Machine Learning-Based Traffic flow Prediction and Management

T. Axcellin
Research Scholar,
Department of Computer Science
School of Computing Sciences
Vels Institute of Science, Technology & Advanced Studies
Chennai, Tamil Nadu,India
axcy.thangadurai@gmail.com

S. Kamalakkannan
Professor
Department of Computer Science
School of Computing Sciences
Vels Institute of Science, Technology & Advanced Studies
Chennai, Tamil Nadu, India
kannan.scs@vistas.ac.in

***Abstract* – Artificial Intelligence (AI) and data analytics are used by a machine learning-based traffic prediction and management to optimize transport networks, improve safety, and improve traffic management. By combining machine learning (ML) with real-time and historical data analysis, ITS makes predictive modeling and automated decision-making possible. This study suggests a hybrid model for traffic prediction based on Random Forest (RF) and Long Short Term Memory (LSTM). The performance is evaluated with Stand alone RF, K-Nearest Neighbors(KNN) and Naïve Bayes(NB) algorithms. The experimental results show that the RF + LSTM hybrid model has better performance than other methods with an accuracy of 95% , precision of 94% , recall of 94% and a F1-score of 93%. The top traffic predicting model is RF because it can handle structured data and LSTM can recognize sequential patterns. The study contributes to the development of more effective traffic management plans in smart cities by highlighting the use of hybrid models in predicting traffic.**

Keywords:* *Traffic Prediction, Machine Learning, Random Forest, Long Short Term Memory

I. INTRODUCTION

Intelligent Transportation Systems (ITS) have gained popularity in the last few years. The increasing rise of urban infrastructures, vehicle sensors, and vehicular computer hardware has led to the presentation of several impressive applications under the ITS theme. These applications have the potential to make our commute safer, more efficient, and more enjoyable. However, a highly accurate and productive traffic flow forecast system is needed to carry out these applications. [1].

It is seen to be one of the most pertinent issues in the modern world to estimate traffic patterns and use them to create solutions for managing transportation systems. In this field, numerous scientists carry out both theoretical and applied research, which results in the creation of novel models, algorithms, and techniques.

The term "traffic management" refers to the coordination of vehicle and pedestrian movement on public roads, roadways, and highways. Fundamentally, traffic management seeks to alleviate traffic congestion, encourage safe and efficient traffic flow, and reduce the frequency of traffic accidents. A range of strategies, including careful transportation planning, transportation engineering, and the employment of congestion control systems, are required to accomplish these goals. These actions all work to regulate traffic flow and ensure everyone's safety while driving. [2].

Traffic signal technologies are installed in urban areas to reduce traffic congestion. However, all roads have the same and uniform traffic light frequency division. Due to resource loss from the unpredictable incoming traffic on the two sides of the road, the signals are not equal. The more automobiles and road infrastructure there are, the more difficult it will be to manage a traffic and transportation network. There is usually a zebra crossing beside the signal on every road, and each signal has a certain amount of time to operate. [3].

The short and long-term predictions are the two types of traffic movement prediction. Based on past and present traffic data, the first category relates to activity in the near future, from just a few seconds to a few hours, and the latter deals with congestion for entire days in the future. Tourists are able to shorten their paths with the application of short-term predictions, while administrative agencies dealing in administrative and signaling control techniques for maximum operations are able to take advantage of long-term forecasts[4].

Network traffic can be classified by employing machine learning (ML) on the basis of user behavior or application type. This can provide priority to high-priority traffic and assist in guaranteeing the necessary QoS levels for critical applications. This can improve user experience and satisfaction, especially in low-latency, high-throughput, or real-time applications [5].

II. RELATED WORKS

An IoT- and ML-oriented adaptive traffic management system (ATM) is developed and deployed (Lihore et al., 2022). The design of the proposed system is based on three essential elements: the events, the infrastructure, and the vehicle. The design uses a range of scenarios to solve all possible issues with the transit system. The proposed ATM system additionally uses the ML-based DBSCAN clustering technique to detect any inadvertent irregularities. The proposed ATM model continuously updates traffic signal schedules depending on traffic flow and projected movements from nearby crossings. By improving the transition and gradually directing automobiles over green lights, it significantly reduces travel time and eases traffic congestion[3]

Tiwari et al., (2024) develop a data-driven approach that uses a ML algorithm to manage traffic management in a smart city. The results show that bringing smart transportation technologies into operation improves air quality and transportation conditions[6]. Kim et al., (2024) suggests SMURP (Simulation and ML Utilisation for Real-time Prediction), a hybrid traffic prediction system that gets over the drawbacks of the current approaches. Any data-based prediction method can be applied with the SMURP framework. The SMURP introduces a second predictor that utilizes simulated traffic data to the prediction outcome when an event is detected during prediction[7]

Sun et al., (2024) elucidates how pedestrian detection and behavior forecasting technologies are being developed and deployed within smart city infrastructure with a focus on enhancing public security and traffic control. The technology employs advanced machine learning models, including LSTM and Transformer architectures, to fuse real-time sensor data, LIDAR, and cameras to predict pedestrian trajectories with 93% accuracy. The device provided vehicles in risky pedestrian scenarios with a mean response time of 1.8 seconds and effectively reduced near-miss incidents by 30%.[8]

Yang et al., (2024) presents traffic flow monitoring and forecasting by cloud data warehousing and ML and their implementation in ITSs. It indicates how technology can assist urban traffic management and services through analysis of in-depth and case studies. The study first emphasizes the significance of traffic prediction and monitoring, pointing towards shortcomings of traditional approaches and promoting ML solutions[9].

Alkarim et al., (2024) offers a new conceptual ITS model that tries to predict intersections based on collective learning for predicting vehicle flow. The three main stages of the proposed method are data acquisition using cameras and sensors, the deployment of machine learning, and development of the DL algorithm. The findings indicated that the proposed method achieved better performance by 93.52%, which was significant[10].

An et al., (2024) introduces the algorithm GC-YOLOv9. By particularly incorporating Ghost Convolution into the YOLOv9 model, scientists have greatly enhanced the detection precision and perception ability of the model. Moreover, this project developed a framework for a comprehensive smart city that includes layers for data centres, edge processing, the IoT and service applications[11].

Rasulmukhamedov et al., (2024) investigates and evaluates various ML techniques for traffic intensity prediction, including gradient boosting, decision trees, and RFs. Video cameras placed at the intersection gave precise and current traffic flow data, which was used to gather data for the models. The comparative evaluation of these methodologies' performance is the primary emphasis of the research. The results have important real-world applications. Road and intersection capacity can be increased by using them to create ITSs. As a result, there may be less traffic, fewer dangerous chemical emissions into the atmosphere, and lower financial expenses related to traffic delays[12]. Table 1 shows the related works in traffic flow management.

Table 1 Related studies in traffic flow management

Author(s) & Year	Methodology	Strengths	Limitations
Lihore et al. (2022)	IoT and ML-based ATM system using DBSCAN clustering algorithm	Reduces congestion, dynamically updates traffic light schedules, improves traffic flow	May require high computational resources, potential scalability issues
Tiwari et al. (2024)	Data-driven ML approach for smart city traffic management	Improves air quality, enhances transportation conditions	Effectiveness depends on data quality and infrastructure availability
Kim et al. (2024)	SMURP (Simulation and ML Utilisation for Real-time Prediction) hybrid system	Adaptable to various prediction techniques	Computationally intensive, requires high-quality simulation data
Sun et al. (2024)	Pedestrian recognition and behavior prediction using LSTM and Transformer architectures	93% accuracy, improves public safety, reduces near-miss events	Requires extensive sensor and camera infrastructure, potential privacy concerns
Yang et al. (2024)	Cloud data warehousing and ML for traffic flow monitoring and prediction	Enhances urban traffic control, in-depth case studies	Traditional methods still in use, requires significant cloud infrastructure
Alkarim et al. (2024)	Conceptual ITS model using collective learning for vehicle flow prediction	93.52% improved performance, effective use of DL algorithms	Dependent on sensor accuracy, high initial setup cost

Current studies on intelligent traffic management systems have witnessed dramatic advancements through simulations, machine learning, and IoT platforms. The methods have despite this yielded improved traffic flow, improved air quality, and public safety. They are, however, numerous areas of research with open areas that remain. They rely most heavily on wide-scale, high-quality sensor networks, cloud, and computational capabilities and therefore have limitations in scalability and affordability for use by the masses. In addition, dependence on data quality and infrastructure conditions limits their use in less developed environments. Privacy concerns, especially with pedestrian behavior observation, and continuous focus on traditional traffic control methods also highlight areas for further investigation and development.

III. METHODOLOGY

The steps in the ML-based traffic prediction and management process, from data collecting to deployment are shown in figure.1.

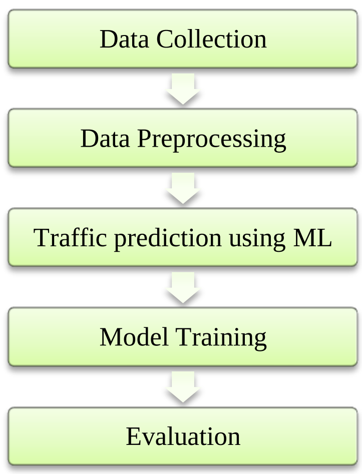

Figure 1 Overall workflow in traffic prediction

Data Collection: Traffic prediction and control rely on heterogeneous data from diverse sources to provide effective and informed decisions. For real-time sources of information, GPS data of cars and mobiles, which provide live traffic information, traffic cameras, and surveillance footage for tracking road conditions, IoT sensors placed on roads that keep track of traffic flow, speed, and lane utilization, and social media feeds that record incidents, weather, and other hazards. Historical data sources are useful in tracking traffic patterns over time, i.e., historical congestion reports, trends in public transport usage, and weather history that could have affected traffic. External data sources like road infrastructure data, i.e., traffic signals, intersections, and lane closures, are used to enhance traffic management measures. Calendar events like sports events, concerts, and public holidays are also taken into account as they have a direct influence on traffic flow and congestion levels. Combining these datasets allows machine learning algorithms to generate more accurate traffic forecasts and enhance urban mobility.

Preprocessing of data: Raw traffic data are not complete and it comes with noisy data. It requires strong preprocessing to improve their quality and value in training the ML models. Data cleaning wherein duplicates are removed, missing values are taken care of, and outliers are removed is the initial step.Following data cleaning is data normalization, wherein such key traffic parameters as speed, density, and volume are normalized to be homogeneous for utilization in model training. Feature engineering is then employed to create informative input features, including time features like rush hour and day of week, weather features like temperature, rain, and fog, and road features like the number of lanes or if a road is a highway or a local street. Finally, in the case of supervised learning models, the data is labeled, where low, medium, and high congestion levels are labeled based on past traffic rates, allowing the model to learn and predict congestion patterns more effectively. Proper preprocessing ensures more reliability and accuracy when it comes to traffic prediction and management systems.

Classification Model: A hybrid approach integrating RF and LSTM utilizes the strengths of both machine learning methods for precise traffic flow forecasting. Random Forest excels in modeling intricate relationships within organized traffic data by extracting prominent features such as speed, congestion rates, weather, and road occupation. It assists with the selection of the most influential variables, denoising, and overall interpretability of models. LSTM, however, is better suited to manage time-series dependencies and also long-term traffic trends through sequential pattern learning of past traffic from historical data. As traffic flow is largely dynamic and based on the past, LSTM's memory enables it well to be placed best for predicting congestion change and travel time change. By combining RF's feature selection property with LSTM's temporal learning capacity, the hybrid model enhances traffic forecasting models to be more dependable and accurate, thereby better applicable.

Random Forest: RF is a ML technique that produces a single outcome by combining the output of several DTs.. Its versatility and ease of use have encouraged its uptake because it is employed the solve regression and classification issues. An ensemble of DTs where each tree is trained using a particular random noise is called a RF. The most common type of DT ensemble is a RF. The most common outcome is identified by aggregating the predictions of a group of classifiers, such as DTs, which makes up ensemble learning techniques.

LSTM: LSTM, is a variant of recurrent neural network (RNN) that can process sequential data and gradually discover dependencies. LSTMs solve the vanishing gradient problem, which prevents typical RNNs from effectively retaining information over lengthy sequences, by employing unique gating methods that control the information flow. over applications like time-series prediction, language processing and speech recognition, these gates allow LSTMs to retain crucial information over extended periods of time.

IV. RESULTS AND DISCUSSION

The study uses the Traffic Flow Dataset on Kaggle, which includes a hourly traffic volume data that was gathered from sensors located in different places. Some of the key features in the dataset includes, timestamp, vehicle count per hour, weather conditions etc., The performance of the suggested hybrid method is evaluated using the evaluation metrics like accuracy, precision, recall and F1-score. Table 1 shows the performance analysis of the suggested method with other state of art methods.

Table 1 Comparison Analysis-Proposed Method

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Random Forest +LSTM	95	94	94	93
Random Forest (RF) [13]	90	88	89	88
K-Nearest Neighbors (KNN)[14]	84	81	79	80
Naïve Bayes (NB)[15]	78	74	72	73

It is clear from table 1 that RF + LSTM gives the best predictions because it has the highest accuracy at 95%. As for false positives (wrong congestion predictions), it is also the best in precision (94%). The high recall (94%) also shows that actual cases of congestion are dealt with very well. The best and most accurate model is the F1-score (93%) that balances precision and recall.

Although 90% accuracy with RF, it is less accurate than the hybrid RF + LSTM model. While RF by itself lacks LSTM's capability for sequential learning, precision and recall are top-notch. Although it might fail with time-series dependencies, it is perfectly fine for predicting average traffic flow.

While having an 84% accuracy rate, KNN is not as accurate as RF-based models. While it is poorer at discriminating among the many different levels of traffic congestion with an 81% precision rate and 79% recall rate, KNN struggles with massive high-dimensional datasets but does more proficiently with small traffic sets.

NB has most commonly misclassified traffic congestion, with worst-case accuracy at 78%. Its low precision and recall demonstrate it cannot determine patterns, possibly due to its assumption of feature independence. It is only suitable for use in simple data distribution systems that involve traffic prediction.

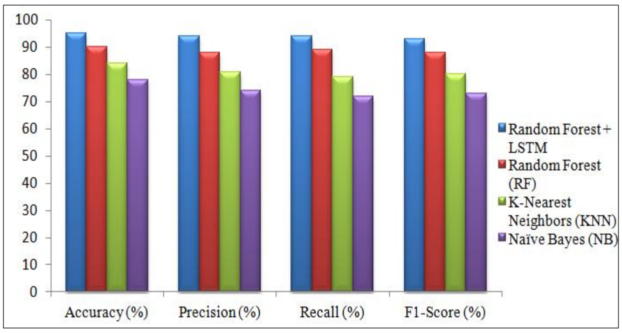

Figure 2 Performance Analysis-Proposed Method

The best results are achieved by Random Forest + LSTM for accuracy, precision, recall, and F1-score. It indicates that both spatial and temporal dependencies in traffic forecasting are effectively captured by the hybrid model. Although Random Forest performs well, it is not as good as the hybrid model. K-Nearest Neighbours performs averagely but lags behind RF-based models. Its distance-based approach likely makes it challenging to handle complex traffic patterns.

V. CONCLUSION

The findings of this study demonstrates that the RF + LSTM hybrid model best fits traffic prediction and is better than single models of RF, KNN, and NB. Harnessing the power of Random Forest in handling structured data and the power of LSTM in recognizing sequential dependency, the proposed hybrid model delivers the best accuracy and reliability in traffic prediction. These results verify the efficiency of ML-based ITS for traffic optimization, alleviation of traffic congestion, and urban mobility. The findings of this study contribute to the establishment of data-driven, AI-oriented traffic policy for smart cities as a foundation for futuristic and optimal transportation networks. Subsequent studies may also explore the integration of other deep learning architectures and real-time adaptive traffic management systems in a bid to predict and systems even more accurately and promptly.

REFERENCES

[1]. Boukerche, Azzedine, and Jiahao Wang. "Machine learning-based traffic prediction models for intelligent transportation systems." Computer Networks 181 (2020): 107530.

[2]. Almukhalfi, Hanan, Ayman Noor, and Talal H. Noor. "Traffic management approaches using machine learning and deep learning techniques: A survey." Engineering Applications of Artificial Intelligence 133 (2024): 108147.

[3]. Lilhore, Umesh Kumar, Agbotiname Lucky Imoize, Chun-Ta Li, Sarita Simaiya, Subhendu Kumar Pani, Nitin Goyal, Arun Kumar, and Cheng-Chi Lee. 2022. "Design and Implementation of an ML and IoT Based Adaptive Traffic-Management System for Smart Cities" *Sensors* 22, no. 8: 2908. https://doi.org/10.3390/s22082908

[4]. Berlotti, Mariaelena, Sarah Di Grande, and Salvatore Cavalieri. "Proposal of a machine learning approach for traffic flow prediction." Sensors 24, no. 7 (2024): 2348

[5]. Serag, R. H., Abdalzaher, M. S., Elsayed, H. A. E. A., Sobh, M., Krichen, M., & Salim, M. M. (2024). Machine-Learning-Based Traffic Classification in Software-Defined Networks. *Electronics*, *13*(6), 1108. https://doi.org/10.3390/electronics13061108.

[6]. Tiwari, Pulkit. "The machine learning framework for traffic management in smart cities." Management of Environmental Quality: An International Journal 35, no. 2 (2024): 445-462.

[7]. Kim, Yeeun, Hye-young Tak, Sunghoon Kim, and Hwasoo Yeo. "A hybrid approach of traffic simulation and machine learning techniques for enhancing real-time traffic prediction." Transportation research part C: emerging technologies 160 (2024): 104490

[8]. Sun, Yuan, and Jorge Ortiz. "Machine Learning-Driven Pedestrian Recognition and Behavior Prediction for Enhancing Public Safety in Smart Cities." Journal of Artificial Intelligence and Information 1 (2024): 51-57.

[9]. Yang, Peiyuan, Zhou Chen, Guangze Su, Han Lei, and Baoming Wang. "Enhancing traffic flow monitoring with machine learning integration on cloud data warehousing." Applied and Computational Engineering 77 (2024): 238-244.

[10]. Alkarim, Anas Saleh, Abdullah S. Al-Malaise Al-Ghamdi, and Mahmoud Ragab. "Ensemble Learning-based Algorithms for Traffic Flow Prediction in Smart Traffic Systems." Engineering, Technology & Applied Science Research 14, no. 2 (2024): 13090-13094.

[11]. An, Ru, Xiaochun Zhang, Maopeng Sun, and Gang Wang. "GC-YOLOv9: Innovative smart city traffic monitoring solution." Alexandria Engineering Journal 106 (2024): 277-287.

[12]. Rasulmukhamedov, Makhamadaziz, Timur Tashmetov, and Komoliddin Tashmetov. 2024. "Forecasting Traffic Flow Using Machine Learning Algorithms" *Engineering Proceedings* 70, no. 1: 14. https://doi.org/10.3390/engproc2024070014

[13]. Navarro-Espinoza, Alfonso, Oscar Roberto López-Bonilla, Enrique Efrén García-Guerrero, Esteban Tlelo-Cuautle, Didier López-Mancilla, Carlos Hernández-Mejía, and Everardo Inzunza-González. "Traffic flow prediction for smart traffic lights using machine learning algorithms." Technologies 10, no. 1 (2022): 5.

[14]. Mrudula, Sunkara Teena, Mahyudin Ritonga, S. Sivakumar, Malik Jawarneh, T. Keerthika, Kantilal Pitambar Rane, and Bhaskar Roy. "Internet of things and optimized knn based intelligent transportation system for traffic flow prediction in smart cities." Measurement: Sensors 35 (2024): 101297.

[15]. Manimaran, A. "Improving the accuracy of predicting congested traffic flow road transport using random forest algorithm and compared with the naives bayes algorithm using machine learning." In AIP Conference Proceedings, vol. 3193, no. 1. AIP Publishing, 2024.