Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606

Volume 08 Issue 01
2024

Research Article
https://doi.org/10.56979/801/2024

Real-Time Traffic Flow Prediction using IoT-Driven Machine Learning

Syeda Sitara Waseem[1], Hina Sattar[1*], Shabana Ramzan[1], Nimra Nasir[1], and Manahil Khan[2]

[1]Department of Computer Science & IT, The Government Sadiq College Women University, Bahawalpur, Pakistan.
[2]Department of Computer Science & IT, The Islamia University of Bahawalpur, Pakistan.
[*]Corresponding Author: Hina Sattar. Email: hinasattar@gscwu.edu.pk

Received: April 11, 2024 **Accepted:** September 19, 2024

Abstract: Road accidents result in deaths, infections, and many injuries. The main causes of these accidents are traffic congestion, road blockages, and traffic anomalies. Several factors, such as busy routes, damaged roads, or incidents, can trigger traffic. Traffic and roadblocks primarily cause time wastage, energy consumption, delays in reaching destinations, and accidents. Tracking traffic patterns may be a solution to this problem. However, IoT has been successful in a wide range of applications, such as healthcare and inventory management; the downside of tracking traffic is that it is difficult to manage. Therefore, to address this issue, we will design and develop a traffic flow forecasting system using an IoT framework and machine learning. To train and test this system, we will use the ANFIS model, which will then integrated into an Android application. This framework will identify the traffic patterns with the help of IoT sensors in real time, taking into account specific origins, times, peak hours, and speeds. It will then display these patterns in a graphical user interface, enabling users to understand traffic flow pattern and select the most efficient route to reach their destination on time.

Keywords: Traffic Flow; Machine Learning; IoT Prediction; Traffic Pattern; IoT Sensors; Smart Traffic.

1. Introduction

Traffic is the movement of vehicles and people on roads and streets. It can range from smooth flow to traffic and is affected by various factors like time, rainfall, and structure. Increase in population is also one of the reasons for traffic. There are several reasons like reasons for traffic blocks including inadequate road capacity, incidents, and weather conditions [6]. Basically traffic or road blocks are the reasons for road accident, wasting time, wasting energy, and reaching late at destination and many more. IoT-enabled vehicles and infrastructure continuously generate and transmit data related to traffic conditions, vehicle movements, and road incidents. This real-time data collection facilitates the detection of traffic patterns, contributing to more accurate traffic prediction models [5].Utilizing IoT technologies to collect real-time traffic data from various sensors and devices. This data is then analyzed to detect patterns in traffic flow, congestion, and other relevant parameters [7].

Previous research in traffic prediction using IoT and machine learning presents numerous challenges. Data gaps and inconsistencies between different sources, such as sensors and cameras, affect model prediction accuracy. Furthermore, managing large-scale datasets and making real-time predictions is difficult for models. One of the model's main problems is adjusting to dynamic traffic conditions and unexpected events. The lack of interpretability makes it difficult to trust the predictions made by applications. The privacy of data collected from IoT devices is another challenge. To increase the efficiency and reliability of traffic flow prediction systems, it is important to address these research gaps.

In our research, we discovered that the earlier work has various flaws, such as:

1. Reliance on air pollution data alone for traffic prediction [1] diminishes the impact of other factors such as roadblocks or road construction.
2. The use of UAV video for real-time date prediction has limitations due to its restricted coverage [4].
3. While using univariate data for traffic prediction, IoT-enabled smart cities may overlook the interactions and correlations between multiple variables that influence traffic dynamics [2].

With the help of previous works, we will design and build traffic flow forecasting using the IoT framework and machine learning. According to our study, the following factors need further research, which includes using a dataset that include various factors of roads like time, speed, density, etc. Predict peak hours of a specific place where the system is present at that time to avoid the usage of the road at a specific time of particular origin. This study will address research gap by specifically considering peak hours based on traffic origin, which most previous studies neglected. Including this aspect aims to enhance the accuracy and reliability of predicting the traffic pattern. This framework will use IoT sensors to detect the traffic pattern of a specific origin, including time, peak hours, and speed, and then display this pattern through a graphical user interface. The traffic pattern will then assist the user in comprehending the flow and choosing the most efficient route to arrive on time.

2. Related Work

In this section, we will discuss the work of various researchers that have already worked on traffic flow prediction. The research scope involves developing an IoT and machine learning-based system to detect traffic patterns with a specific focus on peak hours of specific origin. The literature review will encompass an analysis of existing traffic flow prediction methods, including traditional approaches, weather forecasting approaches, deep learning algorithms, and machine learning algorithms like CNN and SVM. It will also explore IoT applications in real-time data analysis. Furthermore, the review will discuss challenges to traffic prediction, such as variations in weather and the need for large, diverse datasets for training machine learning models. Overall, the literature review aims to provide a comprehensive understanding of the current state of research and identify opportunities for advancements in traffic flow prediction using technology-driven approaches, especially considering IoT and machine learning-based.

The authors Khan et al. (2022) [1] argue that the support of air pollution data through machine learning models can make traffic prediction accurate, thus using it in a bagging ensemble technique. However, their study faces challenges related to data unavailability and Hyperparameters sensitivity. Chahal et al. (2023 propose a hybrid model that fuses classical time series analysis with machine learning to non-intrusively predict traffic congestion in the smart city. However, one of the hindrances to this model is the limited nature of the technology era to huge datasets and rapid changes in vehicular traffic. Wang et al. (2023) introduce a zero-trust-based model for traffic prediction that ensures maximum privacy by sensor data while still providing high accuracy through the proper configurations. In my opinion, the model became possible because of new privacy-aware approaches based on appropriate security measures. However, the practical adoption of this model remains a major stumbling block. In their article, Ke et al. (2018) have used the UAV video data and ensemble classifiers in a real-time traffic analysis to conduct the tests. Using these techniques offered the possibility of achieving better results; however, this work found that one of the required improvements was the quality of the available material. Ejaz and Anpalagan (2019) [5] delve into IoT's influence on smart cities, concentrating on data management and network security. The authors' review, while rich in insights, may become outdated due to the rapid development of IoT technology and the absence of contemporary practical implementation aspects. Besides, Kashyap et al. (2022) [6] do a thorough review on deep learning techniques, namely, CNNs, RNNs, LSTMs, and hybrid models, for traffic flow prediction. Their study certainly enhances the fascination of deep learning by thoroughly reviewing the current state and limitations of various devices. The study may overlook some unpublished studies that could potentially contain important findings, as it does not conduct an in-depth comparison of the approaches and results obtained with different models. Moreover, the quality of the reviewed articles may constrain the validity of their conclusions. Further research in these areas is essential, focusing on enhancing data handling capabilities, ensuring privacy, and developing adaptable traffic prediction models.

Table 1. Related Work

Ref	Technique	Limitations	Future Work
[1]	KNN ensemble bagging method	air pollution data accuracy	Increasing the accuracy of the model
[2]	OWENN (Optimized Weight Elman Neural Network) Back-Propagation Neural Networks (BPNNs)	Limited Feature Consideration	Focus on external factors
[3]	K-nearest neighbor(KNN) algorithm	Privacy issues Focused on prediction	Improving privacy-preserving methods Focus on real time data
[4]	Convolutional neural network (CNN)	Low accuracy of UAV	Enhancing the accuracy of UAV estimation
[5]	CNN	Security concerns	Enhance IoT security
[6]	RNN Hybrid network	Long term dependencies Accurate data needed	Addressing long term dependencies Handle Missing data

3. Research Methodology

The proposed traffic flow prediction system (TFPS) is efficiently designed using IoT and Machine learning approach. The architecture of proposed TFPS is depicted in Figure 1.

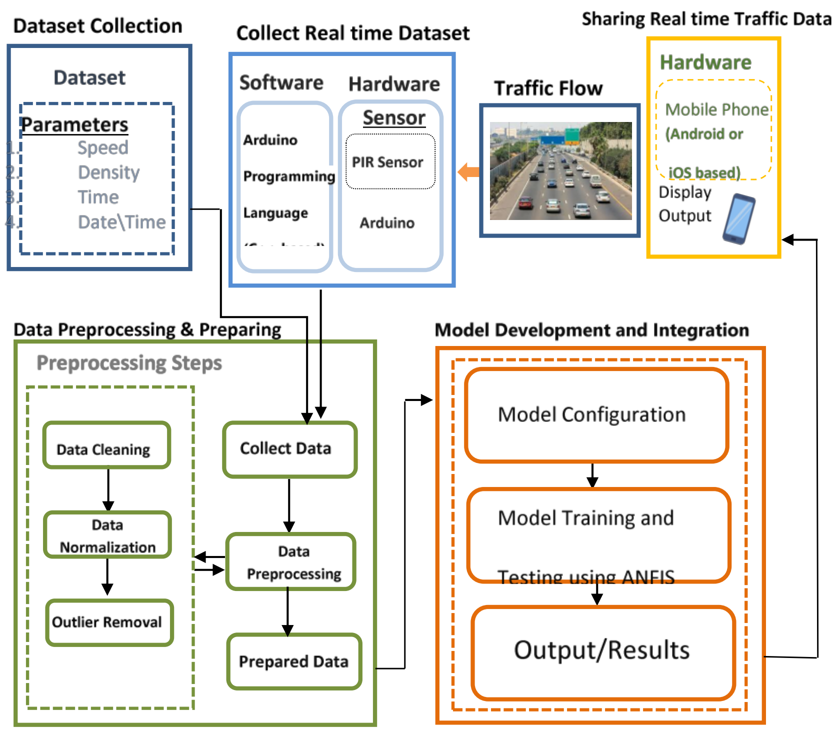

Figure 1. Architecture of TFS

a. Input Module

Input modules in this framework are Arduino UNO-R3, and PIR Motion Sensor. IoT circuit for TFS framework includes Arduino and PIR Motion Sensor. IoT circuit is in figure 2.

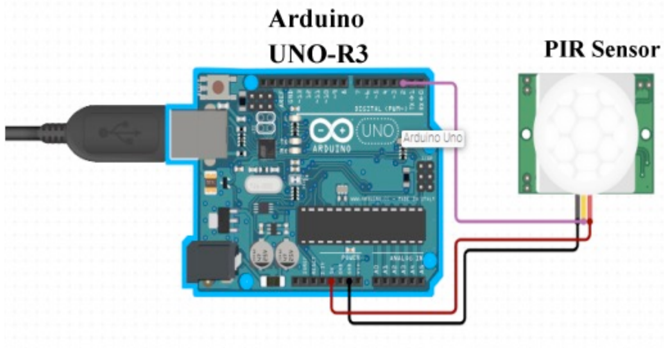

Figure 2. Circuit diagram

b. Dataset

Dataset is taken from [10] and further added some parameters from different databases of traffic. Dataset contains 1501 entries and 8 attributes. Dataset description is stated in Table 2.

Table 2. Dataset Description

Serial No	Parameters	Description	Values
1	Date	Date of the month	Numeric Values
2	Day	Name of the day	String
3	Travel Time	In Hours and minutes	Numeric Values
4	Traffic Situation	In normal, low, heavy	String
5	Time gap	In seconds	Numeric Values
6	Motion Detection	Yes or No	Detected OR Not
7	Speed	Meter/second	Numeric Values
8	Density	Cars per day	Numeric Values
9	Peak Hours	Specific time of day	Numeric Values

c. Data Preprocessing

At this stage, data processing was undertaken to convert the raw data into a usable format for the model, ensuring the relationship between input and output variables was correctly captured. The raw data had issues like outliers and missing values. We used the PANDAS library to identify the column with missing values, index where the data was missing, and then replaced those missing values with either the column's mean or median. To address outliers, we calculated Z-scores to measure each data point's distance from the average. Points with a Z-score greater than 3, were identified using the SciPy library and then removed.

d. Model Configuration

In this framework, ANFIS is used with parameters which are date, day, no of vehicles, speed, peak hours and density while Hyperparameters is traffic situation [13] .ANFIS stands for Adaptive Neuro Fuzzy Inference System. It is kind of artificial neural networks. ANFIS is Neuro-Fuzzy system that uses 5-layer network with supervised learning.

- Input layer: Input layer takes the input and pass it to the next layer.

$$O_{i1} = x_i \tag{1}$$

- Fuzzification layer: Membership value between 0 and 1 represents the assigned fuzzy set with membership function. Membership function determines its size and shape.

$\mu A(x) = \exp(-\frac{(x-c)^2}{2\sigma^2})$ (2)

- Rule based Layer: Membership values from fuzzification layer are combined using fuzzy logic operators (AND or OR) which means product or minimum value.

For an AND operator:

$wi = \mu Ai(x1) \times \mu Bi(x2)$ (3)

For an OR operator:

$wi = max(\mu Ai(x1), \mu Bi(x2))$ (4)

- Defuzzification Layer: In this layer weighted calculation is performed.

$Oi, 5 = w\bar{}i \times fi(x1, x2, \dots, xn)$ (5)

- Output Layer: It sums up all the inputs coming from defuzzification layer into final output.

$Ofinal = i\sum Oi, 5 = i\sum w\bar{}i \times fi(x1, x2, \dots, xn)$ (6)

- Training & Testing: The training dataset is used for training ANFIS model. Model's output is compared with the target output for each entry [12]. Then parameters are adjusted to get expected output. Iteration is done again and again on training and updating parameters multiple times until the criteria met. After training ANFIS model, testing is performed. In testing, trained model is tested on tested dataset to get output. Then model's outputs are compared with ground truth labels of test dataset and find difference between output and truth labels. Then test is analyzed to assess ANFIS model performance. Lower difference represent the better performance while higher represent bad performance. In this testing lower difference is found.
- Integration into Android: After training and testing ANFIS with TensorFlow layer, TensorFlow Lite Converter is used to convert ANFIS model into TensorFlow Lite Format. After integrating, android app is tested multiple times to ensure that model is working correctly in application.
- Prediction: After integrating ANFIS model into Android application, firstly user will enter destination in application, app will collect data of particular origin then application will show three routes to destination. After selecting on of the route, application will predict traffic as high, low, medium and also speed and density of specific route. User can view traffic situation of every route and select any which has low traffic.

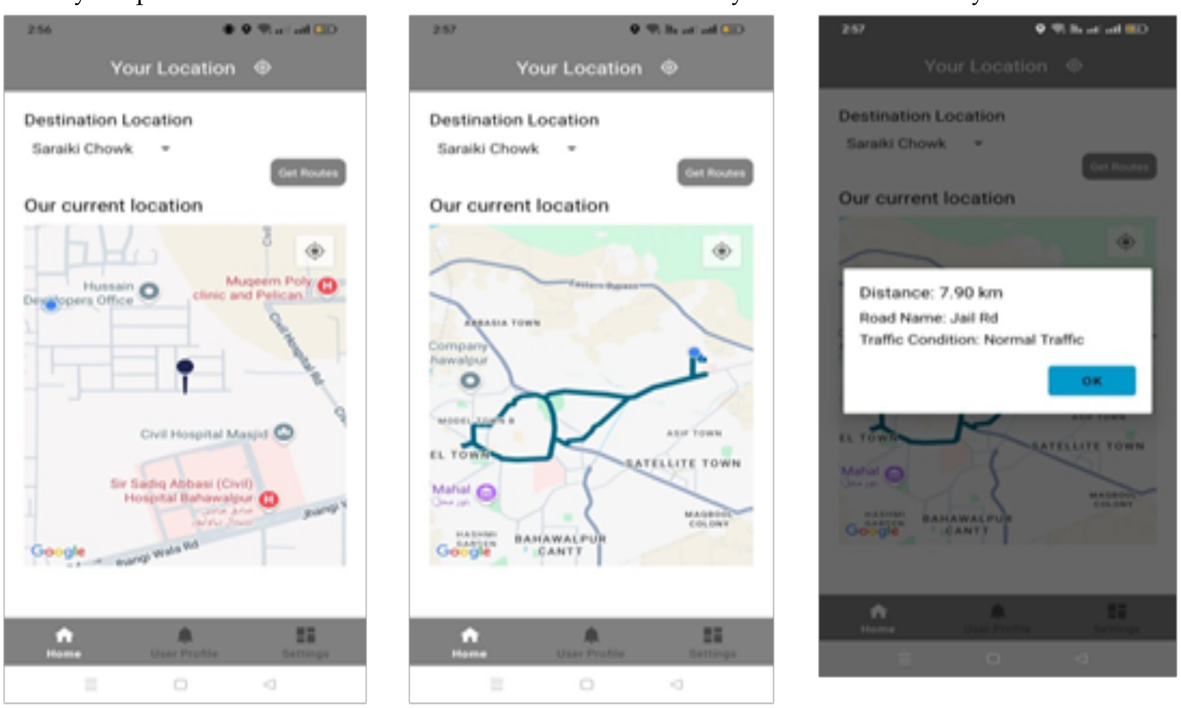

Figure 3. Prediction of application

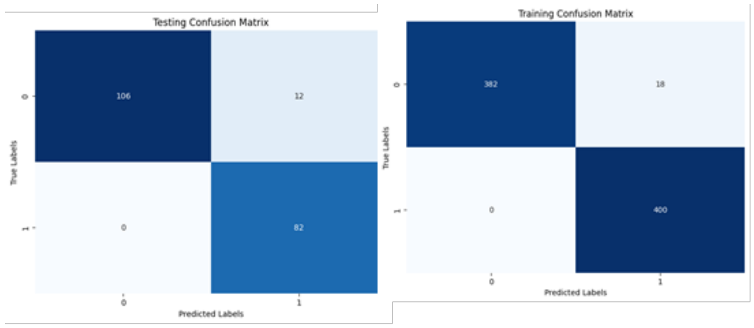

Figure 4. Confusion metrics of training & Testing

4. Performance of Model Training

Performance of the model's training is measured with the help of accuracy metrics and loss metrics. Accuracy metrics provides the number of correct predictions of the model and loss metrics provides the difference in the actual value and the predicted value. Confusion metrics provides a visual representation of model's performance. These provide a critical insight of how well the model is trained and perform on new data. The confusion metrics of training and testing is shown in Figure 4.

The accuracy metrics of training and validation are shown in the Figure 5.

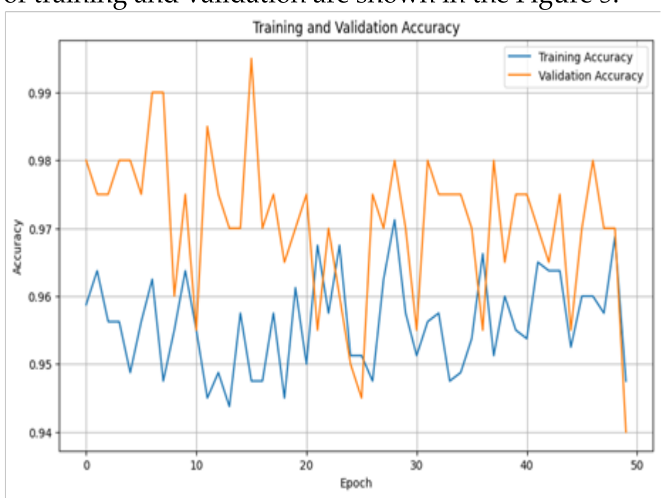

Figure 5. Accuracy metrics

The loss metrics of training and validation are shown in the Figure 6.

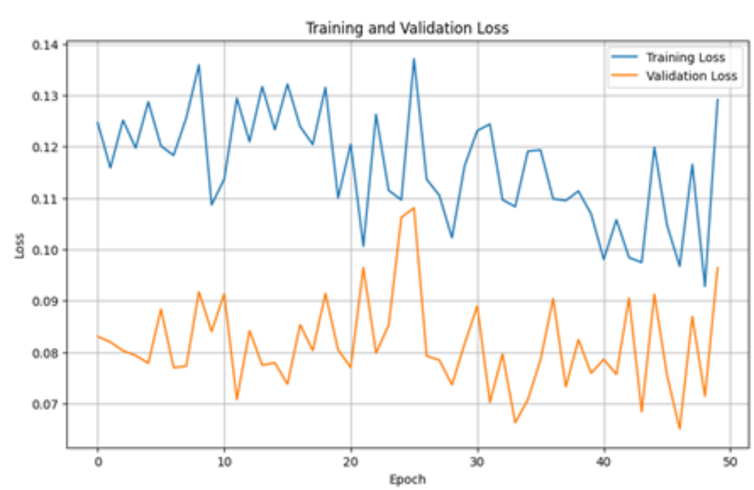

Figure 6. Loss metrics

5. Experiments

In this section of experimental setup, data is collected from different routes in different weeks of July 2024 comprises of required attributes like speed density, peak hours etc. Then extensive experimentation was done to observe the performance of setup. The aim of these experiments was to observe working of proposed system. To get perfect results, evaluation metrics are calculated during each experiment and then compared to one another. A detail of experimentation is listed here in Table 3.

Table 3. Data collection for experiments

No of routes	Instances/ route	Day	Instance Collected	Time Span
13	10	Day 1	130	15 minutes
17	10	Day 2	170	20 minutes
16	5	Day 3	80	30 minutes
20	5	Day 4	100	20 minutes
14	10	Day 5	120	15 minutes
10	9	Day 6	90	30 minutes
15	10	Day 7	150	25 minutes
14	10	Day 8	140	20 minutes
15	5	Day 9	125	25 minutes
20	10	Day 10	200	10 minutes

We have conducted 10 experiments on various routes during different time periods in the evening and morning to assess the accuracy of the sensors and determine whether environmental factors are influencing them. The experimental setup utilizes all the collected data to predict traffic flow. The prediction demonstrated the setup's performance accuracy. This setup showed 96-98% accurate traffic flow with the given data. Table 4 elaborates the performance of the proposed system.

TP stands for True Positive, signifying correctly identified traffic instances. True Negative (TN) represents correctly identified traffic instances that are not present. FP stands for False Positive, which indicates an incorrect identification of a condition that actually belongs to another condition. FN, or false negative, occurs when the system fails to predict a condition that is not present. Table 4 shows all calculated values.

Table 4. Evaluation matrices

Day/Instances	TP	TN	FP	FN	Precision	Recall	Accuracy
Day 1/130	89	32	05	04	94%	95%	93%
Day2/170	100	55	10	05	90%	95%	91%
Day3/80	55	17	04	04	93%	93%	90%
Day4/100	60	30	04	06	93%	90%	90%
Day5/120	76	24	10	10	88%	88%	83%
Day6/90	59	21	05	05	92%	92%	88%
Day7/150	110	25	05	10	95%	91%	90%
Day8/140	86	34	10	10	89%	89%	85%
Day9/125	98	20	02	05	98%	95%	94%
Day10/200	125	50	10	15	92%	89%	87%

In Table 4 True Positives (TP) represents the number of instances that predicted the correct traffic condition in this setup. False Negatives (FN) represents the number of instances that predicted incorrect traffic conditions. False positives (FP) represent the number of instances where a specific traffic condition belongs to another condition. True Negatives (TN) represents the number of instances that predict correctly but not the specific condition. We also calculate the precision recall and accuracy of these instances.

6. Results

The series of experiments are performed to find out the accuracy and performance of the model. After analyzing all the outcomes of the experiments, the graph is generated, which is shown below Figure **7.** The overall accuracy of this model is 96%.

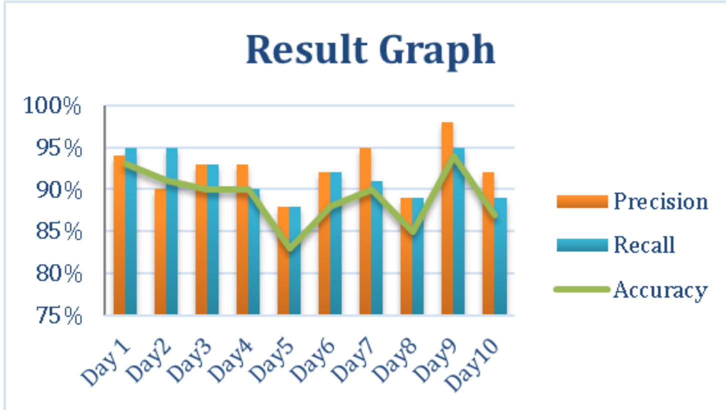

Figure 7. Result Graph

7. Limitations and Future Work

Model performance and prediction are good, but there are some limitations. The models' performance can change based on the features chosen and the dataset used. Also, due to budget issues, the study uses cheap sensors. Problems like poor data, missing values, and outliers can affect how well the models predict. Finally, the model functions optimally under the same training and testing data but not for unseen ones.

There are several areas where future research could improve traffic prediction models.

- First, trying more advanced machine learning and deep learning methods could find hidden patterns and make predictions more accurate.
- Using an inductive loop sensor enables the real-time collection of accurate data and additional factors.
- Combining real-time data with advanced processing techniques like stream processing and online learning could create models that adapt and change dynamically.

8. Conclusions

The study looked at how well machine learning models can predict the number of vehicles using data from Internet of Things (IoT) devices. Among the models tested, the Adaptive Neuro-Fuzzy Inference System (ANFIS) was the best. It made fewer mistakes and was more accurate than the other models. This shows that ANFIS can understand the complex patterns in traffic data, which helps in making better traffic predictions. Combining IoT and machine learning gives us real-time data and a flexible way to predict traffic, making it a useful tool for transportation planning and management.

Despite the positive results, the study highlights some issues that require correction in future research. For instance, the limited location and time of the used data could potentially impact the generalizability of the results. Furthermore, the ANFIS model is difficult to use in real time due to its high computational power requirement. In future research, studies should focus on large and more diverse data sets, use more advanced machine learning algorithms to handle large datasets, and increase accuracy in results. By handling these challenges, future scientists will be able to construct traffic prediction models that are even more accurate, dependable, and applicable in real-life endeavors. The creation of these models will promote traffic management, congestion reduction, modern urban planning, and infrastructure development.

To wrap up, the study demonstrates that using IoT data with machine learning, it is possible to make better traffic predictions. The ANFIS model performed exceptionally well, providing compelling evidence that advanced methods can enhance efficiency. The outcomes are quite hopeful, but there is certainly room for carrying out deeper probing and researching. More research in the field can lead to the creation of a more accurate and efficient traffic prediction method, which in turn will make the transportation systems and city planning more responsive and efficient.

References

1. Khan, N. U., Shah, M. A., Maple, C., Ahmed, E., & Asghar, N. (2022). Traffic flow prediction: an intelligent scheme for forecasting traffic flow using air pollution data in smart cities with bagging ensemble. Sustainability, 14(7), 4164.
2. Chahal, A., Gulia, P., Gill, N. S., & Priyadarshini, I. (2023). A Hybrid Univariate Traffic Congestion Prediction Model for IoT-Enabled Smart City. Information, 14(5), 268.
3. Wang, F., Li, G., Wang, Y., Rafique, W., Khosravi, M. R., Liu, G., ... & Qi, L. (2023). Privacy-aware traffic flow prediction ba sed on multi-party sensor data with zero trust in smart city. ACM Transactions on Internet Technology, 23(3), 1-19.
4. Ke, R., Li, Z., Tang, J., Pan, Z., & Wang, Y. (2018). Real-time traffic flow parameter estimation from UAV video based on ensemble classifier and optical flow. IEEE Transactions on Intelligent Transportation Systems, 20(1), 54-64
5. Ejaz, W., & Anpalagan, A. (2019). Internet of things for smart cities: technologies, big data and security (pp. 1-15). Berlin/Heidelberg, Germany: Springer International Publishing.
6. [Kashyap, A.A.; Raviraj, S.; Devarakonda, A.; Nayak, K.S.R.; Kv, S.; Bhat, S.J. Traffic flow prediction models—A review of deep learning techniques. Cogent Eng. 2022, 9, 2010510.
7. Sarrab, M.; Pulparambil, S.; Awadalla, M. Development of an IoT based real-time traffic monitoring system for city governance. Glob. Transit. 2020, 2, 230–245
8. Feng, S., Ke, R., Wang, X., Zhang, Y., & Li, L. (2017). Traffic flow data compression considering burst components. IET Intelligent Transport Systems, 11(9), 572-580.
9. Lv, Y., Duan, Y., Kang, W., Li, Z., & Wang, F. Y. (2014). Traffic flow prediction with big data: A deep learning approach. IEEE Transactions on Intelligent Transportation Systems, 16(2), 865-873.
10. https://www.kaggle.com/datasets/hasibullahaman/traffic-prediction-dataset
11. Bao-ping, C., & Zeng-Qiang, M. (2009, February). Short-term traffic flow prediction based on ANFIS. In 2009 International Conference on Communication Software and Networks (pp. 791-793). IEEE.
12. Olayode, I. O., Severino, A., Tartibu, L. K., Arena, F., & Cakici, Z. (2021). Performance Evaluation of a Hybrid PSO Enhanced ANFIS Model in Prediction of Traffic Flow of Vehicles on Freeways: Traffic Data Evidence from South Africa. Infrastructures, 7(1), 2.
13. Codur, M. Y., Atalay, A., & Unal, A. (2017). Performance evaluation of the ANN and ANFIS models in urban traffic noise prediction. Fresenius Environ Bull, 26, 4254-4260.
14. Ren, Chuanxiang, Chai, Chunxu, Yin, Changchang, Ji, Haowei, Cheng, Xuezhen, Gao, Ge, Zhang, Heng, Short-Term Traffic Flow Prediction: A Method of Combined Deep Learnings, Journal of Advanced Transportation, 2021, 9928073, 15 pages, 2021. https://doi.org/10.1155/2021/9928073
15. Razali, N. A. M., Shamsaimon, N., Ishak, K. K., Ramli, S., Amran, M. F. M., & Sukardi, S. (2021). Gap, techniques and evaluation: traffic flow prediction using machine learning and deep learning. Journal of Big Data, 8, 1-25.
16. Li, C., & Xu, P. (2021). Application on traffic flow prediction of machine learning in intelligent transportation. Neural Computing and Applications, 33(2), 613-624.
17. Miglani, A., & Kumar, N. (2019). Deep learning models for traffic flow prediction in autonomous vehicles: A review, solutions, and challenges. Vehicular Communications, 20, 100184.
18. Atassi, R., & Sharma, A. (2023). Intelligent traffic management using IoT and machine learning. Journal of Intelligent Systems and Internet of Things, 8(2), 08-19.
19. Wu, Y., Tan, H., Qin, L., Ran, B., & Jiang, Z. (2018). A hybrid deep learning based traffic flow prediction method and its understanding. Transportation Research Part C: Emerging Technologies, 90, 166-180.
20. Sayed, S. A., Abdel-Hamid, Y., & Hefny, H. A. (2023). Artificial intelligence-based traffic flow prediction: a comprehensive review. Journal of Electrical Systems and Information Technology, 10(1), 13.
21. Khedkar, S. P., Canessane, R. A., & Najafi, M. L. (2021). Prediction of traffic generated by IoT devices using statistical learning time series algorithms. Wireless Communications and Mobile Computing, 2021(1), 5366222.
22. Bilotta, S., Collini, E., Nesi, P., & Pantaleo, G. (2022). Short-term prediction of city traffic flow via convolutional deep learning. IEEE Access, 10, 113086-113099.
23. Kaluvana, H. (2022). Intelligent transportation system and smart traffic flow with IOT. Indian Journal of Radio & Space Physics (IJRSP), 50(2), 64-67.
24. Zantalis, F., Koulouras, G., Karabetsos, S., & Kandris, D. (2019). A review of machine learning and IoT in smart

transportation. Future Internet, 11(4), 94.

25. Kumar, B. P., Hariharan, K., Shanmugam, R., Shriram, S., & Sridhar, J. (2022). Enabling internet of things in road traffic forecasting with deep learning models. Journal of Intelligent & Fuzzy Systems, 43(5), 6265-6276.