LSTM-based Resource Prediction for Disaggregated RAN in 5G Non-Terrestrial Networks

Henok B. Tsegaye*, Petro M. Tshakwanda†, Yonatan M. Worku ‡, Claudio Sacchi§,
Christos Christodoulou¶, Michael Devetsikiotis‖
†‡§¶‖Department of Electrical and Computer Engineering, University of New Mexico, USA
*§Department of Information Engineering and Computer Science, University of Trento, Italy
{†pmushidi, ‡yonatanmelese61, ¶christos, ‖mdevets}@unm.edu, {*henokberhanu.tsegaye, §claudio.sacchi}@unitn.it,
§claudio.sacchi.unm@outlook.com

Abstract*—The growing demand for advanced beyond 5G connectivity solutions explores the deployment of end-to-end 5G Non-Terrestrial Networks (NTNs) in cloud-native environments. With the increasing reliance on mobile communications, leveraging next-generation radio access network (NG-RAN) architectures with functional splits has become essential. In the 5G NTN network, some split NG-RAN components can be moved to the satellite node to improve connectivity and resilience. This paper investigates advanced beyond 5G connectivity by deploying end-to-end 5G Non-Terrestrial Networks (NTNs) in cloud-native environments, focusing on Low Earth Orbit (LEO) satellites operating in regenerative mode. Specifically, it explores the implementation of F1 and E1 interface splits within such networks. The first architecture extends the F1 interface over the satellite radio interface (F1 over SRI), linking terrestrial central units (gNB-CU) with satellite-based distributed units (gNB-DU). The second architecture incorporates both F1 and E1 splits, facilitating connections between terrestrial control plane units (gNB-CUCP) and user plane units (gNB-CUUP) on the satellite via the E1 interface over SRI (F1-E1 over SRI). The study's primary goal is to predict the resource utilization—specifically CPU, memory, and bandwidth—of gNB-DU and gNB-CUUP functioning as satellite payloads. Employing Long-Short-Term Memory (LSTM) neural networks, this research aims to enhance network resilience by enabling proactive monitoring and resource allocation decisions, addressing the significant computational and bandwidth demands of payloading gNB-CUUP compared to gNB-DU.

***Index Terms*—5G, NTN, LSTM, Disaggregated NG-RAN.**

I. INTRODUCTION

The rapid advancement and deployment of 5G networks have revolutionized the telecommunications landscape, ushering in an era of ultra-reliable, high-speed, and low-latency communications. This evolution is driven by the increasing demand for high-bandwidth applications such as autonomous driving, smart cities, and the Internet of Things (IoT), which require robust and efficient network infrastructure to handle heterogeneous and dynamic traffic loads [1].

Non-terrestrial networks (NTNs), including satellite communications, complement terrestrial 5G networks by providing extensive coverage, particularly in remote and underserved areas [2]. Integrating NTNs with terrestrial networks promises to enhance global connectivity, enabling seamless communication across diverse environments [3].

In this context, disaggregated radio access networks (RANs) have emerged as a pivotal innovation. Disaggregated RANs separate the traditional monolithic base station architecture into distinct components such as the Central Unit (CU), Distributed Unit (DU), and Radio Unit (RU) [4]. This separation facilitates flexible deployment and efficient management of network resources, thereby enhancing network performance and scalability [5].

One of the primary challenges in managing 5G NTNs with disaggregated RAN architectures is predicting and optimizing resource consumption to maintain Quality of Service (QoS) and minimize Service Level Agreement (SLA) violations. Accurate prediction of network traffic and resource usage is essential for dynamic network slicing, efficient traffic steering, and proactive failure management [6]. In such a framework, Machine Learning can offer valuable solutions. In particular, LSTM networks - a specific typology of recurrent neural network (RNN) - have demonstrated remarkable success in time-series prediction due to their ability to capture long-term dependencies in data [7]–[9]. LSTM models are particularly well-suited for predicting network traffic and resource consumption in 5G networks, where traffic patterns are dynamic and complex [10] [11].

This paper focuses on developing an LSTM-based resource prediction model for disaggregated RAN in 5G NTNs. We propose two architectures: the first involves splitting the gNB into gNB-CU and gNB-DU across the satellite network, extending the F1 interface over the satellite radio interface (SRI). The second architecture implements the F1 and E1 splits, with the gNB-CUUP as a satellite payload. These architectures are evaluated for their resource consumption (CPU, memory, and bandwidth) and prediction accuracy using the LSTM model.

By leveraging LSTM-based predictions, our approach aims to enhance network management by providing insights into future resource requirements, facilitating proactive resource allocation, and improving overall network performance. This research contributes to the broader goal of developing resilient and efficient 5G NTNs capable of meeting the stringent demands of modern communication applications.

The paper is structured as follows: Section II will consider the state-of-the-art background and highlight the novel contribution. Section III will introduce the main concepts of disaggregated RAN networks. Section IV will discuss the proposed network architecture and methodology. Section V will focus on dataset preparation and training. Section VI will show simulation results and related discussion. Finally, Section VII will draw a paper conclusion.

II. Related Work

A. State-of-the-art background

Several studies have explored the use of LSTM NNs for traffic prediction in NTNs, demonstrating its potential for improving resource allocation and network management. [3] provides a comprehensive survey on machine learning techniques for NTNs, emphasizing the application of LSTM models for traffic prediction. This work lays the foundation for employing LSTM in non-terrestrial contexts, showing significant promise in enhancing network efficiency and performance. Similarly, [10] proposes a smoothed LSTM (SLSTM) model for 5G traffic prediction, improving prediction accuracy through adaptive mechanisms and seasonal time difference methods.

An intelligent traffic steering scheme in a disaggregated O-RAN architecture is presented in [5], integrating LSTM-based traffic prediction, flow-split distribution, dynamic user association, and radio resource management. This method enhances resource utilization by predicting dynamic traffic demands, demonstrating the effectiveness of LSTM models in complex RAN architectures. Furthermore, [6] employs LSTM-based traffic prediction within the O-RAN architecture, featuring an LSTM-based prediction rApp at the non-real-time RIC module to enhance decision-making with distributed deep reinforcement learning for network slicing management. This study also illustrates the integration of LSTM and reinforcement learning for adaptive network management, showcasing their combined potential in optimizing network resources and improving performance. [12] explores traffic prediction in 5G using deep learning, underlining the importance of accurate predictions for managing dynamic traffic loads.

The work in [11] proposes an LSTM-autoencoder scheme to predict communication link failures in 5G RAN, accounting for spatial-temporal correlations between radio communication and weather changes. This approach demonstrates LSTM's potential to enhance network reliability by predicting and mitigating failures. Similarly, [13] presents a hybrid model combining LSTM with machine learning to predict and prevent link failures in 5G networks. Additionally, [13] introduces an X-LSTM model for predicting RAN resource usage in 5G, showing the effectiveness of LSTM in resource prediction. Although focused on terrestrial networks, these techniques are also relevant to non-terrestrial networks. Lastly, [14] explores LSTM models for resource prediction in intelligent O-RAN systems, highlighting their potential for optimizing resource allocation.

B. Paper contribution

This work uniquely addresses the prediction of resource consumption for split NGRAN components across 5G NTN networks so that the prediction output will be used by the network management section to decide which network function of the disaggregated NG-RAN component can be considered as a satellite payload based on their previous resource consumption. By considering both F1 and E1 splits and deploying gNB-DU and gNB-CUUP as satellite payloads, our work aims to improve network resilience and readiness for future network management decisions. This approach extends the existing LSTM-based prediction methodologies to NTNs and provides a comprehensive solution for managing the complex resource requirements of disaggregated RAN architectures in 5G NTN environments.

III. Dissagregated NG-RAN Networks

In the evolving landscape of 5G networks, the Next Generation Radio Access Network (NGRAN) disaggregation has emerged as a key architectural innovation to modularize and simplify network complexity and convert the centralized NGRAN functionality into a disaggregated function. This will benefit network service providers and operators by simplifying network monitoring tasks to maintain the required Quality of Service (QoS). The 3GPP TS 38.401 specification [15] discusses how the NGRAN can be functionally split into distinct units, namely the Central Unit (CU) and Distributed Unit (DU), which can be further divided into the CU-Control Plane (CU-CP) and CU-User Plane (CU-UP).

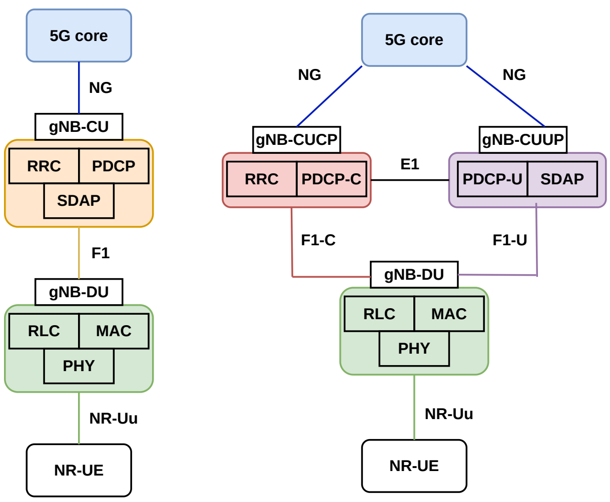

Fig. 1: NGRAN Functional splitting architecture

This disaggregated approach enables operators to deploy network functions in a cloud-native environment, optimizing resource allocation and reducing latency by strategically positioning these components in the network. The disaggregated RAN supports more efficient network traffic management by separating control and user plane functions. It facilitates the integration of new technologies and services, making it a cornerstone of next-generation 5G NTN networks.

As referenced in [15], [16], and [17], the gNB-CU and gNB-DU are connected via the F1 interface. The gNB-CU consists of three layers: the packet data convergence protocol (PDCP), radio resource control (RRC), and service data adaptation protocol (SDAP). The RRC manages connection, mobility, security, and QoS between user equipment (UE) and the network. PDCP handles data compression, security, sequencing, and reliable transfer, while SDAP maps QoS flows to data radio bearers (DRBs) to prioritize traffic based on QoS requirements. The gNB-DU hosts the radio link control (RLC), medium access control (MAC), and physical layers. The RLC manages data segmentation, reassembly, and retransmission; the MAC layer handles scheduling, error correction, and multiplexing; and the physical layer is responsible for transmitting and receiving data over the radio interface.

Figure 1 also shows a further split of the gNB-CU into gNB-CUCP and gNB-CUUP, where the gNB-CUCP handles the RRC and the control plane tasks of PDCP (PDCP-C) and the gNB-CUUP handles the SDAP and PDCP user plane (PDCP-U) tasks [18], [19]. The PDCP-C handles control plane tasks such as managing signaling messages, while the PDCP-U is responsible for user plane data functions like header compression, encryption, and integrity protection. This split allows for flexible network deployment and efficient resource management to maintain QOS and efficient service level agreements (SLAs).

IV. Proposed Network Architecture and Methodology

The payload capability of a typical LEO satellite in terms of available CPU and memory capacity varies significantly based on the specific design and the intended application. This work considers a typical LEO satellite working in a 5G NTN environment. Two network architectures are considered in this work. The first architecture in Figure 2 shows the 5G NTN network with gNB F1 interface over satellite radio interface (F1 over SRI) where the gNB-DU is moved to the satellite payload.

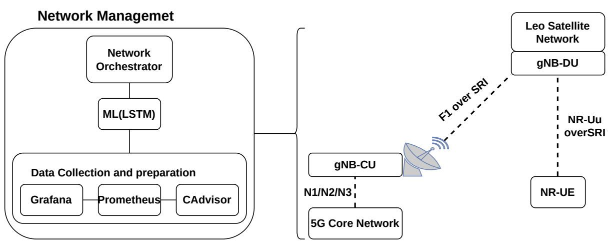

Fig. 2: 5G NTN with F1 split

Figure 3 shows the second 5G NTN network architecture with gNB F1-E1 over SRI (F1-E1 over SRI), where both the F1 and E1 splitting of gNB are deployed across the satellite network. In this scenario, the gNB-CUUP will embark on the payload of the LEO satellite, and the other network components will be placed in the terrestrial network.

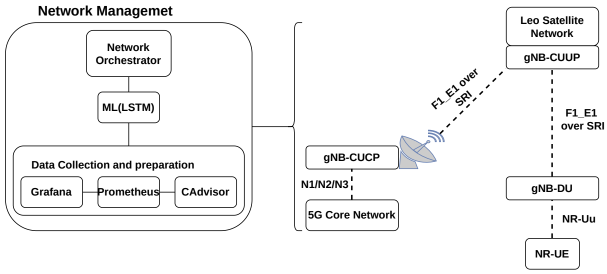

Fig. 3: 5G NTN with F1_E1 split

The network management section in both scenarios above relies on components like data collection and preparation, LSTM-based prediction model, and network orchestrator. The data collection section consists of open-source tools that can be used together to monitor docker container-based networks. These components are used to collect and visualize the resource consumption of the target network function. The first component of this section is CAdvisor [20], which is used to collect, aggregate, and export information about containers running on a host computer. It can collect metrics like CPU usage, memory usage, and bandwidth utilization. The second component is Prometheus [21], which collects and stores metrics as time series data that can be used for visualization. The third component is Grafana [22], which queries and visualizes metrics from Prometheus.

The management section's network orchestrator is assumed to decide on the network functions based on the related historical resource utilization outputs provided by the LSTM prediction model. A real orchestrator has not been implemented in our emulations. We assumed an ideal orchestrator making ideal decisions driven by the prediction outcomes.

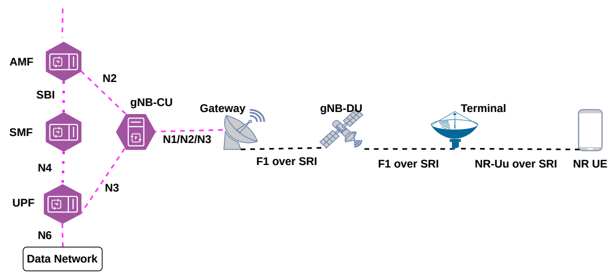

Fig. 4: Emulated 5G NTN with F1 split

The emulated network as shown in Figure 4 and Figure 5 depicts the two architectures with only F1 split considering gNB-DU as the satellite payload and with both F1 and E1 split considering gNB-CUUP as the satellite payload.

Algorithm 1 below describes the resource utilization prediction of the LSTM model used in this work.

V. Dataset Preparation and Training

Data on CPU usage, memory usage, and bandwidth utilization for the gNB-DU and gNB-CUUP components were

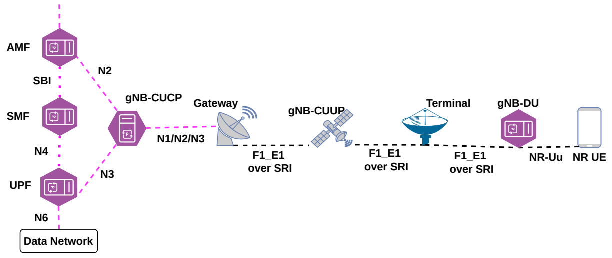

Fig. 5: Emulated 5G NTN with F1_E1 split

Algorithm 1 LSTM for Resource Consumption Prediction

Input: Data (CPU, Mem), seq_length, epochs
for each epoch **do**
  Load data
  data.columns ← strip names
  data(Bandwidth) ← $\frac{data[CPU] \cdot data[Mem]}{10^6}$
  Initialize: MinMaxScaler
  scaled_data ← scaler.fit(data)
  x, y ← create_sequences(seq_length)
  x_train, x_test, y_train, y_test ← split(scaled_data, 0.2)
  Initialize: Sequential(LSTM, Dropout, Dense)
  Initialize Adam
  model.compile(loss='mse', optimizer=Adam)
  early_stopping, reduce_lr ← set callbacks
  model.fit(x_train, y_train, callbacks=[early_stopping, reduce_lr])
  test_loss, test_mae ← model.evaluate(x_test, y_test)
  Predictions ← model.predict(x_test)
  y_test_original ← inverse_transform(y_test)
end for
Output: Trained model with metrics (test_loss, test_mae) and reconstructed labels (y_test_original)

---

collected over 10 hours of simulation from Figures 2 and 3 using Prometheus and Cadvisor. This data was normalized with a MinMaxScaler to scale features between 0 and 1, essential for LSTM model performance. The normalized data was transformed into sequences of 10 consecutive time steps, facilitating the learning of temporal dependencies. The dataset was split 80-20 into training and testing sets, with the training set further divided to include a validation subset to prevent overfitting.

An enhanced LSTM model, incorporating layers such as LSTM, BatchNormalization, Dropout, and Dense, was trained for 300 epochs, employing early stopping and learning rate reduction for optimization. Model performance was assessed on the test set using metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2).

VI. Result and Discussion

This section presents simulation results, focusing on the resource utilization of the LEO satellite payload components gNB-CUUP and gNB-DU within the disaggregated NG-RAN 5G NTN network. The network simulation uses OpenAirInterface for the disaggregated NG-RAN and free5GC as the 5G core, while OpenSAND emulates the satellite network's gateway, satellite, and terminal components. The entire network is emulated in a Docker Compose environment, equipped with data collection and visualization monitoring tools, which are integral to the LSTM model. The experiment is conducted on a Linux OS laptop with an Intel(R) Core(TM) i7-7500U CPU @ 2.70GHz, 4 allocated CPUs, and 16 GB of RAM. The complete code for the experiment is available at https://github.com/HenokBerhanu/disag_vcc.

A. Analysis on the resource consumption of gNB-DU on F1 split

This subsection considers the architecture shown in Figure 4 where the gNB-DU is the payload of the LEO satellite and its resource consumption will be analyzed. Video traffic of 4 Mbits/s is generated across the network using *iperf3* to collect data.

Figure 6 compares the actual and predicted CPU usage of the gNB-DU component over a specific time window. This plot demonstrates how the LSTM model can effectively capture the temporal patterns in CPU usage, with a mean absolute percentage error (MAPE) of 12.24% showing acceptable prediction accuracy. An average predicted value of 0.5% CPU utilization is recorded.

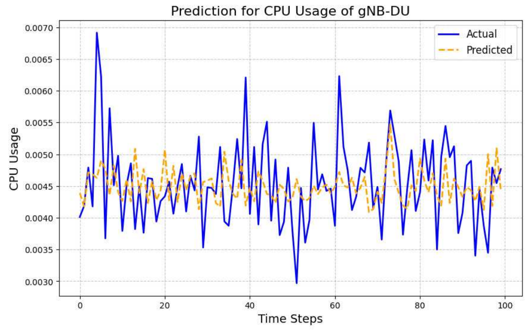

Fig. 6: CPU Usage for gNB-DU F1 split

Figure 7 illustrates the memory usage of the gNB-DU, following the same format as the CPU usage plot. It compares the actual memory usage to the related predicted values, with a MAPE of 0.72% with a highly reliable prediction performance. The average predicted memory usage by the LSTM model is around 20 MiB.

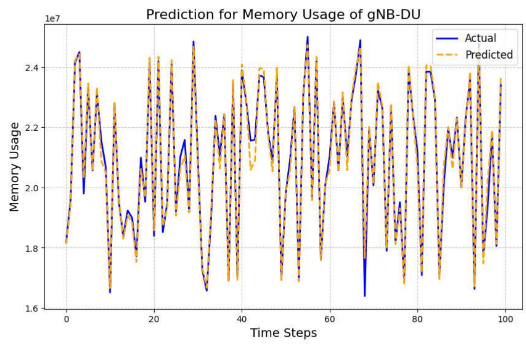

Fig. 7: Memory Usage in byte for gNB-DU F1 split

Figure 8 shows the bandwidth utilization, comparing actual versus predicted values with a MAPE of 11.96%, an MSE of 0.0002, MAE of 0.0111, and an R-squared value of 0.3743 showing acceptable prediction accuracy of satellite gNB-DU.

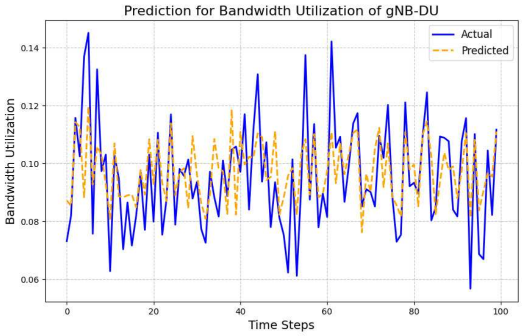

Fig. 8: Bandwidth utilization for gNB-DU F1 split

The training vs. validation Mean Absolute Error (MAE) plot in Figure 9 compares the model performance on the training and validation sets across different epochs. The training MAE indicates how the model error decreases on the training data as it learns over epochs.

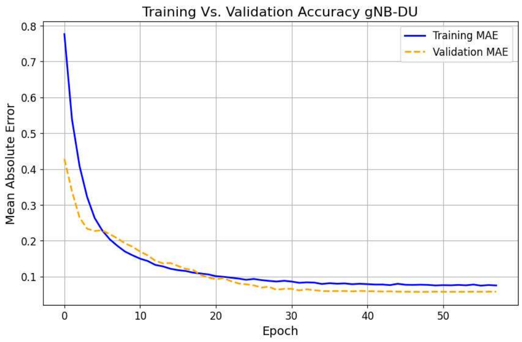

Fig. 9: Training vs Validation Accuracy for gNB-DU F1 split

B. Analysis on the resource consumption of gNB-CUUP on F1-E1 split

Figure 10 compares the actual and predicted CPU usage of the gNB-CUUP component over a specific time window. The model has moderate predictive accuracy with an ideal MSE of 0, MSE of 0.0004, and an R-squared value of 0.0126, indicating that the model has efficiently learned the provided data pattern.

As can be seen from Figure 6 and Figure 10, there is a higher CPU demand for the gNB-DU, as compared to the CPU demand of gNB-CUUP of the F1-E1 split. This indicates that it is technically better to consider the gNB-CUUP of the disaggregated NG-RAN as moved to the LEO satellite payload.

Figure 11 shows the memory usage of the gNB-CUUP of the LEO satellite payload. With a MAPE of 0.9% and an R-squared value of 0.9985, the prediction accuracy of the proposed model exhibits better performance in memory prediction.

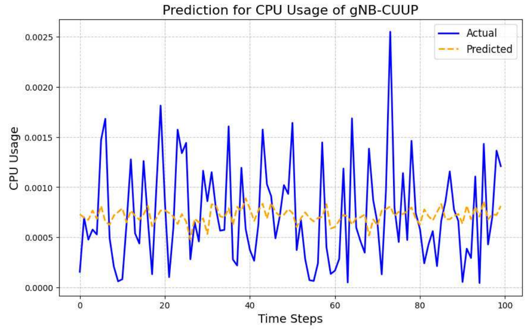

Fig. 10: CPU Usage of gNB-CUUP for F1-E1 split

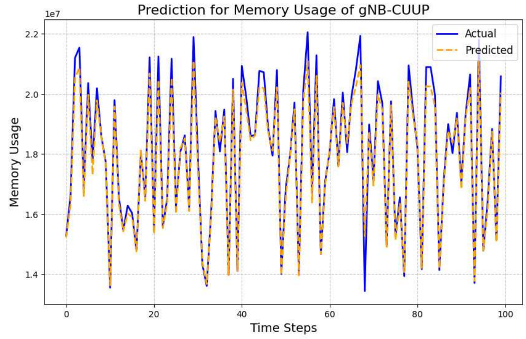

Fig. 11: Memory Usage for gNB-CUUP F1-E1 split

Figure 12 shows bandwidth utilization, comparing actual versus predicted values. The MAPE is 13% which shows acceptable prediction accuracy with an MSE of 0.0007, MAE of 0.0214, and R-squared error of 0.5428.

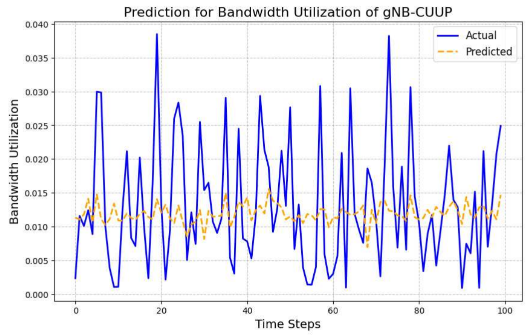

Fig. 12: Bandwidth utilization for gNB-CUUP F1-E1 split

VII. Conclusion and Future Works

This study presents an LSTM-based approach for predicting resource utilization in disaggregated RAN architectures within 5G Non-Terrestrial Networks (NTNs). By comparing two configurations—one with a gNB-CU and gNB-DU split over the F1 interface and another adding F1 and E1 splits with gNB-CUUP as a satellite payload—we demonstrate LSTM's effectiveness in predicting CPU, memory, and bandwidth usage. These predictions enable proactive resource management, ensuring efficient network utilization and high QoS. The combined F1 and E1 split configuration offers greater flexibility and resource efficiency while integrating critical network functions into satellite payloads, which enhances network resilience, particularly in remote areas.

This work underscores LSTM's potential in improving 5G NTN management, especially in handling dynamic traffic patterns. Future work will scale these models for larger networks and incorporate additional ML techniques for fault detection and energy efficiency.

ACKNOWLEDGMENT

The research activities presented in this paper fall within the field of interest of the IEEE AESS technical panel on Glue Technologies for Space Systems and IoT Lab of The University of New Mexico, Department of Electrical and Computer Engineering. This work was also supported by the European Union under the Italian National Recovery and Resilience Plan (NRRP) of NextGenerationEU partnership on "Telecommunications of the Future" (PE00000001 – program "RESTART").

REFERENCES

[1] A. Ghosh, A. Maeder, M. Baker, and D. Chandramouli, "5g evolution: A view on 5g cellular technology beyond 3gpp release 15," *IEEE Access*, vol. 7, pp. 127 639–127 651, 2019.

[2] H. B. Tsegaye and C. Sacchi, "Mec-based experimental framework for service availability in 3d non-terrestrial networks," in *2024 IEEE Aerospace Conference*, 2024, pp. 1–10.

[3] R. Giuliano and E. Innocenti, "Machine learning techniques for non-terrestrial networks," *Electronics*, vol. 12, no. 3, 2023. [Online]. Available: https://www.mdpi.com/2079-9292/12/3/652

[4] H. B. Tsegaye and C. Sacchi, "Graph neural network-based c-ran monitoring for beyond 5g non-terrestrial networks," in *2024 11th International Workshop on Metrology for AeroSpace (MetroAeroSpace)*, 2024, pp. 338–343.

[5] F. Kavehmadavani, V.-D. Nguyen, T. X. Vu, and S. Chatzinotas, "Intelligent traffic steering in beyond 5g open ran based on lstm traffic prediction," *IEEE Transactions on Wireless Communications*, vol. 22, no. 11, pp. 7727–7742, 2023.

[6] F. Lotfi and F. Afghah, "Open ran lstm traffic prediction and slice management using deep reinforcement learning," in *2023 57th Asilomar Conference on Signals, Systems, and Computers*, 2023, pp. 646–650.

[7] S. T. Arzo, P. M. Tshakwanda, Y. M. Worku, H. Kumar, and M. Devetsikiotis, "Intelligent qos agent design for qos monitoring and provisioning in 6g network," in *ICC 2023-IEEE International Conference on Communications*. IEEE, 2023, pp. 2364–2369.

[8] P. M. Tshakwanda, S. T. Arzo, and M. Devetsikiotis, "Advancing 6g network performance: Ai/ml framework for proactive management and dynamic optimal routing," *IEEE Open Journal of the Computer Society*, 2024.

[9] P. M. Tshakwanda, H. Kumar, S. T. Arzo, and M. Devetsikiotis, "Unveiling the future: A comparative analysis of lstm and sp-lstm for network traffic prediction in 6g networks," in *ICC 2024-IEEE International Conference on Communications*. IEEE, 2024, pp. 2919–2924.

[10] Z. Gao, "5g traffic prediction based on deep learning," *Computational Intelligence and Neuroscience*, vol. 2022, no. 1, p. 3174530, 2022. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1155/2022/3174530

[11] M. A. Islam, H. Siddique, W. Zhang, and I. Haque, "A deep neural network-based communication failure prediction scheme in 5g ran," *IEEE Transactions on Network and Service Management*, vol. 20, no. 2, pp. 1140–1152, 2023.

[12] J. Su, H. Cai, Z. Sheng, A. Liu, and A. Baz, "Traffic prediction for 5g: A deep learning approach based on lightweight hybrid attention networks," *Digital Signal Processing*, vol. 146, p. 104359, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1051200423004542

[13] C. Gutterman, E. Grinshpun, S. Sharma, and G. Zussman, "Ran resource usage prediction for a 5g slice broker," in *Proceedings of the Twentieth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, ser. Mobihoc '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 231–240. [Online]. Available: https://doi.org/10.1145/3323679.3326521

[14] S. Niknam, A. Roy, H. S. Dhillon, S. Singh, R. Banerji, J. H. Reed, N. Saxena, and S. Yoon, "Intelligent o-ran for beyond 5g and 6g wireless networks," 2020. [Online]. Available: https://arxiv.org/abs/2005.08374

[15] 3GPP, "NG-RAN:Architecture description," 3rd Generation Partnership Project (3GPP), Technical Specification Group Radio Access Network 38.401, 06 2024, version 18.2.0. [Online]. Available: https://portal.3gpp.org/ChangeRequests.aspx?q=1&versionId=88269&release=193

[16] N. Makris, C. Zarafetas, P. Basaras, T. Korakis, N. Nikaein, and L. Tassiulas, "Cloud-based convergence of heterogeneous rans in 5g disaggregated architectures," in *2018 IEEE International Conference on Communications (ICC)*, 2018, pp. 1–6.

[17] F. Arnhold, S. S. Anbazhagan, L. R. Prade, J. M. Nogueira, A. Klautau, and C. B. Both, "Network slicing support by fronthaul interface in disaggregated radio access networks: A survey," *IEEE Transactions on Network and Service Management*, pp. 1–1, 2024.

[18] R. Khouli, L. Frank, and A. Hofmann, "Functional split evaluation in ntn for leo satellites," in *40th International Communications Satellite Systems Conference (ICSSC 2023)*, vol. 2023, 2023, pp. 1–9.

[19] M. Rihan, T. Düe, M. Vakilifard, D. Wübben, and A. Dekorsy, "Ran functional split options for integrated terrestrial and non-terrestrial 6g networks," in *2023 11th International Japan-Africa Conference on Electronics, Communications, and Computations (JAC-ECC)*, 2023, pp. 152–158.

[20] Github, "Analyzes resource usage and performance characteristics of running containers (cadvisor)," https://github.com/google/cadvisor, 03 2024, last version 0.49.1.

[21] dockerhub, "Prometheus," https://hub.docker.com/r/prom/prometheus, 2024, docker image and installation guidelines.

[22] ——, "Grafana," https://hub.docker.com/r/grafana/grafana, 2024, docker image and installation guidelines.