Iraqi Journal of Information and Communications Technology(IJICT)
Vol. 8, Issue 3, December 2025
www.ijict.edu.iq
ISSN:2222-758X
e-ISSN:2789-7362

TRAFFIC CONGESTION PREDICTION USING BI-LSTM WITH ATTENTION MECHANISM

Mustafa H. Abdulkareem[1], Mohammed I. Aal-Nouman[2], Omar Bouzid
[1,2] Department of Information and Communication Engineering, College of Information Engineering
Al-Nahrain University, Jadriya, Baghdad, Iraq
[3] Department of Electrical Engineering, University of Gharyan, Gharyan, Libya
Mustafa.hamdi@coie-nahrain.edu.iq[1], m.aalnouman@nahrainuniv.edu.iq[2], drombouzid@gmail.com[3]
Corresponding Author: **Omar Bouzid**
Received:08/10/2025; Revised:24/11/2025; Accepted:21/12/2025
DOI:10.31987/ijict.8.3.356

***Abstract*- Modern transportation systems are severely hampered by urban traffic congestion, which causes delays and fuel consumption. Proactive control techniques and intelligent traffic management depend on accurate congestion prediction. In order to predict congestion across urban edge networks, current study presents a deep learning-based framework that combines an attention mechanism with a bidirectional Long Short-Term Memory (LSTM) network with custom learnable attention layer, flowed by focal loss for addressing the data imbalance. A numerous dataset was generated by using SUMO, therefore over 2 million sequence was generated, including 12 spatiotemporal features that were extracted from the dataset. A large scale map was used and the prediction was based on edge level. The model can efficiently learn temporal dependencies and spatial patterns thanks to our preprocessing pipeline, which consists of temporal windowing, edge ID encoding, and cyclical time transformations. Trained with a 30-step sliding window, the model achieved low error metrics (MAE: 0.0744, RMSE: 0.2728), an F1-score of 0.90, and a classification accuracy of 92.56%. Our architecture performs better at detecting congestion events than recent state-of-the-art models. Thus the potential for scalable implementation in urban traffic forecasting systems of deep spatiotemporal learning models trained on realistic but synthetic simulation data.**

keywords: **Traffic congestion prediction, Bi-LSTM, Machen Learning, Attention Mechanism.**

I. INTRODUCTION

Traffic congestion is a major problem in cities, resulting in longer travel times, higher fuel use, and pollution of the environment. Intelligent Transportation Systems (ITS) are now crucial to improving urban mobility and efficiency because cities are growing rapidly and the number of cars is increasing. By facilitating improved traffic management techniques like adaptive signal control, route planning, and emergency response, accurate traffic congestion prediction supports the objectives of smart cities and sustainable transportation systems [1]. Recent advancements in machine learning, particularly deep learning, have dramatically improved the accuracy of traffic forecasting. For example, Recurrent Neural Networks (RNNs), LSTMs, and their variants have been shown to capture well the temporal dependencies of sequential traffic data [2]. Furthermore, attention mechanisms have been incorporated into LSTM networks to improve predictive performance for nonstationary time series-by dates to egress such as traffic flow-by allowing the models concentrate on the most relevant time steps [3]. Optimization-driven AI techniques have grate effect of increase the prediction accuracy based on recent studies, Pelican Optimization Algorithm (POA) and Particle Swarm Optimization (PSO) are grate examples duo to their efficiency in calibrating the models, feature selection, and hyper parameter tuning. POA can achieve reliable and quick exploration of complex search space by mimicking the hunting behavior of pelican [4]. While the PSO can efficiently optimize the model parameters by employing the swarm-based cooperative learning [5]. However, most of the studies in

ISSN:2222-758X
e-ISSN:2789-7362

the literature are based on actual traffic data, which is expensive to collect, sensitive to privacy issues, and has limited coverage both in scale and location. As a result of these limitations, researchers have turned to simulation tools like SUMO and beyond for creating synthetic traffic datasets. SUMO is a free and open-source microscopic traffic simulator that can accurately replicate complex urban topologies along with intricate traffic behaviors. The efficacy of SUMO in studying vehicular networks has been demonstrated through large-scale simulated scenarios known as the Monaco SUMO Traffic (MoST) and Luxembourg SUMO Traffic (LuST) [6]. Only a few studies have been conducted regarding traffic congestion categorization, which is a process to define traffic conditions, for example, congested or non-congested. It has very direct implications for control strategies in the road network. Most of the previous works utilize deep learning models to predict traffic flow data, including speed and vehicle numbers. In addition, in these classification problems where instances of congestion are rare compared to normal flow, class imbalances form a significant problem. Standard cross-entropy loss functions fail most often due to this imbalance. The focus loss was introduced as a solution to this problem and was first proposed for object detection tasks focusing on hard-to-classify minority samples. Thus, it poses as an interesting solution for detecting traffic congestion [7][8]. In this paper the novel approach proposed for traffic congestion prediction in this research using an LSTM-Attention network trained on a large, SUMO-generated urban traffic dataset. The following contributions in current study were:

1) OpenStreetMap allows us to build a realistic urban road network. SUMO then enables us to generate an exhaustive traffic dataset in simulation, which reflects the spatial and temporal dynamics of traffic occupancy over several days.
2) For this, a neural network architecture is designed, operating on sequences, named Att-BiLSTM, so that bidirectional LSTMswith attention can be better utilized for temporal feature representation and interpretation.For this, a neural network architecture is designed, operating on sequences, named Att-BiLSTM, so that bidirectional LSTMswith attention can be better utilized for temporal feature representation and interpretation.
3) Focal loss is also introduced to address the class imbalance, which helps to increase the accuracy of congestion event detection.
4) A detailed analysis performed using standard evaluation metrics (accuracy, precision, recall, F1-score), confusion matrices, and ROC/PR curves, among others, reveals our model's superior performance as compared to well-known classical base models.

II. SIMILAR WORKS

Predicting congestion can be obtaining in a several ways, recently researchers shift their focus on the Attention mechanism with neural networks, in [9] proposed Attention Temporal Graph Convolutional Network (A3T-GCN) for traffic forecasting the algorithm combine the Attention mechanism with graph convolutional network to efficiently model the spatial and temporal dependencies in traffic prediction, resulting improvement in accuracy.

In [10] proposed a graph multi attention network which include an encoder and decoder comprise of multiple spatiotemporal attention blocks, makes it able to capture the complex relationships between traffic conditions over time and space, they used the encoder to extract and encode the features from historical traffic data, and the decoder uses these features to

predict the congestion in the future. In [11] proposed a transformation-based architecture that takes the propagation delays of the traffic system into consideration, it relies on graph masking matrices and spatial self-attention modules to collect long-range spatial data as well as dynamic spatial dependencies, makes it able to improve accuracy over six existing real word datasets.

In [12] proposed hierarchical attention LSTM model, that combines hierarchical architecture with long short term memory networks, that uses a dual pooling mechanism to processes hidden and cell states on several layers, makes it more efficient to capture temporal patterns in traffic data. In [13] proposed a combination of multi scale convolutional neural networks and a transformer based attention mechanisms to gather long and short temporal features in traffic data. The aim of that study was focusing on reducing the MAE and RMSE parameters and it was quite efficient.

In [14] proposed BiGRU Attention LSTM model it is an algorithm followed by a bidirectional gate recurrent unit to enhance the models ability to understand the time series in both directional forward and backward when processing the data, making the model able to capture more complex pattern of the traffic data, the study also proposes a sliding window strategy that helps in capturing long term dependencies.

In order to improve urban traffic prediction, recent state-of-the-art models have addressed spatiotemporal complexity, dynamic patterns, and congestion forecasting accuracy. Table I show a clear comparative in strength and weaknesses. These models include A3T-GCN, GMAN, PDFormer, Hierarchical-Attention-LSTM, RMSCNN-Trans, and BiGRU-Attention LSTM. Although these models employ different methods, such as transformer mechanisms (PDFormer, RMSCNN-Trans), recurrent attention structures (Hierarchical-Attention-LSTM, BiGRU-Attention LSTM), or graph convolutions (A3T-GCN, GMAN), they all aim to enhance predictive performance on intricate urban traffic datasets. The proposed BiLSTM-Attention

TABLE I
Comparative performance and strength of proposed model vs existing models

Model	**Year**	**Key strength**	**Weaknesses**	**our model**
A3T-GCN	[9]	Use graph convolution, good spatial awareness	Lower accuracy 89%, higher RMSE 0.32	Has 3.56% Higher accuracy
GMAN	[10]	Multi-attention, strong spatial-temporal modeling	Very high RMSE ∼5.38 with unstable performance	far more stable
PDFormer	[11]	Strong feature extraction, transformer based	MAE close to zero, overfit suspicious, lower F1-score 0.88	Higher F1-score and provides mor realistic MAE
Hierarchical-attention-LSTM	[12]	Good temporal attention, balanced accuracy	Accuracy limited to 90%, RMSE 0.29	2.56% more accurate and 6% lower RMSE
RMSCNN-Trans	[13]	Hybrid CNN-transformer, strong spatial extraction	Higher MAE 0.07	Better F1-score and overall accuracy
BiGRU-Attention LSTM	[14]	High R2 96%, GRU efficiency	F1-score not reported, Very high RMSE 4.478	Our model far surpasses error metrics

approach maintains computational efficiency and architectural simplicity while achieving comparable or better predictive performance (accuracy: ∼92%, F1-score: 0.90, MAE: 0.0744, RMSE: 0.2728) than these models. Our model is especially useful for large-scale simulated datasets produced by SUMO because it focuses on bidirectional temporal patterns and congestion dynamics. Our approach provides a simplified solution that is simpler to scale and implement in real-time ITS

ISSN:2222-758X
e-ISSN:2789-7362

systems, in contrast to some of these models that call for external graph structures, road maps, or large transformer layers.

III. METHODOLOGY

In this paper, an approach was developed to predict a traffic congestion in a large city based on a large dataset [15] [16], for this purpose the SUMO simulation was used [5], the data collected from the simulation were used to train the model, A BiLSTM Attention architecture for predicting bidirectional temporal patterns were developed that achieved superior performance for edge network congestion prediction. The whole current training was done using 12th generation Intel(R) core (TM) i7-12700F 2.10GHz with RTX 3080, The system was built using python 3.

A. Data collection

The dataset was created by using SUMO simulation. The map of Luxembourg was used [6], it contains (2247) nodes and (5779) edges, the total edge length is (904.81km) and the total lane length is (1532.23km) Due to these large parameters the dataset contains 10 features: edge id, time, traffic flow, avg speed, waiting time, queue length, density, travel time, occupancy, collision count with over 2 million samples. These samples were splatted into 2 classes: non-congested 63.6% (class 0), congested 36.4% (class 1).

Figure 1: SUMO simulation / map of Luxembourg [6].

B. Data Preprocessing

The preprocessing pipeline was designed to transform raw edge network telemetry into a structured format suitable for temporal deep learning while preserving critical spatiotemporal patterns, by converting the timestamps into a human readable date time (hour, weekday) Unix timestamps were converted into cyclical features (hour, weekday) in order to

ISSN:2222-758X
e-ISSN:2789-7362

document weekly and daily congestion trends. In order to preserve memory efficiency and avoid high-dimensional one-hot representations, edge id was label-encoded to numeric IDs. By setting the normalized queue-length threshold at 0.5, a binary congestion label (is- congested) was produced to describe operational congestion scenarios. Min-Max normalization was used to scale all features, excluding the target, to [0, 1] in order to stabilize gradient descent. Tabular data was converted into input-output pairs using a sliding window of 30 timestamps in order to meet the LSTM's temporal dependency requirements. The dataset was split into training (77%) and test (23%) sets, while preserving the class distribution in order to lessen the bias brought on by the imbalance [17][18].

C. Development of Prediction Model

The proposed model in this paper integrates Bidirectonal LSTMs, Attention Mechanism flowed by Focal Loss to overcome the difficulties which were faced in congestion prediction at the edge networks. The idea behind using the BiLSTM is to capture temporal patterns from both future and past contexts within the input sequence [19][20], the BiLSTM uses two parallel LSTMs processes: forward LSTM and backward LSTM in our case there is two stacked LSTM layers, each LSTM contain 128 hidden units per layer the final output of this stage (forward and backward) generating a tensor of shape (batch size of 256, sequence length of 30 timesteps, 256 features) [21]. And it could be expressed in Eqs. (1),(2)and, (3) in [22][23]. So that, for each timestamp t:

$$\overrightarrow{h}_t = \text{LSTM}(x_t, \overrightarrow{h}_t - 1) \tag{1}$$

$$\overleftarrow{h}_t = \text{LSTM}(x_t, \overleftarrow{h}_t + 1) \tag{2}$$

$$h_t = [\overrightarrow{h}_t \,||\, \overleftarrow{h}_t] \tag{3}$$

Where, $\overrightarrow{h}_t$ represents the (past $\rightarrow$ future) forward, $\overleftarrow{h}_t$ represents (future $\leftarrow$ past) backward,and $||$ is the vector concatenation.

An Attention mechanism was added to dynamically weights each timesteps using linear transformation to address the importance of different timestamps of the input sequence [24], due to the learnability of the weights of the timestamps it focusses on the important timestamps because the LSTM treats all equally in [25] as obvious in Eq. (4) below:

$$e_t = v^\top \tan(W h_t + b) \tag{4}$$

Where, W is the learnable weight matrix, v is the learnable context vector, b represents the bias term, and h_t represents hidden state (concatenation of the forward/backward LSTM output).

Furthermore, the softmax function was used to change the Attention score by making it a probabilistic distribution over timestamps [26], the mathematical representation of the softmax adapted from [27], and it can be calculated by Eq. (5) as follows:

$$\alpha_t = \frac{\exp(e_t)}{\sum_{k-1}^{T} \exp(e_k)}, \qquad \sum_{t-1}^{T} \alpha_t = 1 \tag{5}$$

ISSN:2222-758X
e-ISSN:2789-7362

Where, α_t is the normalized attention weight for timestamp t, and T is the total timestamps numbers in the sequence

$$c = \sum_{t-1}^{T} \alpha_t h_t \tag{6}$$

Where, c is the weighted sum of hidden states.

For the final step and to reduce the dimensionality of the context vector, a fully Connected Layers Non-linear Mapping was used in order to extract a hierarchical pattern that maybe missed by the attention mechanism [28], by activating the ReLU allowing the model to learn complex non-leaner relationships between the features, and to solve the overfitting problem in the training part, a 40% of the neurons were randomly deactivate by regulating the dropout, therefore, and due to applying Eq. (7) the features number is reduced from 256 to 128 [29].

$$z = \mathrm{ReLU}(W_1 c_{\text{drop}} + b_1) \tag{7}$$

Where, W_1 is the weight vector, b_1 is the bias vector, and ReLU prevents vanishing gradients and adds sparsity.

For the binary classification, sigmoid activation was used to provide a probabilistic output for the congestion [30], creating edge binary classification decreasing dense layer from 128 to 1. The mathematical representation as in [31].

$$p = (w_2 z + b_2) \tag{8}$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{9}$$

Where, w_2 is weight vector, b_2 is scalar bias, and Sigmoid σ is to ensure the output $p \in [0, 1]$.

Finally, the model was trained using Adam optimizer with learning rate of 0.0003. and the class imbalance was handled by using Focal Loss with ($\alpha = 0.25, \gamma = 2$) which were calculated according to Eq. (10) as in [32], class 0 (non-congested): 63.6% of samples, class1(congested):36.4% of samples, Focal Loss improves the minority class performance by modifying the cross entropy to focus on hard to classify examples [33].

$$FL(p, y) = \begin{cases} -\alpha(1-p)^{\gamma} \log(p), & \text{if } y = 1 \\ -(1-\alpha)p^{\gamma} \log(1-p), & \text{if } y = 0 \end{cases} \tag{10}$$

Where, α: class weighting / assign higher weights to the minority class, and $\gamma$is a hard example focusing / higher γ model focus more on misclassified examples.

However, The procedure steps of the proposed model illustrated in Fig.2 and in Algorithm 1.

ISSN:2222-758X
e-ISSN:2789-7362

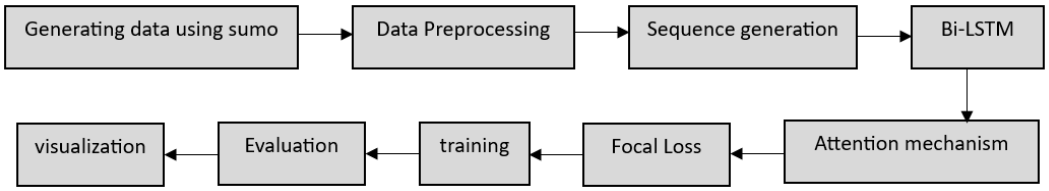


Figure 2: The procedure steps of the proposed model.

Algorithm 1: Predict Congestion (X)

Input: ***Sequence*** X
Output: ***Prediction*** $\hat{y}$
$H_{forward} = Forward\_LSTM(X)$
$H_{backward} = Backward\_LSTM(X)$
$H = Concatenate(H_{forward}, H_{backward})$
for $t = 1$ ***to*** 30 ***do***
  $score[t] = \tanh(W_a \cdot H[t] + b_a)$
end for
$\alpha = Softmax(score)$
$C = \sum(\alpha[t] * H[t])$ ***for*** $t = 1$ ***to*** 30
$C = Dropout(C, 0.4)$
$z = ReLU(W_d \cdot C + b_d)$
$\hat{y} = Sigmoid(W_o \cdot z + b_o)$
return $\hat{y}$

IV. EVALUATION AND RESULTS

According to the use of BiLSTM-Attention, the following findings were obtained, for the evaluation setup, a dataset of more than 2 million temporal sequences taken from a large-scale SUMO simulation of urban traffic was used to train the proposed BiLSTM-Attention model. Class balance was preserved when the dataset was split into 77% training set and 23% testing set, and applied a sliding window of 30 timesteps, the model trained using binary cross entropy loss and Adam optimizer, the learning last for 15 epochs. For the training performance, the loss dropped from 0.0370 in epoch 1 to 0.0118 in epoch 15 showing steady discretion as shown in Fig. 3, indicating a successful learning without overfitting. A 5,10,15,20, and 25 epochs were used in the initial experiments, the improvement in loss was marginal as mentioned (0.0370 to 0.0118) after nearly 12 epochs which indicates a diminishing returns, the reason behind choosing 15 epochs is to prevent the overfitting problem and to ensure an efficient computational cost for the sequences that was generated that exceed the 2 million sequence, the test loss showed a subsequent evaluation with similar downward trend without diverging from training loss, providing an evidence that the model did not suffer from overfitting problems. In order to test the robustness of the suggested architecture, a sensitivity tests were used on three dimensions (sequence length, train-test split ratio, and model parameters), the 30-time step scored the highest F1-score among 15, 30, and 60 time steps, meanwhile shorter windows failed to capture the required temporal context, and the longer windows created noise and raised the computational cost. And for the data split a comparison between three test ratios (70-30, 77-23, and 80-20), the 77-23 split produced the greatest balanced between model stability and evaluation robustness. A set of hidden size (64, 90, 128, and

ISSN:2222-758X
e-ISSN:2789-7362

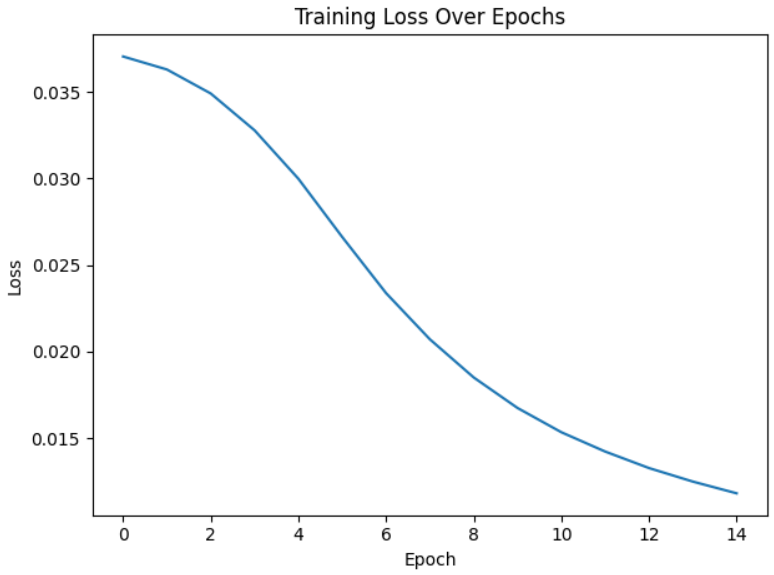


Figure 3: Training loss over epochs.

256 units) were tested, as well as the number of layers (1, 2, and 3), and multi dropouts of the hidden layers (0.2, 0.3, 0.4, and 0.5), the best results we got by using the parameter that mentioned before which is 2 layers each with 128 units, and dropout of 0.4 which is the number of the neurons that were drooped during the training. And for the evaluation part was based on accuracy of 92.56%, mean absolute error (MAE) of 0.0744, and a root mean squared error (RMSE) of 0.2728 were obtained from evaluation on the test set of 478,495. F1-score maximization was used to establish the ideal decision threshold (0.47). With precision and recall scores of 0.94 and 0.91 for the non-congested and congested classes, respectively, the model obtained a macro-average F1-score of 0.92 at that level. As noticeable in Fig. 4. the Confusion Matrix indicates strong classification performance for both class congested and non-congested. As shown in Table II, the first class (0) which is the congested class contain 304,229 samples the model predicted 284524 samples correctly congested, while only 19705 samples are missed to predicted as an actual congestion, and from 174,266 samples which the second class (1) not congested, 158371samples are predicted no congestion correctly, and only 15895 samples are predicted congestion where there was none. The majority of false negative events happened on edges with very little congestion spikes according to an error inspection, meaning that there is no sufficient evidence in the 30-step window that previously applied to the model to classify them as congested, and for the false negative predictions, happened on edges where density or queue length was high bot not high enough to reach the threshold of congestion labels, even with such miss classes the model obtained a high recall of 0.91 for the congested class indicating that the congested events are successfully identified.

The evaluation of the model depended on the accuracy, Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). A 0.47 threshold was chosen to balance precision and recall, due to the imbalance that occurred in the dataset.

ISSN:2222-758X
e-ISSN:2789-7362

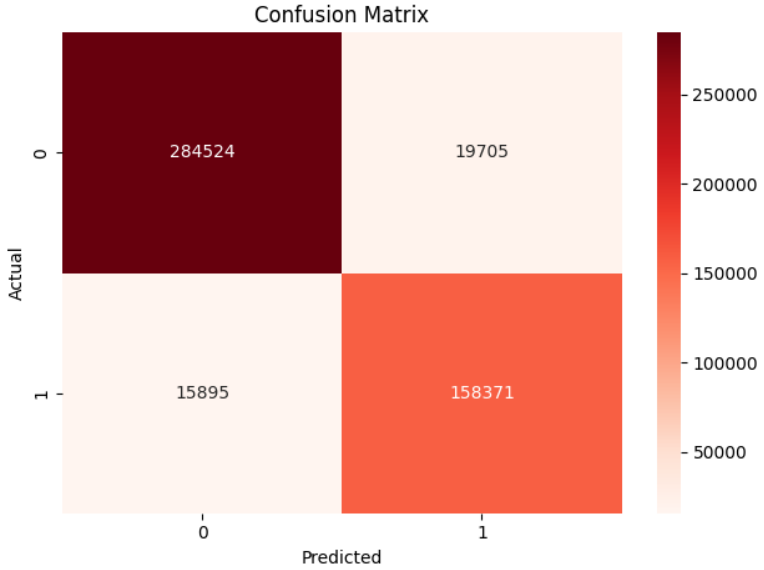


Figure 4: Confusion matrix.

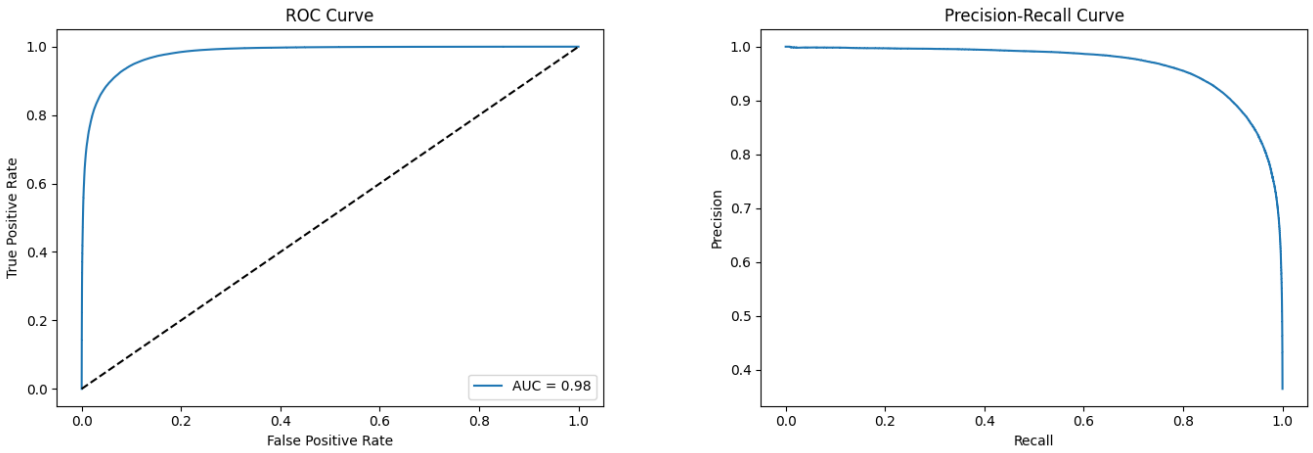
TABLE II
Classification Report

Class	Precision	Recall	F1-score	Support
Non-congested (0)	0.95	0.94	0.94	304,229
Congested (1)	0.89	0.91	0.90	174,226
Macro Avg	0.92	0.92	0.92	478,495
Weighted Avg	0.93	0.93	0.93	478,495

The model performed well even when there was a class imbalance, as evidenced by its high AUC score. Its efficacy in detecting actual congestion scenarios is further supported by the precision-recall curve (see Fig. 5). The ROC curve (receiver operating characteristic) is a plot of true positive rate and false positive rate and the area under the curve indicates the model quality. The precision-recall curve plots the precision and recall, the area under the curve indicates better precision and recall trade off, it is useful when having an imbalance classes. Table III is a comparison between the proposed model and some of the related models according to the accuracy, F1-score, MAE, and RMSE.

Consistent hotspots in particular network regions were identified by analyzing the simulation's most congested edges. The model works on edge level, Fig. 6 shows the edges that experienced recurrent patterns of congestion in the SUMO simulation, that indicate that the proposed model is not just detecting a random congestion, but it is able to detect the congestion in the edges around the network.

ISSN:2222-758X
e-ISSN:2789-7362



(a) (b)

Figure 5: (a) Precision-Recall (b) ROC.

TABLE III
comparisons with related works

Model	Year	Accuracy	F1-Score	MAE	RMSE
A3T-GCN	2020	89%	0.87	0.09	0.32
GMAN	2020	∼88%	0.82–0.88	∼2.77	∼5.38
PDFormer	2023	90%	0.88	∼0.08	0.30
Hierarchical-Attention-LSTM	2024	90%	0.88	0.08	0.29
RMSCNN-Trans	2025	91%	0.89	0.07	0.28
BiGRU Attention LSTM	2025	$R^2 = 96\%$	–	3.609	4.478
proposed Model (BiLSTM + Attention)	**2025**	**92.56%**	**0.90**	**0.0744**	**0.2728**

V. CONCLUSION

In this work, it was combined between memory-efficient BiLSTM-Attention architecture with synthetic data generation using the SUMO simulation framework to offer a novel deep learning-based method for traffic congestion prediction in large-scale urban road networks. A sizable dataset that captured intricate spatiotemporal traffic behaviors across multiple edges was created from a complex, realistic urban mobility scenario. In order to preserve congestion dynamics and facilitate effective learning, the preprocessing pipeline was meticulously crafted to convert unstructured telemetry into temporally ordered sequences. The proposed BiLSTM-Attention model achieved high performance in binary classification comparing to alternative studies as Table. III were previously showed successfully learning bidirectional temporal dependencies. With an accuracy (92.56%), F1-score (0.92), low error rates MAE (0.0744), and RMSE (0.2728). The efficiency of the model improved after applying the threshold optimization strategy under the traffic conditions (congestion and non-congestion)

ISSN:2222-758X
e-ISSN:2789-7362

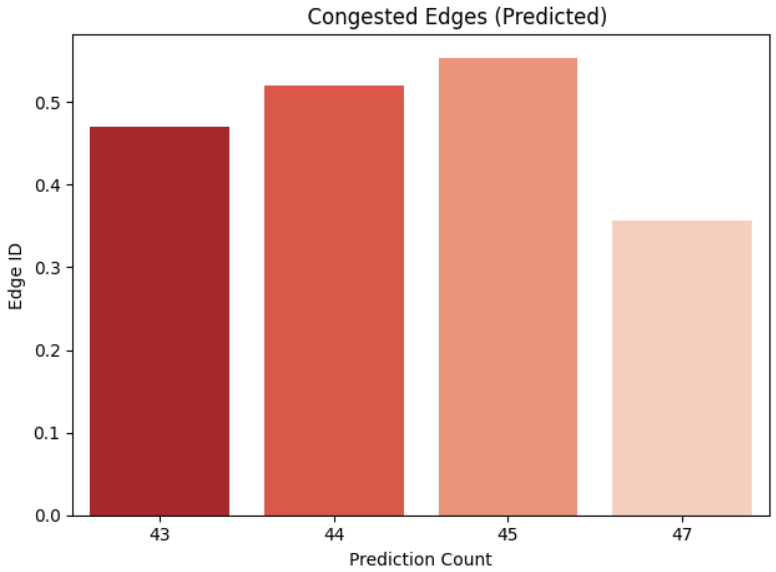


Figure 6: Top Congested Edges in the network.

which offered a balance performance.

FUNDING

None.

ACKNOWLEDGEMENT

The author would like to thank the reviewers for their valuable contribution in the publication of this paper.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

REFERENCES

[1] F. Wang, X. Xu, W. Feng, J. A. Bueno-Vesga, Z. Liang, and S. Murrell, "Towards an immersive guided virtual reality microfabrication laboratory training system," in Proc. IEEE Conf. Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Atlanta, GA, USA, Mar. 2020, pp. 796–797.

[2] B. Yang, S. Sun, J. Li, X. Lin, and Y. Tian, "Traffic flow prediction using LSTM with feature enhancement," Neurocomputing, vol. 332, pp. 320–327, Mar. 2019.

[3] N. S. Chauhan and N. Kumar, "Traffic flow forecasting using attention enabled Bi-LSTM and GRU hybrid model," in Proc. Int. Conf. Neural Information Processing (ICONIP), Singapore, Nov. 2022, pp. 505–517.

[4] M. H. Khoshdel, M. A. Balafar, and M. Gharehchopogh, "Pelican Optimization Algorithm: A Novel Nature-Inspired Algorithm," Journal Européen des Systèmes Automatisés, vol. 57, no. 4, pp. 515–526, 2024. doi: 10.18280/jesa.570408.

[5] O. M. O. Khalaf et al., "An Improved PSO-Based Optimization Framework for Engineering Applications," Journal of Engineering Sciences and Applied Design, vol. 4, no. 2, pp. 150–161, 2024. doi: 10.31272/jeasd.2454.

[6] L. Codeca, R. Frank, S. Faye, and T. Engel, "Luxembourg SUMO Traffic (LuST) scenario: Traffic demand evaluation," IEEE Intell. Transp. Syst. Mag., vol. 9, no. 2, pp. 52–63, Summer 2017. doi: 10.1109/MITS.2017.2666585.

[7] Y. Xie and T. Mallick, "A comparative study of loss functions: Traffic predictions in regular and congestion scenarios," arXiv preprint arXiv:2308.15464, Aug. 2023.

[8] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in Proc. IEEE Int. Conf. Computer Vision (ICCV), Venice, Italy, Oct. 2017, pp. 2980–2988.

[9] J. Bai, J. Zhu, Y. Song, L. Zhao, Z. Hou, R. Du, and H. Li, "A3T-GCN: Attention temporal graph convolutional network for traffic forecasting," ISPRS Int. J. Geo-Inf., vol. 10, no. 7, p. 485, Jul. 2021.

[10] C. Zheng, X. Fan, C. Wang, and J. Qi, "GMAN: A graph multi-attention network for traffic prediction," in Proc. AAAI Conf. Artif. Intell., vol. 34, no. 1, New York, NY, USA, Apr. 2020, pp. 1234–1241.
[11] J. Jiang, C. Han, W. X. Zhao, and J. Wang, "PDFormer: Propagation delay-aware dynamic long-range transformer for traffic flow prediction," in Proc. AAAI Conf. Artif. Intell., vol. 37, no. 4, Washington, DC, USA, Jun. 2023, pp. 4365–4373.
[12] T. Zhang, "Network level spatial temporal traffic forecasting with hierarchical-attention-LSTM," Digital Transp. Saf., vol. 3, no. 4, pp. 233–245, Dec. 2024.
[13] Z. Xijun and Y. Si, "Traffic flow prediction based on MSCNN and attention mechanism," Proc. Inst. Mech. Eng., Part D: J. Automob. Eng., early access, Feb. 2025, doi: 10.1177/09544070241311663.
[14] W. Xu, E. Blancaflor, and M. Abisado, "Performance and improvement of deep learning algorithms based on LSTM in traffic flow prediction," Discover Appl. Sci., vol. 7, no. 4, p. 278, Apr. 2025.
[15] A. Mustafa, M. I. Aal-Nouman, and O. A. Awad, "Cloud-based vehicle tracking system," Iraqi J. Inf. Commun. Technol., vol. 2, no. 4, pp. 21–30, Dec. 2019, doi: 10.31987/ijict.2.4.81.
[16] M. Aal-Nouman, H. Takruri-Rizk, and M. Hope, "Efficient message transmission method for in-vehicle emergency service," in Proc. 6th Int. Conf. Inf. Commun. Manage. (ICICM), Hatfield, U.K., Sep. 2016, pp. 193–196. doi: 10.1109/INFOCOMAN.2016.7784241.
[17] A. Mustafa, M. I. Al-Nouman, and O. A. Awad, "A smart real-time tracking system using GSM/GPRS technologies," in Proc. 1st Int. Conf. Comput. Appl. Sci. (CAS), Baghdad, Iraq, Dec. 2019, pp. 169–174. doi: 10.1109/CAS47993.2019.9075657.
[18] H. Yan, X. Ma, and Z. Pu, "Learning dynamic and hierarchical traffic spatiotemporal features with transformer," IEEE Trans. Intell. Transp. Syst., vol. 23, no. 11, pp. 22 386–22 399, Nov. 2021.
[19] T. Li, A. Ni, C. Zhang, G. Xiao, and L. Gao, "Short-term traffic congestion prediction with Conv–BiLSTM considering spatio-temporal features," IET Intell. Transp. Syst., vol. 14, no. 14, pp. 1978–1986, Dec. 2020.
[20] Y. Zhai, Y. Wan, and X. Wang, "Optimization of traffic congestion management in smart cities under bidirectional long and short-term memory model," J. Adv. Transp., vol. 2022, Art. no. 3305400, Jan. 2022.
[21] W. Zhuang and Y. Cao, "Short-term traffic flow prediction based on CNN-BiLSTM with multicomponent information," Appl. Sci., vol. 12, no. 17, p. 8714, Sep. 2022, doi: 10.3390/app12178714
[22] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," IEEE Trans. Signal Process., vol. 45, no. 11, pp. 2673–2681, Nov. 1997.
[23] A. Graves, "Generating sequences with recurrent neural networks," arXiv preprint arXiv:1308.0850, Aug. 2013.
[24] M. M. Rahman and N. Nower, "Attention based deep hybrid networks for traffic flow prediction using Google Maps data," in Proc. 8th Int. Conf. Mach. Learn. Technol. (ICMLT), Macau, China, Mar. 2023, pp. 74–81. doi: 10.1145/3591196.3591211.
[25] C. Li, B. Zhang, Z. Wang, Y. Yang, X. Zhou, S. Pan, and X. Yu, "Interpretable traffic accident prediction: Attention spatial–temporal multi-graph traffic stream learning approach," IEEE Trans. Intell. Transp. Syst., early access, Jan. 2024, doi: 10.1109/TITS.2024.3356782.
[26] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," arXiv preprint arXiv:1409.0473, Sep. 2014.
[27] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.
[28] S. R. Dubey, S. K. Singh, and B. B. Chaudhuri, "Activation functions in deep learning: A comprehensive survey and benchmark," Neurocomputing, vol. 503, pp. 92–108, Sep. 2022.
[29] A. Labach, H. Salehinejad, and S. Valaee, "Survey of dropout methods for deep neural networks," arXiv preprint arXiv:1904.13310, Apr. 2019.
[30] N. Pu, Z. Wu, A. Wang, H. Sun, Z. Liu, and H. Liu, "Arrhythmia classifier based on ultra-lightweight binary neural network," in Proc. 15th Int. Conf. Electron., Comput. Artif. Intell. (ECAI), Pitesti, Romania, Jun. 2023, pp. 1–7. doi: 10.1109/ECAI57955.2023.10199630.
[31] C. M. Bishop, Pattern Recognition and Machine Learning. New York, NY, USA: Springer, 2006.
[32] O. H. Abdulganiyu, T. A. Tchakoucht, Y. K. Saheed, and H. A. Ahmed, "XIDINTFL-VAE: XGBoost-based intrusion detection of imbalance network traffic via class-wise focal loss variational autoencoder," J. Supercomput., vol. 81, no. 1, pp. 1–38, Jan. 2025, doi: 10.1007/s11227-024-06509-5.
[33] M. Miryahyaei, M. Fartash, and J. A. Torkestani, "Focal causal temporal convolutional neural networks: Advancing IIoT security with efficient detection of rare cyber-attacks," Sensors, vol. 24, no. 19, p. 6335, Oct. 2024, doi: 10.3390/s24196335.