Predicting Network Traffic Flow with a Multi-Layer Short-Term Memory Model

Taran Puvvala[1]✉

[1]North Carolina School of Science and Mathematics ROR

Abstract

In the dynamic realm of digital communication, managing and optimizing network traffic is critical. Efficient forecasting of network traffic volumes, crucial for maintaining network quality, poses significant challenges due to the volatile nature of data flow. This study introduces a novel approach using Long Short-Term Memory (LSTM) networks to enhance forecasting accuracy. We compare a multi-layer LSTM, designed to capture complex dependencies within the data, with a traditional single-layer LSTM, offering insights into their respective capabilities. The multi-layer LSTM's architecture allows it to excel in understanding deeper temporal patterns, making it a significant advancement over simpler models. By employing these two architectures, our research aims to improve predictive performance in network traffic forecasting, crucial for optimized network management and planning. The findings reveal that the multi-layer model significantly outperforms its single-layer counterpart, demonstrating the potential of sophisticated LSTM networks in practical applications.

1. INTRODUCTION

The forecasting of network traffic volume stands as a cornerstone in the management and optimization of modern data networks [1]. In a digital landscape where data consumption and transmission have surged exponentially, the ability to predict network traffic is paramount. This importance stems from the increasing reliance on online services and cloud computing, with networks serving as the backbone of digital communication [2]. Accurate forecasting enables service providers to maintain network efficiency and quality of service, ensuring that the burgeoning demands of users are met effectively and reliably.

Network traffic volume forecasting is a complex endeavor, often formulated as a Time Series Forecasting (TSF) problem [3]. This task involves constructing models that estimate future traffic volumes based on historical data patterns. However, the inherent challenges of this task arise from the unpredictable and dynamic nature of network traffic, influenced by user behavior, technological shifts, and socio-economic factors. Understanding these dynamics is crucial for creating effective predictive models and maintaining robust network operations [4].

The challenge of forecasting network traffic is amplified by the high dimensionality and large data volumes typical of modern networks. Factors such as emergency situations, varying user demands, and technological changes add layers of complexity to the forecasting process. Accurate predictions require models capable of understanding and adapting to these multifaceted and often nonlinear dynamics [5]. Technologies such as Vehicle-to-Everything (V2X) communications enhance traffic prediction by providing real-time data from vehicles and infrastructure, which can be integrated to forecast traffic flows more accurately [6]. Similarly, advancements in artificial intelligence (AI) facilitate the analysis of complex data patterns, enabling more precise predictions of network traffic behaviors. These technologies contribute significantly to the mapping and predictive capabilities necessary for dynamic network management [6].

Despite its critical importance, network traffic forecasting remains a challenging task. Current methods often struggle to accurately capture the intricate and volatile patterns of network data. Traditional models may not fully account for the nonlinear temporal dependencies and are generally ill-equipped to handle sudden fluctuations in traffic. This limitation can lead to suboptimal network resource allocation and management, potentially impacting service quality and user experience [7], [8].

Given these complexities, this study aims to explore the application of Long Short-Term Memory (LSTM) networks in predicting network traffic volumes.

Multi-layer Long Short-Term Memory (LSTM) networks offer significant advantages in this context over traditional single-layer LSTMs or other simpler models. By utilizing multiple layers, multi-layer LSTMs can capture more complex and subtle patterns in data, which are often missed by simpler approaches. This capability is especially critical given the complex, nonlinear dynamics of network traffic, which include interactions across various time scales.

Our research focuses on whether LSTM models, renowned for their capability to process time-series data with intricate temporal patterns, can enhance the accuracy of network traffic forecasts. The study specifically examines and compares the efficacy of multi-layer and single-layer LSTM architectures in this context. The overarching goal is to contribute to more efficient and effective network traffic management strategies, aligned with the evolving demands of the digital era [9].

2. Computational Approach / Methods

In the realm of time-series forecasting, the meticulous design and implementation of computational models play a pivotal role in uncovering underlying patterns and predicting future trends. The methods adopted in this study reflect a harmonious blend of advanced data pre-processing techniques, the intricate architecture of neural networks, and strategic model training and testing procedures. By delving into the specifics of Long Short-Term Memory (LSTM) networks and their implementation, we aim to provide a transparent and thorough understanding of the methodologies that underpin our analysis. This exposition not only serves to clarify the technical intricacies of our approach but also to underscore the robustness and reliability of the derived results.

2.1. Preprocessing

The foundational step in our analysis was the meticulous preprocessing of a comprehensive traffic dataset. Our database used is a publicly available computer traffic set from Stanford University [10]. This set comprises of 8 different servers with over 500,000 rows of information regarding them. The data provides daily traffic metrics, was initially subjected to a thorough cleansing process. Key to this phase was the conversion of date entries into a standardized chronological format, facilitating temporal analyses. Focus was particularly directed towards the total traffic variable, representing daily traffic volumes, which was extracted for subsequent processing Recognizing the sensitivity of neural networks to data scale, we normalized these traffic volume figures using min-max scaling:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \qquad (1)$$

2.2. LSTM Overview

Long Short-Term Memory (LSTM) networks, an advanced variant of recurrent neural networks (RNNs), are specifically designed to address the challenges of learning long-range temporal dependencies [11]. Traditional RNNs, while theoretically capable of handling such dependencies, often fall short in practice due to issues like vanishing or exploding gradients. LSTMs overcome these hurdles with their unique architecture, making them highly effective for timeseries analysis and forecasting.

Core Components of LSTM Networks At the heart of an LSTM network are its LSTM units, each comprising four key elements:

- **Forget Gate:** This gate determines which information should be discarded from the cell state. It takes the previous hidden state and the current input, passes them through a sigmoid function, and outputs a number between 0 and 1 for each number in the cell state. A value close to 1 indicates "keep this information," while close to 0 signifies "discard this information." [12]
- **Input Gate:** This component decides what new information is to be stored in the cell state. It involves two parts: a sigmoid layer which decides which values will be updated, and a tanh layer that creates a vector of new candidate values that could be added to the state [12].
- **Cell State:** The cell state is the key to LSTMs' ability to retain long-term dependencies. It runs straight down the entire chain of the network, with only minor linear interactions. This design allows information to flow unaltered and mitigates the risk of vanishing gradients [12].
- **Output Gate:** Finally, the output gate determines what the next hidden state should be. The hidden state contains information about previous inputs, and is used for predictions. The sigmoid layer decides which parts of the cell state to output, and then a tanh layer creates a vector of all the cell state values, which is multiplied by the sigmoid output, producing the final output of the LSTM unit [12].

2.2.1. LSTM in the Context of Traffic Forecasting::

In our study, the LSTM's ability to capture temporal dynamics is harnessed for predicting traffic volumes. Traffic data, inherently sequential and time-dependent, presents an ideal case for applying LSTM networks [2]. By learning from past traffic patterns and their progression over time, the LSTM model can forecast future traffic trends with a high degree of accuracy. This predictive capability is particularly crucial for short-term forecasting, where understanding immediate trends is vital for effective traffic management and planning.

2.3. Data-Reshaping for LSTM

A crucial step in our implementation process was the preparation of the traffic volume data to make it suitable for LSTM processing. LSTM networks require the input data to be in a specific format, typically a 3-dimensional array. This format is essential because LSTMs are designed to process data in sequences and need to

understand the order and context of the input data. The three dimensions required for LSTM input are:

- **Samples:** Each sample represents an independent instance from the dataset. In our case, each sample corresponded to a specific day's data.
- **Time Steps:** This dimension represents the number of time intervals we consider for each sample. In our implementation, we used the look back parameter, set to 1, indicating that the model uses the traffic data from one previous time step (i.e., the previous day) to make a prediction for the current time step.
- **Features:** This dimension corresponds to the number of attributes or variables used for each time step. In our traffic dataset, this involved the normalized traffic volume data.

2.4. Our LSTM Implementation

In our research, the utilization of Long Short-Term Memory (LSTM) networks was a deliberate choice, aimed at leveraging their renowned capability for modeling time-series data, particularly in capturing temporal dependencies and patterns inherent in traffic volume data. To explore the full potential of LSTM networks in this context, we employed two distinct architectures: a multi-layer LSTM and a single-layer LSTM model. This dual-model approach was instrumental in comprehensively assessing the advantages and limitations of varying LSTM complexities in traffic forecasting.

The multi-layer LSTM model, characterized by its layered structure of LSTM units, was designed to capture more complex and nuanced patterns in the data. This model consisted of two LSTM layers: the first with 30 units and the second with 10 units. The rationale behind this specific architecture was to enable the model to learn higher-level temporal features in the first layer and then refine and distill these features in the subsequent layer, thus potentially enhancing the accuracy of the predictions [13]. The multi-layer approach was hypothesized to be particularly adept at handling the intricate temporal relationships and potential nonlinearities present in traffic data.

In contrast, the single-layer LSTM model, with its singular LSTM layer of 30 units, served as a comparative baseline. This simpler model was intended to provide a benchmark against which the performance of the more complex multilayer model could be evaluated. The choice of 30 units for this model was a balance between computational efficiency and the model's capacity to capture significant temporal dependencies in the data.

For both models, the look back parameter was set to 1, indicating that each prediction was based on the data from the immediately preceding day. This parameter choice was driven by our focus on short-term forecasting, aiming to understand and predict day-to-day fluctuations in traffic volumes. The models were also configured to output a single value, representing the predicted traffic volume for the next day. Additionally, We compiled the multi- layer and single-layer LSTM models using the mean squared error (MSE) as the loss function, a common choice for regression problems. The Adam Optimizer was selected for its efficiency in handling sparse gradients and adaptive learning rate adjustments, which are crucial for managing time-series data. The optimizer's role was pivotal in iteratively updating the network weights based on the training data,

aiming to minimize the loss function throughout the training process.

The utilization of both a multi-layer and a single-layer LSTM model provided us with a nuanced understanding of how different LSTM architectures influence forecasting accuracy in the context of traffic data. This approach allowed us to not only assess the effectiveness of LSTM networks in traffic volume prediction but also to explore the trade-offs between model complexity and predictive performance. By comparing the outcomes of these two models, we aimed to derive insights that could inform the optimal use of LSTM networks in similar time-series forecasting tasks, particularly those involving daily data with significant short-term temporal dependencies.

2.5. Training and Data Splitting

The training of our LSTM models was executed with a meticulous strategy, tailored to optimize their performance for short-term traffic forecasting. This process involved training the models over 100 epochs with a batch size of 1. We intentionally opted for a small batch size, as it allows for more frequent updates to the model weights, a technique that can be particularly advantageous in the context of time-series data. Frequent updates enable the model to adjust more dynamically to the nuances and variations in the data, thus enhancing its predictive accuracy.

A key aspect of our training regimen was the implementation of a training-testing split in the dataset. 80% of our data was used for testing while the remaining 20% was tested on. This split was crucial for evaluating the models' performance on data that they had not encountered during the training phase. We allocated a significant portion of our dataset for training, while reserving a smaller, yet adequate, portion for testing. This division ensured that the models were exposed to a comprehensive range of data scenarios during training, while still preserving a representative sample for unbiased evaluation during the testing phase.

Incorporated into our training strategy was the early stopping mechanism, a powerful tool in combating overfitting—a common challenge in machine learning where a model becomes overly attuned to the training data and performs poorly on new, unseen data. The early stopping callback was configured to monitor the training loss, ceasing the training process if no improvement in this metric was observed for a consecutive span of 10 epochs. This approach not only helped in conserving computational resources but also played a pivotal role in ensuring that the models preserved their ability to generalize to new data. The objective was to find an optimal balance between underfitting and overfitting, thereby achieving the best possible model performance on both the training and testing datasets.

During the training phase, a critical focus was on the learning process of the LSTM units, particularly on their ability to utilize their inherent gates—forget, input, and output. These gates are instrumental in the LSTM's capability to process and retain relevant information while discarding what is deemed irrelevant. The effective utilization of these gates is what empowers LSTMs to excel in modeling time-series data, such as our traffic volume predictions. It is this sophisticated mechanism of information processing and retention that enables LSTMs to capture and learn from the

	Average MAE	Average MSE	Average RMSE	Overall Average Accuracy
Multi-Layer LSTM	0.1005286	0.02618802	0.1549513	89.5%
Single-Layer LSTM	0.1145322	0.0426393	0.1921245	88.55%

Table 1. *Results of Both LSTM Models*

temporal dependencies inherent in sequential data, a characteristic pivotal to their success in our traffic forecasting task.

3. RESULTS

As shown in Table 1, for performance evaluation, a four-pronged approach was adopted, involving Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). These metrics, central to understanding the models' accuracy, are defined as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | \; y_i - \overset{*}{y_i} \; | \tag{2}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n} \left(y_i - \overset{*}{y_i} \right)^2 \tag{3}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left(y_i - \overset{*}{y_i} \right)^2} \tag{4}$$

$$OverallActualAccuracy = (1 - MAE) * 100 \tag{5}$$

Building upon our established framework for performance evaluation, the outcomes of our LSTM models were thoroughly examined through Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Overall Actual Accuracy. These metrics collectively offered a comprehensive view of the models' predictive capabilities.

3.1. Multi-layer Performance

The performance of the Multi-Layer LSTM model was notably promising. The model achieved an average MAE of 0.1005286, indicating that the predictions were, on average, approximately 10.05% off from the actual traffic volumes. This degree of accuracy is particularly significant in the realm of traffic forecasting, where precision can have substantial implications. The predicted vs actual network flow is depicted in Figure 1 where the predicted values seem to stay primarily around the same area. As the prediction is heavily based on the previous value (look back = 1) the model struggled with the predicting peaks.

In terms of MSE, the model exhibited a value of 0.02618802. This figure suggests a consistent performance across predictions with minimal occurrences of extreme errors. The RMSE, standing at 0.1549513, further substantiates the model's accuracy, offering a realistic measure of the error magnitude in the same units as the predicted variable.
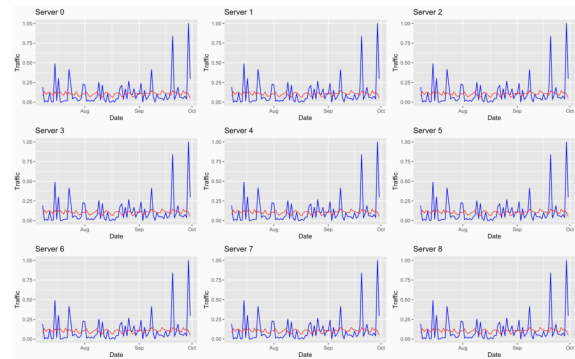

Figure 1. *Predicted Network Traffic (Red) vs Actual Network Traffic (Blue)*

A key highlight was the model's Overall Actual Accuracy, which was calculated to be approximately 89.95%. This high level of accuracy underscores the model's effectiveness in capturing the complex temporal patterns within the traffic data, affirming its suitability for short-term forecasting applications.

3.2. Comparative Analysis

Contrasting these results, the Single-Layer LSTM model, with its simpler architecture, yielded slightly different outcomes. The model recorded an average MAE of 0.1145322, suggesting a deviation of about 11.45% from the actual values. Though marginally higher than the Multi-Layer model, this still signifies a commendable level of accuracy.

The MSE and RMSE values for the Single-Layer model were 0.0426393 and 0.1921245, respectively. These metrics, being higher than those of the Multi-Layer LSTM, point to a greater variance in the predictions and a larger error magnitude. This variance is depicted in Figure 2. The model's Overall Actual Accuracy stood at 88.55%, slightly lower than that highlights the enhanced capability of the multi-layer architecture in handling complex time-series data like traffic volumes.

3.3. Statistical Significance

The approximately 1.5% increase in accuracy from the single-layer to the multi-layer LSTM model, though seemingly modest, is statistically significant and carries practical implications for network traffic management. This improvement is particularly notable given the complexity and variability of network traffic data, where even slight enhancements can lead to substantial benefits in operational efficiency and service quality.

Statistically, the increase in accuracy indicates a more reliable model that is better at capturing the non-linear and dynamic patterns of network traffic. This is crucial for applications where forecasts directly influence real-time decisions,such as dynamic bandwidth allocation and predictive maintenance. In such contexts, even a small percentage improvement can translate into significant operational advantages, such as reduced downtime and optimized resource allocation.

Moreover, from a statistical perspective, the consistency of improved performance across multiple metrics (MAE, MSE, RMSE) suggests that the multi-layer LSTM is not only occasionally outperforming the single-layer model by chance but is systematically more effective. This consistency reinforces the argument for the
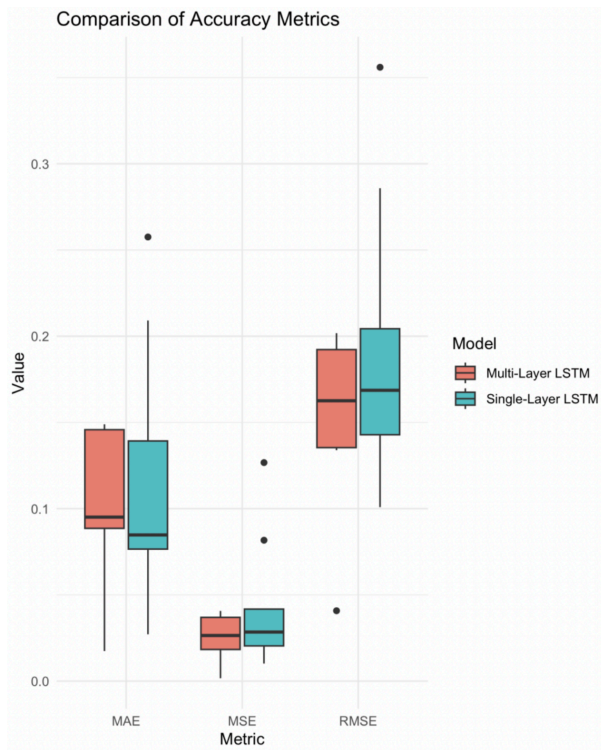
Figure 2.   *Box Plots of Accuracy Metrics Comparing Single & Multi-Layer LSTM Models*

robustness of the multi-layer architecture in handling complex datasets typical of network environments.

4. Discussion

This research aimed to explore the efficacy of Long Short-Term Memory (LSTM) networks in the context of network traffic volume forecasting, a domain where accurate predictions can significantly impact the Cyber-Space. Our investigation was grounded in the application of two LSTM architectures - a Multi-Layer LSTM and a Single-Layer LSTM - each offering different levels of complexity and potential for capturing temporal dependencies in the data.

4.1. LSTM Model Performance

The Multi-Layer LSTM model's superior performance, as evidenced by its lower MAE, MSE, and RMSE, and higher overall accuracy, underscores the value of complex neural network architectures in handling intricate patterns in time-series data. The additional layers in this model provided a more nuanced understanding of the temporal sequence, enabling it to capture subtleties in traffic trends that a simpler model might overlook. This finding aligns with existing literature, which suggests that deeper neural network architectures often yield better results in complex data scenarios, albeit at the cost of increased computational resources and potential risks of overfitting [13].

Conversely, the Single-Layer LSTM model, while not outperforming the Multi-Layer LSTM, demonstrated respectable accuracy, highlighting its utility as a more straightforward and computationally efficient alternative. This model's slightly higher error metrics could be attributed to its limited capacity to model

complex temporal relationships, a trade-off that comes with reduced model complexity.

4.2. Implications for Traffic Forecasting

The application of Long Short-Term Memory (LSTM) models for network data volume forecasting, as explored in our study, carries profound implications for network data management and planning. The ability of LSTM models to accurately predict network traffic volumes is particularly crucial in the current era, where the efficient handling of data is integral to network operations.

The enhanced predictive capabilities of LSTM models, as evidenced by their performance in our study, mark a significant advancement in network data forecasting. These models, adept at capturing complex temporal dependencies, allow for a more accurate anticipation of network loads. This is especially relevant given the current landscape of rapid data generation and consumption. Accurate forecasting enables network administrators to proactively manage network resources, allocate bandwidth efficiently, and optimize overall network performance. This proactive approach to network management is invaluable, especially in preventing congestion during peak traffic periods and maintaining high-quality service for users.

Furthermore, the insights gained from LSTM-based forecasting facilitate strategic network planning and informed decision-making. Network operators and service providers can leverage this data to plan infrastructure expansions and technology upgrades, ensuring that networks are equipped to handle future data demands. This strategic foresight aids in aligning network capabilities with projected requirements, thereby enhancing network efficiency and reducing the risk of over or under-provisioning network resources.

Cost efficiency and operational sustainability are additional benefits that stem from improved network data forecasting. By accurately aligning network capacity with actual data demands, network operators can avoid unnecessary expenditures on infrastructure while still satisfying user needs. Additionally, more precise data forecasting contributes to sustainability by optimizing the use of energy resources within network operations [6].

The study's findings also have significant implications for managing the complexities introduced by emerging technologies such as IoT, cloud computing, and 5G networks. These technologies, characterized by their high data throughput and diverse traffic patterns, present new challenges in network management. The ability of LSTM models to process and forecast large volumes of complex data makes them particularly well-suited for navigating these challenges. Accurate forecasting of data traffic from a myriad of devices and services underscores the potential of LSTM models in ensuring the efficient integration and management of these emerging technologies within network infrastructures.

4.3. Considerations for Model Selection

In choosing between a Multi-Layer and Single-Layer LSTM model, one must consider factors beyond just predictive accuracy. The complexity of a multi-layer model, for instance, necessitates greater computational resources and may pose challenges in terms of training time and model interpretability. Additionally issues with resource efficiencies can arise. Multi-layer LSTMs, due to their

complexity, require more computational resources, including processing power and memory. This can lead to increased operational costs and longer training times, which may not be feasible for all organizations or in scenarios where rapid deployment of models is necessary. Furthermore, the complexity of multi-layer LSTMs also poses a higher risk of overfitting, especially when the available data is limited or the feature space is not sufficiently diverse. Overfitting occurs when a model learns the details and noise in the training data to an extent that it negatively impacts the performance of the model on new data. This makes multi-layer LSTMs less suitable for smaller datasets or less complex problem domains where a simpler model could generalize better. On the other hand, a single-layer model, while less computationally demanding, might not adequately capture the complexity of certain datasets.

4.4. Future Direction

Building upon the insights gained from our study on LSTM models for network traffic forecasting, several promising avenues emerge for future research in this domain. These directions aim to expand the scope and efficacy of forecasting models, catering to the evolving needs of network management and optimization.

The integration of additional data sources stands as a critical area for enhancement. By incorporating diverse data sets, such as network usage during special events, cybersecurity incident data, and user behavior patterns, future models can gain a more holistic understanding of network dynamics. This comprehensive approach could significantly improve the accuracy of traffic predictions and offer a nuanced understanding of various factors influencing network load.

Advancements in model architecture also present exciting possibilities. Experimenting with hybrid models that combine the temporal learning capabilities of LSTMs with spatial analysis techniques could yield more sophisticated forecasting tools. For instance, integrating LSTMs with CNNs or graph neural network might provide a deeper insight into both the temporal and spatial aspects of network traffic. Fine-tuning LSTM configurations, exploring different layer structures, and adapting the models to specific network types are key areas that could lead to more optimized and effective forecasting solutions.

The development of real-time forecasting capabilities is another critical frontier. Models that can offer accurate predictions in near real-time would be invaluable for dynamic network management, allowing for rapid responses to emerging traffic patterns and potential network issues. Such capabilities would significantly enhance the responsiveness and adaptability of network operations.

Additionally, the scalability and generalizability of LSTM models across various network architectures warrant further exploration. Testing these models in diverse network environments, such as cloud networks, IoT infrastructures, and 5G networks, would provide insights into their versatility and applicability. This research could help ascertain the models' effectiveness across different settings and identify any necessary adaptations or optimizations.

Finally, the practical application of LSTM-based forecasting models in real-world network management systems is a crucial step towards their widespread adoption. Integrating these models into existing network infrastructure and evaluating their performance in live environments would not only validate their practical utility but also highlight areas for improvement in real-time network traffic management.

5. Conclusions

This study has successfully demonstrated the potential of LSTM networks in the realm of network traffic forecasting, a critical component in the efficient management of modern digital networks. By employing and comparing both multi-layer and single-layer LSTM models, our research has shed light on the capabilities and limitations of these models in handling the complex task of predicting network traffic volumes. The findings from this investigation underscore the effectiveness of LSTM models in capturing intricate temporal dependencies inherent in network data, thus offering a significant contribution to the field of network data management.

Our research confirms that LSTM models, particularly those with multiple layers, can provide a high degree of accuracy in forecasting network traffic. This accuracy is crucial for network operators to proactively manage network resources, ensuring optimal performance and quality of service. The study also highlights the importance of careful data preprocessing and the selection of appropriate model parameters in enhancing the performance of LSTM models. Moreover, the comparative analysis of different LSTM architectures provides valuable insights for researchers and practitioners in choosing the right model based on the specific requirements and constraints of their network environment

If given the opportunity to further advance this research, several enhancements could be pursued. First, integrating additional types of data, such as real-time traffic updates, weather conditions, and special event schedules, could refine the models' predictive capabilities. This would allow for a more comprehensive understanding of the factors influencing network traffic and could lead to even more robust forecasting models. Additionally, exploring the application of hybrid models that combine the strengths of LSTM with other advanced machine learning techniques like convolutional neural networks or graph neural networks could address some of the limitations observed in traditional LSTM models. These hybrid models could potentially offer a deeper insight into both the temporal and spatial aspects of network traffic.

Moreover, developing real-time forecasting capabilities would be a significant step forward. This would not only enhance the responsiveness of network management systems but also allow for more dynamic adjustments to network operations, ensuring high efficiency and user satisfaction. Implementing these models in a simulated environment to test their real-world applicability and performance before full-scale deployment could mitigate risks and refine the models' effectiveness.

Looking forward, the application of LSTM models in network traffic forecasting opens up exciting possibilities for the development of more intelligent, efficient, and adaptive network management systems. The integration of LSTM-based forecasting models into real-world network operations could mark a significant step towards

more proactive and data-driven network management. As the digital landscape continues to evolve, with increasing data traffic and the integration of emerging technologies, the role of accurate and reliable traffic forecasting will become ever more critical.

References

[1] S. Nihale, S. Sharma, L. Parashar, and U. Singh, "Network Traffic Prediction Using Long Short-Term Memory," in *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, 2020, pp. 338–343. doi: 10.1109/ICESC48915.2020.9156045.

[2] A. Khodadadi, "Traffic Forecasting Using Graph Neural Networks and LSTM." [Online]. Available: https://keras.io/examples/timeseries/timeseries_traffic_forecasting/

[3] V. Kramar and V. Alchakov, "Time-Series Forecasting of Seasonal Data Using Machine Learning Methods," *Algorithms*, vol. 16, no. 5, p. 248–249, 2023, doi: 10.3390/a16050248.

[4] I. Lohrasbinasab, A. Shahraki, A. Taherkordi, and A. Delia Jurcut, "From statistical to machine learningbased network traffic prediction," *Transactions on Emerging Telecommunications Technologies*, vol. 33, no. 4, p. e4394, 2022, doi: 10.1002/ett.4394.

[5] W. Jiang and J. Luo, "Graph neural network for traffic forecasting: A survey," *Expert Systems with Applications*, vol. 207, p. 117921–117922, 2022, doi: 10.1016/j.eswa.2022.117921.

[6] K. Zhang, X. Zhao, X. Li, X. You, and Y. Zhu, "Network Traffic Prediction via Deep Graph-Sequence Spatiotemporal Modeling Based on Mobile Virtual Reality Technology," *Wireless Communications and Mobile Computing*, vol. 2021, pp. 1–12, 2021, doi: 10.1155/2021/2353875.

[7] J. Liu, N. Wu, Y. Qiao, and Z. Li, "A scientometric review of research on traffic forecasting in transportation," *IET Intelligent Transport Systems*, vol. 15, no. 1, pp. 1–16, 2021, doi: 10.1049/itr2.12024.

[8] A. Mozo, B. Ordozgoiti, and S. Gómez-Canaval, "Forecasting short-term data center network traffic load with convolutional neural networks," *PLOS ONE*, vol. 13, no. 2, p. e191939, 2018, doi: 10.1371/journal.pone.0191939.

[9] E. Haugsdal, S. Malacarne, and M. Ruocco, "Circle Attention: Forecasting Network Traffic by Learning Interpretable Spatial Relationships from Intersecting Circles," in *Machine Learning and Knowledge Discovery in Databases: Applied Data Science and Demo Track*, vol. 14175, G. De Francisci Morales, C. Perlich, N. Ruchansky, N. Kourtellis, E. Baralis, and F. Bonchi, Eds., Cham: Springer Nature Switzerland, 2023, pp. 106–121. doi: 10.1007/978-3-031-43430-3_7.

[10] C. Sabatti, "Computer Network Traffic Data." [Online]. Available: https://chiarasabatti.su.domains/data.html

[11] 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.

[12] F. Gers, "Learning to forget: continual prediction with LSTM," in *9th International Conference on Artificial Neural Networks: ICANN '99*, 1999, pp. 850–855. doi: 10.1049/cp:19991218.

[13] H. N. Bhandari, B. Rimal, N. R. Pokhrel, R. Rimal, K. R. Dahal, and R. K. Khatri, "Predicting stock market index using LSTM," *Machine Learning with Applications*, vol. 9, p. 100320–100321, 2022, doi: 10.1016/j.mlwa.2022.100320.