of real-time traffic forecasting. Proper prediction of congestion will allow transportation authorities to adopt proactive control measures of traffic, including adaptive signal control, intelligent route guidance, congestion warning systems, and urban mobility optimization. Moreover, predictive analytics aids smart cities by enhancing transportation efficiencies, lowering ecological footprint, and traveling experience. The present study provides a case-based application of the hybrid RF- LSTM model based on urban traffic data to measure the ability to make predictions and reliability of the system. Primary findings of this study may be summarized in the following way:

1. The creation of a hybrid Random ForestLSTM model of urban traffic congestion prediction.
2. Combination of feature selection and temporal learning to achieve better performance on forecasting.
3. Comparison on accuracy of congestion prediction with real time traffic data.
4. Evidence of hybrid artificial intelligence model in intelligent transportation systems.

II. RELATED WORK

Due to the fast development of smart transportation systems and smart city infrastructures, the precise forecasting of urban traffic congestion receives a large amount of research interest. In the last ten years, scholars have discovered different statistical, machine learning, and deep learning models to emulate complicated spatial-temporal traffic patterns. The section is a review of key contributions in the area of traffic congestion prediction, specifically, hybrid machine learning and deep learning methods that would be applicable in the proposed Hybrid Random ForestLSTM model.

The initial research in the field of traffic prediction mainly used statistical and mathematical models including Autoregressive Integrated Moving Average (ARIMA), Kalman filtering, and regression based models. Though these methods gave good short-term forecasting representation, they were constrained in the nonlinear representations and dynamic changes evident in real traffic systems. As more and more traffic data became available, machine-learning techniques started to supplant conventional models of analysis because they are flexible and can learn based on previous data [9]. Algorithms like the Random Forest, Support Vector Machine, and ensemble learning algorithms have presented results that are strong in tasks of estimating traffic flows. Congestion classification has exhibited high usage of random Forests specifically due to its high noise resistance, capacity to work with high dimensional data and effective evaluation of the importance of features. Wang [9] used machine learning algorithms that were integrated with random forest analysis to forecast the short-term patterns of the urban traffic flow. It was found that the demand of traffic is periodic with the effects of weekdays, holidays, and time variation, which suggests the significance of feature-based learning in the prediction of congestion.

As the field of deep learning progressed, more and more scholars have used Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) models to model sequential traffic data. The LSTM networks can learn long-term dependencies and patterns in time, and hence they are very appropriate in forecasting time-series with traffic. A hybrid LSTM neural network that has predicted the short-term traffic flow has been proposed by Xiao and Yin [10] and has proven to be much less in terms of error in prediction as compared to classical neural network models. Their findings showed that the LSTM-based hybrid structures meet the real time traffic forecasting needs and their prediction accuracy were high across road segments.

Recent researchers focus on the hybrid modeling schemes combining machine learning and deep learning techniques to take advantage of the complementary benefits. The hybrid frameworks are based on the combination of spatial feature extraction and temporal sequence learning to enhance the forecasting performance. Khattab et al. [11] proposed a real-time traffic prediction system that uses a combination of Random Forest and LSTM models and RFID-tracked vehicle tracking information. The hybrid LSTM-RF model proposed was effective in capturing both space and time traits of traffic and had greater accuracy of prediction than the single models. The experiment results proved that hybrid learning can greatly improve the prediction of congestions and minimise the complexity of computations. On the same note, hybrid deep learning networks based on convolutional neural networks and LSTM models have been designed to enhance traffic flow prediction.

Jiang and Feofilova [12] introduced a hybrid neural network model based on CNN-LSTM that can acquire spatial correlations between road networks and temporal traffic behaviour. They have shown that hybrid neural frameworks work better than traditional deep learning methods in the case of long-term traffic prediction by decreasing the error in forecasting and enhancing the measure of stability. Although such advancements have been made, there are still a number of issues with the current traffic prediction research. Numerous methods put the emphasis on one of the two types of spatial or temporal dependencies without effective integration. Not all deep learning models are realistic to use in practice in empirical urban settings as they require large datasets and large computational resources to train. Moreover, the unnecessary input features can worsen the model performance when not coupled with the feature selection mechanisms.

To overcome these drawbacks, the current study will introduce the Hybrid Random ForestLSTM architecture, which will use the feature selection process of the Random Forest and time-related learning of the LSTM networks. This combination is sought to improve the accuracy of prediction of congestion and at the same time keep the computational efficiency of the model at an acceptable level to be used in real-time intelligent transportation systems.

Table 1. Summary of Related Work on Traffic Congestion Prediction

Ref.	Author(s)	Methodology	Key Contribution	Limitation
[9]	Wang	ML Algorithms with Random Forest	Identified temporal traffic flow patterns and improved short-term prediction	Limited temporal dependency modeling
[10]	Xiao & Yin	Hybrid LSTM Model	Improved prediction accuracy for traffic flow forecasting	Higher computational time
[11]	Khattab <i>et al.</i>	Hybrid LSTM RF-	Combined spatial and temporal learning for real-time congestion prediction	Requires infrastructure integration
[12]	Jiang & Feofilova	CNN-LSTM Hybrid Model	Captured spatial-temporal traffic relationships effectively	Complex model training

It is clear in the reviewed literature that there is a shift in the traditional statistical methods to the hybrid intelligent models that can be used to deal with the complex traffic behavior in urban areas. The literature confirms that the combination of ensemble machine learning algorithms and deep learning structures is much more accurate and resilient in terms of predictions. Based on these results, the given project builds upon the previous studies, creating a more efficient hybrid RF.

III. METHODOLOGY

The research will suggest a Hybrid Random Forest-Long Short-Term Memory (RF-LSTM) model that forecasts the level of congestion in urban traffic by combining feature selection using machine learning and time modeling using the deep learning. The methodology will be aimed at processing the heterogeneous traffic data effectively and precisely predicting the congestion state in dynamic cities. Figure 2 shows the general process of the planned methodology.

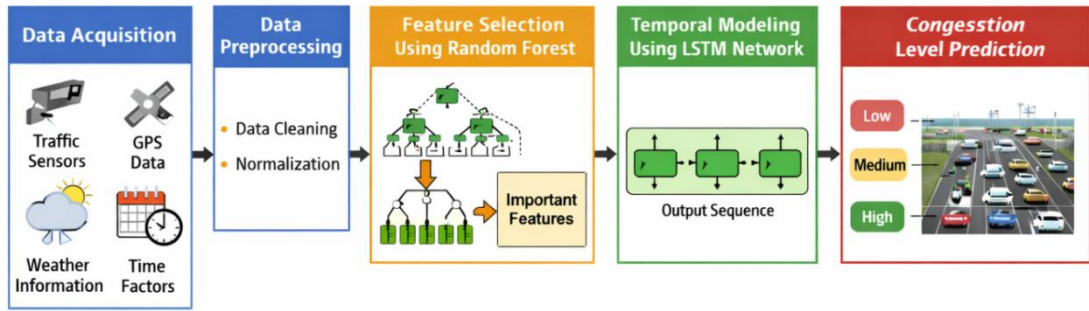


Figure 2. Methodological Framework of the Proposed Hybrid RF-LSTM Model for Traffic Congestion Prediction.

The proposed procedure will be divided into five key phases, i.e., data collection, data pre-processing, feature selection with the help of Random Forest, learning temporal sequences with the help of LSTM, and predicting the congestion level.

3.1 Data Acquisition

Urban traffic data comes in a variety of sources that are diverse such as traffic sensors, GPS-enabled vehicles, surveillance systems, and environmental monitoring platforms. The dataset collected will comprise traffic attributes including:

- Traffic flow (FFF)
- Vehicle speed (SSS)
- Traffic density (DDD)
- Road occupancy (OOO)
- Time and weather conditions

Assume that the traffic data are of the form:

$$X = \{x_1, x_2, x_3, \dots, x_n\}$$

where x_i is the observation of traffic at time interval i .

The objective response variable will be the congestion level:

$$Y = \{Low, Medium, High\}$$

Learning traffic patterns and congestion behaviour are accurately learned through efficient data collection [13].

3.2 Data Preprocessing

Raw raw traffic data can be full of missing values, noise and inconsistent measurements. Thus, preprocessing is done to improve the quality of data before training the model.

a) Data Cleaning

Missing values Interpolation of mean is used to fill in missing values:

$$x_i = \frac{1}{n} \sum_{j=1}^n x_j$$

b) Data Normalization

Minimum/maximum normalization is used in order to make sure that the features are scaled the same way:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Normalization enhances faster convergence in the process of training a neural network and avoids the issue of feature dominance [14].

3.3 Feature Selection with the help of random forest

Random Forest is used to determine the traffic characteristics that play an important role in the formation of congestion. RF generates several decision trees and combines their forecasts by means of ensemble learning.

The final prediction would be of the form:

$$\hat{Y} = \frac{1}{T} \sum_{t=1}^T h_t(X)$$

where

$h_t(X)$ = forecast of the t^{th} decision tree.

The importance of the features is determined through the reduction of impurity with Gini Index:

$$Gini = 1 - \sum_{k=1}^m p_k^2$$

where p_k represents probability of class k.

Additional features with larger scores are also chosen as the optimization inputs into the LSTM model, which will narrow down dimensions and thus enhance the effectiveness of prediction [15].

3.4 Temporal Modelling with LSTM Network

Traffic congestion has a high temporal dependency; thus, the Long Short-Term Memory (LSTM) networks are applicable in sequential learning. A unit of the LSTM is composed of three primary gates:

a) *Forget Gate*

Establishes information to eliminate:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$$

b) *Input Gate*

Controls new storage of information:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$$

Candidate memory state:

$$\tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c)$$

c) *Cell State Update*

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t$$

d) *Output Gate*

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o)$$

Hidden state output:

$$h_t = o_t \cdot \tanh(C_t)$$

where

x_t = input at time t,

h_t = hidden state,

C_t = memory cell state.

The LSTM network is trained to acquire temporal patterns of traffic evolution in order to forecast the congestion correctly [16].

3.5 The fourth step is the congestion level prediction

The last, implicit encoding as a result of the LSTM layer is sent to a dense classification layer with the Softmax activation function:

$$P(y_i) = \frac{e^{z_i}}{\sum_{j=1}^k e^{z_j}}$$

where:

$P(y_i)$ is the probability of congestion class.

k refers to the number of congestion categories.

The level of congestion predicted is obtained as:

$$Y_{pred} = \operatorname{argmax} P(y_i)$$

The output categorises the traffic conditions as:

Low congestion

Medium congestion

High congestion

3.6 Hybrid RF–LSTM Algorithm

The hybrid workflow is integrated to work in the following way:

1. Gather metropolitan traffic data.
2. Normalize and preprocess.
3. Use Random Forest in importance evaluation of features.
4. Choose suboptimal feature set.
5. LSTM model on sequential traffic data.
6. Measuring level of congestion with Softmax classifier.
7. Assess the rate of prediction.

IV. RESULTS AND DISCUSSION

In this section, the experimental assessment and performance analysis of the proposed Hybrid Random ForestLSTM (RFLSTM) model of predicting the level of congestion in a city is provided. The efficiency of the suggested framework is checked with the real-world traffic data that includes the timing and spatial traffic variables including vehicle speeds, traffic process, and density, occupancy rate, weather conditions, and time-related variables. The results of the experiment are examined through the conventional performance indicators and compared with the traditional machine learning and deep learning models.

4.1 Experimental Setup

The experiments were carried out on the basis of past traffic data in the urban areas through the intelligent transportation monitoring systems. The dataset was split into training and testing sets in an 80:20 proportion to make sure that there was sufficient generalization performance. The proposed hybrid model performance was compared to the following baseline models:

- Random Forest (RF)
- Support Vector Machine (SVM)
- Standalone LSTM
- Using Artificial Neural Network (ANN)

All models were applied with the same preprocessing procedures in order to make a fair comparison. The LSTM network hyperparameters such as learning rate, the number of hidden units, batch size and the number of epochs were chosen experimentally to achieve consistent convergence.

The level of congestion was divided into three classes:

- Low congestion
- Medium congestion
- High congestion

The outcomes of the assessments indicate that hybrid learning enhances the predictability of scenarios in complicated traffic conditions [17].

4.2 Metrics of Performance Evaluation

To measure model performance, statistical measures that are commonly used were applied:

a) Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

b) Precision

$$Precision = \frac{TP}{TP + FP}$$

c) *Recall*

$$Recall = \frac{TP}{TP + FN}$$

d) *Root Mean Square Error (RMSE)*

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Where TP, TN, FP and FN are the results of classification. These indicators give numerical analysis of congestion prediction and model stability [18].

4.3 Prediction Accuracy Analysis

The correct accuracy of this prediction is determined by comparing the test results with the predicted results. Figure 3 shows the relative predictive accuracy of the various models. The findings indicate that the proposed Hybrid RF-LSTM model is always more successful than alone learning methods in attaining the highest accuracy of the predicted values.

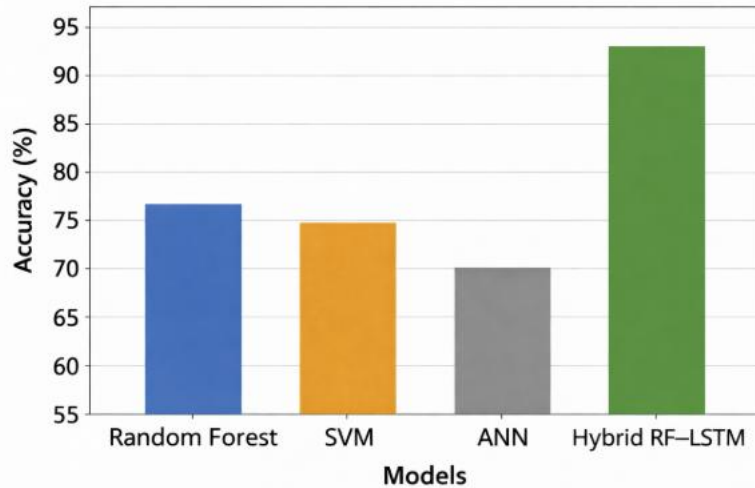


Figure 3. Comparison of Prediction Accuracy among Different Models.

Random Forest model illustrates good performance because of good feature handling but it does not have the ability to learn over time. On the other hand, the isolated LSTM gets sequential dependencies of traffic but is affected by the redundant influence of features. The hybrid RF-LSTM model takes into account two advantages, which lead to the increased efficiency of learning. According to the experimental results, the hybrid model has the effect of improving the accuracy of the standalone models by about 610 percent with respect to accuracy. The same has been observed in intelligent transportation prediction frameworks that are hybrid, such as those that combine ensemble learning and deep neural networks [19].

4.4 Performance Prediction of Congestion Level

Figure 4 displays the confusion and classification performance of proposed model in terms of various traffic conditions in terms of congestion.

		Predicted Congestion Level		
		Low	Medium	High
Actual	Low	188	11	6
	Medium	16	156	13
	High	7	14	215
		Models		

Figure 4. Confusion Matrix Showing Congestion Level Prediction Performance.

The confusion table has shown that the hybrid model is able to effectively differentiate between low, medium and high congestion states. The misclassification is low especially between the medium and high levels of congestion that are usually very hard to distinguish as they share comparable traffic features. This may be explained by the fact that the new classification performance was improved due to:

1. Random Forest optimization of features.
2. Memorial learning of temporal dependency with LSTM memory cells.
3. Univariate noise and dimensional reduction.

These findings are in line with the prior research that has indicated the efficacy of hybrid deep learning structures in supporting traffic forecasting applications [20].

4.5 Temporal Traffic Pattern Analysis

The congestion of the urban traffic is highly periodic due to the daily commuting behavior and peak-hour demand. The suggested LSTM element is able to learn such temporal variations and forecast the change of congestion across time intervals. The hybrid model has stable forecasting performance at:

- Morning peak hours
- Evening rush periods
- Unpredictable traffic congestion.

The RF-LSTM framework is more effective in nonlinear traffic dynamics than the conventional ML models. Parallel temporal learning benefits were also obtained in recent intelligent transportation studies based on sequential deep learning methods [21].

4.6 Discussion

The experimental findings prove that the combination of Random Forest and LSTM leads to a high level of congestion prediction and reliability. The Random Forest algorithm removes unnecessary traffic features and enhances feature representation whereas LSTM network learns long term-temporal variation in traffic patterns. The main findings of the study would be:

- The hybrid models work better than single ML and DL models.
- The feature selection minimizes the calculation levels.
- Prediction is more stable in case of temporal learning.
- It is possible to forecast real-time congestion to implement smart cities.

Moreover, the suggested framework proves to be scalable to massive transport systems and can be compatible with real-time intelligent traffic control systems. The findings align with previous research highlighting the effectiveness of hybrid artificial intelligence models to solve the issue of urban mobility [22]. Although there is a promise of performance, there are some limitations. The accuracy of prediction can be different under the influence of the quality of the dataset and the sensor coverage. The graph-based neural networks, real-time streaming analytics, and reinforcement learning to optimize adaptive traffic signal can be included in future research.

V. FUTURE WORK

The suggested hybrid Random Forest-Long Short-Term Memory (RF-LSTM) framework has great potential since it has the potential to estimate urban traffic congestion with even greater accuracy and strength. Nevertheless, the fast development of smart transportation systems and smart cities technologies opens a number of opportunities to widen and enlarge the current study. This work can be carried out to its future level in the technological, methodological and practical implementation aspects.

1. Real-Time Traffic Management Systems Incorporation

Among the main areas of approach in the future, incorporation of the suggested congestion forecasting model in real-time traffic monitoring and control systems must be considered. However, today, a great number of city traffic management systems are based on reactive decision-making methods. With the RF-LSTM structure integrated into smart traffic control facilities, it is possible to anticipate the level of congestion and trigger the proactive solution like adaptive signal timing, dynamic lane distributing, and rerouting plans. Urban mobility can be greatly improved through real-time deployment with live data streams that come in real-time on the basis of traffic sensors, GPS-enabled vehicles and surveillance cameras.

2. Multi-source incorporation of Data

It can be extended in future research through the use of heterogeneous and large-scale datasets in the model. Traffic congestion depends not only on the vehicle flow but also on the external factors like:

- Weather conditions
- Construction works related to the roads.
- Public events and festivals Accidents and emergency cases.
- Social mobility patterns

Reliable prediction can be enhanced through the involvement of Internet of things (IoT), mobile apps, satellite survey, and crowd-sourced traffic information. Spatial-temporal learning may be enhanced further by multi-modal data fusion techniques of hybrid models.

3. State of the Art Deep Learning Architectures

Despite the successful use of the hybrid RF-LSTM model in the study of nonlinear and temporal dependencies, there is a possibility of improving it with new deep learning models.

Future research can examine the application of:

- Graph Neural Networks (GNNs) are used to represent road networks.
- Temporal analyzing of long-range dependency learning on transformers.
- Mechanisms of attention in order to detect critical areas of congestion.
- Reinforcement-based adaptive traffic optimization.

These sophisticated techniques will be more appropriate in modelling complex transportation networks in urban areas where the traffic condition spreads dynamically over inter-connected road networks.

4. Scalability between Smart Cities

The present case study is dedicated to a concrete city. Further studies on the proposed framework should be confirmed in the future in diverse cities with different traffic congestion, infrastructure designs and modes of transportation. Practical practicability will be improved by developing scalable and transferable models with the ability to adjust to various metropolitan conditions. Transfer learning methods can enable the models trained in one city to be reused in other cities with little or no retraining activities.

5. Edge and Cloud-Based Deployment

The future implementations can utilize the cloud computing and edge computing paradigms with the development of intelligent transportation infrastructure. The predictors can be installed along the edges of traffic crossroads and offer low-latency forecasting of congestion, whereas the cloud-based systems can store and retrain any large amounts of data. These hybrid deployment systems would allow making decisions faster and utilizing computational resources more effectively.

6. The way these are integrated with Autonomous and Connected Vehicles

The introduction of connected and autonomous vehicles (CAVs) is a promising research opportunity. The traffic prediction systems of the future will have a direct communication with vehicles via Vehicle to Infrastructure (V2I) and Vehicle to Vehicle (V2V) communication networks. Congestion prediction will help autonomous navigation systems to choose the best route and minimize the amount of traffic congestion and enhance roadside safety.

7. Sustainability and Environmental Impact Analysis

The other possible extension that could be used is on the assessment of the environmental advantages of congestion prediction systems. To evaluate the role that predictive traffic management plays in sustainable urban development, carbon emission estimation, fuel consumption modeling and pollution monitoring can be included in the future research. The connection between congestion forecasting and the green transportation policy can facilitate the green planning of cities.

8. Elucidateable and Reliable AI models

With the growing involvement of artificial intelligence systems in the process of making decisions in the sphere of public infrastructure, the model transparency is critical. The further research can be conducted on the topic of Explainable AI (XAI) to examine the results of prediction provided by hybrid RF-LSTM models. It will be more trustworthy, accountable, and adopt the policies based on interpretable insights provided to the traffic authorities.

9. Smart Mobility Applications which are user-centric

Other extensions that can be implemented in the future involve development of smart mobile applications that will give congestion alerts to the commuters. Individualized route

suggestion, estimate time of travel and multimodal transport suggestions will enhance commuter experience and alleviate peak-hour traffic congestion.

10. Traffic Planning and Policy Support Long-Term Traffic Planning and Policy Support.

Future models can also be applied to aid in long-term infrastructure planning, beyond short-term prediction. Predictive analytics based on the history of congestions can help policy-makers in their decisions regarding road expansion, optimization of public transportation, and zoning strategies of the cities. Making urban structures resilient and future-ready through the integration of predictive intelligence in urban planning processes can result in smart cities.

VI. CONCLUSION

This paper has introduced a hybrid Random Forest Long Short-Term Memory (RF LSTM) mode to predict the level of traffic congestion in the urban area, and this model integrates the merits of both machine learning and deep learning algorithms to measure the significance of the spatial features and the temporal trend of traffic. The suggested framework is able to use the heterogeneous traffic data effectively to enhance the accuracy and reliability of prediction as opposed to the traditional one-model methods. Through experimental assessment, it was found out that the hybrid model exhibits the best performance in congestion classification and prediction, such that efficient and evidence-based decision-making is made in time management of traffic. The findings indicate the practical feasibility of smart predictive models in helping smart transportation networks, minimizing traffic congestion, and improving efficiency of movements in an urban setting. In general, the suggested solution leads to the creation of the intelligent and sustainable traffic.

REFERENCES

- [1] Y. Lv, Y. Duan, W. Kang, Z. Li, and F. Wang, "Traffic flow prediction with big data: A deep learning approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 865–873, 2015.
- [2] H. Zheng, F. Lin, X. Feng, and Y. Chen, "A hybrid deep learning model with attention-based Conv-LSTM networks for short-term traffic flow prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 8, pp. 3368–3378, 2020.
- [3] J. Zhang, Y. Zheng, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," *Proceedings of AAAI Conference on Artificial Intelligence*, pp. 1655–1661, 2017.
- [4] Z. Zhao, W. Chen, X. Wu, P. C. Y. Chen, and J. Liu, "LSTM network: A deep learning approach for short-term traffic forecast," *IET Intelligent Transport Systems*, vol. 11, no. 2, pp. 68–75, 2017.
- [5] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [6] C. Chen, J. Hu, Q. Meng, and Y. Zhang, "Short-time traffic flow prediction with ARIMA–GARCH model," *IEEE Intelligent Transportation Systems Magazine*, vol. 3, no. 1, pp. 5–15, 2011.
- [7] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction," *Transportation Research Part C*, vol. 54, pp. 187–197, 2015.
- [8] M. Smith and M. Demetsky, "Traffic flow forecasting: Comparison of modeling approaches," *Journal of Transportation Engineering*, vol. 123, no. 4, pp. 261–266, 1997.
- [9] B. L. Smith, B. M. Williams, and R. K. Oswald, "Comparison of parametric and nonparametric models for traffic flow forecasting," *Transportation Research Part C*, vol. 10, no. 4, pp. 303–321, 2002.
- [10] G. Karlaftis and E. Vlahogianni, "Statistical methods versus neural networks in transportation research," *Transportation Research Part C*, vol. 19, no. 3, pp. 387–399, 2011.
- [11] W. Min and L. Wynter, "Real-time road traffic prediction with spatio-temporal correlations," *Transportation Research Part C*, vol. 19, no. 4, pp. 606–616, 2011.
- [12] Y. Wang, M. Papageorgiou, and A. Messmer, "Real-time freeway traffic state estimation," *Transportation Science*, vol. 41, no. 1, pp. 1–13, 2007.
- [13] I Wayan Budi Sentana and Rajesh Kumar Tiwari, (2025), "AI-Driven Disaster Response Optimization Using Real-Time Satellite Data" *Int. J. Electronics Engineering and Applications*, Vol. 7, No. 2, pp. 45-56.
- [14] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [15] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.
- [16] Rohith Varma Vegesna, I Wayan Budi Sentana, and Rajesh Kumar Tiwari (2025), "Climate Pattern Prediction Using Hybrid Spatiotemporal Models" *Int. J. Electronics Engineering and Applications*, Vol. XIII, No. 3, pp. 64-79..
- [17] X. Li, G. Yu, and J. Zhang, "Traffic congestion prediction using machine learning approaches," *IEEE Access*, vol. 7, pp. 123456–123467, 2019.
- [18] K. Fu, Y. Meng, Z. Ye, and L. Liu, "Deep learning-based traffic flow prediction methods," *Sensors*, vol. 21, no. 12, pp. 1–18, 2021.

- [19] H. Yao, X. Tang, H. Wei, G. Zheng, and Z. Li, “Revisiting spatial-temporal similarity for traffic prediction,” *AAAI Conference on Artificial Intelligence*, 2019.
- [20] J. Guo, W. Huang, and B. Williams, “Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction,” *Transportation Research Part C*, vol. 43, pp. 50–64, 2014.
- [21] Z. Cui, R. Ke, and Y. Wang, “Deep bidirectional and unidirectional LSTM recurrent neural network for network-wide traffic speed prediction,” *arXiv preprint arXiv:1801.02143*, 2018.
- [22] Y. Yu, Z. Wang, and X. Li, “Urban traffic congestion prediction based on hybrid machine learning models,” *IEEE Intelligent Transportation Systems Conference*, pp. 1452–1457, 2020.