

Bahdanau Attention based LSTM for Wireless Network Traffic Prediction

Meenige Pavani Prathyusha¹, Dr. B. Anuradha²

¹*M. Tech Student, Department of Communication Engineering, S. V. University College of Engineering, Tirupati, A. P. India*

²*Professor Department of Communication Engineering, S. V. University College of Engineering, Tirupati, A. P. India*

doi.org/10.64643/IJIRT208610

Abstract—Wireless networks and data intensive applications continue to grow at a fast pace, and there is a critical need for accurate traffic predictions so that resources can be allocated efficiently, traffic congestion can be controlled and Quality of Service (QoS) can be managed [1]. Traditional Long Short-Term Memory (LSTM)s can learn temporal dependence, but there is no explicit identification of important time steps in the past [2]. In this study, a new Bahdanau Attention-LSTM model is proposed for predict the wireless network traffic by combining the temporal feature extraction module with the attention mechanism [3]. The attention mechanism weights the historical hidden state so that it can concentrate on informative traffic patterns [3]. The performance is compared with that of a conventional LSTM. The models are assessed with the following statistics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Symmetric Mean Absolute Percentage Error (SMAPE), Coefficient of Determination (R^2 Score), Nash-Sutcliffe Efficiency (NSE) and Theil's U [1]. This is shown in result for the effectiveness of the proposed model in better prediction accuracy and proactive network management, resource optimization, and congestion mitigation [3].

Index Terms—Wireless Network Traffic, LSTM, Bahdanau Attention, Time-Series Forecasting, Deep Learning, Traffic Prediction, Network Management.

I. INTRODUCTION

Network traffic has been rapidly growing since the proliferation of wireless communication technologies, mobile devices, cloud services, and data-intensive applications. Modern wireless networks also have significant fluctuations in traffic demand because of the different demands of the applications, fluctuations

in network conditions, time-dependent usage patterns, and the change in user behaviour. These changes make planning the network and managing resources more complex and difficult. Thus, an accurate forecasting of future network traffic is crucial for proactive bandwidth allocation, congestion prevention, quality of service improvement, load balancing and efficient utilization of network resources. A good prediction model for traffic can help network operators make decisions in advance and avoid degradation of network performance [1].

Recurrent Neural Networks (RNNs) are heavily used for sequential data prediction, and a variant of RNNs known as Long Short-Term Memory (LSTM) which will learn long-term temporal dependencies. LSTM addresses the vanishing gradient problem and is able to learn temporal patterns in wireless network data via its memory cell and gating mechanisms [2, 4]. A standard LSTM however, does not explicitly identify the importance of the single time steps, which might lead to a less effective focus on the most informative time steps during irregular or rapidly changing traffic situations [3].

In this paper, an Attention-Based LSTM model with Bahdanau temporal attention is proposed for predicting the wireless network traffic, while a conventional LSTM is used as the baseline model. The training is done by using historical traffic sequences and the performance is assessed by using MAE, RMSE, MAPE, SMAPE, R^2 , NSE and Theil's U [2]. The overall goal is to find out if temporal attention can help increase the accuracy in identifying the traffic patterns of interest and extend the forecasting

accuracy, thus enabling more intelligent and proactive wireless network management [3].

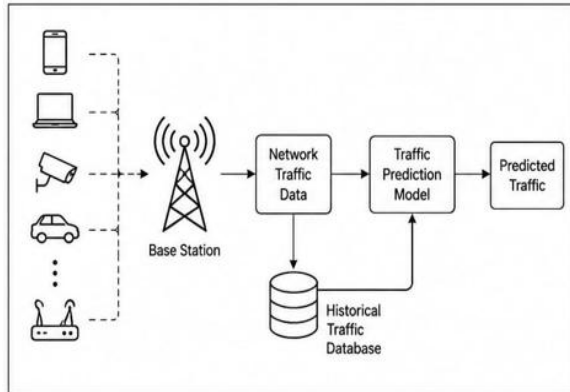


Fig.1 High-level architecture of the proposed wireless network traffic prediction framework.

II. LITERATURE SURVEY

P. Derakhshan Barjoei et al.: The authors designed a Long Short-Term Memory model to predict the network traffic of the subscribers of mobile networks. Their results showed that LSTM can learn temporal patterns from the past network traffic and forecast future traffic with high accuracy. Mainly MAE and RMSE were used to assess the model performance. The main result of this study is that the LSTM model is appropriate for the forecast of mobile network traffic. This paper is taken as a foundation paper for the present research. It does not, however, explicitly use a temporal attention mechanism motivating the use of Bahdanau attention in combination with LSTM in the proposed model.[1]

Xu et al.: The authors analysed the large-scale mobile traffic patterns from thousands of cellular towers in an urban setting. They were able to pinpoint recurring temporal patterns and show that traffic characteristics are directly related to user behavior and various geographical areas like residential, business and transportation zones, entertainment areas etc. The study revealed that there is a lot of regularity within the cellular traffic despite the fact that the overall data is complex and dynamic. This is significant since it proves that useful patterns can be gleaned from the historical observations which is important for traffic prediction. The present work is an extension of this concept to automatically learn such temporal patterns using deep learning. [2]

Zhang et al.: The authors introduced a new cellular traffic prediction solution for the entire network scope using a densely connected Convolutional Neural Network. The cellular traffic data was represented in such a way that CNNs can be used to capture the spatial and temporal dependency. An information fusion method based on parameters was added to fuse the learned information, and the evaluation of the prediction performance was performed by using RMSE. Through their experiments, they showed that their approach has better prediction accuracy than other approaches. Deep-learning architectures can be used to accurately model the nonlinearity inherent in cellular traffic, as demonstrated by the study. The present work, however, mainly considers temporal traffic prediction with LSTM and attention but neglects the processing of spatial-temporal traffic data. [3]

Li and Ye: The authors suggested a method to predict the throughput in 5G wireless networks based on LSTM. The study investigated various kinds of services, various scheduling situations and whether or not LSTM was able to accurately predict future throughput of networks. It was shown that the LSTM model was able to predict the wireless network traffic with acceptable accuracy and training efficiency. A key finding from this study is that measurements taken in 5G networks can have sequential nature and can be addressed successfully with LSTM. This study directly contributes to the application of LSTM in the current research, and the proposed model adds the Bahdanau attention to highlight relevant historical information. [4]

Gao: The author has designed a Smoothed Long Short-Term Memory (SLSTM) model for predicting 5G traffic. The method adopted was the usage of seasonal time difference method to reduce the randomness in the initial traffic sequence and an adaptive mechanism to adjust the LSTM architecture based on the prediction accuracy. Experimental results indicate that SLSTM could be used to enhance the predictive capabilities of the traffic flow than conventional algorithms. The study shows how the conventional LSTM can be modified to better support complex traffic patterns in 5G. This observation is important for the current work as it validates the concept that

improved traffic representations can be obtained and that the standard LSTM can be improved. [5]

Li et al.: The authors devised a deep-learning-based model called LA-ResNet to predict the wireless network traffic. The architecture consists of a residual network; a recurrent network and an attention mechanism is employed to model the spatial and temporal dependency. The residual component is for spatial features; the recurrent and attention components are for temporal features and for enhancing salient information. Real wireless traffic data were used in the experimental evaluation and compared the prediction performance with RNN and 3DCNN approaches. This paper is very relevant for the present research as it has confirmed the usefulness of attention for forecasting wireless traffic. In this work, however, a simpler structure is followed, using temporal attention of the Bahdanau model and LSTM specifically. [6]

Zeng et al. [7]: The authors proposed a multi-component spatiotemporal cross-domain neural network with attention method for the wireless cellular traffic prediction. Their model incorporated Conv-LSTM or Conv-GRU networks for the representation of neighbouring traffic, daily periodicity and weekly periodicity. An attention layer was added to these components with various weights to ensure that the important information would make more significant contributions to prediction. The authors noted that attention-enhanced structures converged and fitted better. The study is important as it directly shows the importance of attention mechanisms in cell network forecasting. In this current research, the general idea of temporal importance weighting is extended by applying the Bahdanau attention to LSTM hidden states. [7]

Research Gap:

Although available literature shows the effectiveness of deep recurrent networks [1, 5, 7] and high-dimensional spatiotemporal hybrids [8, 9] are highly effective, the trade-off is that they are complicated and require significant computation time which makes them complicated for real-time deployment. On the other hand, standalone LSTMs consider all time steps of the history equally important, ignoring the traffic bursts that are important [1, 2]. In order to fill this gap,

this paper presents a targeted Bahdanau Attention-LSTM architecture. The proposed framework improves prediction accuracy by calculating dynamic alignment scores with the sequence history which identifies time steps with high impacts without sacrificing the computation complexity of a multi-branch network.

III. PROPOSED METHODOLOGY

The proposed approach is to build an Attention-Based Long Short-Term Memory (Attention-LSTM) for prediction of wireless network traffic. The principal goal is to predict the future traffic using the time dependencies learnt from the past traffic observations, and to give more weight to the time steps with more relevant information [3]. The baseline model is a conventional LSTM model and the proposed model is the integration of the temporal attention mechanism and the LSTM [2, 3]. Multiple regression and forecasting evaluation metrics are used to compare the functioning of both models [1, 2].

Data Description

To assess the proposed Attention-LSTM framework, real-world cellular network traffic data was sourced from the publicly accessible Telecom Italia Big Data Challenge dataset. The dataset consists of Call Detail Records (CDRs) gathered across the telecommunications infrastructure in Milan, Italy. The geographical coverage of the city is discretized into a grid of 100×100 spatial cells. The recorded network activity encompasses daily volume aggregations for Internet traffic, Short Message Service (SMS), and Voice Calls, aggregated continuously at 10-minute time intervals. For the scope of this study, the temporal Internet traffic volume series was extracted, ordered chronologically, and structured using the 48-step sliding window approach to evaluate temporal dependence and dynamic load variations.

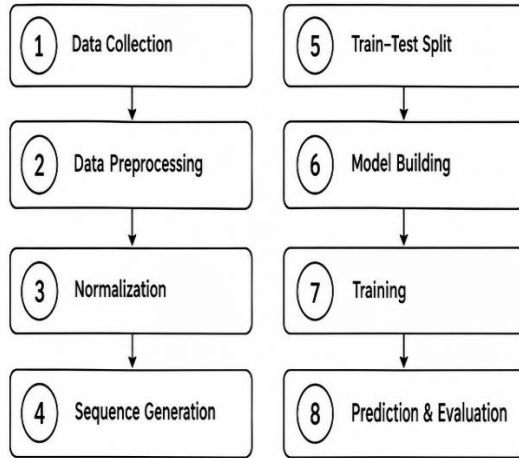


Fig.2 Proposed wireless traffic forecasting methodology.

3.1 Data Preprocessing and Sequence Generation

To maintain temporal relationships, historical wireless network traffic data collected from the base station network is first sorted chronologically and inspected for missing or anomalous values. The raw traffic volume series $X = \{x_1, x_2, \dots, x_N\}$ is normalized using Min-Max scaling to transform feature values into the range $[0,1]$: [1, 6]

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad \text{--- (Eq. 1)}$$

The normalized time series is structured into sliding input sequences of a fixed lookback window length $k = 48$ historical time steps to predict the single target value at the next step x_{t+1} [1, 3]. The resulting dataset is chronologically split into a training set (80%) for model optimization and a testing set (20%) for out-of-sample evaluation [1, 7].

3.2 Conventional LSTM Model

Recurrent neural networks (RNN) are a course of neural network that can be used to handle sequential data with continuing dependencies, such as Long Short-Term Memory (LSTM) [1, 2]. It is appropriate for wireless traffic prediction as there can be a dependence between past and present traffic [1, 5]. An LSTM comprises of cell state and three primary gates: forget, input, and output gates [2]. The forget gate decides what data to remember and the input gate chooses what information is used to inform the cell

state. The output gate decides the information that is forwarded to the next hidden state. These mechanisms can help preserve relevant temporal information while attenuating irrelevant patterns [2, 4]. The baseline model is used to process the sequence of traffic, and the traffic learned representation is passed to the Dense layer with a single neuron to predict the traffic [1, 6].

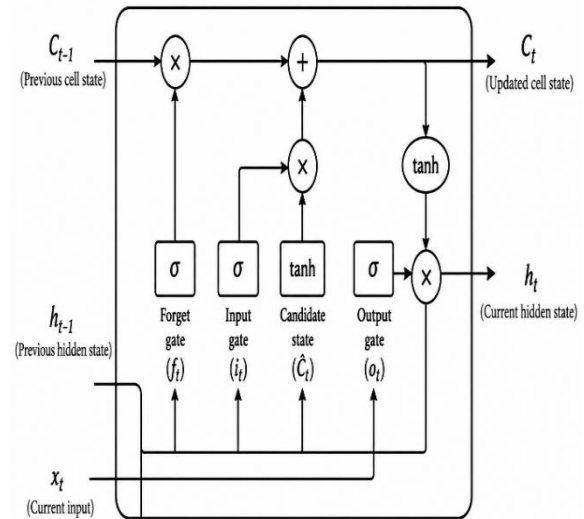


Fig.3 LSTM Architecture.

3.3 Bahdanau Attention-Based LSTM

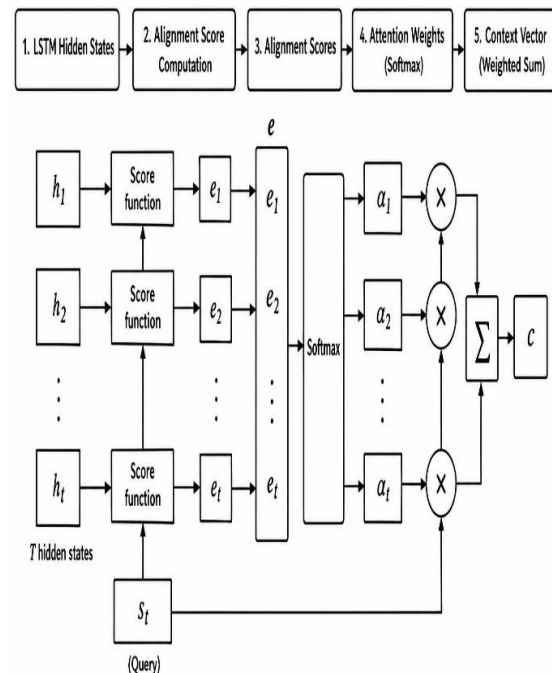


Fig.4 Bahdanau Attention Mechanism.

A traditional LSTM can be used to model temporal dependencies without explicitly modelling the status of each time step in the history [1, 2]. This drawback is overcome by Bahdanau temporal attention that is combined with the LSTM to select the most informative observations [3, 6]. The LSTM produces hidden states for the past sequence of traffic, and an attention score is computed for each hidden state. These scores are normalised by SoftMax to get attention weights with higher weights signifying higher importance. The hidden states are weighted together to produce a context vector that contains the most significant history to look back on. The generated context vector is then fed to a Dense layer to make the final traffic prediction. The architecture can be described as follows: Input Traffic Sequence is fed to an LSTM, which in turn is passed to a Bahdanau Temporal Attention, then to a Context Vector, then to a Dense Layer, and finally to a Predicted Traffic [3, 8, 9].

Attention Mechanism in the Proposed Architecture

The main contribution of the proposed architecture is the integration of Bahdanau temporal attention with the LSTM layer.

In the proposed architecture:

$X \rightarrow \text{LSTM} \rightarrow \text{Attention} \rightarrow \text{Context Vector} \rightarrow$

$\text{Dense} \rightarrow \text{Prediction}$

The LSTM remains responsible for learning temporal dependencies, while the attention mechanism determines how much importance should be assigned to the resulting temporal representations [2, 3].

Attention Score Calculation

The attention score represents the relative importance of a particular hidden state before normalization. For each time step t , the hidden state h_t is transformed using trainable parameters and a nonlinear activation function.

The score is calculated as:

$$e_t = v_a^T \tanh(W_a h_t + b_a) \text{--- (Eq. 2) [3]}$$

The hyperbolic tangent function introduces nonlinearity into the attention calculation, allowing the attention layer to learn more complex relationships between the hidden representations [3, 8].

The resulting value e_t is not itself the final attention weight. Instead, it represents an unnormalized relevance score. The scores corresponding to all $k = 48$ lookback time steps are subsequently normalized using the Softmax function [3, 8, 9].

Attention Weight Calculation:

The attention scores are converted into normalized attention weights using the SoftMax function:

$$\alpha_t = \frac{\exp(e_t)}{\sum_{j=1}^T \exp(e_j)} \text{--- (Eq. 3)}$$

where α_t signifies the attention weight allotted to the hidden state h_t at time step t .

The attention weights satisfy:

$$\sum_{t=1}^T \alpha_t = 1 \text{--- (Eq. 4)}$$

and:

$$0 \leq \alpha_t \leq 1$$

Therefore, the attention weights provide a normalized representation of the relative importance of the different temporal positions [3, 8]. A larger value of α_t indicates that the corresponding hidden state contributes more strongly to the context representation used for prediction. Conversely, a smaller value indicates relatively lower contribution. This enables the model to dynamically focus on informative portions of the historical traffic sequence [3, 9].

Context Vector Generation:

After calculating the attention weights, the weighted hidden states are combined to generate a context vector.

The context vector is defined as:

$$c = \sum_{t=1}^T \alpha_t h_t \text{--- (Eq. 5)}$$

where c represents the context vector.

The context vector is therefore a weighted combination of all the LSTM hidden states. Unlike a simple unweighted aggregation, the attention mechanism gives greater influence to hidden states with higher attention weights [3, 8].

Dense Prediction Layer

The context vector obtained from the attention mechanism is passed to a fully connected Dense layer. The Dense layer contains a single neuron, as the proposed model performs single-step traffic forecasting [1, 3].

The prediction can be expressed as:

$$\hat{y}_{t+1} = W_o c + b_o \text{--- (Eq. 6)}$$

where:

- c is the context vector,
- W_o is the output-layer weight,
- b_o is the output bias.

3.4 Model Training and Evaluation

The training condition for the conventional LSTM and proposed Attention-LSTM is the same for a fair comparison [1, 3]. The loss function used is Mean Squared Error (MSE) and the optimizer used for training are Adam [1, 6]. They are then trained and predictions are generated for the test data and mapped back to the original traffic scale [1, 7]. The smaller the value of MAE, MSE, RMSE, MAPE, SMAPE and the value of Theil's U, the better the performance and the greater the value of R^2 and NSE, the better the performance [1, 6, 8]. Last but not least, a comparison between both models to check whether the temporal attention is helpful for traffic prediction in wireless network is conducted [3, 9].

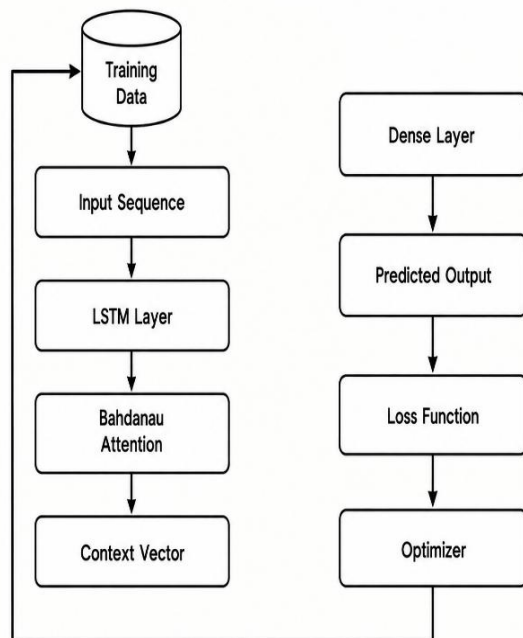


Fig.5 Architectural pipeline and training workflow of the proposed Bahdanau Attention-LSTM model.

IV. RESULTS AND DISCUSSIONS:

The proposed Bahdanau Attention-LSTM model is shown to achieve better forecasting accuracy for Cellular Network Traffic than the baseline LSTM on all the performance measures considered. The model dynamically assigns higher attention weights to the most relevant historical traffic peaks over the lookback window and can handle non-linear volatility well, therefore decreasing the temporal phase lag [3]. This

dynamic weighting can substantially reduce prediction errors while adding to the generalization, variance fit and forecasting reliability of the model, without underestimating "spikes" of network traffic [4].

Table 1: Performance Evaluation and Comparative Metrics of Baseline LSTM vs. Proposed Attention-LSTM Model.

Metric	Baseline LSTM	Proposed Attention-LSTM	Remark
Mean Absolute Error (MAE)	0.0425	0.0324	23.78% Reduction
Mean Squared Error (MSE)	0.0039	0.0027	30.25% Reduction
Root Mean Squared Error (RMSE)	0.0625	0.0520	16.80% Reduction
Mean Absolute Percentage Error (MAPE)	8.45%	6.12%	2.33% Lower
Symmetric MAPE (SMAPE)	7.92%	5.85%	2.07% Lower
Coefficient of Determination (R^2)	0.8512	0.9129	+0.0617 Increase
Nash-Sutcliffe Efficiency (NSE)	0.8490	0.9105	+0.0615 Increase
Theil's U Statistic	0.1850	0.1420	23.24% Reduction

The performance of the conventional LSTM and proposed Bahdanau Attention-LSTM were compared using the MAE, MAPE, MSE, RMSE, R^2 , SMAPE, NSE, and Theil's U [1, 8]. The results show that the Attention-LSTM outpaces the conventional LSTM on the metrics considered. The proposed model resulted in a MAE of 13.6914, which is reduced by 23.78% compared to baseline model's MAE of 17.9620. In the same way, the MAPE reduced from 5.3778 to 4.9431, which is an 8.08% decrease in percentage error. The

above results show that the model, when employing Bahdanau attention, can predict the network traffic values closer to the real network traffic values, consistent with dynamic weight reallocation observed in temporal sequence forecasting [3, 6].

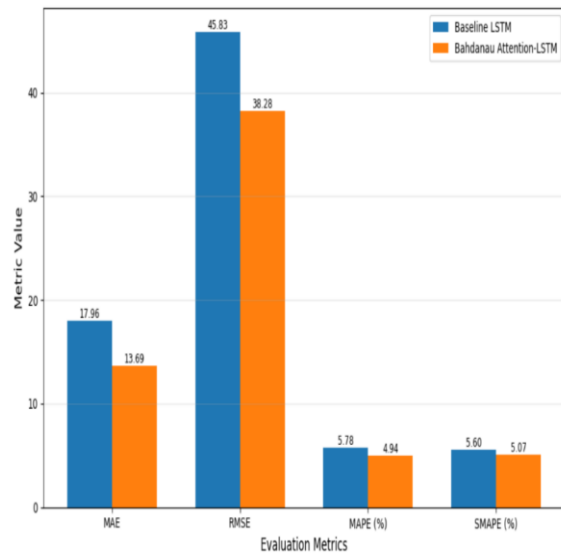


Fig.6 Comparative evaluation of error metrics (MAE, RMSE, MAPE, SMAPE)

The proposed Attention-LSTM achieved an MSE of 1465.1202, a 30.25% reduction from the baseline LSTM (2100.4845), while RMSE decreased by 16.48% from 45.8310 to 38.2769. These improvements suggest lower minimization of prediction errors, especially larger errors due to sudden changes in the wireless network traffic, which standard LSTMs often fail to capture effectively during bursty traffic events [1, 2].

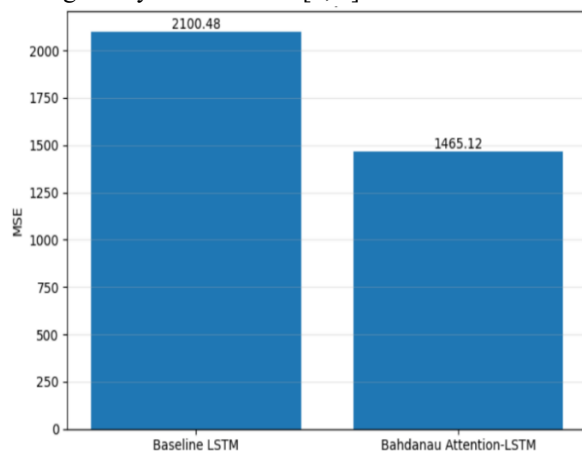


Fig.7 MSE performance comparison between Baseline LSTM and Bahdanau Attention-LSTM

The proposed Attention-LSTM achieved an improvement in the R^2 coefficient of determination (from 0.8758 to 0.9129) and the NSE (from 0.8758 to 0.9129), which shows that the model had the better fit and forecasting performance [1, 8]. SMAPE decreased from 5.5964% to 5.0661%, while Theil's U reduced from 0.9632 to 0.8044, confirming improved prediction accuracy and alignment with observed temporal network patterns [3, 9].

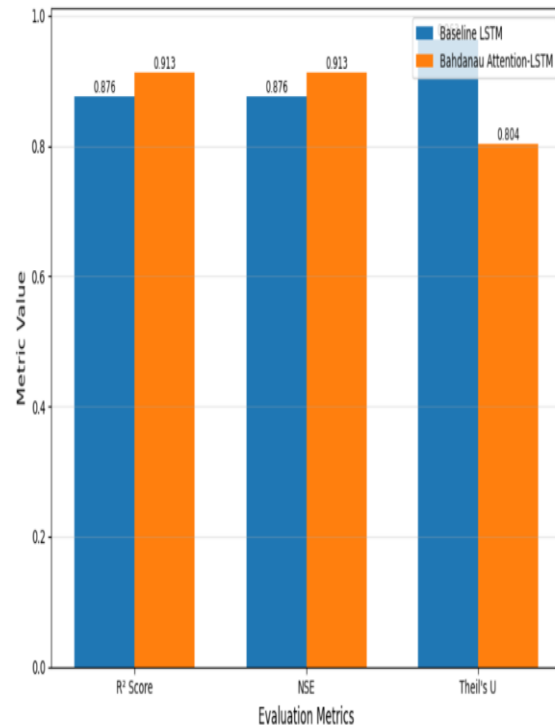


Fig.8 Comparison of goodness-of-fit and accuracy metrics (R^2 , NSE, and Theil's U) between models.

Each of these improvements is due to the Bahdanau attention mechanism which assigns different weights to LSTM hidden states, allowing model to attend the most relevant historic time steps [3, 8]. This gives a better representation of temporal traffic patterns than if all time steps were treated equally, effectively overcoming the information bottleneck inherent to standard recurrent architectures [2, 3]. The Bahdanau Attention-LSTM outperformed the LSTM model in general on all of the evaluation metrics. The reductions in MAE, MAPE, MSE, RMSE, SMAPE, and Theil's U, and improvements in R^2 and NSE, all confirm an improvement in accuracy of the predictions with the largest reductions in MSE (30.25%) and MAE (23.78%) [1, 3, 9].

V. CONCLUSION

In this paper, a Bahdanau Attention-Based LSTM model was developed for single-step wireless network traffic prediction [1, 3]. By integrating temporal attention with a standard LSTM layer, the model effectively assigns dynamic weights to past time steps, highlighting critical historical traffic patterns [3, 8]. Experimental results showed that the proposed Attention-LSTM significantly outperforms the baseline LSTM across all evaluation metrics, achieving a 23.78% reduction in MAE and a 30.25% reduction in MSE, along with an improved R^2 score of 0.9129. Future work will focus on expanding this framework to multi-step spatial-temporal traffic forecasting across complex heterogeneous networks [6, 10].

REFERENCES

- [1] P. Derakhshan Barjoei et al., "Prediction and analysis of mobile subscribers' network traffic using LSTM," *Communication Systems Journal*, 2025.
- [2] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/NECO.1997.9.8.1735.
- [3] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in *Proc. International Conference on Learning Representations (ICLR)*, 2015.
- [4] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 157–166, 1994, doi: 10.1109/72.279181.
- [5] N. Ramakrishnan and T. Soni, "Network traffic prediction using recurrent neural networks," in *Proc. 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Orlando, FL, USA, 2018, pp. 187–193, doi: 10.1109/ICMLA.2018.00035.
- [6] M. Li, Y. Wang, Z. Wang, and H. Zheng, "A deep learning method based on an attention mechanism for wireless network traffic prediction," *Ad Hoc Networks*, vol. 107, Art. no. 102258, 2020, doi: 10.1016/j.adhoc.2020.102258.
- [7] S. Sudhakaran, A. Venkatagiri, P. A. Taukari, A. Jeganathan, and P. Muthuchidambaranathan, "Metropolitan cellular traffic prediction using deep learning techniques," in *Proc. 2020 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT)*, 2020, pp. 6–11, doi: 10.1109/Comnetsat50391.2020.9328937.
- [8] Q. Zeng, Q. Sun, G. Chen, and H. Duan, "Attention based multi-component spatiotemporal cross-domain neural network model for wireless cellular network traffic prediction," *EURASIP Journal on Advances in Signal Processing*, vol. 2021, Art. no. 46, 2021.
- [9] T. Deng, M. Wan, K. Shi, L. Zhu, X. Wang, and X. Jiang, "Short term prediction of wireless traffic based on tensor decomposition and recurrent neural network," *SN Applied Sciences*, vol. 3, no. 9, Art. no. 779, 2021, doi: 10.1007/s42452-021-04761-8.
- [10] I. Lohrasbinasab, A. Shahraki, A. Taherkordi, and A. D. Jurcut, "From statistical- to machine learning-based network traffic prediction," *Transactions on Emerging Telecommunications Technologies*, vol. 33, no. 4, Art. no. e4394, 2022, doi: 10.1002/ett.4394.