AMERICAN JOURNAL OF AI CYBER COMPUTING MANAGEMENT
Peer Reviewed, Referred & Indexed Journal
E-ISSN:3069-0102
VOL.6, NO. 2(2)(2026)

Hybrid LSTM–Transformer Model for Network Traffic Prediction and Intelligent Congestion Control in 6G Networks

Munay Said Ambani[1], Abdulhamid Eltheeb[2]

Faculty of Engineering, Ajdabiya University Ajdabiya, Libya[1]
Faculty of Electrical & Electronic, Engineering, Bright Star University AlBurayqah, Al-Burayqah, Libya[1]
Department of Information Technology Faculty of Computing and Information Technology, Ajdabiya University, Ajdabiya, Libya[2]

Abstract

The development of sixth-generation (6G) communication networks will ensure ultra-reliable low-latency communication, massive machine-type communication, intelligence at the edge, and connectivity ubiquity. Nonetheless, there emerge major challenges associated with traffic prediction and traffic congestion control owing to the growing heterogeneity and dynamism of network traffic. Currently, the congestion control algorithms are designed in a reactive fashion and lead to packet drops, latency, and poor resource utilization in cases of traffic increases. In this regard, this paper proposes a Hybrid LSTM-Transformer (HLT) solution for traffic prediction and intelligent traffic congestion control for 6G networks. The proposed solution merges the temporal learning capabilities of LSTM algorithms with the capability of Transformers to model long-range dependencies. The traffic patterns that are obtained by prediction will be used within an artificial intelligence (AI)-based congestion control module. Various experiments are performed on benchmark datasets of network traffic such as MAWI, CAIDA, UNSW-NB15, and CICIDS dataset. Through the performance analysis, it is found that the proposed HLT framework performs with more accuracy than other models such as ARIMA, RNN, GRU, LSTM, and individual Transformer. Additionally, it is found that proactive congestion control not only decreases packet losses and end-to-end delay but also increases throughput and utilization of the network.

Keywords: 6G Networks, Network Traffic Prediction, Congestion Control, Deep Learning, LSTM, Transformer, Artificial Intelligence, Quality of Service.

Introduction

Exponential growth in terms of devices, IoT ecosystems, driverless cars, industrial automation systems, and immersive applications is pushing towards sixth-generation communication networks [1]. Contrary to what is observed with other communication generations, 6G networks have the objective of delivering data transfer rates greater than 1 Tbps, latencies lower than 1 ms, as well as artificial intelligence embedded in the operation of the network. Network traffic is becoming increasingly complex, thereby bringing about challenges for operators of such networks. The nature of the traffic created by diverse devices is highly dynamic and not predictable using traditional statistical methods. Such an unpredictable nature brings about congestion, lost packets, and decreased QoS [2].

However, existing congestion control strategies, like TCP Congestion Avoidance and AQM, work in a reactive way once the congestion problem appears. Thus, the inefficiency of the existing methods for traffic congestion control can be explained by the dynamic nature of 6G [3]. In the recent years, deep learning techniques have shown impressive performance in traffic forecasting tasks [4][5][6]. Nevertheless, each technique has its own limitations:

LSTM neural networks perform well in modeling short-term dependencies but poorly in handling long sequences, and the Transformer method models long-term dependencies while ignoring short-term ones [7].

To eliminate the disadvantages of the abovementioned methods, a Hybrid LSTM–Transformer architecture is proposed, which will learn to model both short- and long-term traffic patterns. The obtained traffic data are used to perform proactive congestion control.

The contributions of this study include the following:

- Novel Hybrid LSTM-Transformer based network traffic prediction model.
- Congestion control algorithm designed using predictive intelligence.
- Prediction of network resources using AI.
- Performance comparison with advanced forecasting algorithms.
- Experiments conducted using real world network traffic data sets.

Related Works

Predicting network traffic is one of the core topics in current research activities on modern communication networks since accurate traffic prediction can contribute towards more effective utilization of resources and improving Quality of Service (QoS). The initial attempts in predicting network traffic focused on using statistical modeling techniques such as ARIMA, Seasonal ARIMA [8], Exponential Smoothing and Kalman Filtering. Such methods were efficient enough in modeling the stationary network traffic with linear properties but failed to properly cope with the nonlinear and bursting behavior of traffic in modern wireless and Internet network environments. Moreover, the increasing diversity of network traffic produced by different types of devices, applications and services has substantially decreased the efficiency of classical statistical prediction methods. Recent studies show that statistical methods cannot adjust to fast changing environment in future Beyond-5G/6G communications networks [9].

In order to address these issues, traffic forecasting using machine learning algorithms was suggested. Various algorithms such as Support Vector Regression (SVR), Random Forest (RF), Decision Trees, Gradient Boosting Machines, and Artificial Neural Networks (ANN) have been proven to provide better performance in terms of predicting by utilizing non-linear models trained directly on traffic datasets [10]. The application of ensemble learning techniques has also been considered in network traffic forecasting due to the capability of combining several weak learners into one powerful predictor. Recent research performed by Kim (2025) showed that Gradient Boosting Ensemble methods could be used successfully for modeling network traffic in the context of AI-supported 6G networks. Nevertheless, this approach also demands extensive feature engineering and is not capable of capturing temporal dependencies in traffic data [11].

The advent of deep learning technologies has greatly impacted the research on network traffic prediction. The Recurrent Neural Network (RNN) was one of the early deep learning models utilized in network traffic prediction since it is capable of processing data over time using recurrence. However, the RNN model has the drawback of suffering from vanishing and exploding gradients when dealing with long-term dependencies. In order to solve this problem, the Long Short-Term Memory (LSTM) networks came to be used, which rapidly

became the most popular model for network traffic prediction [12]. The LSTM network makes use of the memory cell and gating mechanism, which enables the storage of historical data for a long period. Many researchers have proven the effectiveness of LSTM models in dealing with temporal traffic data. Recent studies on Beyond-5G and 6G traffic prediction revealed that more advanced LSTM variations, such as Bi-LSTM and stacked LSTM, are superior to traditional ML techniques [13].

Even though they have been successful, LSTM models show certain drawbacks. The larger the sequence, the greater the difficulty in training them, and they may be unable to adequately model very long-range dependencies. Furthermore, the sequential approach to processing sequences makes LSTM models computationally inefficient and unscalable when dealing with extensive traffic data. In order to overcome such difficulties, more and more research is done using Transformer architectures for network traffic forecasting [14]. These architectures were first used in natural language processing tasks but make use of self-attention mechanism that allows them to process sequences in parallel and capture the relationships between different observations globally. Studies conducted recently have proved that the models based on Transformers are capable of detecting long-range temporal dependencies and outperform recurrent models in forecasting accuracy. In mobile communication networks, specialized TFT models have been proved to efficiently forecast traffic demand and enhance Quality of Experience (QoE) [15].

3. Methodology

The structure of the proposed system includes a traffic prediction module and an intelligent congestion control module tailored to the 6G network environment as shown in figure 1. In the general architecture of the system, the initial step entails gathering of traffic data using publicly available traffic data sets and performing some pre-processing of the data in order to maintain quality and consistency of the data set. The data is transformed into sequential

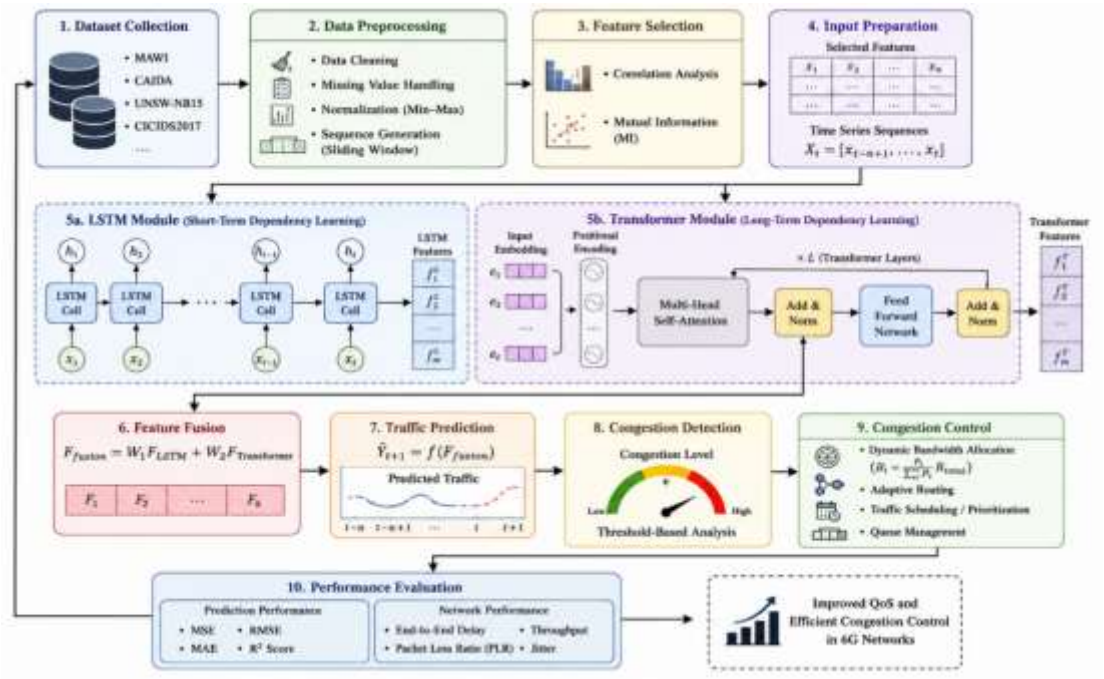


Fig.1 Proposed Methodology for Network traffic prediction

samples appropriate for deep learning. The main advantage of the proposed framework is that it uses the combination of two approaches: LSTM and Transformer architectures. The LSTM architecture detects the short-term dynamics in traffic changes (sudden traffic bursts, local variations). Meanwhile, the Transformer approach identifies the long-term dependencies in the traffic data using the self-attention mechanism. Then, the results of the two approaches are combined together to make accurate traffic predictions. Finally, the traffic predictions are fed into the intelligent congestion control engine to prevent any traffic congestion.

3.1 Data Collection

The Hybrid LSTM–Transformer model proposed above will be tested on the available public dataset of network traffic data, which includes real traffic measurement data and various network cases.The MAWI Traffic Archive provides backbone Internet traffic traces gathered from trans-Pacific networks. The available data include information about traffic volumes, packets flows, and bandwidth usage, thus, being appropriate for traffic prediction. The CAIDA Dataset provides large-scale measurements of the Internet traffic including packets traces, traffic flows, and performance data related to networks. The dataset is commonly used in traffic analysis and network optimization research. The UNSW-NB15 Dataset contains both normal and abnormal network traffic with such characteristics as number of packets, flow duration, and protocol. Diverse traffic patterns in the dataset can be used for traffic prediction purposes. The CICIDS2017 Dataset contains realistic network traffic created by normal activities and attacks. Such flow-based characteristics as throughput, packet rate, and traffic volume are included in the dataset.

3.2 Data Pre-processing

Data pre-processing is an important process in the suggested Hybrid LSTM–Transformer architecture since the quality of the data that goes into the system plays a crucial role in determining the level of prediction accuracy. The dataset used in network traffic prediction is normally gathered from sources like MAWI Traffic Archive, CAIDA Internet Traffic Dataset, UNSW-NB15, and CICIDS2017, among others. Such a dataset contains noise, missing data, duplicated records, inconsistencies, and even some irrelevant features. Failure to take care of these problems will result in the model learning wrong patterns, thus performing poorly on the task of forecasting and controlling congestion. These features provide valuable information about network behavior and are used as inputs to the prediction model.

3.2.1 Data Cleaning

Raw network traffic datasets frequently contain incomplete, noisy, or redundant records. Data cleaning is performed to improve dataset quality and ensure reliable model training. Missing values may occur due to packet loss during monitoring, sensor failures, or incomplete traffic logs. Several techniques is applied to address missing values. Mean Imputation is given by eqn 1.

$$x_i = \left(\frac{1}{n}\right) \sum_{j=1}^{n} x_j \quad (1)$$

where x_j represents the missing value and n denotes the number of available observations. For time-series traffic data, interpolation estimates missing values using neighbouring observations:

3.2.2 Missing Value Handling: This method preserves temporal continuity and is often preferred for traffic forecasting applications.

$$x_t = x_{t-1} + \left(\frac{x_{t+1} - x_{t-1}}{2}\right) \quad (2)$$

Duplicate Record Removal*:* Duplicate entries can introduce bias into model training. Duplicate traffic flows are identified and removed using unique identifiers such as timestamps, source IP addresses, destination IP addresses, and flow IDs.

Noise Reduction: Network traffic often contains abnormal spikes caused by measurement errors or temporary disturbances. Noise reduction techniques such as moving average smoothing are applied:

$$MA_t = \left(\frac{1}{k}\right) \sum_{i=0}^{(k-1)}{}_{t-i} x \quad (3)$$

where k denotes the smoothing window size. This process helps the model focus on meaningful traffic trends rather than random fluctuations.

3.2.3 Feature Selection

Network traffic datasets may contain dozens or even hundreds of attributes, many of which contribute little to traffic prediction. Including irrelevant features increases computational complexity and may lead to overfitting. Feature selection techniques are employed to identify the most informative traffic variables.

3.2.4 Mutual Information

Mutual Information (MI) is a feature selection method that can be employed for assessing dependence between two variables. Mutual Information is a measure that can detect linear and nonlinear relationship between variables which cannot be measured using correlation. MI is useful in network traffic prediction to quantify the amount of information that a traffic feature (such as packet arrival rate, throughput, or queue length) carries with respect to the target traffic (traffic volume). The ones with high Mutual Information values are deemed to be more informative and are used to train the model to make predictions while minimizing the computational complexity.

$$MI(X, Y) = \sum_x \sum_{(y) p(x,y) \log} \left(\frac{p(x,y)}{p(x)p(y)}\right) \quad (4)$$

This approach helps identify traffic attributes that provide significant predictive information. Selected features are then forwarded to the normalization stage.

3.2.5 Data Normalization

In network traffic datasets, different features often have significantly different numerical ranges. For example, packet counts may range from a few hundred to several hundred

thousand packets, while delay values may only vary between a few milliseconds and a few hundred milliseconds. Similarly, throughput may be measured in megabits per second and queue lengths may have entirely different scales. These large variations can negatively affect the training process of deep learning models because features with larger numerical values tend to dominate the learning process, causing the model to assign disproportionate importance to them. As a result, features with smaller numerical ranges may contribute less to the prediction process, even if they contain valuable information about network behavior.

Table II. Numerical ranges Traffic features

Feature	**Range**
Packet Count	0–100000
Delay	0–500 ms
Throughput	0–1000 Mbps
Queue Length	0–10000

Large numerical disparities can negatively affect neural network convergence. Therefore, feature scaling is applied.

3.3 LSTM Module

The LSTM architecture constitutes the first part of the suggested Hybrid LSTM–Transformer model. The network traffic dataset includes temporal dependencies in which current network traffic is dependent upon prior network traffic observation values. In comparison to classical RNNs, the LSTM model resolves the issue of vanishing gradients by means of memory units and gates. In the presented framework, the LSTM architecture learns the short-term traffic patterns like traffic spikes, oscillations, and temporary traffic congestions.

3.3.1 Forget Gate

The forget gate decides what information from the earlier time steps needs to be stored or forgotten. Not all past information about traffic is pertinent; therefore, the forget gate helps to forget the irrelevant information while retaining significant past traffic information. The forget gate is given by

$$ft = \sigma(Wf[ht-1, xt] + bf)f_t = \backslash sigma(W_f[h_\{t-1\}, x_t] + b_f)ft = \sigma(Wf[ht-1, xt] + bf) \quad (5)$$

Where, ft is the forget gate output, $ht-1$, $h_\{t-1\}$, is the previous hidden state, and xtx is the current input.

3.3.2 Input Gate

The input gate regulates how much new information will enter the memory cell. In this way, the model can learn about significant events happening in the traffic, like sharp increases in the number of vehicles or new traffic congestion patterns. The input gate is determined by the following formula

$$it = \sigma(Wi[ht-1, xt] + bi)i_t = \backslash sigma(W_i[h_\{t-1\}, x_t] + b_i)it = \sigma(Wi[ht-1, xt] + bi) \quad (6)$$

where it represents the input gate activation.

3.3.3 Cell State

The cell state serves as the memory for the LSTM cell. It incorporates the useful information contained in previous states together with the newly learned information concerning traffic. The cell state formula is defined as follows

$$Ct = ftCt - 1 + itCt\sim C_t = f_tC_\{t - 1\} + i_t\backslash tilde\{C_t\}Ct = ftCt - 1 + itCt\sim \quad (7)$$

where Ct is the updated cell state and $Ct\sim \{C_t\}Ct\sim$ represents candidate information.

3.3.4 Output Gate

The output gate determines the information transferred from the memory cell to the hidden state. The generated hidden state summarizes important traffic features and serves as input to the Transformer module. The output gate is defined as

$$ot = \sigma(Wo[ht - 1, xt] + bo)o_t = \backslash sigma(W_o[h_\{t - 1\}, x_t] + b_o)ot = \sigma(Wo[ht - 1, xt] + bo) \quad (8)$$

where ot denotes the output gate activation.

The LSTM module effectively captures short-term traffic dynamics, including sudden traffic changes and local temporal dependencies. These features are essential for accurate network traffic prediction. By learning local traffic behavior, the LSTM complements the Transformer module, which focuses on long-range dependencies. Together, they provide a comprehensive representation of network traffic, improving prediction accuracy and enabling proactive congestion control in 6G networks.

3.4 Transformer Module

Even though LSTM models can well describe local traffic dynamics, they are prone to miss dependencies between traffic observations in longer time periods. To tackle the above issue, the framework integrates the Transformer architecture after the LSTM component. The Transformer is a model that makes use of self-attention mechanism, which helps to examine dependencies between all components in the sequence.

The Transformer encoder processes traffic feature representations that have been created by the LSTM model and calculates attention weights that describe the significance of a particular traffic observation in comparison to other traffic observations. With multi-head attention mechanism, the model can identify various relationships in the traffic sequence. Consequently, the Transformer model can find recurring traffic behaviors, seasonality, and other dependencies in the traffic observations that cannot be found only through local analysis.

3.4.1 Self-Attention

Self-Attention is the feature that allows the Transformer to recognize the significant traffic observations in the sequence. Self-Attention is performed by analyzing all the traffic

observation points at once while taking into account long dependencies, which affect future network traffic.

The Attention function is defined as:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\text{sqrt}(d_k)}\right)V \quad (9)$$

where Q, K, and V are the Query, Key, and Value vectors, respectively.

3.4.2 Multi-Head Attention

Multi-head Attention further expands the concept of self-attention by enabling the Transformer to capture multiple traffic connections at once. Each individual head learns about some specific characteristics of network traffic features, like short-term dynamics, regularities, and long-term dynamics. All attention heads' outputs are summed up into a feature representation that helps improve understanding of traffic dynamics.

is the output weight matrix. This mechanism enhances prediction accuracy by capturing diverse traffic patterns within 6G networks.

$$MHA = Concat(head_1, head_2, \dots, head_h)W^O \quad (10)$$

The Transformer processes LSTM-generated features to enhance contextual understanding.

3.5 Feature Fusion Layer

The Feature Fusion Layer combines the outputs generated by the LSTM and Transformer modules to create a unified representation of network traffic behavior. While the LSTM network captures short-term temporal patterns and local traffic fluctuations, the Transformer learns long-range dependencies and global traffic trends through its attention mechanism. Individually, each model provides valuable information; however, combining their outputs allows the framework to leverage the strengths of both architectures. The outputs from LSTM and Transformer are combined:

$$F_{fusion} = W_{1F_{LSTM}} + W_{2F_{Transformer}} \quad (11)$$

Feature fusion provides the ability to consider both local and global traffic information in one single feature set representation. The feature set formed by the combination of two different features is used for fully connected layers that produce future traffic predictions. In this way, the Feature Fusion Layer increases prediction accuracy by using complementary features of both networks. The fused representation is passed through fully connected layers for prediction.

3.6 Traffic Prediction Output

Predicted traffic is defined as the predicted future network traffic produced using the Hybrid LSTM–Transformer model. The model predicts the future network traffic using the past traffic pattern along with present network condition.

$$\widehat{\{Y\}_{t+1}} = f\left(F_{fusion}\right) \quad (12)$$

is the predicted traffic volume at the next time interval. The predicted traffic is obtained using the combined features of the LSTM and Transformer networks. This predicted traffic is utilized by the congestion control module for proactive resource management, route optimization, and traffic management, which helps in reducing the congestion in 6G networks.

4. Intelligent Congestion Control Mechanism

4.1 Congestion Detection

The congestion detection component predicts the occurrence of network congestion prior to its actual occurrence. The predicted traffic from the Hybrid LSTM-Transformer model is constantly checked against a predetermined congestion threshold that acts as an indicator of future network load. This threshold value is set as a percentage of the total link capacity. If the predicted load crosses this threshold value, the network enters into a warning stage, implying that there is a high probability of network congestion in the coming time interval.

4.2 Dynamic Resource Allocation

Dynamic allocation of resources makes optimal use of the available network resources by allocating them depending on the forecasted traffic demand. In case of any forecasted congestion, allocation of bandwidth takes place according to the priorities of traffic classes. The more important an application, the more bandwidth it will get from the total available bandwidth in order to ensure its proper functioning.

$$B_i = \left(\frac{P_i}{\sum P_i}\right) * B_{total} \quad (13)$$

where P_i is traffic priority B_{total} is the available bandwidth

4.3 Adaptive Routing

The adaptive routing component picks the most appropriate way to facilitate communications taking into account the prevailing network circumstances and anticipated traffic. Rather than always choosing the shortest way, this system looks for the best available routes by comparing their congestion costs depending on criteria such as delays and link usage. In doing so, it ensures that the traffic load is evenly distributed on the network, avoiding any possible bottlenecks. Thus, it increases the reliability and efficiency of communication and reduces latencies.

$$Cost = \alpha D + \beta U \quad (14)$$

Where, D is the delay and U is the link utilization

4.4 Traffic Scheduling

Traffic scheduling determines the order in which network traffic is transmitted. In the proposed framework, a priority-based scheduling mechanism is employed to support different

6G service categories. Ultra-Reliable Low-Latency Communication (URLLC) traffic receives the highest priority due to its strict latency requirements, followed by Enhanced Mobile Broadband (eMBB) traffic and Massive Machine-Type Communication (mMTC) traffic. This scheduling strategy ensures that critical applications receive timely network access while maintaining Quality of Service (QoS) for all users in the 6G environment.

5. Result and Discussion

5.1 Hardware Configuration

The designed Hybrid LSTM–Transformer architecture is deployed and tested on a high-performance computer to provide efficient training and testing processes. Specifically, the computational system includes an Intel Core i9-13900K processor, a NVIDIA RTX 4090 graphics card with 24 GB VRAM, 64 GB DDR5 memory, and Windows 11 OS. Deep learning libraries like TensorFlow and PyTorch are utilized for model design and training. In particular, the GPU provides fast processing of the big datasets and complex computations of the neural networks, whereas the high memory configuration facilitates efficient work with time series and model parameters. Table 1 shows the hardware configuration.

Table 1: Hardware Configuration

Component	**Specification**
CPU	Intel Core i9-13900K
GPU	NVIDIA RTX 4090 (24GB)
RAM	64 GB DDR5
OS	Windows 11
Framework	TensorFlow/PyTorch

5.2 Hyper parameters

Hyperparameters refer to pre-specified parameters that define how learning can be done in the Hybrid LSTM-Transformer model. Appropriate choice of hyperparameters is necessary in order to get accurate predictions and consistent performance of the model. In this research, 60 is the value of the sequence length to allow learning the traffic pattern through previous time steps. The LSTM layer uses 128 units to help the model learn short-term traffic relationships. The Transformer layer includes four encoder layers and eight attention heads for learning long-range traffic relationships. The batch size selected is 64 to strike a balance between efficiency and effective utilization of memory. 100 epochs are used in training the model using 0.001 learning rate in order to converge and minimize prediction errors.

Table 2: Hyper-parameters

Parameter	**Value**
Sequence Length	60
LSTM Units	128
Transformer Layers	4
Attention Heads	8

Batch Size	64
Learning Rate	0.001
Epochs	100

5.3 Correlation Matrix

Correlation matrix demonstrates the associations between different network traffic attributes. Traffic Volume is strongly positively associated with Packet Arrival Rate (0.93) and Throughput (0.91), and that means that an increase in traffic volume is directly connected with network load and data transfer rate. Queue Length has a strong connection with Delay (0.91), and that means that long queues increase the time spent by packets to get through the queue. The feature Bandwidth Utilization has a strong connection with Traffic Volume (0.88), and that means that bandwidth is effectively utilized in case of high traffic.

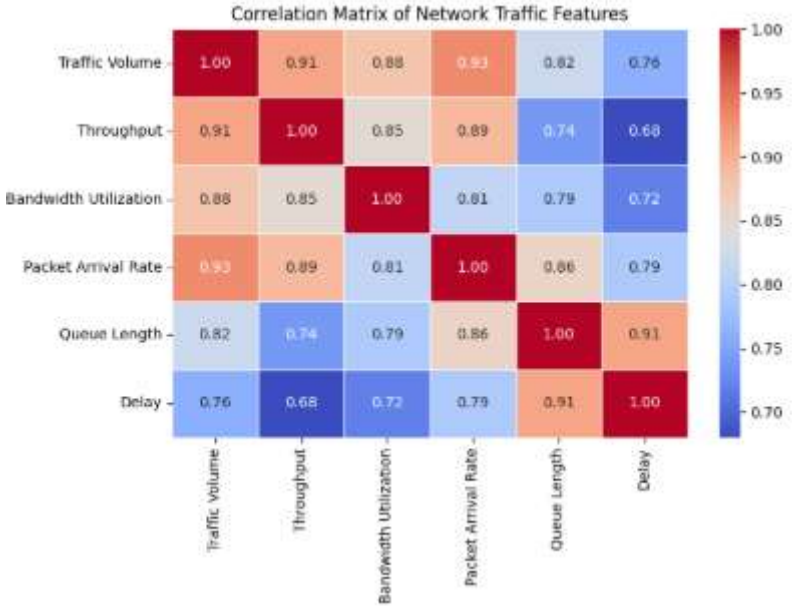


Fig 2. Correlation Matrix

5.4 Performance metrics

Figure 3 compares the prediction performance of ARIMA, RNN, GRU, LSTM, Transformer, and the proposed Hybrid LSTM–Transformer (HLT) model using MSE, RMSE, MAE, and R^2 Score. Lower MSE, RMSE, and MAE values indicate better prediction accuracy, while a higher R^2 score indicates a better model fit. The results show that ARIMA has the highest prediction errors, whereas deep learning models achieve better performance. LSTM and Transformer significantly improve prediction accuracy by learning temporal and long-range traffic patterns. The proposed HLT model achieves the best results, with the lowest MSE (0.005), RMSE (0.071), and MAE (0.058), and the highest R^2 score (0.97). These findings demonstrate that combining LSTM and Transformer architectures improves traffic prediction accuracy and supports effective congestion control in 6G networks [16],[17,[18]. Fig 3 explains the performance metrics.

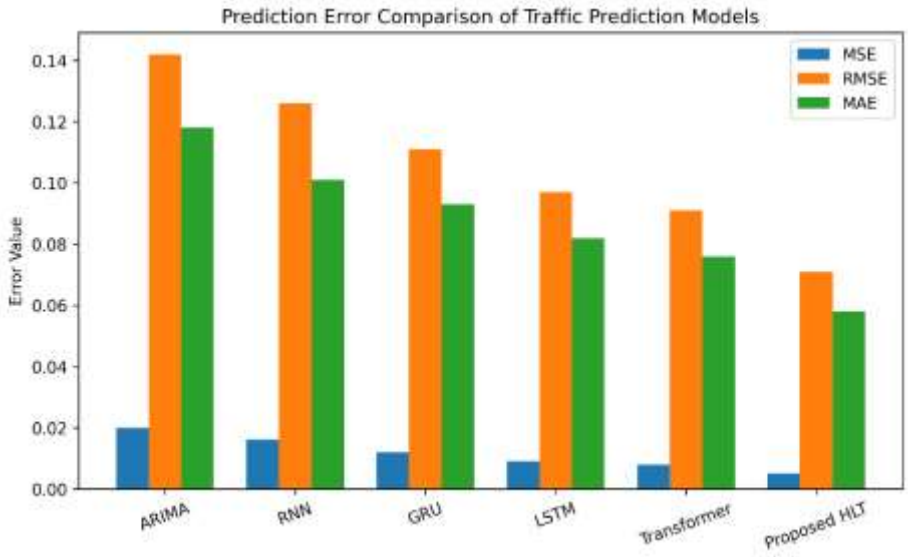


Fig 3. Prediction Error Comparison

5.5 Actual vs Predicted Network Traffic

The actual network traffic is compared to the traffic predicted by the proposed Hybrid LSTM–Transformer model in Figure 4. The accuracy of the prediction is high as the predicted values are close to actual traffic pattern for each time step. These minor discrepancies between the actual and predicted traffic show that the model can capture both short and long-term variations in traffic. This tight fit validates the proposed model of the proactive congestion control and resource management model in 6G networks.

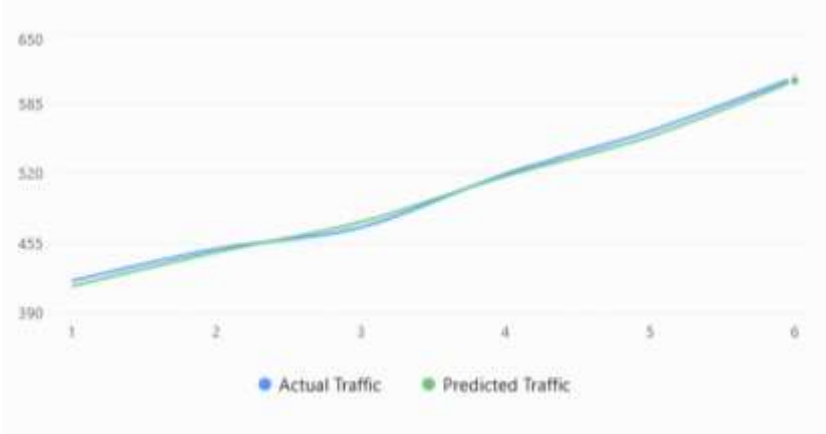


Fig. 4. Actual vs Predicted Network Traffic

5.6 Training and Validation Loss Curve

Both the training and validation loss curves show a smooth reduction with the number of epochs, suggesting that the model learns well and converges. The validation loss is very close to the training loss throughout the training process, indicating that the model is not overfitting and is learning the data. The loss of the training set at epoch 100 is 0.010, and the loss of the validation set is 0.012, indicating good training stability and prediction ability of the proposed Hybrid LSTM–Transformer model.



Fig. 5. Training and Loss Hybrid LSTM–Transformer model

5.7 Network Performance Comparison

Figure X, the performance comparison of the existing model and the proposed Hybrid LSTM–Transformer (HLT) framework. The proposed model significantly reduces the packet loss, delay, congestion rate, but simultaneously increases the network utilization and throughput. Specifically, the network utilization is improved from 73% to 92%, the delay is reduced from 24 to 11ms, the throughput is increased from 87 Mbps to 112 Mbps, and the packet loss is decreased from 5.8% to 1.7%. The proposed framework is found to be effective in improving the QoS and providing proactive congestion control in 6G networks through these improvements.

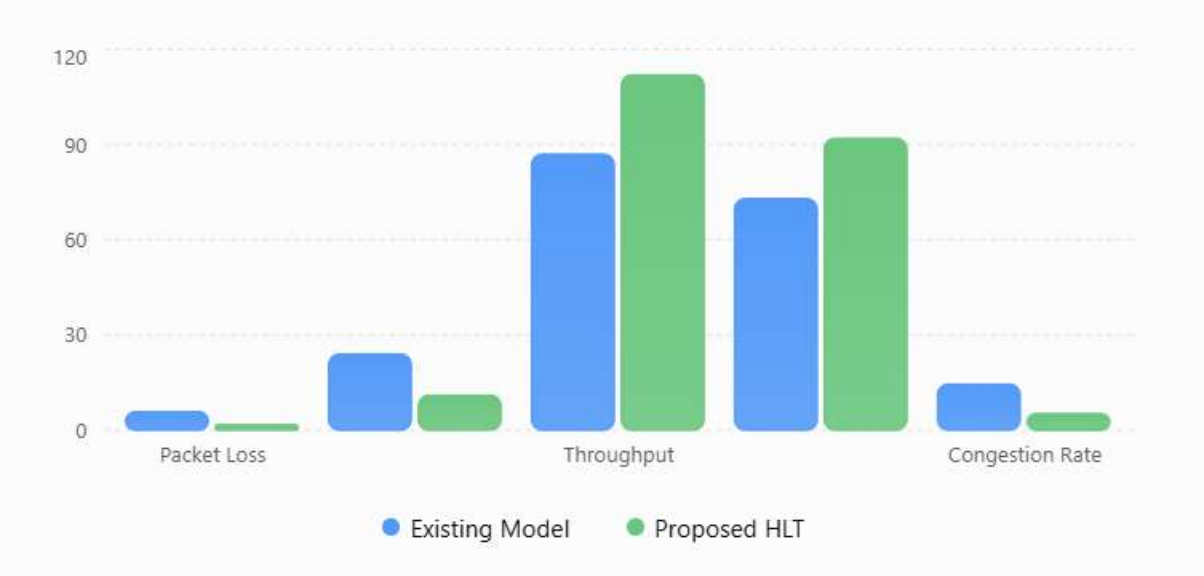


Fig 6. Network Performance Comparison

5.8 Discussion

From the results above, it is evident that the proposed Hybrid LSTM–Transformer (HLT) algorithm can be efficiently used to predict network traffic and implement intelligent

congestion control in 6G. It recorded the least MSE (0.005), RMSE (0.071), and MAE (0.058) values and highest R^2 score (0.97) compared to other models such as ARIMA, RNN, GRU, LSTM, and Transformer models. The good performance can be explained by the use of the combination of the LSTM and Transformer architecture. In particular, the LSTM helps in capturing the short-term traffic variations, whereas the attention mechanism of the Transformer makes it possible to learn long-range traffic dependencies. This means that the fusion layer makes the predictions better because of the complementary nature of these two features. Predicting traffic with accuracy leads to a proactive approach in dealing with network congestion. Using the predicted traffic data, the congestion control algorithm allocates resources, optimizes routing, and priorities traffic flows. The result of this process is the minimization of packet loss and delay and maximizing throughput and network utilization. Therefore, the above results prove that the HLT framework has proven to be efficient in traffic prediction and congestion control.

6. Conclusion

In this study, a Hybrid LSTM-Transformer (HLT) architecture was developed for traffic prediction and intelligent congestion control in 6G communication networks. The hybrid model takes advantage of temporal learning of LSTM and attention learning of Transformer to be able to identify short and long dependencies of the network traffic. In experiments conducted, the HLT model showed the best results among other models like ARIMA, RNN, GRU, LSTM and Transformer, in terms of low prediction errors and high prediction accuracy. Traffic prediction results have been used for intelligent congestion control in the network via dynamic allocation of resources, adaptive routing, and traffic scheduling. This approach has led to significant improvements in the performance of networks in terms of reduction of packet losses, delays, and congestion, as well as increases in throughput and network utilization. The combination of traffic forecasting and intelligent congestion control allows optimizing the use of network resources and improving its Quality of Service (QoS). Thus, the proposed hybrid architecture is useful for managing 6G networks in an autonomous manner and makes contributions to the development of self-optimizing 6G communication systems. Possible future research directions include integration of federated learning, reinforcement learning, and edge intelligence.

References

[1] Y. Wang and P. Chen, “Network traffic prediction based on transformer and temporal convolutional network,” *PLOS ONE*, vol. 20, no. 4, 2025.

[2] E. Lykakis, “Data Traffic Prediction for 5G and Beyond: Methods, Challenges, and Future Directions,” *Electronics*, vol. 14, no. 23, p. 4611, 2025.

[3] X. Sun et al., “Enhancing Network Traffic Prediction by Integrating Graph Transformer with Temporal Models,” in *Proc. ACM International Conference*, 2025.

[4] S. Jayashree, E. FARHAT, and E. ABDULHAMID, “Deep Learning-Driven Forecasting Models for Iot Data in Cloud Computing Environments Leveraging Temporal Convolutional Networks,” *J. Theor. Appl. Inf. Technol., vol. 103, no. 6, pp.* 2108–2122, 2025.



[5] A. O. Alsayed, N. A. Ismail, L. Hasan, M. Binsawad, and F. Embarak, "Leveraging a hybrid convolutional gated recursive diabetes prediction and severity grading model through a mobile app(PDF)," *PeerJ Comput. Sci., vol. 11, p. e2642, 2025, doi: 10.7717/PEERJ-CS.2642/SUPP-3.*

[6] F. EMBARAK, A. ELTHEEB, A. O. ALSAYED, and D. N. P. DEVADHAS, "Designing Personalized Assistive Technologies for Elderly Users in Human-Computer Interaction for Haptic Recognition Using Gnn," *J. Theor. Appl. Inf. Technol., 2025, Accessed: Dec. 29, 2025.*

[7] G. Kougioumtzidis et al., "Mobile Network Traffic Prediction Using Temporal Fusion Transformers," *IEEE Intelligent Systems*, 2025.

[8] R. A. J. Al Hamadani, "Improvement of Network Traffic Prediction in Beyond 5G Networks Using BiLSTM," *Qubahan Academic Journal, vol. 5, no. 2, 2025.*

[9] A. A. Hussien, H. Nashaat, and R. F. Abdel-Kader, "Machine Learning Techniques for Spatiotemporal Traffic Prediction in 5G Cellular Networks," *Discover Applied Sciences*, vol. 7, 2025.

[10] H. Kim, "Network Traffic Prediction Using Gradient Boosting Ensemble Method," in *Proc. Artificial Intelligence and Cloud Computing Conference (AICCC)*, 2025, pp. 608–614.

[11] P. K. Patidar, N. Pal, M. Patidar, and N. C. Gediya, "Spatio-Temporal Graph Neural Network for Real-Time Network Traffic Prediction and Adaptive Resource Allocation in 6G Networks," in *Proc. IEEE Conference on Advanced Computing*, 2025.

[12] M. A. Habib, P. E. Rivera, Y. Ozcan, and M. Elsayed, "Transformer-Based Wireless Traffic Prediction and Network Optimization in O-RAN," in *IEEE ICC Workshops*, 2024.

[13] C. Hu, H. Zhou, D. Wu, X. Chen, J. Yan, and X. Liu, "Self-Refined Generative Foundation Models for Wireless Traffic Prediction," *arXiv preprint arXiv:2408.10390*, 2024.

[14] Q. Luo, S. He, X. Han, Y. Wang, and H. Li, "LSTTN: A Long-Short Term Transformer-Based Spatio-Temporal Neural Network for Traffic Flow Forecasting," *arXiv preprint arXiv:2403.16495*, 2024.

[15] K. Yalda, "Network Traffic Prediction Performance Using LSTM," *Romanian Journal of Information Science and Technology*, 2024.

[16] Y. Wang et al., "A Comparative Study of SVR, LSTM, GRU, and Bi-LSTM for Network Traffic Forecasting," *ITM Web of Conferences*, 2025.

[17] A Framework for User Traffic Prediction and Resource Allocation in 5G Networks, *Preprints*, 2025.

[18] K. I. M. Ata et al., "A Multi-Layer CNN-GRUSKIP Model Based on Transformer for Spatial-Temporal Traffic Flow Prediction," *arXiv preprint arXiv:2501.07593*, 2025.