SAJSET
ISSN: 2992-3581

Savannah Journal of Science and Engineering Technology
Volume: 03; Issue: 04; Pages: 270-278; 2025.
www.sajsetjournal.com.ng

DOI: xx.xxxx.x/sajset-03-2025-0036

SAJSET
ISSN: 2992-3581

A framework for traffic flow forecasting using graph attention networks with external factor integration

[1]Jauro, S. S., [1,2]Suleiman, Z., [2]Baba, U. A. and [2]Abdullahi, M. A.
[1]Department of Computer Science, Gombe State University, Gombe, Gombe State Nigeria
[2]Department of Computer Science, North-Eastern University, Gombe State Nigeria
ssjauro@gsu.edu.ng, usman.baba@neu.edu.ng, mahmoud.abdullahi@neu.edu.ng

Paper History
Received: 21st July, 2025
Accepted: 19th August, 2025
Published: August, 2025

Corresponding author
Suleiman, Z.
zakari.suleiman@neu.edu.ng

Abstract:
External factors against traffic patterns are one of the most important parameters that influences the prediction of traffic flow rates. The study aimed to offer advancement over Graph Convolutional Networks (GCNs) by overcoming the limitation of GCNs, which depend on node degrees to assess importance. The study proposes traffic forecasting framework using Graph Attention Networks (GATs), incorporating external factors such as weather and public events, which significantly influence traffic patterns. Recent advancements in machine learning, particularly deep learning and graph neural networks, have significantly improved traffic forecasting accuracy. However, these models often neglect these external factors. The framework using deep learning techniques, with a focus on Graph Attention Networks (GATs) integrates spatial, temporal and contextual data. This led to achieving forecasting accuracy with superior performance in terms of metrics like Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). The study achieved 7.017 × 10 ⁻7 MSE, 0.0010RMSE and 0.176 MAPE. Experimental results demonstrate the effectiveness of the proposed approach in handling real-world traffic complexities. In addition, the study evaluates the effect of external factors on traffic predictions, and the findings emphasize the importance of incorporating contextual data in traffic forecasting to enhance prediction reliability. Comparative analysis shows that the proposed GAT-based framework outperforms existing methods, including STGCN and GRAM-ODE. However, this reveals new challenges and research opportunities for future studies.

Keywords: *Flow rates, GATs, Public event, Traffic, Patterns, Weather*

1. Introduction

One of the newest developments in machine learning is deep learning. Using multilayered non-linear functions, it can capture the abrupt discontinuities in traffic flows. Deep learning model applications in the transportation sector will enable us to handle larger and more complicated data sets [1] deep neural networks have become increasingly prevalent in resolving traffic forecasting issues. One method entails utilizing multilayered deep neural networks (MLDNN) coupled with a congestion index (CI) grounded on traffic density to directly project traffic congestion [2]. Another technique is to apply deep learning approaches to estimate future traffic patterns based on past and current data [3]. Graph neural networks (GNNs) have also been implemented by Liu, *et al.* [4] to acquire representations of road networks and capture spatiotemporal correlations within traffic. Furthermore, deep learning networks founded on transformers was proposed by Hu, *et al.* [5] for precise prediction.

Traffic congestion, due to its significant impact on the economy has been a serious challenge as it causes increase in fuel consumption and waste of time. The advancement in machine learning techniques and the increasing availability of sensor data have motivated novelties in traffic forecasting. Traffic congestion is a significant problem with negative economic and environmental impacts, and one effective solution is future traffic prediction using deep neural network models [6]. The work presents a comprehensive survey of deep neural network architectures used in traffic flow prediction, categorizes and describes the literature, and discusses the challenges and future directions in this field. The authors emphasize the need for a standardized approach in the field of traffic flow prediction, including the establishment of a comprehensive benchmark dataset that incorporates various traffic factors and external data, such as social media, weather, and accidents data.

Troia, *et al.* [7] focuses on using deep learning methods, specifically Gated Recurrent Units (GRU), to predict traffic matrices in order to optimize resource allocations in optical backbone networks. The GRU achieved high accuracy in predicting traffic, and the predictions were used to dynamically allocate network resources, resulting in a significant saving of available capacity. Despite considering extra capacity, the results reported some traffic loss over the testing set, indicating a limitation.

Existing models such as STCNN, STGCN, and GRAM-ODE have demonstrated the capability to capture spatiotemporal features in traffic data. However, most methods lack incorporation of external variables or rely deeply on stationary symbols of the road network [8]. A lot of research has been done on traffic flow rate prediction. However, several important aspects remain inadequately covered in the literature [9]. These include the need to better explore the effect of external data like weather, special event, condition of the road, etc., as well as the inadequate study of the best training epochs and sensor density within the datasets that are currently accessible. Hence the need to further research to consider and incorporate some factors in addition to other traffic patterns to improve accuracy in prediction. Thus, these gaps informed the idea to propose framework that takes into account external factors that may influence traffic to enable proactive decision-making in order to enhance traffic operation and management. Graph Attention Networks (GATs) can be a sophisticated method for traffic forecasting that takes outside factors into account [10]. GATs discourse these limitations by applying learnable attention mechanisms to dynamically weigh the significance of nodes and their features, including external inputs like weather and event data [11].

Lv, *et al.* [12] presented a novel method to forecast traffic flow using deep learning methods. Stacked auto encoder model was used by the authors to capture spatial and temporal relationships, allowing the model to gain a thorough comprehension of traffic flow patterns. The training of the model was carried out in an acquisitive layer-wise manner.

Kang, *et al.* [13] introduced a Long Short-Term Memory (LSTM) recurrent neural network for the purpose of traffic flow prediction. This network takes into consideration various inputs including speed, flow, as well as occupancy at individual sensor. Additionally, the authors incorporate the variables from the sensors located upstream as well as the downstream in order to capture the spatial effects. The findings of the study demonstrate the enhanced accuracy in prediction achieved by including this spatial information.

Troia, *et al.* [7] focuses on using deep learning methods, specifically Gated Recurrent Units (GRU), to predict traffic matrices for resource allocations optimization in an optical backbone network. The GRU achieved high accuracy in predicting traffic, and the predictions were used to dynamically allocate network resources, resulting in a significant saving of available capacity. The work also discusses the unpredictability of network traffic and the challenges in accurately predicting it. Despite considering extra capacity, the results reported some traffic loss. However, the overall performance of the deep learning-based prediction model and resource allocation was promising, and it represents one of the first attempts to optimize network operations using machine learning. The predictions obtained from the deep learning model were used to dynamically and proactively allocate network resources, significantly saved 66.3% of the capacity available in the network. This dynamic allocation helped manage unexpected traffic peaks. The proposed deep learning algorithm showed promising performances in both traffic prediction and resource allocation, indicating its effectiveness in optimizing the operator's network. The work highlights the potential of using machine learning methods, particularly deep learning, for network optimization, representing one of the first attempts to optimize network operations with machine learning.

Qu, et al. [14] employed deep neural network for the purpose of predicting long-term daily traffic flows. This was accomplished by utilizing both temporal correlation and related factors, such as day of the week, season and weather conditions. In order to explore relationship between those factors, the authors opted to train a predictor by using multi-layer supervised learning approach. In order to reduce the time required for training, they introduced batch training method. By conducting a case study in Seattle, the researchers were able to demonstrate that their proposed method outperformed conventional forecasting methods in terms of prediction accuracy.

He, *et al.* [8] proposed a model called STCNN (Spatiotemporal Convolutional Neural Network) for accurate prediction of long-term traffic, which has been challenging because of the dynamic nature of traffic. STCNN captures spatiotemporal correlations and periodic traffic forms to forecast long-term traffic. The results show that STCNN surpass other state-of-the-art models.

GRAM-ODE (Graph-based Multi-ODE Neural Networks) is a new architecture proposed by Liu, *et al.* [4] for spatiotemporal traffic forecasting. It combines the graph neural networks (GNNs) and the neural ordinary differential equations (ODE) to overcome the over-smoothing problem in deep architectures and capture both the complex spatiotemporal dependencies at the local and global levels. This proposed method was evaluated on six different real-world datasets; extensive experiments were conducted to demonstrate its superior performance compared to the state-of-the-art baselines. The work addresses challenge of long-range traffic forecasting by proposing this novel architecture called GRAM-ODE.

Belt, *et al.* [15] focuses on using STGNN (Spatial Temporal Graph Neural Network) to forecast traffic flow rates on an hourly basis in Amsterdam, based on data collected from 34 traffic sensors. While generally, the STGNN model did not perform better than a baseline seasonal naive model, it showed better performance for sensors situated on the road network near to one another. The model was trained using the collected data, which includes information about the traffic flow rates from the sensors. The model takes into account spatial and temporal connections between sensors and uses graph neural networks to record the dependencies and patterns in the data. Performance of the STGNN model was evaluated based on its capacity to predict traffic flow rates with accuracy and it was in contrast to the baseline model's performance.

2. Methodology

The Proposed GAT-based Traffic Forecasting Framework employs a Graph Attention Network (GAT) to forecast traffic, involving data exploration through visualization, feature engineering, and exploratory data analysis. External factors are integrated using data augmentation, while MinMax Scaling is applied for data transformation. The GAT-based model is proposed for prediction, and its performance is assessed using Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). Figure 1 described a step-by-step approach ensuring consideration of external factors in traffic forecasting.

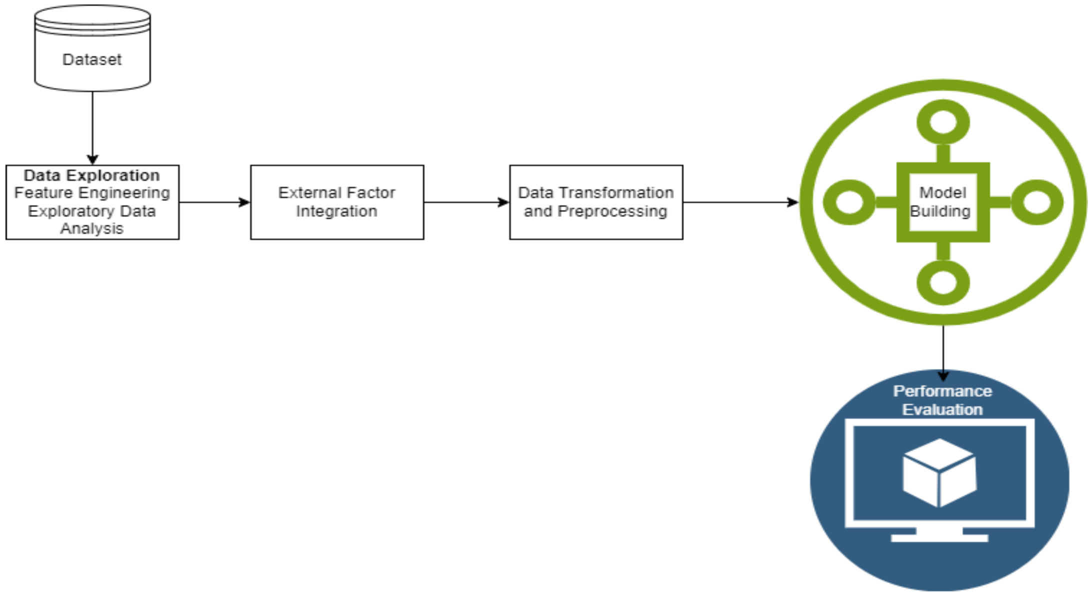

Figure 1: GAT-based Framework for Traffic Forecasting

2.1 Dataset description

Three datasets were sourced from Kaggle as reported in Fedesoriano [16], Manga [17] and Zia [18]:

a. Traffic Prediction Dataset: hourly vehicle counts at junctions comprises 48,120 observations
b. WeatherHistory Dataset: meteorological data containing 96,453 observations
c. Traffic Time Series Dataset: synthetic data including special events and weather variables containing 8,736 observations

2.2 External factor Integration

External factors were incorporated into the traffic dataset using feature engineering and label encoding. MinMax normalization was applied to scale features. The graph representation included traffic nodes, external factor nodes, and weighted edges based on spatial proximity and traffic correlation.

In order to get insights right away and pinpoint regions or patterns to further investigate, data exploration was done in order to examine and visualize the data. The exploration was carried out using techniques such as Magdum [19]:

a. Time-Series data visualization: for Trend Analysis (to identify long-term patterns), Seasonality Detection (to check for repeating patterns over time such as daily, weekly, yearly) and Autocorrelation (to discover dependencies between observations over time)
b. Feature Engineering: for Transformations (scaling and normalization), Creation (creating new features using domain knowledge or feature extraction techniques)
c. Feature Correlation, Grouping and Aggregation.

2.3 Data transformation and preprocessing

The study employs the MinMax Scaler as its preprocessing technique. MinMax Scaler transforms data by scaling features within a defined range. The MinMax scaling is performed using equation 1 to 3 de Amorim [20]:

$$X_{std} = \frac{X - X_{min}(x,\ axis{=}0)}{X_{max}(x,\ axis{=}0) - X_{min}(x,\ axis{=}0)} \quad (1)$$

$$X_{scaled} = X_{std}\ (max\ -\ min)\ +\ min \quad (2)$$

Where: min, max = feature_range, X_{min}(axis = 0): Minimum feature value and X_{max}(axis=0): Maximum feature value. The syntax is: class sklearn. Pre-processing. MinMaxScaler (feature range = 0, 1, *, copy = True, clip = False)

The collected data had undergone cleaning to handle missing values, outliers, and inconsistencies. The normalization technique used is Min-Max Normalization (MinMax Scaling):

$$X_{normalised} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (3)$$

Where: X is the original value, X_{min} is the minimum value in the column and X_{max} is the maximum value in the column. This helps standardize feature values, making them comparable for GAT training.

The data was normalized and synchronized to a common time frame, ensuring consistency. The main techniques used are data aggregation, pivoting, missing value handling, normalization, and feature alignment. These steps help prepare the data as a time-series node-feature matrix, which is crucial for applying a Graph Attention Network (GAT) model to forecast traffic and evaluate external factors.

The data is transformed and prepared for graph representation, which served as input to GAT-based Model

a. Traffic Nodes: Each intersection and road segment in the traffic network was represented as a node in the graph.
b. External Factor Nodes: External factors were incorporated into the graph as additional nodes or as attributes associated with traffic nodes.
c. Edge Definition: The connections between traffic nodes are represented by edges, indicating the flow of traffic. The weight of edges can be determined based on geographical proximity, traffic flow correlation, or other criteria.
d. Graph Representation: The graph represents the spatial and temporal relationships among traffic nodes and external factors, which serves as input to the Graph Attention Networks.

2.4 Model building (GAT architecture)

The model consists of two GAT layers using multi-head attention to capture temporal and spatial connections. Attention weights were learned for neighboring nodes, enabling adaptive focus on relevant traffic and environmental contexts. Figures 2 and 3 break down the Proposed Graph Attention Network model and its operations, describing the flow of information through the layers. In the GAT layers, the model computes attention scores (weights) for each pair of nodes based on their features, and uses these scores to adjust how much influence neighboring nodes should have on each other. The edge connectivity (as defined by the edge_index tensor) dictates which nodes will receive the attention mechanism.

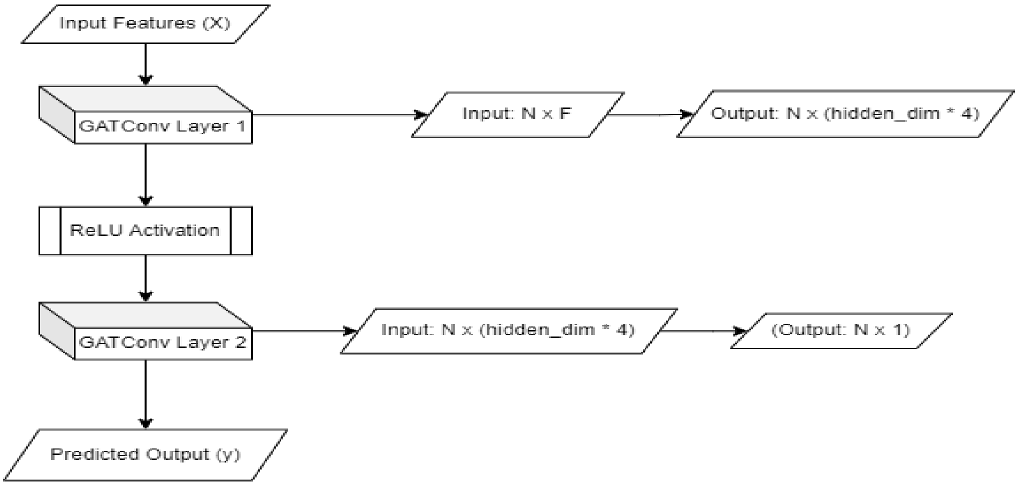

Figure 2: GAT Model Layers Visualization

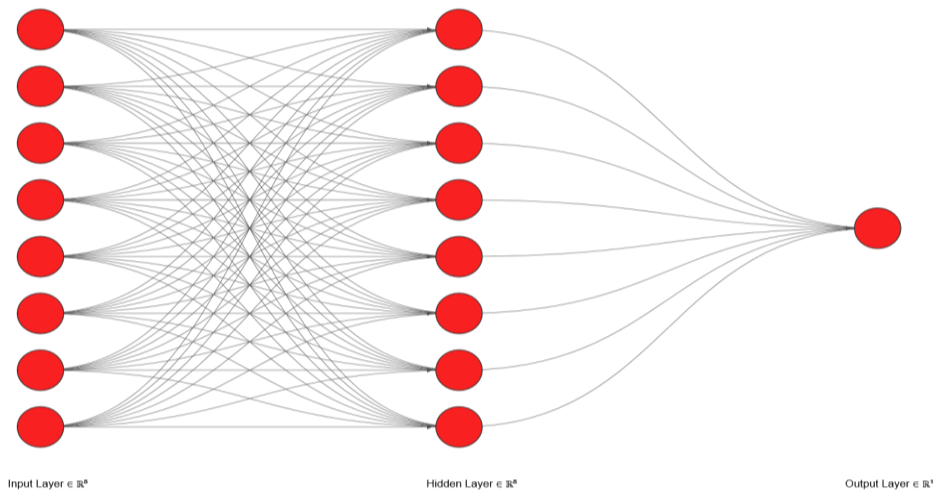

Figure 3: Neural Network Architecture

2.5 Training and optimization

The Adam optimizer was used to train the model, regularization techniques such as dropout, and a learning rate of 0.001. Training was conducted over 5000 epochs on a machine with Intel Core i7 processor and 8GB RAM. The Adam Optimizer (Adaptive Moment Estimation) was chosen in this study over other optimizers because it is a fast, memory-efficient algorithm for training neural networks, combining Root Mean Squared Propagation (RMSprop) and Stochastic Gradient Descent (SGD) with momentum for adaptive learning rates and bias correction. It ensures stable convergence and works well with a default learning rate of 0.001 and epsilon around 1e-7. Regularization techniques like dropout or weight decay help reduce over fitting.

2.6 Performance metrics

Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) were used to assess the performance of the proposed framework:

2.6.1 Root mean squared error (RMSE):

The square root of the average squared discrepancies between the true and forecasted values is known as the RMSE. It is computed using equation 4 [21]:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{true,i} - y_{pred,i}\right)^2} \quad (4)$$

Where $y_{true,i}$ is the true value of the i-th data point, and $y_{pred,i}$ is the predicted value of the i-th data point.

2.7 Mean absolute percentage error (MAPE):

The MAPE measures the average absolute percentage difference between the predicted and true values. It is calculated using equation 5 [21]:

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_{true,i}-y_{pred,i}}{y_{true,i}}\right| \times 100 \quad (5)$$

Where $y_{true,i}$ is the true value of the i-th data point, $y_{pred,i}$ is the predicted value of the i-th data point.

2.8 Mean Squared Error (MSE):

The MSE is the mean of the sum of all squared errors. It is calculated using equation 6 [21]:

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y - \hat{y})^2 \quad (6)$$

Where N is the number of data points, y_i is the actual (true) value, and $\hat{y}_i$ is the predicted value

3. Result and discussion

The result of the proposed GAT-Based framework achieved a Test Lost (MSE) of 0.0094 and RMSE of 0.097. In addition, by incorporating "Weather History", the proposed framework offers an impressive performance based on the values of the performance metrics as shown in Table 4.1. Integrating external factors like weather and events significantly enhanced forecasting performance. Ablation studies showed that removing these factors led to a noticeable drop in accuracy, reinforcing their importance. The model was particularly adept at adjusting to fluctuations caused by rain, traffic incidents, and public events, which existing models struggled to capture. Compared to existing models such as STGCN (RMSE = 7.45, MAPE = 9.73%) and GRAM-ODE (RMSE = 3.34, MAPE = 3.83%), the proposed model greatly achieved better performance in all metrics.

Table 1: Results of the proposed GAT-Based framework

MSE	RMSE	MAPE (%)
0.0094	0.097	0.159
7.017×10^{-7}	0.0010	0.176

3.1 Comparison of MSE of the proposed GAT-based with existing techniques

The performance of the Proposed GAT-Based Framework was compared with existing traffic forecasting techniques in terms of MSE (Mean Squared Error. It achieves 7.017×10^{-7}, while Zhang and Kabuka [22] that reported 3.76×10^{-5}. This suggests that the GAT-based model has significantly lower squared error, indicating better predictive accuracy. This probably could be attributed to the fact that the Proposed GAT-Based Model utilizes adequate training Epochs than D. Zhang and Kabuka [22]. Figure 4 shows the graphical representation of the Mean Squared Error comparison.

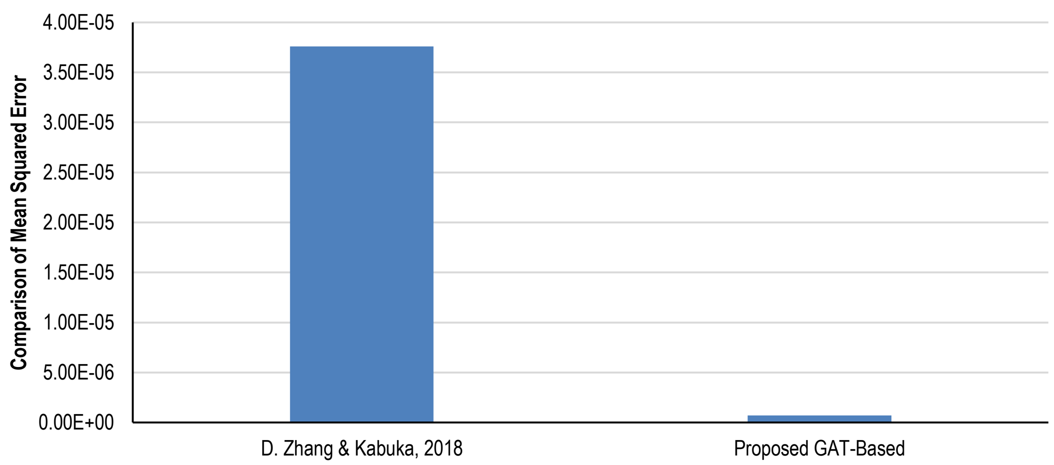

Figure 4: Comparison of Mean Squared Error (MSE)

3.2 Comparison of RMSE of the proposed GAT-based with existing techniques

The proposed GAT-Based Model's output was contrasted with existing techniques in terms of the Root Mean Squared Error. It offers the lowest RMSE of 0.0010, signifying that the model's forecasts closely matched the actual traffic volume values. Thus, the RMSE is much lower than the existing frameworks, indicating highly accurate predictions. Other models, especially STGCN, performed much worse with RMSE of 7.45 while GRAM-ODE reported 3.34 RMSE and STCNN offers 1.27 RMSE. Figure 5 shows a Bar Graph representation of the Root Mean Squared Error comparison.

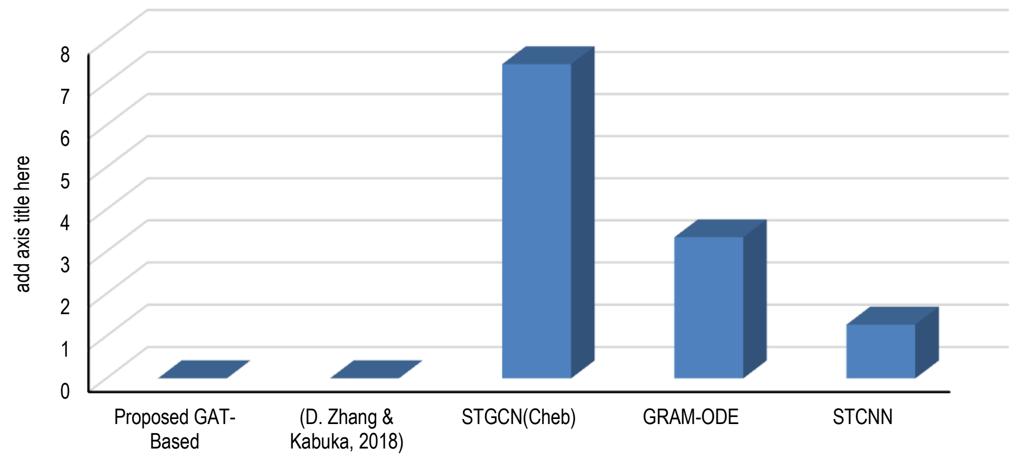

Figure 5: Comparison of Root Mean Squared Error (RMSE)

3.3 Comparison of MAPE of the proposed GAT-based with existing techniques

The Proposed GAT-Based Framework outperformed other existing techniques by achieving an extremely low MAPE of 0.176%, which is much better than STGCN that reported MAPE of 9.73% and GRAM-ODE with 3.83% MAPE. However, other models did not report MAPE. Figure 6 presents the Bar Graph representation of the Mean Absolute Percentage Error comparison. The proposed GAT-based framework achieves lower MSE, RMSE, and MAPE compared to other existing models. This can be attributed to several factors: The GAT-based model surpasses STGCN, STCNN, and GRAM-ODE in traffic forecasting due to its adaptive attention mechanism, which effectively captures spatial-temporal dependencies without relying on fixed adjacency structures. It excels in handling non-stationary traffic patterns, prioritizing high-impact roads, and overcoming limitations of previous models by dynamically learning relationships.

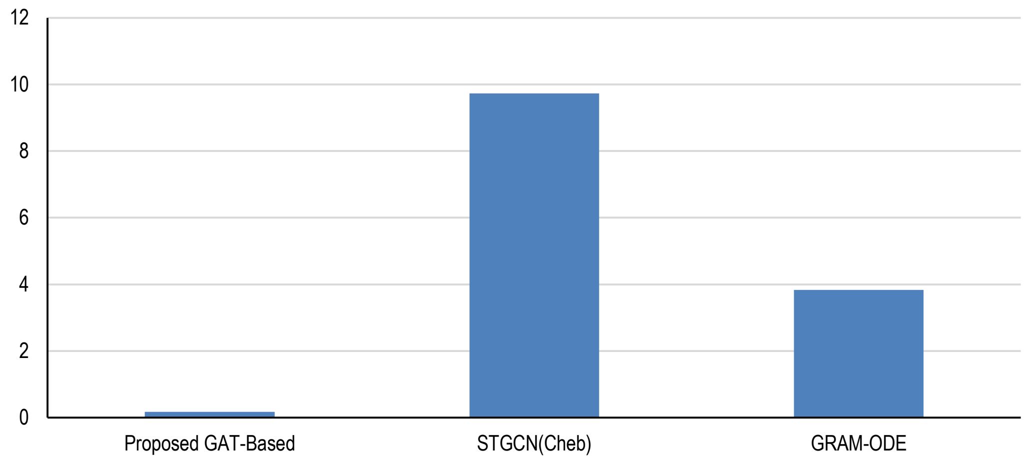

Figure 6: Comparison of Mean Absolute Percentage Error (MAPE)

3.4 Training Output

Figure 7 shows the Training loss and Test/Prediction loss over Epochs while figure 8 denotes the True Values versus Predicted values. The influence of external factors, weather conditions and events on traffic predictions was examined as shown in Figure 10 and Figure 11 respectively. Bar chart analyses revealed that snow increased traffic volume more than rain, while public events significantly raised traffic levels, highlighting the model's sensitivity to external stimuli.

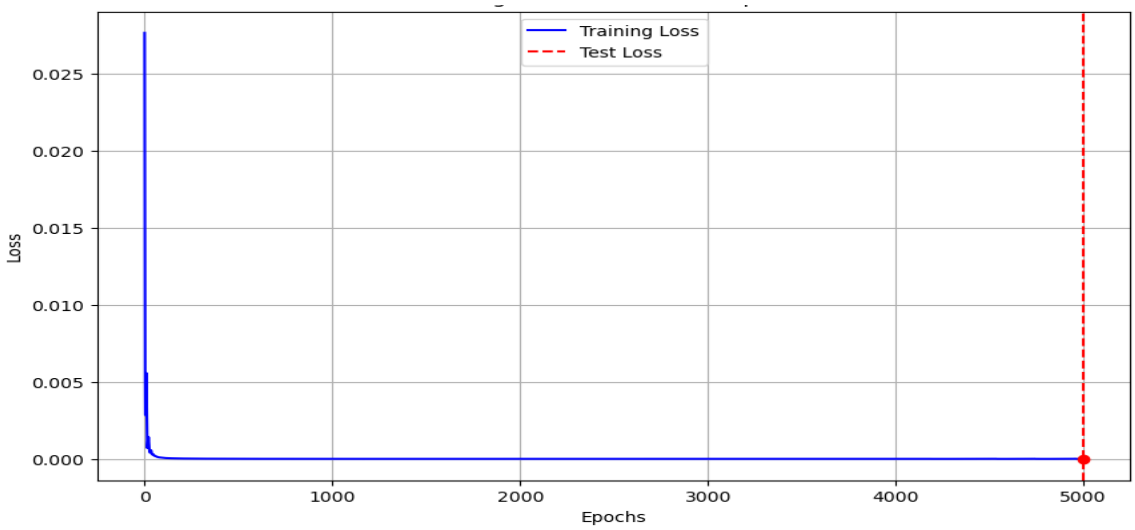

Figure 7: Training and Test Lost Over Epochs

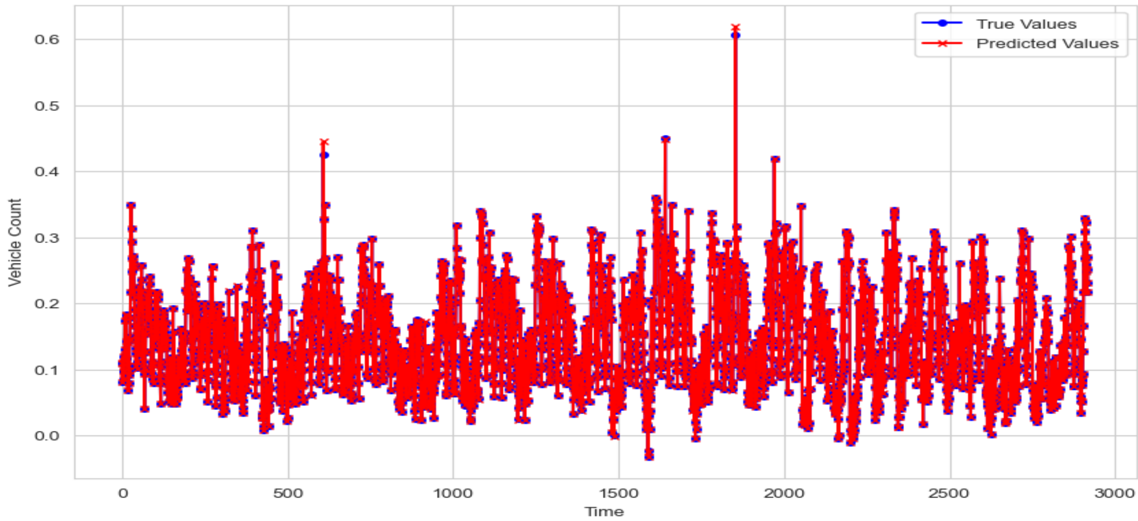

Figure 8: True vs Predicted Values

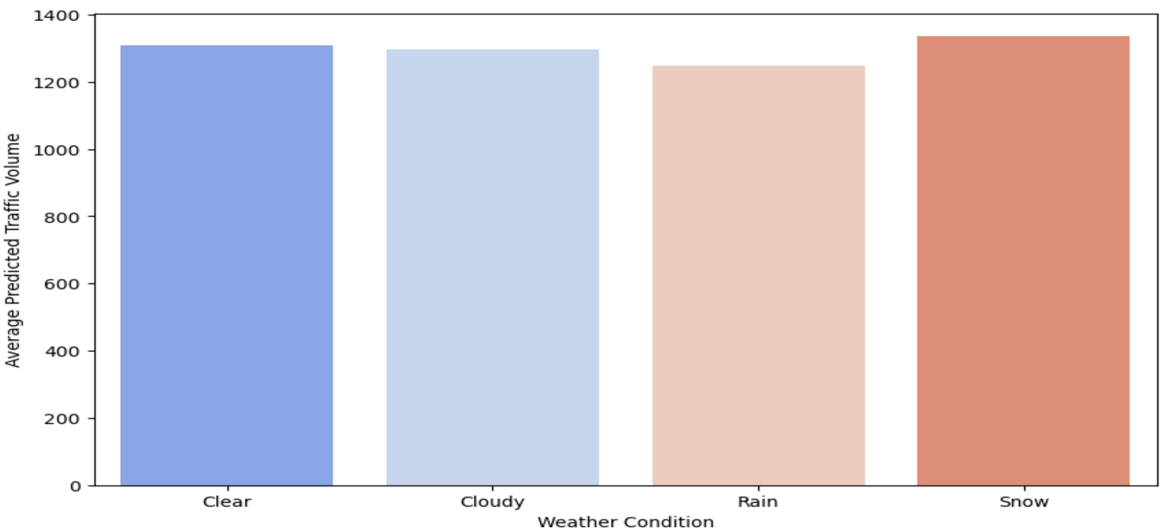

Figure 9: Impact of different weather conditions on traffic volume

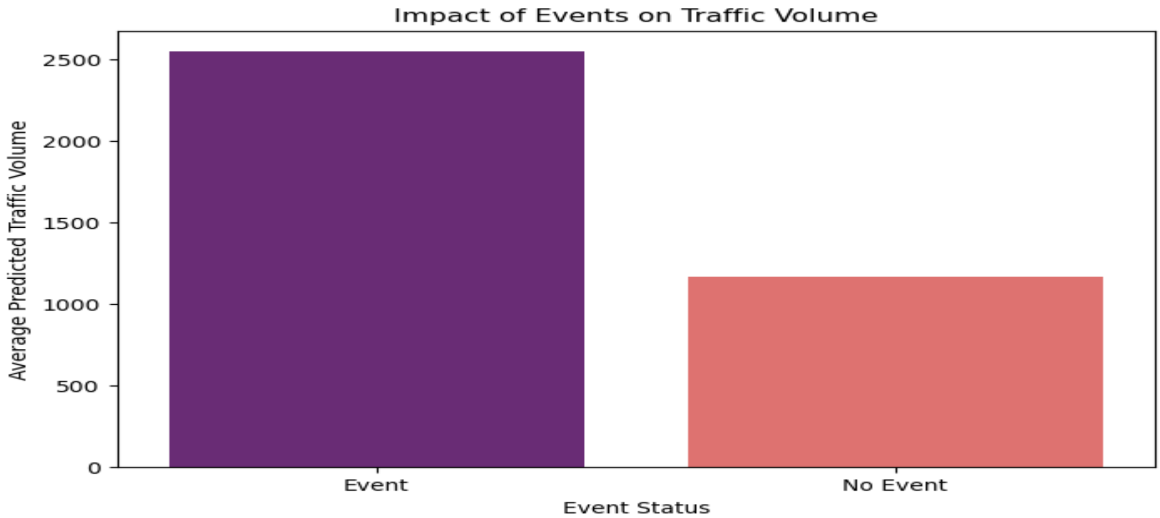

Figure 10: Impact of events on traffic volume

4. Conclusion

The study investigates gaps in existing deep learning models related to traffic forecasting, particularly in their inability to effectively incorporate external influences or manage sensor sparsity. Using three Kaggle-sourced datasets: Traffic Prediction, Weather History, and Traffic Time Series, the study proposes a framework that combines spatial, temporal, and other related data. The study concludes that Graph Attention Networks (GATs) offer a strong and dynamic solution for traffic forecasting, especially in environments influenced by variable external conditions. The proposed framework offers better performance compared other models like STGCN, GRAM-ODE, and STCNN, achieving very low MSE (7.017×10^{-7}), RMSE (0.0010), and MAPE (0.176%). The proposed traffic forecasting framework (GAT-based) effectively addresses the limitations of the existing models by integrating external factors into traffic predictions, which enabled the model to better adapt to real-world complexities such as weather disruptions and public events. The attention mechanism used by GATs allows the model to focus on the most influential data points, enhancing prediction accuracy. This highlights the effectiveness of GATs in capturing spatial-temporal dependencies and external influences. However, this uncovers new challenges and opens up avenues for further research exploration.

Recommendation

Incorporation of More External Factors: Future research should explore inclusion of additional circumstantial data including accident reports, social media sentiment, and construction zones to further improve forecasting accuracy.

Geographical Expansion: While this study focused on a limited dataset, testing the model on diverse geographical locations would validate its scalability and generalizability.

Real-Time Implementation: The model's real-time deployment optimization may be the main focus of future research by reducing computational overhead through model compression techniques.

References

[1]. Rahman, R. (2019). *Applications of deep learning models for traffic prediction problems*, *Electronic Theses and Dissertations*. 6286. Retrieved from https://stars.library.ucf.edu/etd/6286

[2]. Kumar, K., Kumar, M. and Das, P., (2024). Traffic congestion forecasting using multilayered deep neural network, *Transportation Letters*, *16*(6), 516–526. https://doi.org/10.1080/19427867.2023.2207278

[3]. Dharani, D. L., Bindu, B. H., Chaitanya, D., Yamuna, G. and Reshwanth, M., (2023). Website Traffic Forecasting Using Deep Learning, *International Journal for Research in Applied Science and Engineering Technology*, *11*(4), 2654–2658. https://doi.org/10.22214/ijraset.2023.50712

[4]. Liu, Z., Shojaee, P. and Reddy, C. K., (2023). *Graph-based Multi-ODE Neural Networks for Spatio-Temporal Traffic Forecasting*, *Transactions on Machine Learning Research*, 2305.18687v2, 1-22. https://doi.org/10.48550/arXiv.2305.18687

[5]. Hu, Y., Zhou, Y., Song, J., Xu, L. and Zhou, X., (2023). Citywide Mobile Traffic Forecasting Using Spatial-Temporal Downsampling Transformer Neural Networks. *IEEE Transactions on Network and Service Management*, *20*(1), 152–165. https://doi.org/10.1109/TNSM.2022.3214483

[6]. Tedjopurnomo, D. A., Bao, Z., Zheng, B., Choudhury, F. M., and Qin, A. K. (2022). A Survey on Modern Deep Neural Network for Traffic Prediction: Trends, Methods and Challenges. *IEEE Transactions on Knowledge and Data Engineering*, *34*(4), 1544–1561. https://doi.org/10.1109/TKDE.2020.3001195

[7]. Troia, S., Alvizu, R., Zhou, Y., Maier, G., and Pattavina, A. (2018). Deep Learning-Based Traffic Prediction for Network Optimization. *2018 20th International Conference on Transparent Optical Networks (ICTON)*, 1–4. IEEE. https://doi.org/10.1109/ICTON.2018.8473978

[8]. He, Z., Chow, C. Y., and Zhang, J. D. (2019). STCNN: A spatio-Temporal convolutional neural network for long-Term traffic prediction. *Proceedings - IEEE International Conference on Mobile Data Management*, *2019-June*, 226–233. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/MDM.2019.00-53

[9]. Alexander, T. D., Bao, Z., Murtaza, C. F., Qin, K., Alexander, D., Murtaza, F. and Murtaza Choudhury, F. (2022). A survey on modern deep neural network for traffic prediction: In *IEEE Transactions on Knowledge and Data Engineering* 34, *1544-1561*. Retrieved from https://ink.library.smu.edu.sg/sis_research/5995

[10]. Vrahatis, A. G., Lazaros, K., and Kotsiantis, S. (2024). Graph Attention Networks: A Comprehensive Review of Methods and Applications. *Future Internet*, *16*(9), 318. 1-34. https://doi.org/10.3390/fi16090318

[11]. Wang, Y. and Luo, Y., (2023). WeaGAN: Weather-Aware Graph Attention Network for Traffic Prediction," in Frontiers in *Artificial Intelligence and Applications*, IOS Press BV, Sep., 2591–2598. https://doi.org/10.3233/FAIA230565.

[12]. Lv, Y., Duan, Y., Kang, W., Li, Z. and Wang, F. Y., (2015). Traffic Flow Prediction with Big Data: A Deep Learning Approach. *IEEE Transactions on Intelligent Transportation Systems*, *16*(2), 865–873. https://doi.org/10.1109/TITS.2014.2345663

[13]. Kang, D., Lv, Y. and Chen, Y., (2017). Short-term traffic flow prediction with LSTM recurrent neural network. *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, 1–6. IEEE. https://doi.org/10.1109/ITSC.2017.8317872

[14]. Qu, L., Li, W., Li, W., Ma, D., and Wang, Y. (2019). Daily long-term traffic flow forecasting based on a deep neural network. *Expert Systems with Applications*, *121*, 304–312. https://doi.org/10.1016/J.ESWA.2018.12.031

[15]. Belt, E. A., Koch, T., and Dugundji, E. R. (2023). Hourly forecasting of traffic flow rates using spatial temporal graph neural networks, *Procedia Computer Science*, *220*, 102–109. https://doi.org/10.1016/j.procs.2023.03.016

[16]. Fedesoriano, S. (2021). Traffic Prediction Dataset. Retrieved March 20, 2024, from Kaggle website: https://www.kaggle.com/datasets/fedesoriano/traffic-prediction-dataset

[17]. Manga, P. K., (2021). Weather History, Retrieved January 22, 2025, from Kaggle website: https://www.kaggle.com/datasets/pavankumarmanga/weatherhistory

[18]. Zia, U., (2024). Traffic Time Series Dataset, Retrieved January 22, 2025, from Kaggle website: https://www.kaggle.com/datasets/stealthtechnologies/traffic-time-series-dataset

[19]. Magdum, R., (2022). What is Data Exploration? and its Importance in Data Analytics. *International Research Journal of Engineering and Technology*, 09(01), 1482- 1485. www.irjet.net

[20]. de Amorim, L. B. V., Cavalcanti, G. D. C. and Cruz, R. M. O., (2022). The choice of scaling technique matters for classification performance, *Applied Soft Computing*, 133(2–3), 109924. https://doi.org/10.1016/j.asoc.2022.109924

[21]. Schlosser, T., Dresden, T. U., Friedrich, M., Meyer, T. and Kowerko, D. (2024). A Consolidated Overview of Evaluation and Performance Metrics for Machine Learning and Computer Vision, *ResearchGate*. https://doi.org/10.13140/RG.2.2.14331.69928

[22]. Zhang, D., & Kabuka, M. R. (2018). Combining weather condition data to predict traffic flow: A GRU-based deep learning approach. IET Intelligent Transport Systems, 12(7), 578–585. https://doi.org/10.1049/iet-its.2017.0313