World Journal of Advanced Engineering Technology and Sciences

eISSN: 2582-8266

Cross Ref DOI: 10.30574/wjaets

Journal homepage: https://wjaets.com/

(RESEARCH ARTICLE)

Application of deep learning models for traffic flow prediction based on time-series data

Bich Ngoc Nguyen Thi *

Faculty of Information Technology, Sao Do University, HaiDuong, Vietnam.

World Journal of Advanced Engineering Technology and Sciences, 2024, 13(01), 655–661

Publication history: Received on 17 August 2024; revised on 26 September 2024; accepted on 29 September 2024

Article DOI: https://doi.org/10.30574/wjaets.2024.13.1.0467

Abstract

This paper investigates and applies the Long Short-Term Memory (LSTM) deep learning model for traffic flow prediction based on time-series data. The model is trained and tested with a large dataset comprising 26,497 records of vehicle counts at a specific observation point over four months. To evaluate the performance of the LSTM model, we conduct experiments and compare it with other popular machine learning methods. The results demonstrate that the LSTM deep learning model achieves an accuracy of 88.91%, outperforming traditional machine learning techniques. These results promise to support traffic flow prediction and provide reliable data for managers, helping them make accurate decisions in traffic coordination, thereby reducing congestion and enhancing the efficiency of urban traffic systems.

Keywords: Traffic flow prediction; LSTM model; Machine Learning; Time-series Data

1. Introduction

Traffic prediction is a key factor in smart transportation, particularly important for urban infrastructure planning and management. Understanding and forecasting vehicle movement trends provides strategic insights, supporting the development of transportation infrastructure. Early and accurate predictions will help optimize traffic flow and minimize congestion while improving traffic safety and saving resources, contributing to the sustainable development of cities. Therefore, researching and developing traffic prediction methods is essential.

There have been many methods proposed for traffic prediction. In paper [1], to address traffic congestion in large cities, the authors propose using deep learning and machine learning models to predict time-series vehicle traffic, thereby adjusting the traffic light waiting times to optimize vehicle flow through intersections. The authors experimented with five different machine learning and deep learning models using data from Huawei Munich's research center and the Multilayer Perceptron Regressor model achieved the best performance with an EV value of 0.93. For the most accurate traffic prediction when affected by factors such as weather, accidents, etc., papers [2, 4] propose a model based on four methods: DAN (Deep Autoencoder), DBN (Deep Belief Network), RF (Random Forest) and LSTM (Long Short-Term Memory) to predict traffic conditions using data collected from traffic sensors. The authors of paper [3] propose a wide-range traffic prediction model called NN-ARIMA, combining the MLP network and the statistical time series model ARIMA, in which MLP captures the movement patterns of the entire traffic system and ARIMA analyzes specific local traffic conditions from the MLP results. The model has a MAPE rate of 11.11% using six months of traffic data from the Ayalon Highway, Tel Aviv, Israel. Combining meteorological data with traffic indicators, the authors of paper [5] propose a convolutional neural network model for predicting traffic conditions, tested with data from the coastal region of Barra, Costa Nova, Portugal. The conclusion shows that weather conditions greatly affect traffic within a short time frame (1 hour) and this data can be used in deep learning prediction models. With the fast-paced updates of social media platforms, such as accidents, blockages, or bad weather, the authors of paper [6] propose a deep learning model that

* Corresponding author: Bich Ngoc Nguyen Thi

predicts traffic conditions by integrating weather, traffic and social media data from Twitter/X. The model uses a Bi-directional LSTM combined with Stacked Autoencoder (SAE) architecture to perform multi-layer predictions across different datasets. When tested with data from Greater Manchester, UK, the model achieved MAE of 5.5049, RMSE of 6.8579 and sMAPE of 0.1743%. Paper [7] proposes a machine learning model for traffic prediction based on real-world data from LTE telecommunications towers, testing techniques such as Linear Regression, Gradient Boosting, Random Forest, Bootstrap Aggregation (Bagging), Huber Regression, Bayesian Regression and Support Vector Machines (SVM). Results show that the training speed with SVM was the fastest, while the best-performing technique in the paper was Gradient Boosting. The authors of paper [8] suggest a new direction for traffic prediction models, not only focusing on traffic flow prediction but also expanding to predict vehicle speed and the likelihood of accidents. To optimize the performance and accuracy of traffic condition prediction solutions, the authors of paper [9] propose a machine learning model combined with an improved Bayesian model based on real-world traffic data from Beijing. Their IBCM-DL model yielded promising prediction results.

In this paper, we first study the deep learning model LSTM and analyze the collected data. Next, the implementation and training of the model are performed on a large dataset. Finally, we conduct testing to evaluate the model's effectiveness. To affirm the superiority of LSTM, we compare the results with three popular machine learning models: SVR, KNN and Decision Tree. The experimental results show that the LSTM model achieves higher accuracy than traditional machine learning techniques, demonstrating the strong potential of LSTM in traffic prediction.

The next section presents the LSTM model. Section 3 covers the collected data used in the paper and the experimental results and finally, Section 4 provides the conclusion.

2. Material and method

2.1. Long Sort-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is an advanced variant of the Recurrent Neural Network (RNN), introduced by Hochreiter and Schmidhuber in 1997 to address the limitations of processing long and complex data sequences [10]. For the dataset comprising over 26,000 records in this study, LSTM is a suitable choice due to its superior ability to learn and model temporal dependencies. Traffic data contains repeating patterns by hour, day and week, requiring a model capable of efficiently storing and utilizing information from previous time points to predict future outcomes. With its unique mechanism, LSTM can retain important information over long periods while discarding irrelevant details, thereby improving the accuracy of predictions. In the collected data, temporal dependencies and cyclical patterns play a crucial role, making LSTM an ideal model due to its ability to store and analyze past information to forecast future trends. Figure 1 illustrates the architecture of the LSTM network.

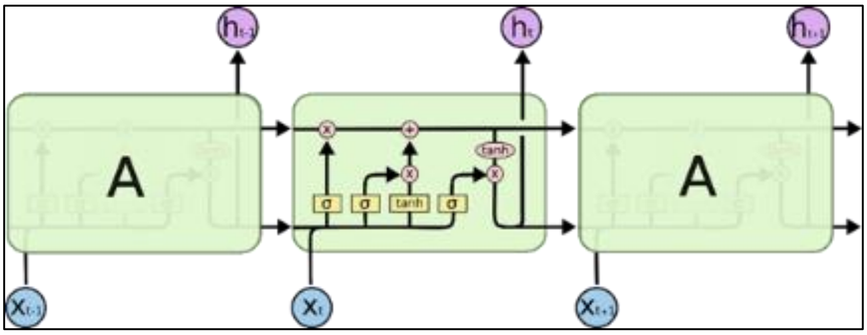

Figure 1 LSTM network structure

2.2. Dataset

In this paper, we use the Coplin dataset with 26,479 records representing the number of vehicles on the streets, updated every five minutes [11]. The data was collected continuously over four months, from 00:00 on October 4, 2015, to 23:55 on January 3, 2016, with two columns of information: the first column records the data collection time and the second column is the corresponding number of vehicles at that moment. Figure 2 presents the data distribution chart in the study, where the x-axis represents the time of the records and the y-axis represents the number of vehicles passing through the observation point.

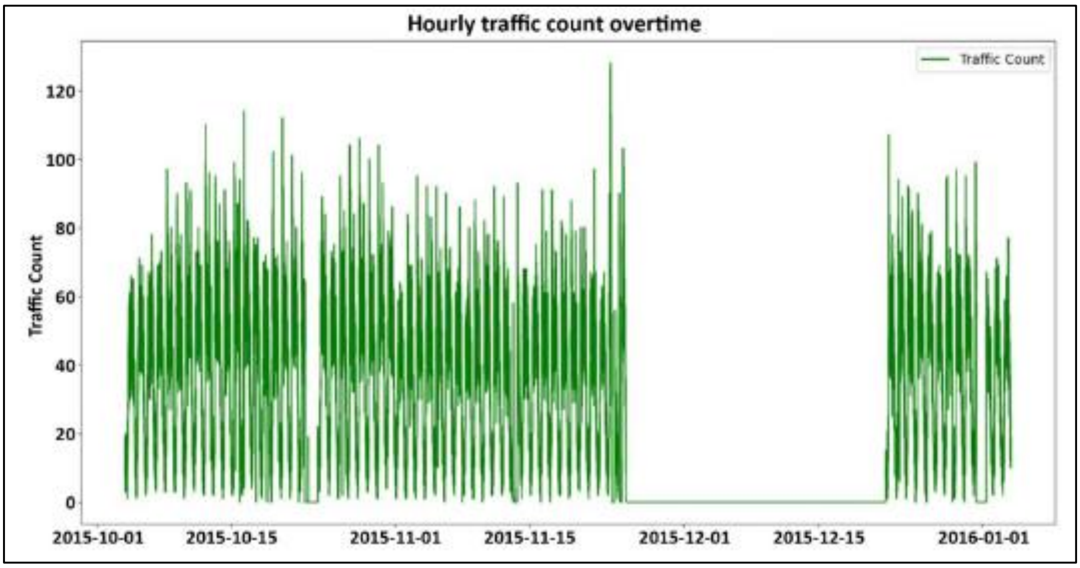

Figure 2 The distribution of data in the paper

The data was pre-processed and split into two parts: training and testing, with a ratio of 80:20. To remove test data with a vehicle count of zero, we limited the time range from December 27, 2015, to December 30, 2015. The remaining data was used for training. Figures 3 and 4 illustrate the distribution of the training and testing data. After splitting, the data was pre-processed using Min/Max scaling to normalize and ensure that all features are within the same range, from 0 to 1, thereby speeding up the training process and improving the model's learning ability.

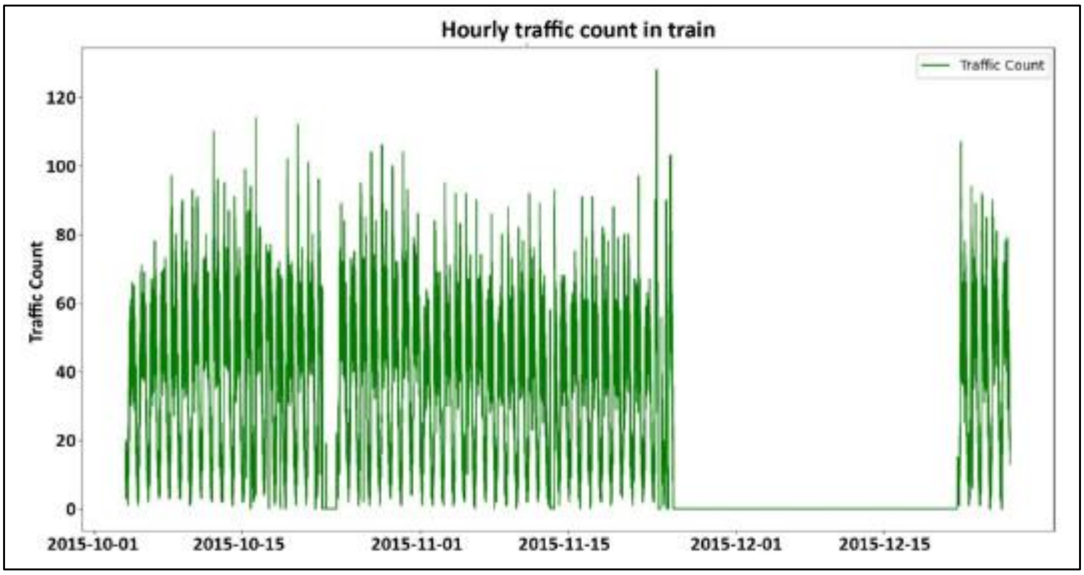

Figure 3 Training data distribution

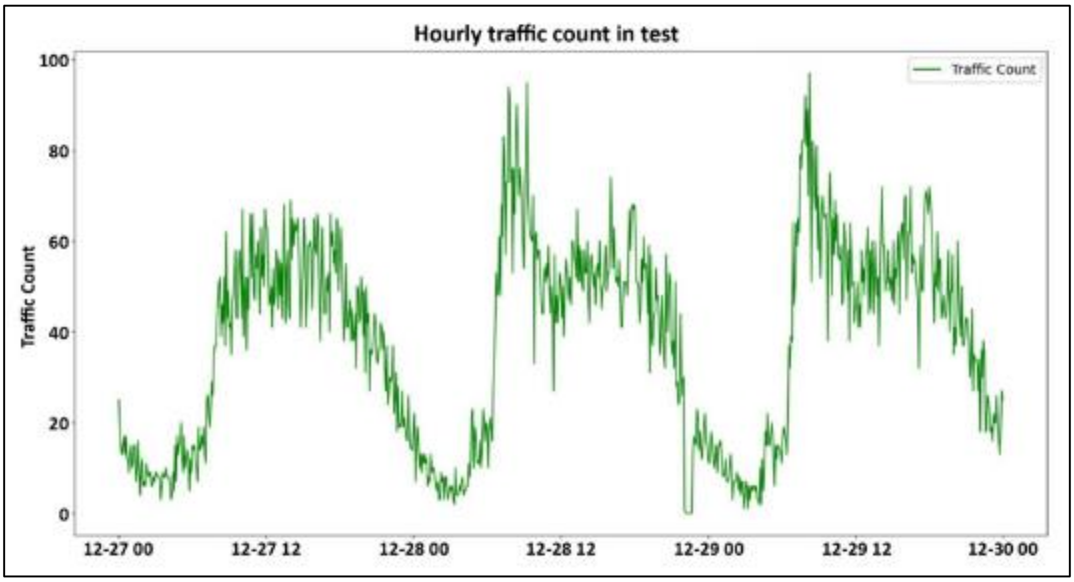

Figure 4 Testing data distribution

3. Experimental results

In this paper, we conducted experiments with the deep learning model LSTM and three machine learning models (SVR, KNN, Decision Tree) on a computer with the following configuration: CORE I7-10700 2.9GHz, 16GB RAM, Windows 10 OS, development environment Python 3.6 and parameters time_step = 24, train_epochs = 100, batch_size = 32 (for the LSTM model), C = 100, Gamma = 0.01 and Epsilon = 0.35 (for the SVR algorithm). After running the experiments, the traffic prediction performance of the four methods, LSTM, SVR, KNN and Decision Tree, is illustrated respectively in Figures 5, 6, 7 and 8. In these figures, the red line represents actual data, while the blue line shows the model's predictions. The smaller the difference between the two lines, the higher the model's accuracy.

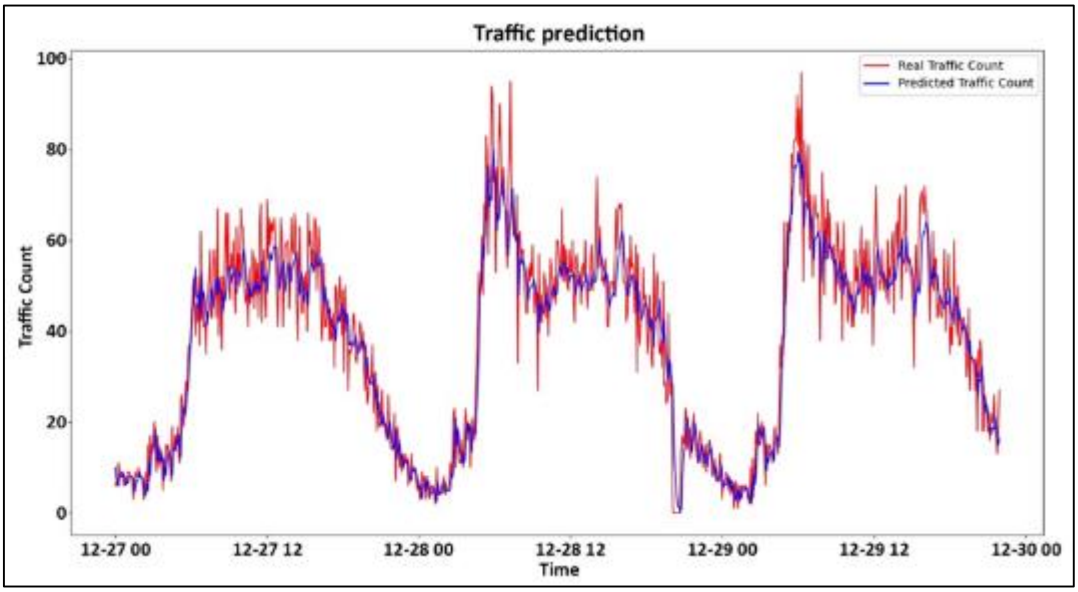

Figure 5 The prediction process of the LSTM model

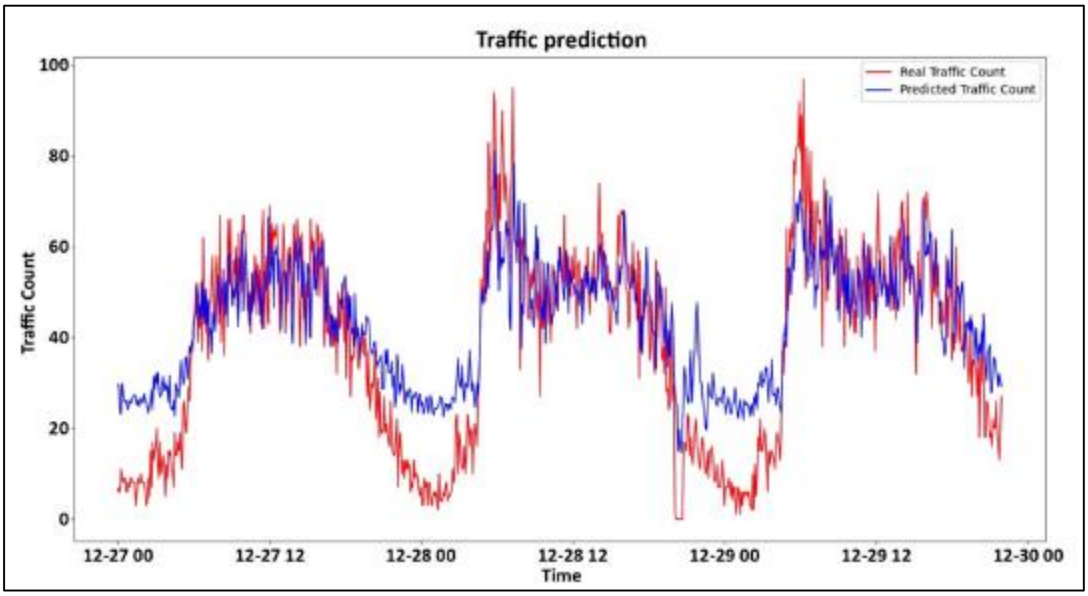

Figure 6 The prediction process of the SVR model

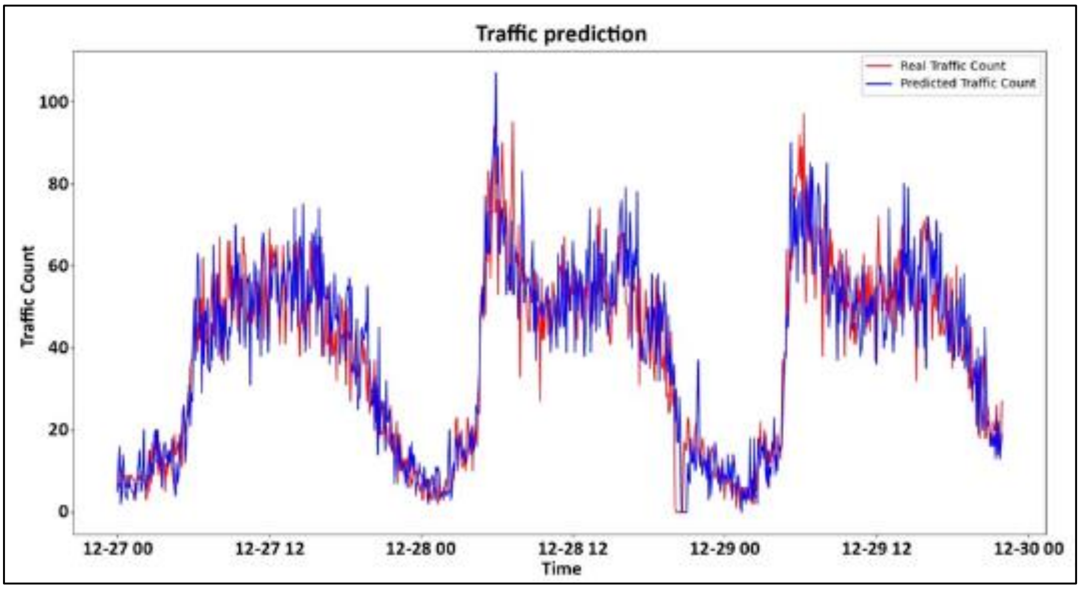

Figure 7 The prediction process of the KNN model

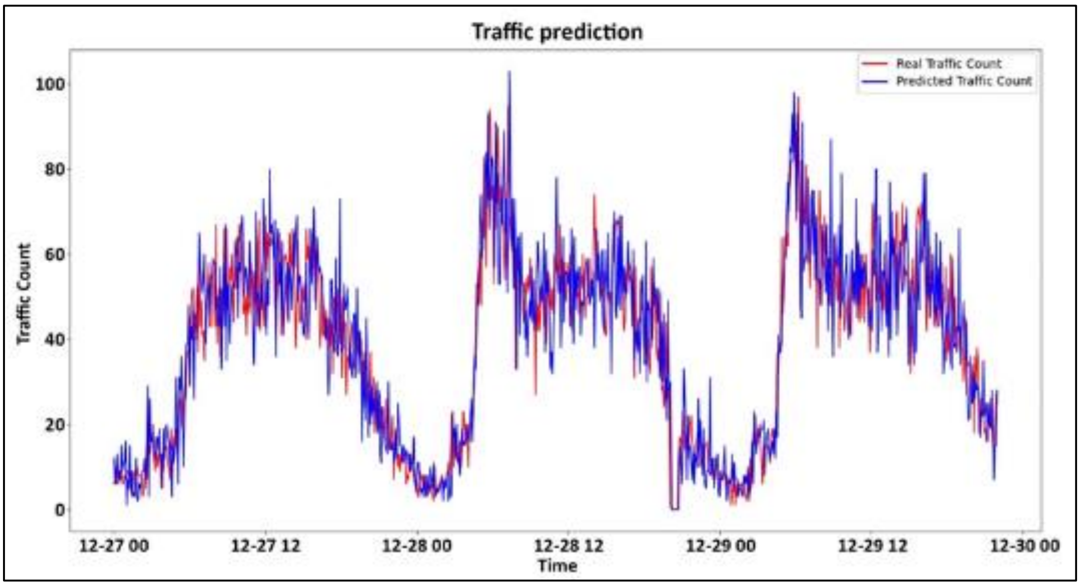

Figure 8 The prediction process of the Decision Tree model

For the LSTM model, the chart is shown in Figure 5, where the prediction line (blue) closely follows the actual data line (red) over an extended period, indicating that the LSTM model performs quite well. However, there are certain periods where the predicted line significantly diverges from the actual data, suggesting that the model needs improvement for specific scenarios. For the machine learning models, the charts are shown in Figures 6, 7 and 8. It can be observed that these methods show considerable deviations between the predicted and actual data, especially at high and low peaks, indicating that they may struggle to predict rapid fluctuations in the dataset.

To further clarify the effectiveness of the proposed method, we performed statistical comparisons of three values after testing the models: R^2, MSE and RMSE. Specifically, R^2 is a coefficient used to evaluate the accuracy of a regression model, representing the percentage of total variance in the predicted values compared to the actual values, calculated according to formula (1). The closer the R^2 value is to 1, the higher the model's accuracy. Mean Squared Error (MSE) is a statistical index used to measure the average error between predicted and actual values in a regression model. MSE evaluates the performance of a forecasting model by averaging the squared errors between predicted and actual values, calculated by formula (2). A lower MSE value indicates a more accurate prediction model, with predicted values closer to actual ones. RMSE is the square root of Mean Squared Error (MSE), representing the average of MSE, calculated by formula (3).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(yi - \hat{y}i)^2}{\sum_{i=1}^{n}(yi - \bar{y})^2} \ldots\ldots\ldots\ldots\ldots (1)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(yi - \hat{y}i)^2 \ldots\ldots\ldots\ldots\ldots\ldots..(2)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(yi - \hat{y}i)^2} \ldots\ldots\ldots\ldots\ldots\ldots.(3)$$

In the above formulas, yi is the actual value at the ith data point, ŷi is the predicted value of the model at the ith data point and ȳ is the average of the actual values yi. The R^2, MSE and RMSE values of the four models are compared in Table 1.

Accordingly, the LSTM deep learning model achieved the highest R^2 at 88.91%, significantly higher than traditional machine learning methods. The error rate between predicted and actual values was also much lower than the machine learning models, demonstrating the effectiveness of the LSTM deep learning model in predicting traffic based on time-series data.

Table 1 Compare the accuracy of models

STT	Model	R^2	MSE	RMSE
1	LSTM	88.91%	53.19	7.29
2	SVR	64.52%	170.19	13.04
3	KNN	77.01%	110.26	10.5
4	Decision Tree	76.36%	113.31	10.64

4. Conclusion

Traffic prediction is becoming increasingly essential in the context of rapid urbanization today. Understanding traffic models and analyzing traffic data can provide valuable insights for planning, infrastructure development and congestion management. This paper has explored and applied the LSTM deep learning model to predict the number of vehicles at a specific location based on time-series data. The experimental results show that the LSTM model achieved an accuracy of 88.91%, outperforming traditional machine learning methods such as SVR, KNN and Decision Tree. In the future, we will continue to research and develop new methods to improve the model's accuracy and apply the research results in building applications aimed at advancing smart traffic systems and urban planning in Vietnam.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Navarro-Espinoza, Alfonso, et al, Traffic Flow Prediction for Smart Traffic Lights Using Machine Learning Algorithms, *Technologies*, 10, 1, ISSN 2227-7080, 2022.

[2] Nazirkar Reshma Ramchandra, C. Rajabhushanam, Machine learning algorithms performance evaluation in traffic flow prediction, *Materials Today: Proceedings*, 51, 1, 1046-1050, ISSN 2214-7853, 2022.

[3] Tao Ma, Constantinos Antoniou, Tomer Toledo, Hybrid machine learning algorithm and statistical time series model for network-wide traffic forecast, *Transportation Research Part C: Emerging, Technologies*, 111, 352-372, ISSN 0968-090X, 2020.

[4] Azzedine Boukerche, Jiahao Wang, Machine Learning-based traffic prediction models for Intelligent Transportation Systems, *Computer Networks*, 181, 107530, ISSN 1389-1286, 2020.

[5] Braz, Fernando José, et al, Road Traffic Forecast Based on Meteorological Information through Deep Learning Methods, *Sensors 22*, 12: 4485, 2022.

[6] Essien, A., Petrounias, et al, A deep-learning model for urban traffic flow prediction with traffic events mined from Twitter, *World Wide Web 24*, 1345–1368, 2021.

[7] D. Alekseeva, N. Stepanov, et al, Comparison of Machine Learning Techniques Applied to Traffic Prediction of Real Wireless Network, *IEEE Access*, 9, 159495-159514, 2021.

[8] Z. Liu, Z. Li, K. Wu, M. Li, Urban Traffic Prediction from Mobility Data Using Deep Learning, *IEEE Network*, 32, 4, 40-46, 2018.

[9] Y. Gu, W. Lu, et al, An Improved Bayesian Combination Model for Short-Term Traffic Prediction With Deep Learning, *IEEE Transactions on Intelligent Transportation Systems*, 21, 3, 1332-1342, 2020.

[10] Hochreiter, S., & Schmidhuber, Long short-term memory, *Neural Computation, 9*(8), 1735-1780, 1997.

[11] Coplin, DataSet(Traffic flow), 2018.