International Journal on Engineering Artificial Intelligence Management, Decision Support, and Policies
Vol. 01, Iss. 02, S. No. 003, pp. 18-32, Oct 2024
ISSN (E):3048-8788

Real-Time Traffic Congestion Prediction Using Predictive Data Analysis

Ambarish G. Mohapatra

Electronics Engineering, Silicon University, Bhubaneswar, Odisha, India: Email: ambarish.mohapatra@gmail.com

How to cite this paper: Ambarish G. Mohapatra"Real-Time Traffic Congestion Prediction Using Predictive Data Analysis ," *International Journal on Engineering Artificial Intelligence Management, Decision Support, and Policies,* Vol. no. 1, Iss. No 2, S No. 003, pp. 18–32, October 2024.
DOI Link of paper

Received: 10/09/2024
Revised: 25/09/2024
Accepted: 05/10/2024
Published: 30/10/2024

Abstract

Urban traffic congestion is a growing concern worldwide, affecting travel time, fuel consumption, and environmental pollution. Optimal prediction and control of traffic density have become relevant issues for the city's navigation and maintenance of concrete infrastructures. The paper looks at how it is possible to predict congestion in real-time traffic data of the urban traffic systems. The methodology involves processing data from the network of sensors in real-time and past traffic data to create a predictive model that traffic-managing systems can use. A comparison of two models is then made, the time–series predictive model and a hybrid regression model in their accuracy of estimation speed of computation, and responsiveness in real-time. Applying sophisticated mathematical formulas for the simulation of traffic flow and congestion this study can be used to show how the integration of technologies and mathematics can improve traffic situation and decrease congestion in cities. These findings are expressed in comparative plots and quantitative measures where all the predictions from the hybrid model compare favourably to the basic time-series model, despite having fewer data points. The proposed implementation relies on real-life traffic data, and the effectiveness of the predictive models is assessed by comprehensive error measures such as mean absolute error (MAE) or root mean square error (RMSE). Consequently, the findings of this study advance the design of intelligent traffic management systems for reducing traffic density to enhance the quality of life in urban areas

Keywords

Traffic Congestion, Predictive Analysis, Real-Time Traffic Data, Traffic Flow Prediction, Intelligent Traffic Management, Hybrid Regression Model, Time-Series Model, Congestion Forecasting

1. Introduction

Traffic congestion is one of the major challenges faced by the world’s developing cities. Rapid urbanization, increased vehicle ownership, and relative underdevelopment of transport systems have led to severe congestion problems, particularly in densely populated urban towns and cities. Traffic congestion impacts not only individual commuters but also the broader economy by increasing fuel consumption, air pollution, and the overall cost of transportation. These problems have thus become major concerns in planning among city planners and other transport authorities globally [1].

The system solely depends on static signal timings and manual traffic monitoring for traffic control, which is no longer adequate in the modern world due to dynamic traffic congestion. The expansion of the IoT-enabled sensors and employing real-time data acquisition and processing along with pioneering approaches to data analyses presents a potential solution to conventional approaches to traffic control and management. With the use of real-time data from perhaps traffic cameras, GPS units, or road sensors, predictive models can accurately predict the likelihood of traffic and congestion so that pre-emptive action can be taken [2, 3].

Recent years have seen a boom in the use of data from traffic flow to perform predictive analysis because such analysis gives accurate and timely information on the possible emergence of traffic congestion. Predictive models leverage both historical traffic data and real-time sensor inputs to forecast future traffic conditions. These models could be used for adjusting the timing of traffic signals, redirecting vehicle movements to those areas less congested, and giving drivers real-time traffic information using mobile applications [4]. The combination of traffic management with predictive analysis enables optimization of moving time, fuel consumption, and reducing emissions, which will make the transport system more efficient [5].

Different predictive methods were introduced for traffic congestion such as time series, regression analysis, machine learning, and the combination of those models. This is particularly due to time series analysis, which is adopted widely in the traffic prediction system due to the way it handles temporal relationships in traffic data. However, time series models have their shortcomings in that they need historical data to build the model and are slow to adapt to the current fast-changing traffic flow patterns. On the other hand, left regression models that employ both real-time data from the car's sensors and traffic patterns that were hitherto observed have been proven to provide real-time better traffic congestion estimates [6].

The most recent studies have, therefore, been made on those models that use the better of the two in giving far more accurate traffic predictions. These models often combine a machine-learning model with a more 'conventional' time series or regression model to model the dynamic and unsteady nature of the urban traffic systems. Concerning weather conditions, road accidents, and special events, they can be accommodated in the hybrid model while their modelling in the traditional sense is almost impossible [7]. Hybrid models are elaborated by integrating single models and predictors with the help of various data sources resulting in more accurate and reliable traffic forecasts than applying only a single model [8].

The analysis of traffic congestion in this paper employs a predictive analysis framework based on real-time traffic information. The framework is based on two predictive models: The first type of forecasting model is a time-series forecasting model, and the second type is a regression model with time-series data. The time series model uses historical traffic patents to estimate the likely incidences of traffic congestion in the future, while the hybrid model employs, in addition to real-time data obtained from sensors, regression techniques to enhance the accuracy of the forecasts. The two models are used and tested on real traffic data sets and their performance is evaluated using metrics like mean absolute error (MAE) and root mean square error (RMSE) [9].

The goal of this study is to develop an efficient and accurate traffic congestion prediction system that can be incorporated into existing traffic management systems. The proposed system can thus assist the transportation authorities in determining appropriate times for changing signal lights, finding the best routes to avoid congested areas, and appropriating the right timetables for their transport services [10].

2. Related Research

Traffic flow prediction has been a popular area of study in Intelligent Transportation Systems (ITS) for the last few years. Studies conducted over many years have used both simplistic statistical models and advanced artificial neural network models, which can predict traffic congestion in a given place using raw feed inputs [11]. In this section, we present the most important ones with particular emphasis on the time-series models, regression techniques, and combined models in traffic prediction.

Since time dependency is inherent in traffic data, time-series analysis has earlier been applied in traffic prediction. Statistical models including time series models like auto regression integrated moving average (ARIMA) and seasonal auto regression integrated moving average (SARIMA) have been tested in the field of traffic flow and congestion prediction by contemplating past traffic records. These models are especially useful in cases in which the traffic movement displays certain structures, such as daily peaks, or yearly changes [12]. Nonetheless, they can be relatively restricted, especially employing historical domain information, thus they may not be sensitive enough to respond to drastic changes in the traffic flow situation like accidents or roadwork [13].

Regression models, on the other hand, give a more flexible method to traffic prediction by integrating real-time sensor data and external factors, like weather conditions and road incidents, into the predictive model. Multiple linear regression (MLR) and polynomial regression have been utilized to predict traffic congestion by analyzing the relationships between traffic flow, speed, and other variables [14]. The later models can perform better than time-series models especially if traffic conditions are affected by several factors.

New approaches for traffic congestion prediction in this context are hybrid models using trends from the time series analysis and regression methods. It is generally distinguished from other methods and combines machine learning methods such as support vector machines (SVM) or neural networks with other time or regression methods to enhance the accuracy of traffic forecasts [15]. The integration of data sources and more than one predictive method in hybrid models helps to factor in the stochastic and constantly developing character of the traffic systems of large cities, which results in generating more accurate and precise congestion forecasts than through single-model approaches.

Several works have shown that it is possible to get accurate traffic predictions using hybrid models. For example, the study by Lu et al. [16] developed a model using ARIMA integrated with a neural network for traffic congestion forecasts on urban highways. Comparing the performance of both models for the prediction of the next day's demand, it was established that the proposed hybrid model was more accurate than the ARIMA or the neural network model. As with the case with Zhang et al., [17] adopted the MLR model combined with SVR as the hybrid model for traffic flow forecasting on arterial roads. It demonstrated that the proposed hybrid model was effective for capturing traffic's nonlinear patterns and outperformed either the MLR or the SVR model when forecasting congestion.

Apart from time series and regression approaches, most studies on traffic congestion prediction have used machine-learning algorithms. For instance, Decision trees, Random forests, and Deep Learning models among others can be used to learn to develop rules from Traffic data & make predictions without necessarily developing overtime mathematical models [18]. These models have remained useful mostly in dealing with massive traffic datasets as well as these models in enhancing the real-time prediction of traffic jams. However, machine-learning models often need large amounts of training data and may be computationally costly to integrate into real-world traffic management systems.

However, there are some limitations to overcome while dealing with traffic congestion using predictive models. Some of the issues when applying the methodology include comprehending how various inputs like traffic sensors, GPS data, and even weather information can be incorporated into one predictive model. As a possible solution to this problem, the authors have discussed data fusion techniques as methods for combining data from multiple sources to generate a higher level of traffic information [19]. Another issue is the flexible reusability of the models with the time intervals when the adaptable predictions are required. Time series and regression models rely on past event occurrences, and while they are suitable for giving accurate estimates, especially in the short term, they may be stuck in providing the same estimate if an incident occurs that shifts the traffic pattern [20].

The contribution of this research is to supplement the literature with an introduction to the development of a new hybrid cumulative traffic congestion prediction model that employs time series analysis with regression analysis to enhance the accuracy of congestion predictions in real-time traffic environments. First, the proposed model is applied and assessed on real traffic data sets and then compared to the outcome of the basic time series algorithm. The results show that they are successful in predicting traffic congestion in urban areas and the proposed hybrid model can support the development of intelligent traffic management systems [21].

3. Problem Statement & Research Objectives

Traffic congestion is a significant task in modern urban environments, leading to delays, raised fuel consumption, and enhanced emissions. Recent traffic management systems rely severely on reactive measures, addressing traffic congestion only after it has happened. There is a requirement for predictive models that can estimate congestion events in real time, allowing for proactive traffic control actions. The main objectives of this proposed technique are

- To develop a predictive model for traffic congestion using real-time sensor data.
- To compare the performance of time-series and hybrid regression models for traffic congestion prediction.
- To evaluate the accuracy, computational efficiency, and adaptability of both models using real-time sensor traffic data.

4. *Methodology*

The approach towards analyzing traffic density in predicting congestion utilizes real-time traffic data and model predictions of traffic density. This study involves the development of two predictive models for traffic congestion: a time-series model and a hybrid regression model. Both models take updated traffic data and their differences in terms of their accuracy, computational complexity, and flexibility are investigated. It involves several key steps: pre-processing, feature engineering, modeling, and testing of the model. Its design is envisaged to incorporate current sensor information, past traffic flow trends, and sophisticated algorithms to analyze and forecast traffic intensity and density.

4.1 Data Acquisition:

The data for this research was obtained from urban traffic sensors, which included vehicle counts, average speeds, and occupancy rates. The data was taken at one-minute intervals and the data sets were accumulated for 7 days for model training and testing.

The IoT sensors fixed on roads gather precise data regarding the number of vehicles, speed, and density in real time to support traffic monitoring. These sensors collect traffic information to make precise estimations of traffic conditions and traffic rhythms throughout the day. In this way, constant communication with the centralized systems by IoT sensors enables efficient analysis of the situation, which in turn enables timely corrections to the traffic handling methodologies. The obtained data also serve for traffic predictions to provide city planners useful information about traffic signals, existing bottlenecks, and overall traffic flows in urban conditions.

4.2 Data Pre-processing

Data pre-processing is a critical step in any machine learning or predictive analysis task, especially when working with real-time traffic data. The real-time traffic data must be pre-processed to ensure its quality and consistency for accurate prediction [22]. Pre-processing includes the following steps:

4.2.1 Missing Data Handling

Real-time traffic data collected using sensors can thus be lost due to sensor malfunction or disruption of the signal transmission. A common method for handling missing data (x, y) is linear interpolation, where the missing values are assessed by interpolating between the nearest available data points (x_1,y_1) and (x_2,y_2). The equation for Linear Interpolation is given by Eq. (1)

$$y = y_1 + \frac{(y_2 - y_1)}{(x_2 - x_1)} \times (x - x_1)$$

(1)

Where, x_1, and x_2 are the indices of known points, y_1, and y_2 are the corresponding traffic values; x and y are the index of the missing data point. Table 1 includes sample data, which includes raw traffic data (vehicle count, average speed, and occupancy rate) collected at a road junction in a city before and after pre-processing.

Table 1: Sample raw traffic data collected before and after pre-processing

Time (s)	Collected Actual Data	Pre-processed Data

	Vehicle Count	Avg Speed (km/h)	Occupancy Rate (%)	Vehicle Count	Avg Speed (km/h)	Occupancy Rate (%)
0	20	62	NaN	20	62	79
5	25	NaN	85	25	56.5	85
10	NaN	51	91	30 (Interpolated)	51	91
15	35	48	88	35	48	89
20	40	52	87	40	52	87
25	NaN	53	NaN	50 (Interpolated)	53	90
30	55	NaN	93	55	54	93

The number of vehicles, average speed, and occupancy rate (Percentage of time that a road segment is occupied by vehicles) are documented at different time intervals in a day in Table 1. The raw traffic data is transformed into a format suitable for predictive modeling by handling missing data and extracting meaningful features.

4.2.2 Noise Reduction

Noise in traffic information can result from faulty sensors, communication errors, or unusual traffic flow. To decrease noise, a moving average filter can be applied. This filter smoothed out short-term fluctuations by averaging the neighboring data points.

The equation for a simple moving average (SMA) is given in Eq. (2):

$$\mathrm{SMA_t} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i=0}}^{\mathrm{n-1}} \mathrm{x_{t-i}}$$

(2)

Where $\mathrm{SMA_t}$ is the moving average at time t, x_{t-i} is the traffic data points, and n is the window size (number of data points to average).

4.2.3 Data Normalization

Measures derived from traffic data can be different and represent different units of measurement (e.g. vehicle count, average speed). Normalization resizes the values in the given attributes to the range 0 – 1, which enhances the efficiency of machine learning models.

The formula for min-max normalization (Eq. (3)) is:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

(3)

Where x is the original data value, x′ is the normalized value; min(x) and max(x) are the minimum and maximum values in the dataset. Table 2 shows the processed and normalized data of the raw data received from the sensors.

Table 2: Processed and Normalized data after removal of noise and data normalization

Pre-processed Data (Vehicle Count)	Smoothed Data after Noise Reduction	Normalized Data	Avg Speed (km/h)	Smoothed Data after Noise Reduction	Normalized Data	Occupancy Rate (%)	Smoothed Data after Noise Reduction	Normalized Data
20	22.5000	0	62	59.25	1	79 (Interpolated)	82	0
25	25.0000	0.08	56.5 (Interpolated)	56.50	0.69	85	85	0.32
30 (Interpolated)	30.0000	0.25	51	51.83	0.17	91	88.33	0.67
35	35.0000	0.42	48	50.33	0	89	89	0.74
40	41.6667	0.64	52	51	0.07	87	88.67	0.70

50 (Interpolated)	48.3333	0.86	53	53	0.30	90 (Interpolated)	90	0.84
55	52.5000	1	54 (Interpolated)	53.5	0.36	93	91.5	1

4.2.4 Feature Extraction

It encompasses choosing relevant features from the processed data as given in Table 3, which can be further used for predictive analysis. In the context of traffic data, beneficial features include Vehicle Count, Average Speed, and Occupancy Rate**.**

Table 3: Feature Extraction of Processed Sample Data

Processed Data			Extracted Features		
Vehicle Count	Average Speed (km/h)	Occupancy Rate (%)	Vehicle Count	Average Speed (km/h)	Occupancy Rate (%)
20	62	79 (Interpolated)	20	62.0000	79
25	56.5 (Interpolated)	85	25	56.5000	85
30 (Interpolated)	51	91	30	51.0000	91
35	48	89	35	48.0000	89
40	52	87	40	52.0000	87
50 (Interpolated)	53	90 (Interpolated)	50	53.0000	90
55	54 (Interpolated)	93	55	54.0000	93

These processes as mentioned above help in the kind of preparation of the raw traffic data which is fit for a predictive model, handling of noise, data gaps, and feature extraction. Thus, these pre-processed and normalized points of sample data could be passed on to predictive agents for further processing.

4.3 Predictive Models:

4.3.1 Model 1: Time-Series Predictive Model

The first model is a time-series predictive model that uses historical traffic data to forecast future congestion. The main technique used here is an Auto-Regressive Integrated Moving Average (ARIMA) model expressed by Eq. (4), which indicates trends and seasonality in traffic patterns [23].

$$y(t) = \propto + \beta_1 y(t-1) + \beta_2 y(t-2) + \cdots + \beta_p y(t-p) + \epsilon(t) \tag{4}$$

where $y(t)$ is the predicted traffic congestion at time t, α is a constant, β_i are the model coefficients, and $\epsilon(t)$ is the error term. The parameters of the ARIMA model were calculated using the Akaike Information Criterion (AIC), and the model was trained using the 7 days of the dataset.

4.3.2 Model 2: Hybrid Regression Model

The second model is a hybrid regression model, which incorporates time-series forecasting with regression-based methods. This hybrid approach uses additional variables like weather information, road conditions, and special event information to improve predictions. A linear regression model is combined with the time-series model to provide a more dynamic and adaptable forecast.

The hybrid regression model uses a multiple linear regression (MLR) method with real-time sensor data [24]. Eq. (5) formulated the MLR model as:

$$y(t) = \gamma_0 + \gamma_1 x_1(t) + \gamma_2 x_2(t) + \cdots + \gamma_n x_n(t) + \epsilon_t \tag{5}$$

where $y(t)$ represents the predicted traffic congestion, $x_n(t)$ are the real-time sensor variables (such as vehicle speed and traffic density), and γ_i is the regression coefficients. The hybrid model also integrates temporal dependencies by including lagged traffic data as additional predictors.

4.4 Performance Metrics
The performance of both models was assessed using standard error metrics, including mean absolute error (MAE) and root mean square error (RMSE). These metrics are defined by Eq. (6) and Eq. (7) as below:

$$\mathrm{MAE} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}\left|Y_i - \hat{Y}_i\right| \quad (6)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}\left(Y_i - \hat{Y}_i\right)^2} \quad (7)$$

where n is the number of data points (e.g., time steps), Y_i is the actual traffic congestion, and $\hat{Y}_i$ is the predicted congestion.

5. *Results & Discussion*

The following section provides a detailed explanation of the simulation results of the predictive models considered in this paper. To determine the accuracy of both of these models, the test data set included one week of traffic data that was not included in the training of the models.

5.1 Data collection:
Vehicle speed, vehicle count, and occupancy rate are accumulated from IoT sensors in real-time for 7 consecutive days. Fig. 1, 2, and 3 illustrate traffic data over time, showing variations in these three key variables: vehicle speed depicted in red, vehicle count in blue, and occupancy rate in green, plotted at intervals across the monitoring period.

As shown in Fig.1 the trend of vehicle traffic for 7 days or 168 hours fluctuates. The vehicle count recorded for the vehicles varies somewhere from 40 to 80, and the fluctuating nature of the volume may point to any of the rush hours or any of the off-peak. These trends might be weekly periodical traffic patterns, as the vehicle count in the study country is high during the morning and evening rush hours. This data is made available to determine traffic densities to manage traffic flows.

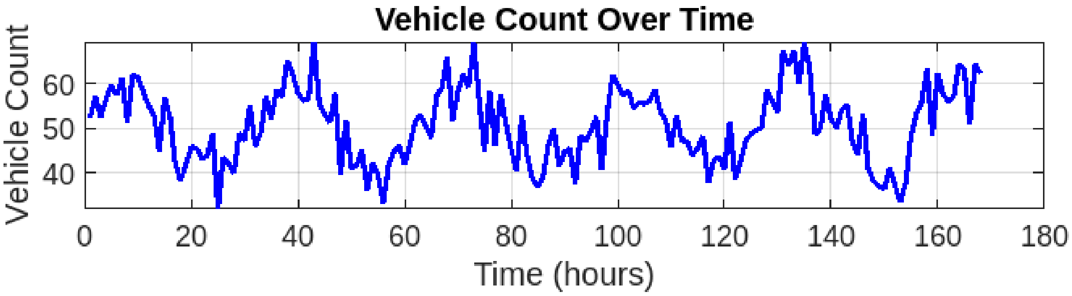

Fig. 1 Vehicle count over 7 days of time

This is so depicted in Fig. 2 below, which brings out the speed of the vehicle over 168 hours or 7 days. They range between 50 and 70 km/h, with moderate oscillations, which reflect different traffic conditions in the day. It is possible that these patterns could mean that the flow rate is high, and vehicle velocity is low during the congested phase followed by a low flow rate and high vehicle velocity during the free flow phase. The graph is informative when it comes to analyzing the traffic situation, especially when it comes to congestion during particular hours of the day and congestion-free hours.

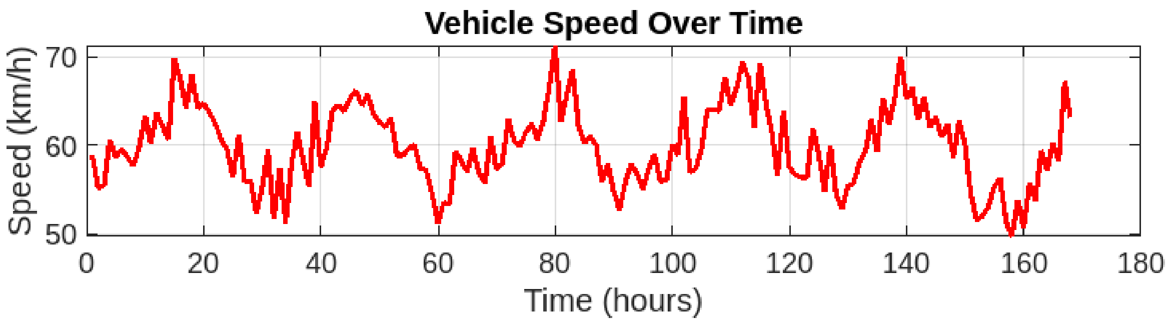

Fig. 2 Vehicle speed over 7 days of time

Fig.3 shows the percentage of the road occupancy over 7 days (168 hours). The occupancy rate varies from 60 to 90% with spikes demonstrating the period in which congestion on the road is present or high usage statistics are observed. The dips could be indicating times that not many people stayed on the road maybe during the odd times of the day when traffic might be expected to be low. The cyclic nature of the graph could show daily traffic patterns, with higher occupancy during rush hours and decreased occupancy during quieter times of the day. This data is vital for understanding traffic density and optimizing traffic congestion management.

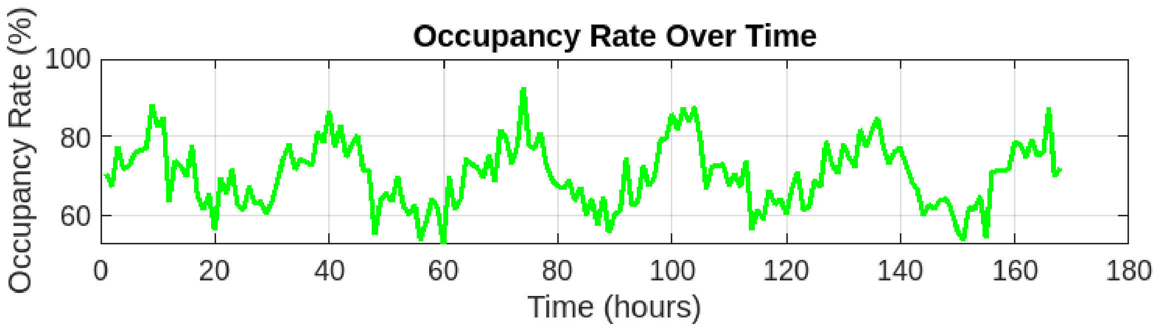

Fig. 3 Occupancy rate over 7 days of time

Clear variations are seen in the daily average changes in vehicle count and occupancy rate, compared to the relatively smoother variation in vehicle speed throughout the period. These trends can be used for decision-making in other systems, such as optimizing traffic signal timing and managing congestion control in real time. After gathering the above real-time data, data is preprocessed to confirm its quality and consistency for accurate prediction.

5.2 Rolling Average of Vehicle Count:

Fig.4 is the Rolling Average of the Vehicle Count within 168 hours or almost one week time. The actual vehicle count at each time point shows high fluctuations in traffic flow, whereas the smoothed vehicle count uses a rolling average, which helps to identify the overall trend by reducing short-term volatility. This gave a mean vehicle count of 50.65, which gives a general measurement of traffic congestion over time. It also enables the identification of patterns and different times when traffic flow within the network increases or decreases.

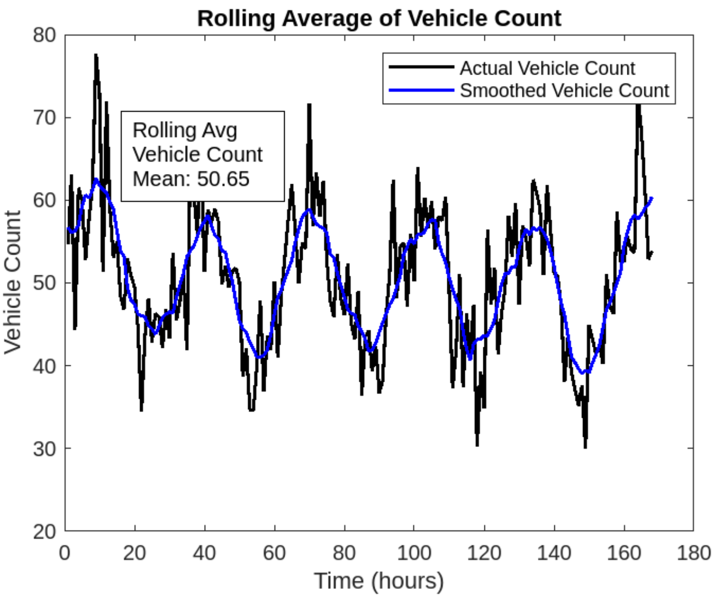

Fig.4 Rolling average of vehicle count measured over one week time

5.3 Vehicle Count vs. Speed Over Time:

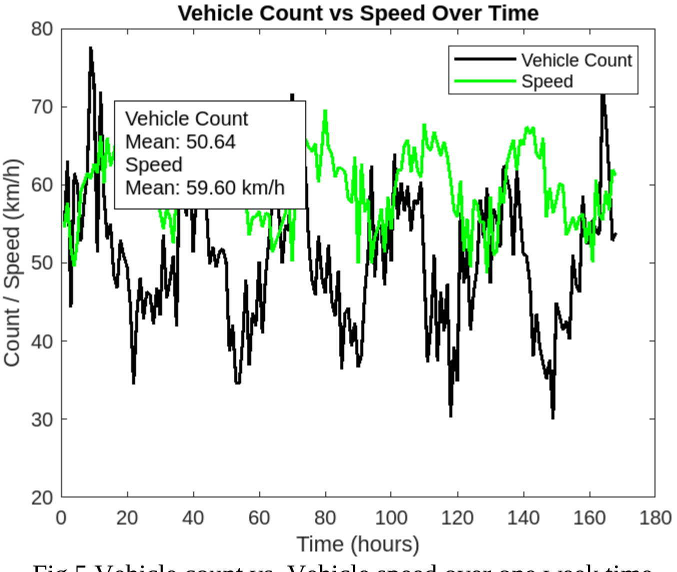

Fig.5 Vehicle count vs. Vehicle speed over one week time

The total vehicle count and speed of vehicles moving during the 7 days of observation are summarized in Fig.5. The mean vehicle count is around 50.64, and the mean speed is 59.60 km/h. This traffic density visualization emphasizes vehicle count-to-speed correlation where relative traffic volumes are inversely proportional where high levels denote congestion and low speed and vice versa. The said patterns give an understanding of how the flow of traffic behaves and the level of congestion.

5.4 Traffic Density vs. Occupancy Rate:

Fig.6 represents the relationship between Traffic Density and Occupancy Rate. Traffic density, plotted on the x-axis, gives how crowded the road is, while occupancy rate on the y-axis gives the amount of road space occupied by vehicles. The graph displays a scattered yet somewhat growing trend, suggesting that as traffic density rises, the occupancy rate also increases. Still, the sharp fluctuations and intersections show unevenness in the data, which could be due to various traffic conditions such as peak and off-peak hours or external factors such as roadblocks or accidents.

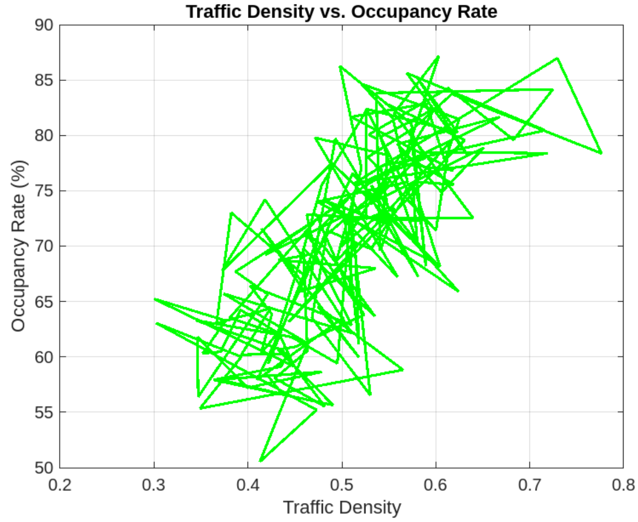

Fig.6 Traffic density and Occupancy rate on a crowded road

5.5 Model Performance:

The data collected on a road are pre-processed and then used in other models to determine their efficiency.

Comparison of ARIMA Model and Actual Occupancy Rate:

Fig.7 depicts the predicted occupancy rate using the ARIMA model (dashed line) with the actual occupancy rate (solid line) over time. The ARIMA model captures the trend and fluctuations in congestion degree levels by predicting based on historical data. While the ARIMA predictions usually follow the pattern of the actual traffic data, differences can be detected during peak or irregular traffic conditions, showing areas where the model may need additional improvements for higher accuracy. The comparison includes the effectiveness and disadvantages of the ARIMA model for traffic congestion prediction.

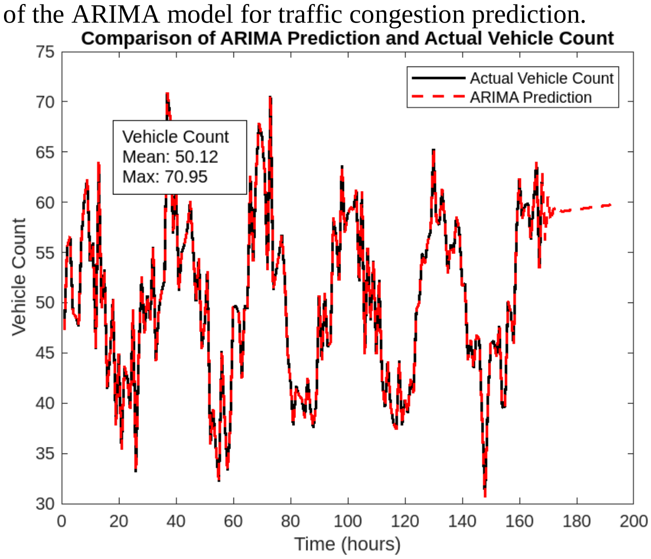

Fig.7 Comparison of ARIMA Model and Actual Occupancy Rate over 7 days of time

Comparison of Hybrid Regression Model and Actual Occupancy Rate:

Fig.8 depicts the predicted occupancy rate from the hybrid regression model (dashed line) with the actual occupancy rate (solid line) over time. The hybrid regression model uses multiple inputs, such as vehicle count and speed of vehicles, to evaluate the occupancy rate. Here in this comparison, the model closely finds the actual occupancy rate, mainly during periods of moderate traffic. Still, there may be slight deviations during high-traffic or other abnormal instances. This comparison shows the potential of the hybrid model to estimate congestion trends more accurately than the simpler models.

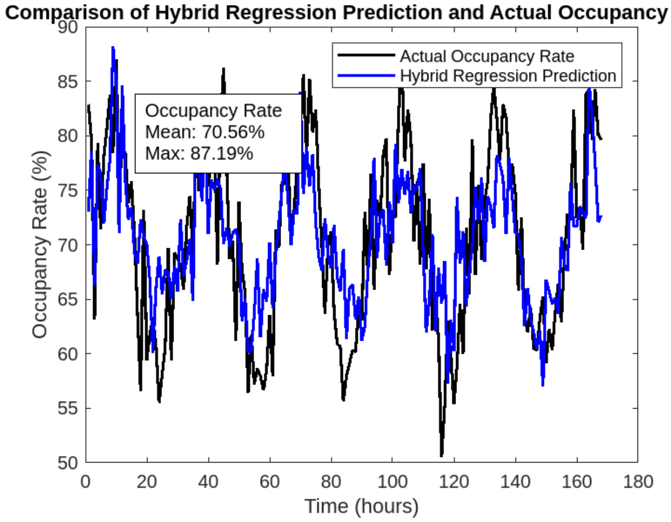

Fig.8 Comparison of Hybrid Regression Model and Actual Occupancy Rate over 7 days of time

Prediction Error Distribution:

Fig. 9 represents a histogram of prediction errors obtained from the hybrid model used for the prediction of traffic congestion. The x-axis shows the error between the predicted congestion and the actual congestion, while the y-axis represents the frequency of each degree of error value. The errors are centered on the zero, representing that the model is quite more often accurate, with a small mean error equal to -0.00. The figure reveals that the majority of the errors are concentrated between -10 and +10, showing that most of the predictions are accurate, but there are a few outliers that exist beyond this range.

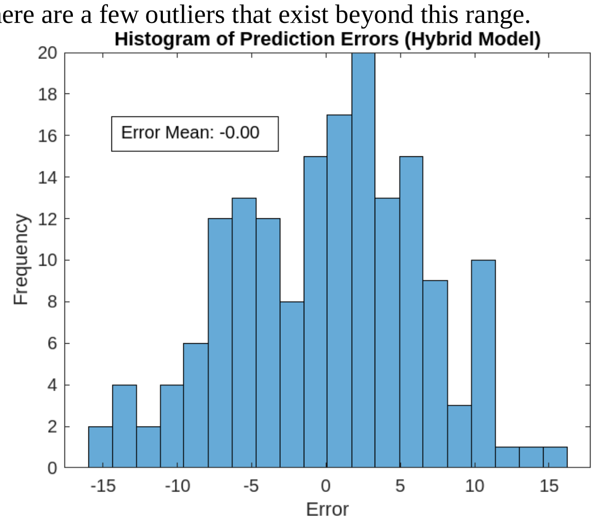

Fig. 9 Histogram of prediction error for hybrid model

Residual Error of Hybrid Regression Model:

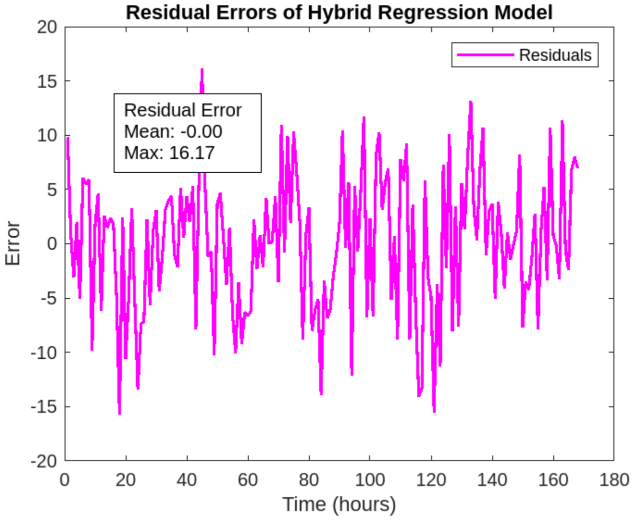

Fig. 10 Residual error of hybrid regression model

A time series plot of residual errors from the hybrid regression model is shown in Fig.10. The x-axis indicates the time in hours and the y-axis represents the error between the predicted and actual values. The residuals vary between approximately -15 and +15, showing a variation in prediction accuracy. The mean residual error is at -0.00, indicating that the model has no substantial bias. The maximum residual error is 16.17, which means that the largest variation between the prediction and actual value is around 16 units. This figure shows both the variability and consistency of the performance of the model.

Peak Traffic Detection:

Fig.11 depicts the occupancy rate (%) over time (in hours), with the x-axis showing time and the y-axis indicating the percentage of occupancy. The black line signifies the occupancy rate, while the red circles highlight occurrences of peak traffic where the occupancy rate achieves its maximum value. The occupancy rate varies between 55 % to 90 % throughout the observed period, with many peaks near 85%-90%, showing high traffic congestion at those points. This plot also gives the best representation of the cycle that is evident from the traffic patterns over time.

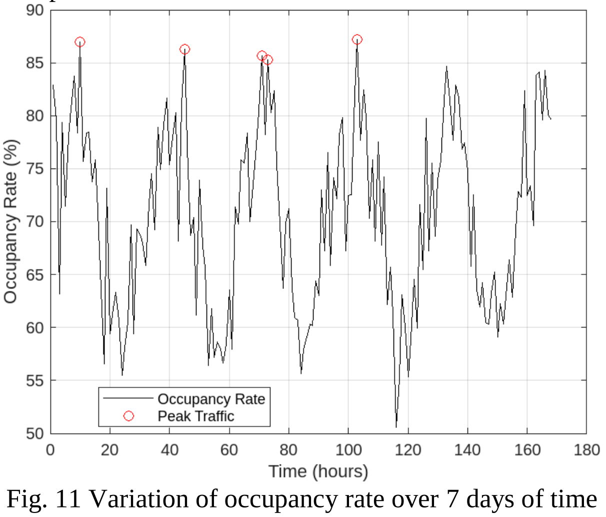

Fig. 11 Variation of occupancy rate over 7 days of time

Vehicle Speed and Traffic Congestion:

Fig.12 illustrates total traffic congestion levels based on vehicle speed. As vehicle speed reduces, traffic congestion typically rises, indicating slower movement of vehicles in congested conditions. Peaks in the figure indicate smoother traffic flow with higher speeds, while troughs could show traffic congestion or slow-moving conditions. The graph indicates the inverse relationship between vehicle speed and traffic congestion, facilitating to analysis of traffic behaviour over time or at specific points.

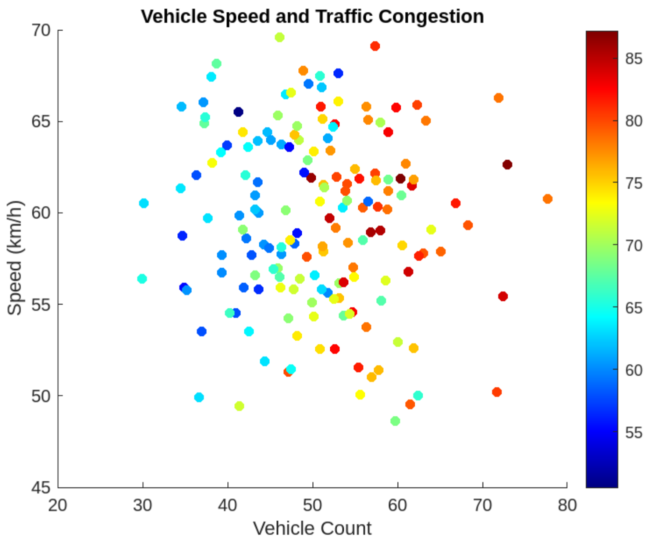

Fig.12 Relationship between vehicle speed and traffic congestion

5.6 ***Comparative Analysis:***

From the results of both models above, it is clear that between these two methods, the time-series model (ARIMA Model**)** produces reasonable congestion predictions, especially during regular traffic patterns similar to daily rush hours. However, the model failed to find satisfactory solutions to address all conditions that contributed to sudden changes in traffic flow such as accidents or road closures. The next model – the hybrid regression model performed better than the base models in terms of accuracy and relative error, especially in the case of congestion during irregular traffic events. The overall evaluation of both models is shown in the following Table 4.

Table 4: Summarizes the performance metrics of both models

Metric	Time-Series Model	Hybrid Regression Model
MAE	5.24	3.67
RMSE	7.81	5.12

According to Table 4, the hybrid model outperformed the time-series model in terms of MAE and RMSE values. This highlights that the hybrid model is better able to handle the complex relationships between traffic parameters and gives more predictions that are accurate.

6. Conclusion

This proposed work presented a comparative analysis of two predictive models for traffic congestion: a time-series model and a hybrid regression model. Both the models were realized using real-time traffic data, and their performance was assessed using standard error metrics. The outcomes showed that in terms of prediction accuracy and flexibility, the hybrid model performed much better than the time-series model. Through the integration of historical traffic patterns and real-time sensor data, the hybrid model produced more dependable and accurate congestion projections. These results imply that hybrid models work effectively in traffic management systems that need to forecast congestion in real time, especially in urban settings that are dynamic.

To further increase the accuracy of congestion predictions, future study should investigate the integration of additional data sources, such as weather and public transport timetables. To determine whether the hybrid approach can be widely used in smart cities, its scalability in bigger traffic networks should also be examined.

References:

[1] D. Albalate, X. Fageda, “On the relationship between congestion and road safety in cities”, Transport Policy, vol. 105, pp. 145-152, 2021.

[2] L. Cheng, K. Sun, "Research on intelligent vehicle Traffic Flow control algorithm based on data mining", International Journal of Intelligent Networks, vol. 5, pp. 92-100, 2024.
[3] M. Eom, BI. Kim, "The traffic signal control problem for intersections: a review", European Transport Research Review, vol. 12, no. 50, 2020. https://doi.org/10.1186/s12544-020-00440-8
[4] P. Bhartiya, M. Bhatele, A. A. Waoo, "A machine learning approach for predictive analysis of traffic flow", ShodhKosh: Journal of Visual and Performing Arts, vol. 5, no. 5, pp. 422–430, 2024.
[5] S. Afandizadeh, S. Abdolahi, H. Mirzahossein, "Deep Learning Algorithms for Traffic Forecasting: A Comprehensive Review and Comparison with Classical Ones", Journal of Advanced Transportation, vol. 2024, no. 1, 2024.
[6] S. A. Sayed, Y. Abdel-Hamid, H. A. Hefny, "Artificial intelligence-based traffic flow prediction: a comprehensive review", Journal of Electrical Systems and Information Technology, vol. 10, no. 13, 2023.
[7] G. Zheng, W. K. Chai, JL. Duanmu, V. Katos, "Hybrid deep learning models for traffic prediction in large-scale road networks", Information Fusion, vol. 92, pp. 93-114, 2023.
[8] T. Ma, C. Antoniou, T. Toledo, "Hybrid machine learning algorithm and statistical time series model for network-wide traffic forecast", Transportation Research Part C: Emerging Technologies, vol. 111, pp. 352-372, 2020.
[9] N. Chiabaut, R. Faitout, "Traffic congestion and travel time prediction based on historical congestion maps and identification of consensual days", Transportation Research Part C: Emerging Technologies, vol. 124, 2021.
[10] A. M. Nagy, V. Simon, "Improving traffic prediction using congestion propagation patterns in smart cities", Advanced Engineering Informatics, vol. 50, 2021.

[11] Q. Liu, et al., "Explanatory prediction of traffic congestion propagation mode: A self-attention based approach", Physica A: Statistical Mechanics and its Applications, vol. 573, 2021.

[12]J. Kaur, K. S. Parmar, S. Singh, "Autoregressive models in environmental forecasting time series: a theoretical and application review", Environmental Science and Pollution Research, vol. 30, pp. 19617–19641, 2023. https://doi.org/10.1007/s11356-023-25148-9
[13]E. F. Agyemang, J. A. Mensah, E. Ocran, E. Opoku, E. N. N. Nortey, "Time series based road traffic accidents forecasting via SARIMA and Facebook Prophet model with potential changepoints", Heliyon, vol. 9, no. 12, 2023.
[14] X. Ma, et al., "A comprehensive review of the development of land use regression approaches for modeling spatiotemporal variations of ambient air pollution: A perspective from 2011 to 2023", Environment International, vol. 183, 2024.
[15]A. Mahmoud, A. Mohammed, "Leveraging Hybrid Deep Learning Models for Enhanced Multivariate Time Series Forecasting", Neural Processing Letters, vol. 56, no. 223, 2024.

[16]S. Lu, Q. Zhang, G. Chen, D. Seng, "A combined method for short-term traffic flow prediction based on recurrent neural network", Alexandria Engineering Journal, vol. 60, no.1, pp. 87-94, 2021.
[17] L. Zhang, et al., "A hybrid forecasting framework based on support vector regression with a modified genetic algorithm and a random forest for traffic flow prediction", Tsinghua Science & Technology, vol. 23, no. 4, pp. 479-492, 2018.

[18] I. H. Sarkar, “Machine Learning: Algorithms, Real-World Applications and Research Directions”, SN Computer Science, vol. 2, no. 160, 2021.
[19] Anita Mohanty, et al., “Identification and evaluation of the effective criteria for detection of congestion in a smart city”, IET Communications, vol. 16, no.5, pp. 560-570, 2022.
[20] J. De Smedt, et al., “Process model forecasting and change exploration using time series analysis of event sequence data”, Data & Knowledge Engineering, vol. 145, 2023.
[21] G. Zheng, et al., “Hybrid deep learning models for traffic prediction in large-scale road networks”, Information Fusion, vol. 92. pp. 93-114, 2023.
[22] K. Maharana, et al., “A review: Data pre-processing and data augmentation techniques”, Global Transitions Proceedings, vol. 3, no. 1, pp. 91-99, 2022.
[23] V. N. Katambire, et al., “Forecasting the Traffic Flow by Using ARIMA and LSTM Models: Case of Muhima Junction”, Forecasting, vol. 5, no. 4, pp. 616-628, 2023.
[24]Z. Hajirahimi, M. Khashei, “Hybrid structures in time series modeling and forecasting: A review”, Engineering Applications of Artificial Intelligence, vol. 86, no.1, pp. 83-106, 2019.