Enhanced Traffic Congestion Prediction Using Attention-Based Multi-Layer GRU Model with Feature Embedding

Sreelekha M[1], Dr. Midhunchakkaravarthy Janarthanan[2]
Research Scholar, Faculty of Engineering, Lincoln University College, Malaysia[1]
Faculty of Computer Science and Multimedia, Lincoln University College, Malaysia[2]

***Abstract*—Intelligent Transportation Systems (ITS) are crucial for managing urban mobility and addressing traffic congestion, which poses significant challenges to modern cities. Traffic congestion leads to increased travel times, pollution, and fuel consumption, impacting both the environment and quality of life. Traditional traffic management solutions often fall short in predicting and adapting to dynamic traffic conditions. This study proposes an efficient deep learning (DL) model for predicting traffic congestion, utilizing the strengths of an attention-based multilayer Gated Recurrent Unit (GRU) network. The dataset used for this study includes 48,120 hourly vehicle counts across four junctions and additional weather data. Temporal and lagged features were engineered to capture daily and historical traffic trends and categorical data were considered by employing feature embedding. The attention-based GRU model integrates an attention mechanism to focus on relevant historical data, improving predictive performance by selectively emphasizing crucial time steps. This model architecture, consisting of two hidden layers and attention mechanisms, allows for nuanced traffic predictions by handling temporal dependencies and variations effectively. The performance was evaluated using various error metrics. The results demonstrate the model's ability to predict traffic congestion with MSE of 0.9678, MAE of 0.4322, R^2 of 0.8686, MAPE of 6% offering valuable insights for traffic management and urban planning.**

Keywords*—Intelligent transportation system; traffic congestion; urban mobility; deep learning; gated recurrent unit

I. INTRODUCTION

Traffic congestion is a persistent and growing problem in urban areas globally. It leads to significant challenges that affect economic productivity, environmental sustainability, and the quality of life [1]. As cities continue to expand and urbanization accelerates, the demand for road space often surpasses the available infrastructure, resulting in slower traffic speeds, longer travel times, and increased vehicular queuing. Several factors contribute to this issue, including the rapid rise in vehicle ownership, insufficient public transportation systems, and inadequate urban planning. The surge in private vehicles, due to the rising incomes and population growth, places immense pressure on existing road networks, intensifying congestion, especially during peak hours. In many cities, the public transportation infrastructure is either lacking or inefficient, prompting people to rely heavily on private car. Additionally, poor urban planning—such as poorly designed road networks, insufficient parking, and a lack of infrastructure for pedestrians and cyclists further intensifies traffic blocks [2].

The impacts of traffic congestion are wide-ranging and severe. Economically, it leads to significant costs, including wasted fuel, vehicle maintenance, and lost productivity as people spend more time in traffic. Businesses suffer due to delays in goods transportation and reduced employee efficiency [3]. Environmentally, traffic congestion contributes to higher emissions of greenhouse gases and pollutants, as vehicles emit more while idling in traffic jams than when traveling smoothly. This not only worsens air quality but also accelerates climate change. Socially, congestion reduces the quality of life, as long travels lead to stress, less time for personal activities, and overall frustration. The safety concerns are also notable, with increased stop-and-go traffic raising the likelihood of accidents and making roads more dangerous for both drivers and pedestrians. One of the most critical consequences is its impact on emergency services [4]. Congested roads can significantly delay ambulances, fire trucks, and police vehicles, potentially leading to life-threatening situations due to increased response times.

Addressing traffic congestion requires a multi-faceted approach. Enhancing public transportation is a key strategy, as efficient and reliable public transit systems can reduce the number of private vehicles on the road. Investments in bus transit systems, light rail networks, and better integration of transport modes can make public transport a more attractive option [5]. Additionally, traffic management schemes that use real-time data to enhance traffic flow, such as adaptive traffic signal controls, can alleviate congestion at critical points. Urban planning also plays a crucial role; cities need to adopt designs that encourage high-density, mixed-use developments that reduce the need for long shuttles. Promoting smart mobility solutions, such as ride-sharing, autonomous vehicles, and Mobility-as-a-Service (MaaS), can also help by optimizing road usage. Behavioral changes, supported by public awareness campaigns and incentives for carpooling or telecommuting, are essential in reducing the number of vehicles on the road during peak times [6]. In summary, while traffic congestion is a complex problem with significant impacts, a combination of improved infrastructure, smart technology, and policy measures can help mitigate its effects and create more sustainable and functional urban environments. A DL model for traffic congestion prediction is proposed in this study. The main contributions of the study are given below:

- To develop a DL-based traffic congestion prediction model.
- To compare the effectiveness of the suggested model with existing models.
- To evaluate the efficiency through various error metrics.

The remaining portion of the paper is organized as: Section II provides a comprehensive literature review emphasizing the need for the current research. Section III details the methodology and the deep learning model architecture for effective traffic prediction. Section IV presents the results and discussion, highlighting the potential of the suggested model. Section V concludes the paper by summarizing the key contributions.

II. Literature Review

Li et al. [7] introduced the AST3DRNet model, which incorporated a 3D residual network with a self-attention mechanism. This approach utilized a 3D convolutional module and employs a spatio-temporal attention module to dynamically adjust the impact of these relationships. Experiments conducted using a real-world traffic dataset from Kunming demonstrated that AST3DRNet outperformed existing baseline methods, achieving accuracy improvements of 59.05%, 64.69%, and 48.22% for short-term predictions at 5, 10, and 15 minutes, respectively. Despite its innovations, the model's dependency on convolutional neural networks (CNN) and residual networks was a limitation.

Tsalikidis et al. [8] evaluated various models for multi-step forecasting of traffic flow, particularly in areas with limited historical data. The methodology involved assessing a range of interpretable predictive algorithms, including Ensemble Tree-Based (ETB) regressors like Light Gradient Boosting Machine (LGBM) and comparing them with traditional deep learning methods. Results indicated that ETB models generally outperformed DL approaches, particularly for longer forecasting horizons, achieving high accuracy even at extended prediction steps. The study demonstrated that feature selection and engineering, incorporating temporal and weather data, improved model performance. The study was limited by its reliance on the statistical characteristics of the specific dataset, which could affect the efficiency of the algorithms. High data volume and complexity also posed challenges, impacting model training and performance.

Jiang et al. [9] introduced Congestion Prediction Mixture-of-Experts (CP-MoE) to improve prediction accuracy for dynamic traffic scenarios. The methodology involved developing a Mixture of Adaptive Graph Learners (MAGLs) with a sparsely-gated mechanism and congestion-aware biases, complemented by two specialized experts designed to identify stable trends and periodic patterns. This model was rigorously tested on real-world datasets, demonstrating its superiority over existing spatio-temporal prediction methods. Notably, CP-MoE was successfully integrated into DiDi's system, enhancing travel time estimation reliability. A key limitation identified was the utility of CP-MoE's application to other aspects of ride-hailing services.

Hao et al. [10] presented a fuzzy logic system based on the Greenshields model, designed to predict highway traffic congestion without requiring extensive training data. The methodology involved processing vehicle speed and traffic flow inputs using specified membership functions and applying fuzzy rules guided by Greenshields theory. The approach was validated through a comparative analysis with a polynomial regression model using real-world data from the Sun Yat-Sen Highway in Taiwan, demonstrating consistent prediction results. The fuzzy logic system proved effective in estimating congestion levels and adapting to various road conditions with minimal data preparation. A noted limitation was the potential for reduced precision in highly dynamic traffic scenarios where the fuzzy logic system's fixed rules not capture complex variations as effectively.

Zhang et al. [11] introduced a deep marked graph process (DMGP) model that combined a spatiotemporal convolutional graph network with a traditional point process model to predict congestion indices and occurrence times for large signalized road networks. This hybrid approach utilized the simplicity of the point process model and the advanced capabilities of graph neural networks to model the evolution of traffic congestion. Experiments using real-world traffic data demonstrated that the DMGP model outperformed existing baseline methods, achieving superior prediction accuracy and computational efficiency. While the model showed promise in supporting advanced traffic management and traveler information systems, a significant limitation was its reliance on high-quality citywide traffic data, which had not been available for all road segments.

Jasim et al. [12] analyzed the efficiency of several machine learning (ML) algorithms for congestion detection and prediction within Vehicular Ad hoc Networks (VANETs). The study focused on Support Vector Machines (SVM), Ensemble Learning classifiers, K-Nearest Neighbors (KNN), and Decision Trees (DT). The methodology involved training these algorithms with historical traffic congestion data and applying advanced feature engineering techniques. The study found that SVM, along with KNN and Ensemble Learning classifiers, achieved high classification accuracies. The study was limited by the dependence on precise feature selection and model optimization techniques, which required careful tuning.

Arabiat et al. [13] addressed the challenge of predicting traffic congestion using data mining and ML techniques. The study compared the performance of two open-source software tools, WEKA and Orange, in predicting traffic congestion in Amman, Jordan. Various classifiers, including SVM, KNN, Logistic Regression (LR), and Random Forest (RF), were tested using data from the Greater Amman Municipality for the year 2018. Results revealed that Orange excelled with high prediction accuracy. The study highlighted the superior performance of Orange over WEKA, particularly in handling different classifiers. A notable limitation was the reliance on specific data mining tools, which are not generalize across all types of traffic data or scenarios, potentially affecting the applicability of the findings.

Chahal et al. [14] tackled the challenge of traffic flow prediction using a hybrid model that combines Seasonal Auto-

Regressive Integrated Moving Average (SARIMA) with Bidirectional Long Short-Term Memory (Bi-LSTM) and Back Propagation Neural Network (BPNN). This approach aimed to address both linear and non-linear components of traffic data from the CityPulse EU FP7 project. The hybrid model demonstrated superior performance with the lowest MAE of 0.499 compared to single SARIMA, LSTM, and other models. The study was limited by the feature set considered, as the model did not account for external factors like weather or peak hours, which could affect traffic predictions.

Jin et al. [15] introduced a spatio-temporal graph neural point process (STGNPP) framework specifically designed to predict traffic congestion events that occur sporadically over time. The model incorporated a spatio-temporal graph learning component to efficiently capture long-term dependencies from historical traffic data and road network information. This information is then processed through a continuous GRU to model congestion evolution patterns, with a periodic gated mechanism enhancing the intensity function to account for periodic variations. Extensive experiments conducted on two large-scale real-world datasets demonstrated that STGNPP significantly outperformed existing methods in predicting both the timing and duration of congestion events. However, the model’s reliance on historical data may limit its adaptability to sudden or unprecedented traffic disruptions.

Pan et al. [16] presented Ising-Traffc, a dual-model framework that employs the Ising model to address traffic management. Unlike conventional approaches that struggle with balancing algorithmic complexity and computational efficiency, Ising-Traffc combines two distinct Ising models: Predict-Ising and Reconstruct-Ising. Reconstruct-Ising utilized advanced Ising machines to handle traffic uncertainties with reduced latency and lower energy consumption, while Predict-Ising uses conventional processors to project future traffic congestion, requiring only 1.8% of the computational resources compared to existing methods. The proposed framework demonstrated an average speed up of 98× and a 5% accuracy improvement over conventional solutions when evaluated on real-world traffic datasets. A notable limitation of this approach was the dependency on specific hardware for Reconstruct-Ising, which affect its scalability and adaptability across different computational platforms.

Zhang et al. [17] explored urban traffic condition prediction and congestion control by integrating improved particle swarm optimization (IPSO) with radial basis function (RBF) networks and a fusion model of LSTM networks and SVM. The proposed feature fusion model demonstrated superior performance in predicting traffic states, validated by experiments using regional traffic data from Shenyang Station, with the model achieving the lowest RMSE compared to other algorithms. For congestion control, a traffic allocation-based method was developed and tested using VISSIM simulation, showing effective congestion management. The primary limitation of the study was the reliance on simulation models to fully capture the complexities of real-world traffic dynamics and thus limit the applicability of the proposed methods in varied urban settings.

Wang et al. [18] addressed urban traffic congestion by developing a prediction model called Spatio-Temporal Transformer (STTF), which utilizes DL techniques. Traditional models struggled with the growing complexity of urban traffic networks, prompting the introduction of STTF. This model integrated traffic speed data, road network structure, and spatio-temporal correlations to enhance prediction accuracy. The STTF employed an information embedding module to convert both spatial and temporal data into feature vectors, which were then processed through spatial and temporal attention modules. The model was tested on real-world datasets, demonstrating substantial improvements in prediction accuracy compared to existing methods. Despite its advancements, the STTF model's main limitation was its dependence on comprehensive feature engineering and attention mechanisms, which increase computational complexity and impact its efficiency in real-time applications. Table I provides the summary of the existing traffic congestion prediction models.

While existing models, such as traditional spatio-temporal graph-based methods and hybrid approaches, have made significant strides, they often struggle with real-time efficiency and adaptability to rapidly changing traffic conditions. Additionally, existing approaches frequently lack the capability to adaptively weigh the importance of different time steps or traffic features, leading to suboptimal predictions. A critical gap in current traffic congestion prediction methods lies in the need for more advanced deep learning models that can seamlessly integrate and process complex spatio-temporal dependencies and dynamic factors. Deep learning models offer promising avenues for capturing intricate patterns in traffic data and improving prediction accuracy. Addressing these gaps requires the development of deep learning frameworks that can balance high accuracy with operational efficiency, effectively handling large-scale, dynamic data while being robust to varying traffic conditions and external influencing factors.

TABLE I. SUMMARY OF EXISTING TRAFFIC CONGESTION PREDICTION MODELS

Ref. No.	Works Carried Out in the Reference Papers	Advantages	Disadvantages
[7]	AST3DRNet model with a 3D residual network and spatio-temporal attention module for traffic prediction.	Achieved accuracy improvements of 59.05%, 64.69%, and 48.22% for 5, 10, and 15-minute predictions.	Dependency on CNNs and residual networks limits the scalability and adaptability of the model.
[8]	Ensemble Tree-Based (ETB) regressors (e.g., LGBM) compared with traditional DL methods for multi-step forecasting.	ETB models outperformed DL approaches, especially for long-term predictions; feature engineering improved performance.	Results relied on statistical characteristics of the dataset; complexity in high-volume data handling.
[9]	CP-MoE with Mixture of Adaptive Graph Learners and congestion-aware biases for dynamic traffic prediction.	Outperformed baseline methods; integrated into DiDi's system to enhance travel time reliability.	Limited applicability to other ride-hailing service aspects; utility depends on specific use cases.
[10]	Fuzzy logic system using Greenshields model for highway congestion prediction with minimal data preparation.	Consistent results; adaptable to various road conditions with minimal data.	Lower precision in dynamic scenarios with complex variations.
[11]	Deep Marked Graph Process (DMGP) combining spatiotemporal graph network and point process model for congestion prediction.	Superior prediction accuracy and computational efficiency.	Reliance on high-quality, citywide traffic data limits scalability to less monitored areas.
[12]	ML algorithms (SVM, KNN, Ensemble Learning, DT) with advanced feature engineering for VANET congestion detection.	High classification accuracy achieved with SVM, KNN, and Ensemble Learning classifiers.	Dependence on precise feature selection and careful optimization of models.
[13]	Comparison of WEKA and Orange tools for traffic prediction using ML classifiers like SVM, KNN, LR, and RF.	Orange achieved superior prediction accuracy over WEKA.	Limited generalizability across different traffic datasets and tools.
[14]	Hybrid SARIMA-Bi-LSTM-BPNN model for traffic flow prediction, addressing both linear and non-linear components.	Lowest MAE (0.499) compared to single models; superior performance on CityPulse EU FP7 data.	Did not account for external factors like weather or peak hours affecting predictions.
[15]	STGNPP framework with spatio-temporal graph learning and periodic gated mechanism for sporadic traffic event prediction.	Outperformed existing methods in predicting timing and duration of congestion events.	Limited adaptability to sudden or unprecedented disruptions due to reliance on historical data.
[16]	Ising-Traffc framework combining Predict-Ising and Reconstruct-Ising models for traffic management.	Achieved 98× speedup and 5% accuracy improvement; reduced latency and energy consumption.	Dependency on specific hardware for Reconstruct-Ising limits scalability.
[17]	IPSO-RBF network fusion model with LSTM and SVM for traffic prediction and congestion control via traffic allocation.	Superior performance in RMSE; effective congestion management in simulations.	Limited real-world validation; reliance on simulations restricts practical applicability.
[18]	Spatio-Temporal Transformer (STTF) integrating traffic speed, road networks, and spatio-temporal correlations.	Substantial accuracy improvements compared to existing methods.	High computational complexity due to feature engineering and attention mechanisms.

III. MATERIALS AND METHODS

Traffic congestion prediction is crucial for optimizing traffic flow and enhancing commuter experience by enabling more efficient traffic management and route planning. It helps reduce economic losses associated with delays, fuel consumption, and vehicle wear, benefiting both individuals and businesses. Accurate predictions also contribute to minimize the idle time and slow-moving traffic, thereby supporting environmental sustainability. Additionally, it provides valuable data for urban planners to design better infrastructure and improve overall urban mobility. Thus, this study proposes an efficient DL technique for traffic congestion prediction. Fig. 1 shows the detailed block diagram of the proposed traffic congestion prediction model.

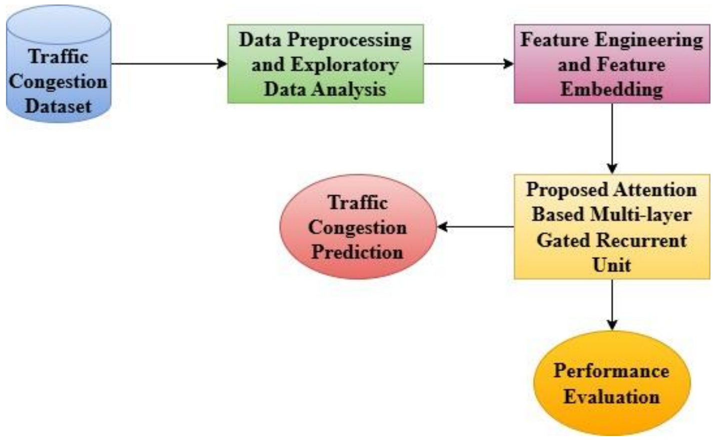

Fig. 1. Block diagram of proposed traffic congestion prediction model.

A. *Dataset Description*

The study utilized traffic data from Kaggle repository [19]. The dataset comprises 48,120 observations of hourly vehicle counts across four different junctions during the periods of November 1, 2015, to July 1, 2017, with observations distributed across different months and years. The data was collected by sensors placed at each junction, though these sensors operated at different times, resulting in traffic data from various time periods. The extensive range of hourly traffic counts across multiple junctions provides a valuable resource for modeling and predicting traffic congestion. The key features in the dataset are tabulated in Table II.

TABLE II. KEY FEATURES IN THE DATASET

Features	Description
Date and Time	The specific date and time of the observation.
Junction	The identifier for the junction where the data was collected.
Vehicles	The number of vehicles counted at the junction during the specified hour.
ID	A unique identifier for each observation.

The study incorporated weather data [20] of same period of time for creating more accurate, responsive, and comprehensive traffic management systems. It helps in optimizing traffic flow, improving safety, enhancing emergency response, and supporting long-term infrastructure planning. By understanding how weather influences traffic patterns, cities can better prepare for and mitigate the impacts of both regular and extreme weather conditions on their transportation networks. The dataset provides the temperature alongside two types of radiation measurements: direct horizontal radiation and diffuse horizontal radiation. These metrics are essential for understanding the overall weather conditions, as direct radiation measures the sunlight that reaches the ground without scattering, while diffuse radiation accounts for sunlight scattered by the atmosphere.

B. *Preprocessing and Exploratory Data Analysis*

Preprocessing and EDA are crucial in the data analysis pipeline, with preprocessing ensuring the data is suitable for analysis and EDA providing insights and understanding of the data [21]. Fig. 2 provides the statistics of the data statistics.

The histogram in Fig. 3 provides a visual representation of the distribution of traffic volumes in the dataset. It displays how frequently different ranges of vehicle counts occur by dividing the data into 30 bins. Each bar in the histogram represents the frequency of vehicle counts falling within a specific range, allowing for easy identification of common traffic volume ranges and patterns. The plot also helps in detecting any skewness in the data, understanding the spread of traffic volumes, and spotting potential outliers.

	Junction	Vehicles	ID
count	48120.000000	48120.000000	4.812000e+04
mean	2.180549	22.791334	2.016330e+10
std	0.966955	20.750063	5.944854e+06
min	1.000000	1.000000	2.015110e+10
25%	1.000000	9.000000	2.016042e+10
50%	2.000000	15.000000	2.016093e+10
75%	3.000000	29.000000	2.017023e+10
max	4.000000	180.000000	2.017063e+10

Fig. 2. Data statistics.

The line plot shown in Fig. 4 illustrates the average traffic volume over time by resampling the data on a daily basis. It shows how the average number of vehicles varies across different days, providing insights into daily traffic trends and fluctuations. The boxplots in Fig. 5 compare traffic volumes across different junctions, highlighting variations in vehicle counts. Each boxplot visualizes the distribution of traffic volumes for each junction, showing median values, interquartile ranges, and any outliers.

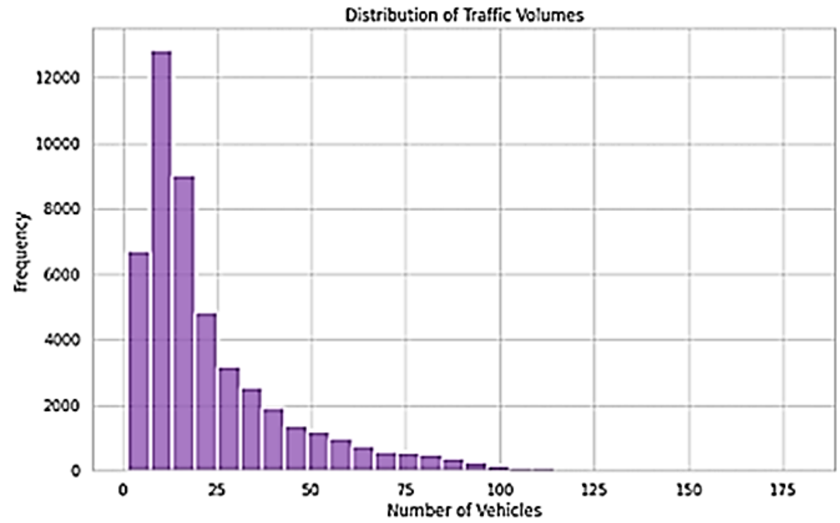

Fig. 3. Distribution of traffic volumes.

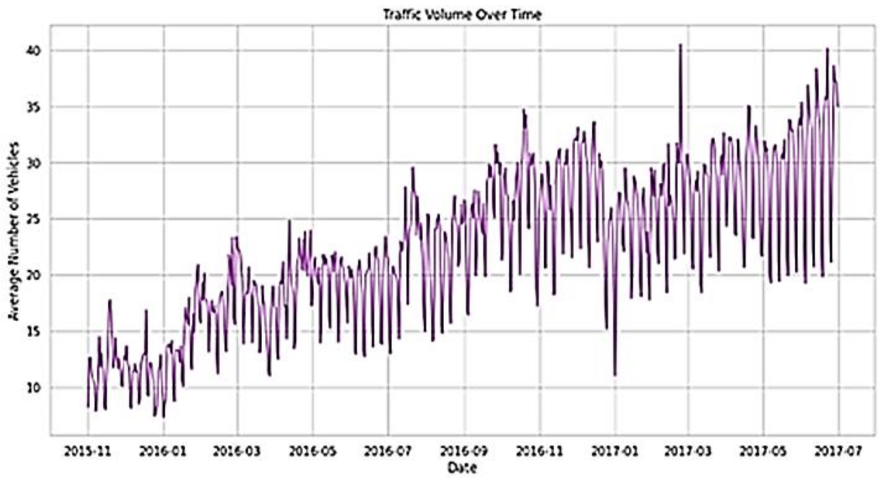

Fig. 4. Traffic volume over time.

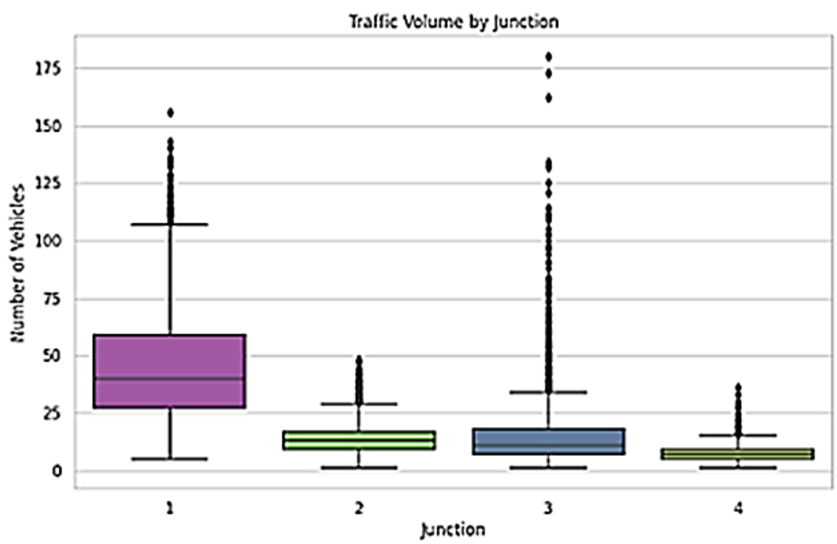

Fig. 5. Traffic volume by junction.

The average traffic volume for each hour and each day of the week is visualized in Fig. 6, revealing variations in traffic patterns across different days. These visualizations help in identifying peak traffic times and understanding daily and weekly traffic patterns.

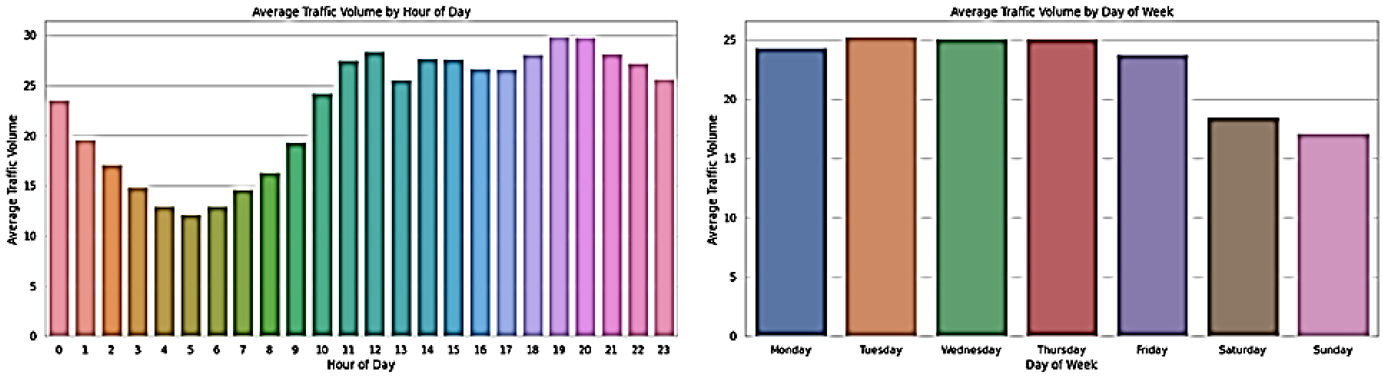

Fig. 6. Average traffic volume for (a) hour of the day (b) day of the week.

The analysis involves identifying peak and off-peak hours for traffic management by calculating the average traffic volume for each hour and each day of the week, highlighting times of high and low traffic as in Fig. 7.

Additionally, time series decomposition is performed to understand the underlying patterns in traffic volume data. By breaking down the data into trend, seasonal, and residual components, this approach reveals long-term trends, recurring seasonal effects, and irregular variations, providing a clear scenario of traffic dynamics over time as in Fig. 8. This comprehensive analysis supports better traffic management and planning by pinpointing peak traffic periods and understanding traffic behavior patterns.

The traffic volume analysis reveals a clear seasonal pattern with fluctuations that repeat on a weekly basis, suggesting regular peaks and troughs corresponding to weekly traffic variations. The trend component shows long-term changes in traffic volume, indicating periods of increase or decrease, but without providing conclusive trends due to its variability. The residuals, representing the noise after accounting for trend and seasonality, are scattered around zero with a few outliers. Using a correlation matrix in Fig. 9, the correlation analysis examines at the associations between traffic volume and variables like the day of the week and hour of the day.

The correlation matrix reveals that there is a moderate positive correlation between traffic volume and the hour of the day, indicating that traffic volume tends to increase with later hours. In contrast, the correlation between traffic volume and the day of the week is weakly negative, suggesting minimal variation in traffic volume across different days. The traffic flow variations throughout the week are illustrated in Fig. 10, which reflects patterns in commuter and commercial traffic.

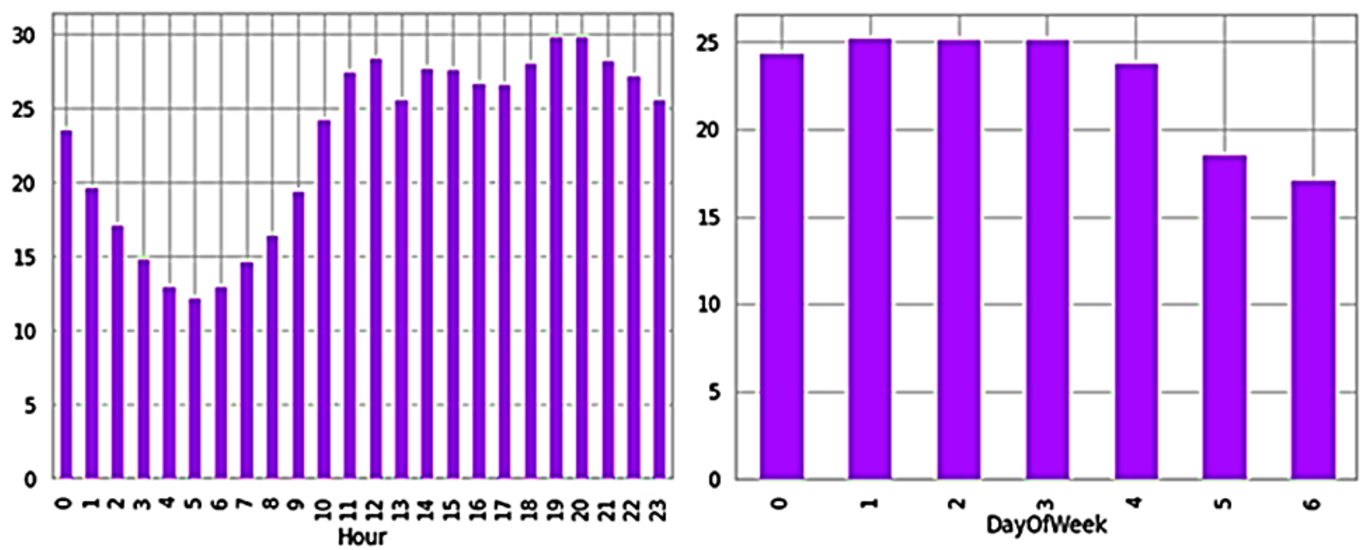

Fig. 7. Peak hour analysis.

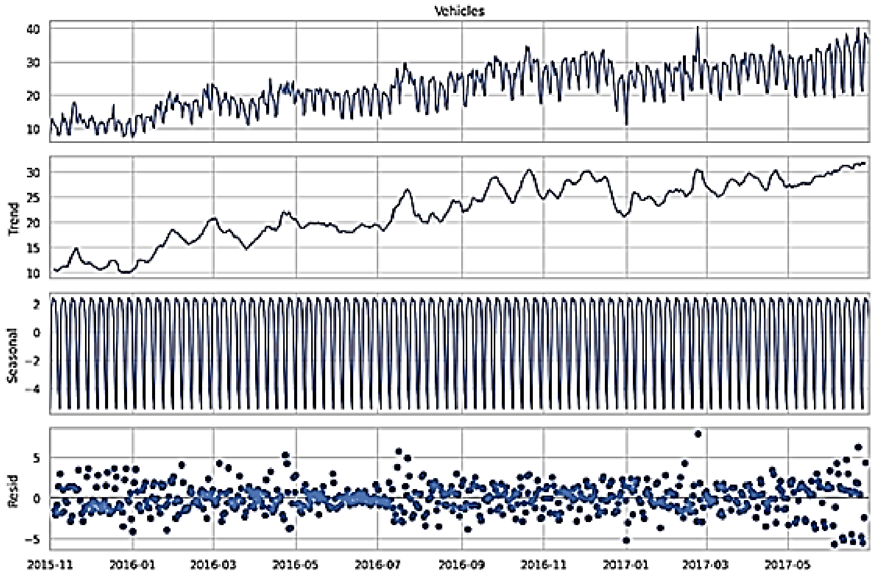

Fig. 8. Time series decomposition.

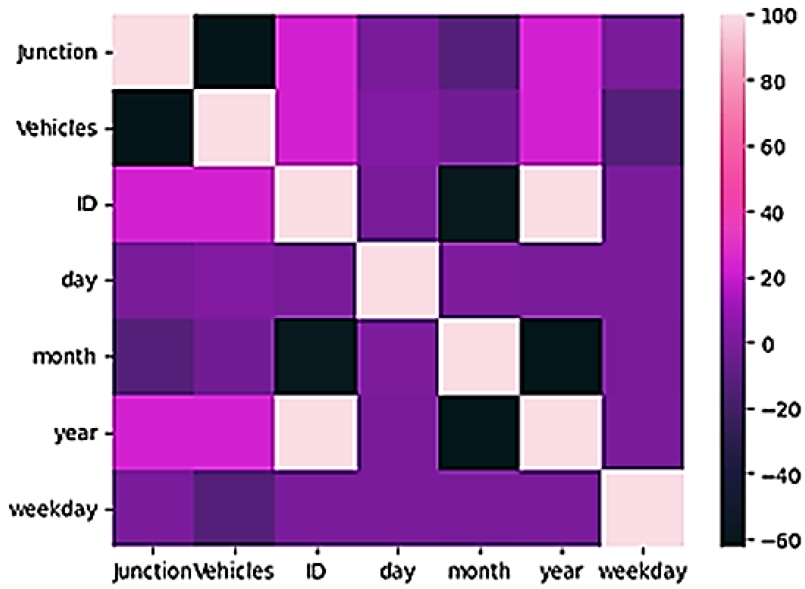

Fig. 9. Correlation matrix.

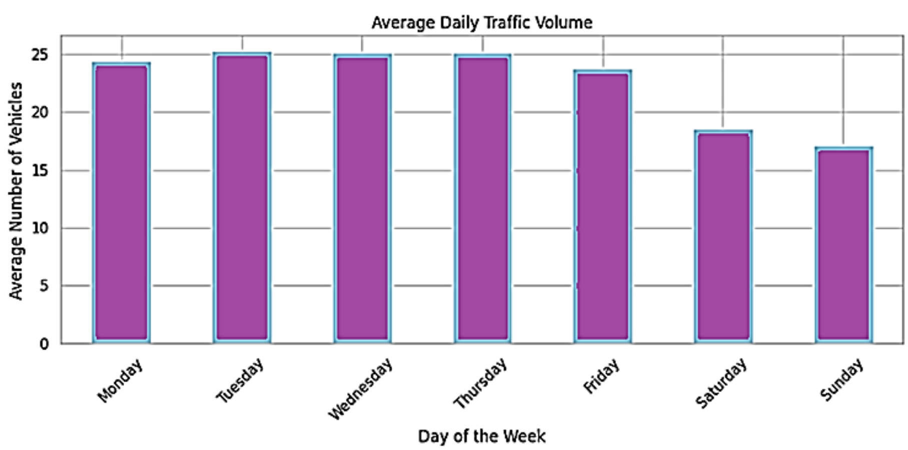

Fig. 10. Average daily traffic rate.

The analysis of holiday effects on traffic volumes, shown in Fig. 11, aims to determine how public holidays impact traffic patterns compared to non-holidays. By marking specific public holidays and comparing the average traffic volumes on these days to those on regular days, the analysis identifies any significant differences in traffic flow.

For long-term trend analysis, monthly traffic volumes were examined to assess any significant changes over an extended period. The monthly average traffic volume data as illustrated in Fig. 12, indicates a general upward trend, suggesting that traffic has been increasing over time. An addition of a trend line to the plot confirmed this long-term upward trajectory. This trend could reflect urban development, population growth, or other factors influencing traffic patterns. Such insights are valuable for traffic management and infrastructure planning, as they highlight the need for adapting strategies to handle increasing traffic volumes.

The junction comparison analysis in Fig. 13 highlights variations in average traffic volumes across different junctions. By grouping the traffic data by junction and calculating the average number of vehicles for each, the analysis reveals that Junction 1 consistently experiences the highest average traffic volume, significantly surpassing the other junctions.

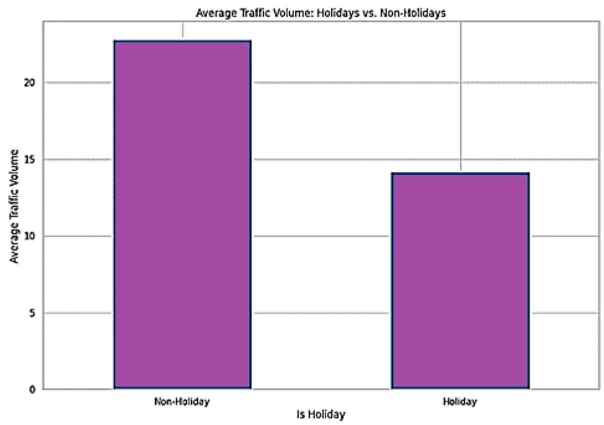

Fig. 11. Holiday effect on traffic volume.

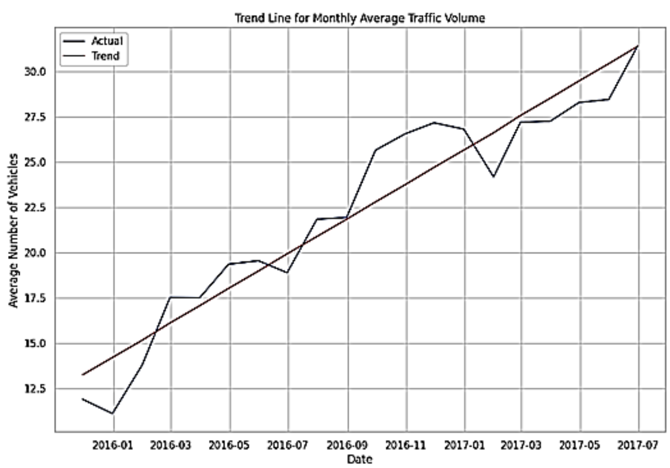

Fig. 12. Monthly average traffic volume over time with a linear trend line.

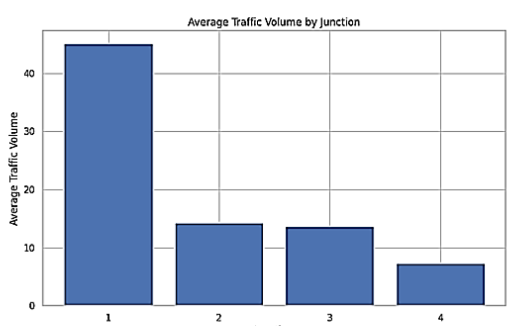

Fig. 13. Average traffic volume by junction.

The analysis of daily traffic volumes through Z-scores has identified specific dates with unusually high traffic levels. Z-scores quantify how far each data point deviates from the average, highlighting days with significantly above-average traffic. These elevated traffic volumes could be attributed to various factors. For instance, there have been special events, like concerts or sports games, that led to increased traffic on these days. Alternatively, temporary disruptions such as road construction or detours have redirected traffic through these areas, causing a spike. Additionally, seasonal patterns or local events also explain the higher traffic volumes observed.

Incorporating the weather data, the scatter plot in Fig. 14 shows that there is no evident correlation between temperature and the number of vehicles, suggesting that temperature has no significant impact on traffic volume.

The analysis of daily traffic volumes included the application of the Augmented Dickey–Fuller (ADF) test to assess the stationarity of the time series data. Stationarity is a critical property for time series analysis, as non-stationary data can lead to misleading results in forecasting models. The ADF test was employed to test the null hypothesis that the traffic volume time series contains a unit root, which would indicate non-stationarity. The results of the ADF test revealed a significantly negative ADF statistic and a p-value much smaller than conventional significance levels. These findings strongly reject the null hypothesis, indicating that the traffic data is stationary. The Auto-Correlation Function (ACF) plot displays the correlation between the data and its lagged values over time, while the Partial Auto-Correlation Function (PACF) plot shows the direct correlation at specific lags, controlling for the effects of intermediate lags as in Fig. 15.

To complement this stationarity check, a Z-score analysis was performed to identify anomalies in daily traffic volumes. By calculating Z-scores, which indicate how many standard deviations a data point is from the mean, the analysis was able to identify days with significantly higher or lower traffic volumes compared to the average. These anomalies were then visualized on a line graph as in Fig. 16, with red dots marking the days where traffic volumes deviated notably from the norm. This visual representation provided insights into trends, potential seasonality, and outlier events that could be linked to external factors such as road closures, construction projects, or special events.

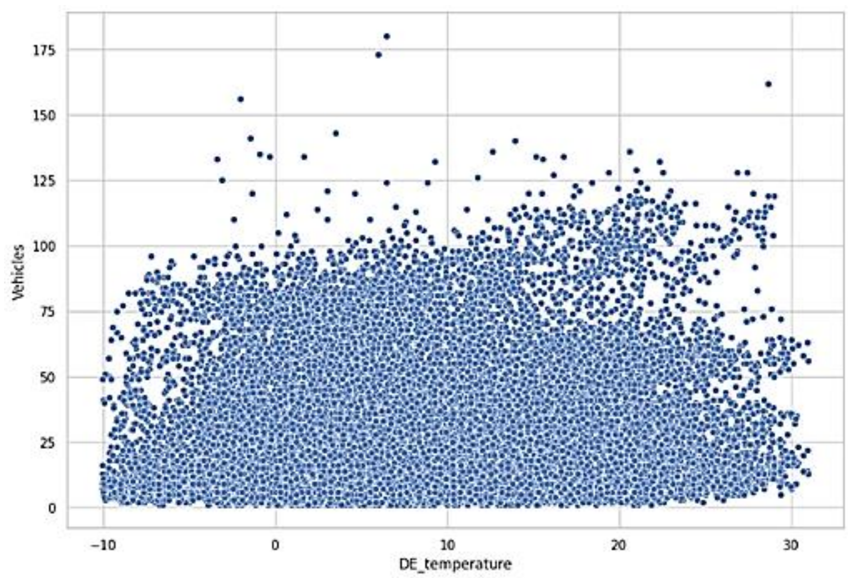

Fig. 14. Scatter plot of temperature vs. number of vehicles.

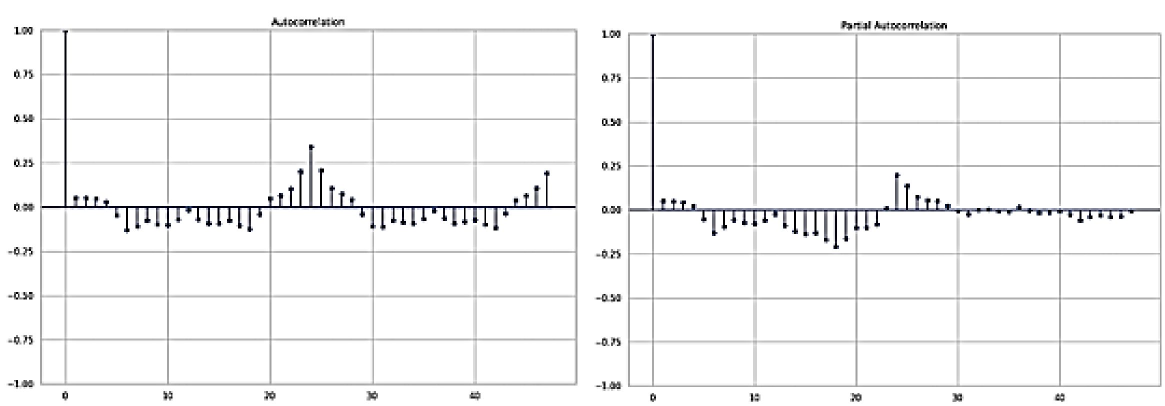

Fig. 15. ACF and PACF plot.

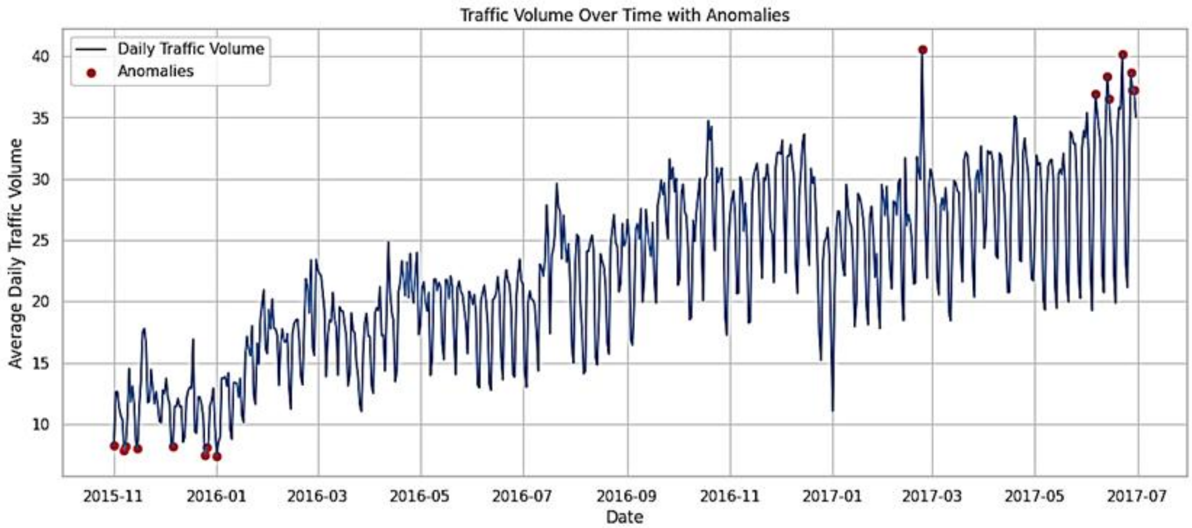

Fig. 16. Traffic volume over time with anomalies.

Additionally, the analysis involved checking for missing data, consistency of data reporting, and potential outliers. Missing timestamps were identified and accounted for, ensuring that the data was complete and accurately represented. The consistency of data reporting was verified by examining the number of records per junction and analyzing the time intervals between records. This step ensured that data collection was uniform across different junctions and time periods. Outlier detection further refined the analysis by identifying traffic volumes that were unusually high or low as depicted in Fig. 17, which could distort the overall findings if not properly addressed.

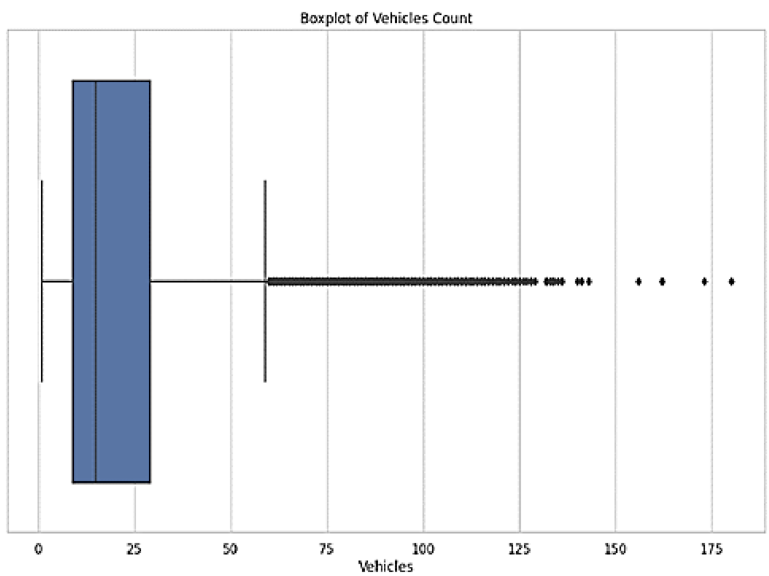

Fig. 17. Box plot of vehicle count.

C. Feature Engineering and Embedding

Feature engineering is a crucial step in improving model performance by creating and transforming features to capture the underlying patterns in traffic congestion data [22]. One of the primary types of features engineered for this purpose is temporal features. These include the hour of the day (extracted from the timestamp) to capture daily traffic variations, the day of the week to distinguish between weekday and weekend traffic patterns, and the month to account for seasonal trends. Additionally, a holiday indicator is used as a binary feature to differentiate between holidays and regular days, which often exhibit different traffic behaviors. In addition to temporal features, lagged features are introduced, such as the previous hour traffic volume, which helps in incorporating short-term historical trends into the model as represented by Eq. (1).

$$traffic\_volume\_lag\_k = traffic\_volume_{(t-k)} \tag{1}$$

where, k denotes the lags in hours, $traffic\_volume_{(t-k)}$ denotes the traffic volume at time $t-k$, indicating the value of the traffic volume variable k periods before the current time t. This is particularly useful for predicting current traffic conditions based on recent patterns. Aggregated features like the daily average traffic volume are also created by averaging traffic data over a day, which helps smooth out short-term fluctuations and captures overall daily trends as illustrated by Eq. (2).

$$daily\_avg\_traffic = \frac{1}{N}\sum_{i=1}^{N} traffic\_volume_i \tag{2}$$

where N is the total number of observations. Furthermore, interaction features are engineered by combining different factors, such as the interaction between the hour of the day and weather conditions, to capture the combined effect on traffic patterns. Normalization standardizes the features by subtracting the mean and dividing by the standard deviation.

$$scaled\_feature = \frac{feature\_mean}{std\_dev} \tag{3}$$

Feature embedding is particularly useful when dealing with categorical features that have a large number of unique values, such as Junction IDs in traffic data. By converting these categorical variables into dense vectors, feature embedding allows the model to learn complex relationships within the data. Junction IDs are categorical variables representing different traffic junctions. Temporal features are represented by time-based encodings. Using an embedding layer, each unique Junction ID is mapped to a continuous vector in a high-dimensional space as Eq. (4). This allows the model to capture similarities between different junctions.

$$Embedding\ matrix, E \in \mathbb{R}^{V\times d} \tag{4}$$

where, V is the number of unique junctions and d is the dimensionality of the embedding vector. Each junction i is represented by Eq. 5.

$$E_i \in \mathbb{R}^d \tag{5}$$

The embedding matrix E is initialized randomly and is learned during the training process. The embeddings are updated to minimize the loss function, allowing the model to capture relevant patterns in the data as Eq. (6).

$$embedded\_vector = E_i \tag{6}$$

D. Proposed Traffic Congestion Prediction Model

The attention-based multilayer GRU model is designed to handle sequential data, such as traffic flow over time, by utilizing both the GRU for capturing temporal dependencies and an attention mechanism to focus on the most relevant time steps. This approach helps in improving the model's predictive performance by selectively concentrating on important historical data.

1) Attention based multi- layer gated recurrent unit: The GRU is a variant of Recurrent Neural Networks (RNNs) designed to handle sequential data effectively [23]. It employs update gates and reset gates to manage the flow of information through the network. Fig. 18 illustrates the GRU cell architecture.

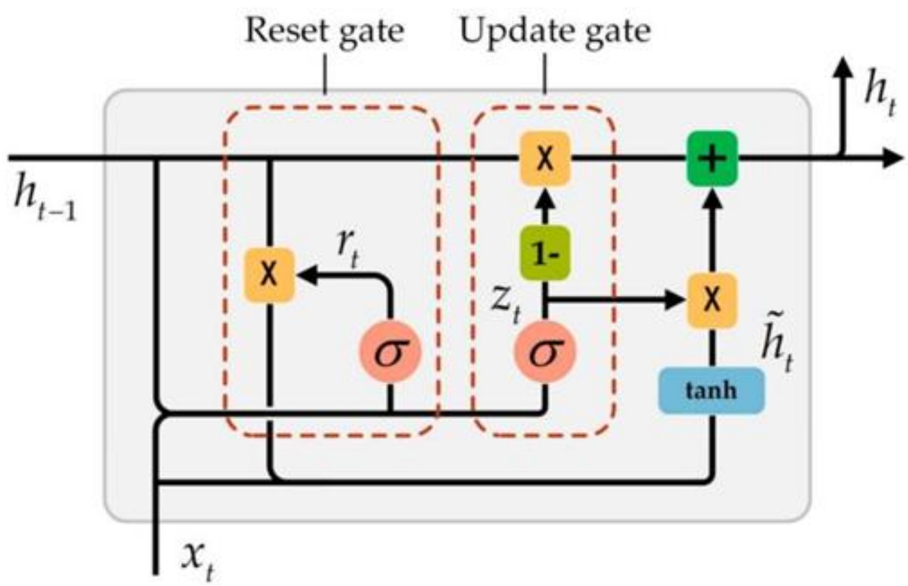

Fig. 18. Cell structure of GRU.

In a GRU cell, a gate controller, represented by z, oversees the operation of both the input and forget gates. When z equals 1, the forget gate is turned off, enabling the input gate to function. On the other hand, when z equals 0, the forget gate is activated and the input gate is turned off. At every time step, the GRU cell maintains the memory from the previous time step (t - 1) while resetting the input for the current step. The operation of the GRU cell is governed by the following equations: Eq. (7) through Eq. (9).

Reset Gate (r_{ti}),

$$r_{ti} = \sigma(W_r.\left[h_{ti-1,}x_{ti}\right] + b_r) \tag{7}$$

Update Gate (z_{ti}),

$$z_{ti} = \sigma(W_z.\left[h_{ti-1,}x_{ti}\right] + b_z) \tag{8}$$

Candidate Activation (h_{ti}),

$$h_{ti} = (1 - z_{ti}) * h_{ti-1} + z_{ti} * \overline{h_{ti}} \tag{9}$$

The GRU network is employed to forecast traffic congestion levels at 24 distinct time intervals, spanning from one hour to one day ahead, for model optimization. This GRU model features two hidden layers, with the input layer having 18 nodes and each hidden layer containing 13 nodes, as determined by the two-thirds rule applied to the input layer size and the inclusion of the output layer size. When predicting traffic congestion, extending the input sequence in a GRU network can reduce prediction accuracy because the model tends to equally weight all input variables despite their varying relevance to the forecasted outcomes. To mitigate this issue, an attention mechanism is incorporated, enabling the model to prioritize the most pertinent input variables.

The attention mechanism comprises an encoder that creates an attention vector from the input data and a decoder that generates a hidden state based on the encoder's output [24]. The encoder produces hidden states h_t for each time step t. These hidden states are segmented, and the encoder calculates an attention score $e_{t'}^{t}$ for each segment's hidden state using the hidden state from the preceding decoder segment. The attention score is calculated as specified in Eq. (10).

$$e_{t'}^{t} = score(h_{t'}, h_t) \tag{10}$$

This process creates an attention vector through a Softmax operation on the attention scores as given by Eq. (11).

$$\alpha_{t'}^{t} = \frac{\exp(e_{t'}^{t})}{\sum_k \exp(e_{t'}^{t})} \tag{11}$$

The context vector $c_{t'}$ is then computed as a weighted sum of the encoder hidden states as in Eq. (12), where the weights are the attention weights.

$$c_{t'} = \sum_t \alpha_{t'}^{t} h_t \tag{12}$$

This method ensures that the encoder concentrates on input variables that are closely related to the predicted value whenever the decoder generates an output. The decoder uses the context vector $c_{t'}$ along with its previous hidden state to generate the next hidden state and output as in Eq. (13).

$$h_{t'} = \tanh(W_h[c_{t'}, h_{t'-1}] + b_h \tag{13}$$

To enhance the accuracy of traffic congestion predictions, the attention mechanism is designed to focus on highly correlated input variables. The size of the attention window is set to 96 for this specific model configuration. The attention based GRU model architecture is illustrated by Fig. 19.

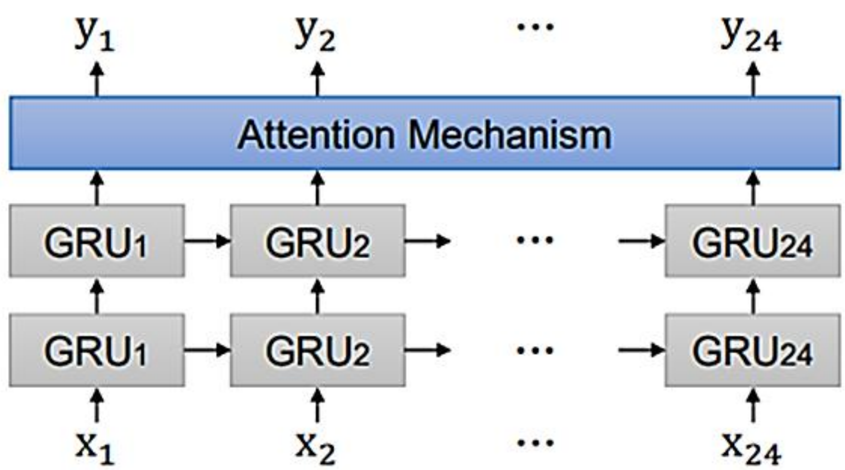

Fig. 19. Attention-based GRU model architecture.

Thus, the attention-based multi-layer GRU captures and processes the temporal dependencies in traffic data, the dense layer integrates and refines these features, and the output layer generates the final traffic prediction based on the transformed data.

E. Hardware and Software Setup

The experimental setup included an NVIDIA GeForce GTX 1080Ti GPU, an Intel Core i7 processor, 32GB of RAM, and utilized python and the Keras library with TensorFlow as the backend. Keras' intuitive interface, combined with the computational power of Google Colab, enabled efficient model training with GPU support. The dataset was split for training and testing to ensure robust evaluation. Table III outlines the hyperparameters chosen for the training phase, which played a critical role in fine-tuning the model's performance on the traffic prediction dataset, ensuring both accuracy and rapid convergence.

TABLE III. HYPERPARAMETER SPECIFICATION

Hyperparameters	Values
Loss function	Mean squared error
Activation function	Sigmoid
Batch size	32
Epochs	150
Optimizer	Adam
Learning rate	0.001

IV. Results and Discussion

The model's performance was assessed by means of various metrics, as detailed in Table IV. Table V shows the model performance assessment using evaluation metrics for traffic congestion prediction.

TABLE IV. Evaluation Metrics

Metric	Equation
Mean Squared Error (MSE)	$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$
Mean Absolute Error (MAE)	$MAE = \frac{1}{n}\sum_{i=1}^{n}\lvert y_i - \hat{y}_i\rvert$
Coefficient of Determination (R^2)	$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$
Mean Absolute Percentage Error (MAPE)	$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left\lvert\frac{y_i - \hat{y}_i}{y_i}\right\rvert \times 100$
n is the number of observations, y_i is the actual value, $\hat{y}_i$ is the predicted value	

TABLE V. Performance Assessment using Evaluation Metrics

Evaluation metrics	Values
MSE	0.9678
MAE	0.4322
R^2	0.8686
MAPE	6%

The evaluation metrics demonstrate that the model excels in predicting traffic congestion with impressive performance. The MSE of 0.9678 indicates that the model generates predictions with minimal squared errors, reflecting a high degree of accuracy in capturing the nuances of traffic patterns. The MAE of 0.4322 underscores the model's strong predictive capability, with average deviations being relatively low and manageable. The R^2 of 0.8686 reveals that the model accounts for approximately 87% of the variability in traffic congestion, showcasing its effectiveness in explaining the observed data. Additionally, the MAPE of 6% demonstrates that the model's predictions are, on average, within 6% of the actual values, highlighting its robustness and reliability. Overall, these results confirm that the model delivers highly accurate and reliable predictions for traffic congestion, marking it as an exceptional tool for traffic forecasting. Fig. 20 illustrates the predicted and actual values of the suggested attention based multilayer GRU model for traffic congestion prediction.

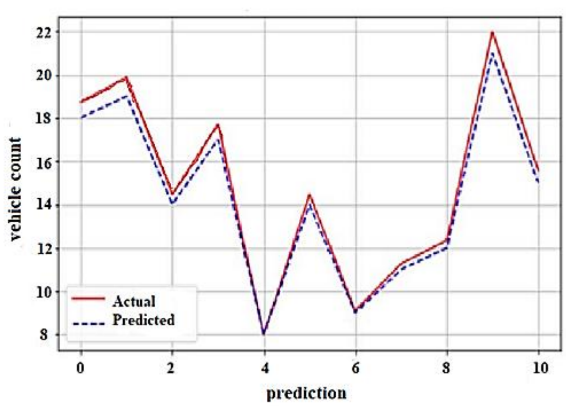

Fig. 20. Predicted vs. actual values of proposed model.

The attention-based multilayer GRU model showed considerable efficiency in predicting traffic congestion. The model demonstrated robust prediction accuracy and reliability, evidenced by an MSE of 0.9678, an MAE of 0.4322, and an R^2 value of 0.8686. A MAPE of 6% further demonstrates the model's resilience while processing real traffic data. The integration of the attention mechanism within the multilayer GRU architecture enables the model to concentrate on the most pertinent temporal patterns and features in traffic data, enhancing prediction accuracy and ensuring effective capture of both short-term fluctuations and long-term dependencies in congestion patterns. The attention-based multilayer GRU model is an exceptionally excellent method for predicting traffic congestion. An attention mechanism in a multilayer GRU aids in the prediction of traffic congestion by allowing the model to concentrate on the most important features and pertinent temporal patterns in the data. The attention mechanism, in contrast to conventional GRU models, gives significant time steps, ensuring that significant congestion-related occurrences or patterns are given priority during prediction. This enhances the model's ability to capture long-term dependencies while reducing the influence of irrelevant or noisy inputs. The multilayer GRU utilizes the attention mechanism to enhance prediction accuracy and interpretability, making it especially suitable for the intricate and variable nature of traffic congestion prediction.

V. Conclusion

The proposed attention-based multilayer GRU model offers a substantial improvement over traditional traffic management methods by addressing the dynamic and complex nature of traffic congestion. The model's ability to capture temporal dependencies and intricate traffic patterns through attention mechanisms enables more accurate predictions of traffic conditions and congestion levels. The model achieved a notable improvement in accuracy, with MAE and MSE values of 0.4322 and 0.9678, respectively. This enhanced predictive capability facilitates timely and efficient traffic management interventions, reducing travel times, minimizing fuel consumption, and lowering emissions. The effectiveness of the model in various urban scenarios demonstrates its potential to significantly improve overall traffic flow and urban mobility. Future research could explore integrating real-time data sources and extending the model's application to different traffic management systems to further enhance its effectiveness. Future studies should investigate transfer learning to adapt models for areas with scarce historical data and reduce dependence on computational resources. More thorough and useful solutions for traffic management systems will be ensured by integrating external factors like weather, road construction, and special events, as well as by creating reliable models for unexpected disruptions.

Acknowledgment

I would like to express my sincere gratitude to all those who contributed to the completion of this research paper. I extend my heartfelt thanks to my supervisor, my family, my colleagues and fellow researchers for their encouragement and understanding during the demanding phases of this work.

References

[1] Akhtar, M., & Moridpour, S. (2021). A review of traffic congestion prediction using artificial intelligence. Journal of Advanced Transportation, 2021(1), 8878011.

[2] Ranjan, N., Bhandari, S., Zhao, H. P., Kim, H., & Khan, P. (2020). City-wide traffic congestion prediction based on CNN, LSTM and transpose CNN. Ieee Access, 8, 81606-81620.

[3] Li, T., Ni, A., Zhang, C., Xiao, G., & Gao, L. (2020). Short - term traffic congestion prediction with Conv‐BiLSTM considering spatio - temporal features. IET Intelligent Transport Systems, 14(14), 1978-1986.

[4] Li, T., Ni, A., Zhang, C., Xiao, G., & Gao, L. (2020). Short - term traffic congestion prediction with Conv‐BiLSTM considering spatio - temporal features. IET Intelligent Transport Systems, 14(14), 1978-1986.

[5] Gollapalli, M., Musleh, D., Ibrahim, N., Khan, M. A., Abbas, S., Atta, A., ... & Omer, A. (2022). A Neuro-Fuzzy Approach to Road Traffic Congestion Prediction. Computers, Materials & Continua, 73(1).

[6] Li, L., Lin, H., Wan, J., Ma, Z., & Wang, H. (2020). MF-TCPV: a machine learning and fuzzy comprehensive evaluation-based framework for traffic congestion prediction and visualization. IEEE Access, 8, 227113-227125.

[7] Li, L., Dai, F., Huang, B., Wang, S., Dou, W., & Fu, X. (2024). AST3DRNet: Attention-Based Spatio-Temporal 3D Residual Neural Networks for Traffic Congestion Prediction. Sensors, 24(4), 1261.

[8] Tsalikidis, N., Mystakidis, A., Koukaras, P., Ivaškevičius, M., Morkūnaitė, L., Ioannidis, D., ... & Tzovaras, D. (2024). Urban traffic congestion prediction: a multi-step approach utilizing sensor data and weather information. Smart Cities, 7(1), 233-253.

[9] Jiang, W., Han, J., Liu, H., Tao, T., Tan, N., & Xiong, H. (2024, August). Interpretable cascading mixture-of-experts for urban traffic congestion prediction. In Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (pp. 5206-5217).

[10] Hao, M. J., & Hsieh, B. Y. (2024). Greenshields Model-Based Fuzzy System for Predicting Traffic Congestion on Highways. IEEE Access.

[11] Zhang, T., Wang, J., Wang, T., Pang, Y., Wang, P., & Wang, W. (2024). A deep marked graph process model for citywide traffic congestion forecasting. Computer - Aided Civil and Infrastructure Engineering, 39(8), 1180-1196.

[12] S Jasim, M., Zaghden, N., & Salim Bouhlel, M. (2024). Improving Detection and Prediction of Traffic Congestion in VANETs: An Examination of Machine Learning. International Journal of Computing and Digital Systems, 15(1), 947-960.

[13] Arabiat, A., & Altayeb, M. (2024). Assessing the effectiveness of data mining tools in classifying and predicting road traffic congestion. Indonesian Journal of Electrical Engineering and Computer Science, 34(2), 1295-1303.

[14] 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.

[15] Jin, G., Liu, L., Li, F., & Huang, J. (2023, June). Spatio-temporal graph neural point process for traffic congestion event prediction. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 37, No. 12, pp. 14268-14276).

[16] Pan, Z., Sharma, A., Hu, J. Y. C., Liu, Z., Li, A., Liu, H., ... & Geng, T. (2023, June). Ising-traffic: Using ising machine learning to predict traffic congestion under uncertainty. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 37, No. 8, pp. 9354-9363).

[17] Zhang, T., Xu, J., Cong, S., Qu, C., & Zhao, W. (2023). A hybrid method of traffic congestion prediction and control. IEEE Access, 11, 36471-36491.

[18] Wang, X., Zeng, R., Zou, F., Liao, L., & Huang, F. (2023). STTF: An efficient transformer model for traffic congestion prediction. International Journal of Computational Intelligence Systems, 16(1), 2.

[19] https://www.kaggle.com/datasets/fedesoriano/traffic-prediction-dataset.

[20] https://www.kaggle.com/datasets/alioraji/weather-data-nov-2015

[21] Jawad, Y. K., & Nitulescu, M. (2024). Improving Driving Style in Connected Vehicles via Predicting Road Surface, Traffic, and Driving Style. Applied Sciences, 14(9), 3905.

[22] Liu, Y., Lyu, C., Liu, X., & Liu, Z. (2020). Automatic feature engineering for bus passenger flow prediction based on modular convolutional neural network. IEEE Transactions on Intelligent Transportation Systems, 22(4), 2349-2358.

[23] Mahjoub, S., Chrifi-Alaoui, L., Marhic, B., & Delahoche, L. (2022). Predicting energy consumption using LSTM, multi-layer GRU and drop-GRU neural networks. Sensors, 22(11), 4062.

[24] Zou, X., Zhao, J., Zhao, D., Sun, B., He, Y., & Fuentes, S. (2021). Air quality prediction based on a spatiotemporal attention mechanism. Mobile Information Systems, 2021(1), 6630944.