Article

Transformer-Based Beam Prediction for THz Drone Communication with RIS Integration

Md Nurul Absar Siddiky [*] and Muhammad Enayetur Rahman

Article

Transformer-Based Beam Prediction for THz Drone Communication with RIS Integration

Md Nurul Absar Siddiky [1],* and Muhammad Enayetur Rahman [2]

[1]   Electrical and Computer Engineering, University of North Carolina at Charlotte, NC, 28223, USA
[2]   Electrical and Computer Engineering, Old Dominion University, Norfolk, VA, 23509, USA
*   Correspondence: msiddiky@uncc.edu

Abstract

This study addresses beam prediction in Terahertz (THz) Unmanned Aerial Vehicle (UAV) communication, a critical component of sixth-generation (6G) wireless networks. While the THz spectrum enables high-speed data transmission, UAV mobility introduces challenges like signal impairments and blockages. To overcome this, a deep learning model using self-attention networks is proposed, integrating Reconfigurable Intelligent Surfaces (RIS) to enhance coverage and adapt to channel dynamics. The approach predicts optimal beams for UAVs based on sequential data, including beam trajectories, position, and Line of Sight (LoS) information. Compared to Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRUs), the model demonstrates superior performance in prediction accuracy, as validated by metrics like Root-Mean-Square Error (RMSE) and Mean Absolute Error (MAE).

Keywords: terahertz; 6G; beam prediction; RIS; drone; unmanned aerial vehicle communication

1. Introduction

1.1. Motivation and Challenges

Terahertz (THz) communication, operating in the 0.1–10 THz range, is a promising technology for 6G networks due to its high-speed, high-capacity capabilities [1,2]. THz communication supports applications like immersive Virtual Simulation (VS), smart cities, and autonomous vehicles by offering ultra-high data rates and precise localization [3,4]. For example, experiments have achieved data rates of 100 Gb/s using THz wireless links [5].

However, THz communication faces significant challenges, including signal attenuation from molecular absorption and sensitivity to physical barriers, resulting in limited range and non-line-of-sight (NLoS) issues [6]. Addressing these limitations requires advanced techniques like Reconfigurable Intelligent Surfaces (RIS), which enhance coverage and mitigate attenuation by reflecting and modulating electromagnetic waves [7].

Incorporating drones and RIS in THz systems adds complexity, requiring sophisticated beam-forming and channel estimation algorithms to handle dynamic environments. The mobility of drones further complicates accurate tracking and beam steering. This research proposes a time-series-based transformer Deep Learning model for beam prediction in THz UAV systems, leveraging RIS to address these challenges. By combining RIS with Deep Learning, the aim was to improve coverage and beam prediction accuracy in dynamic environments, offering a practical solution for enhancing THz communication.

1.2. Contribution

- Proposed an attention-based Deep Learning model for predicting 3D RIS beams and user locations.
- Compared transformer model with LSTM and GRU models, evaluating performance using RMSE and MAE. The proposed model achieves competitive results.
- Demonstrated that transformer model requires fewer parameters and learns faster than LSTM and GRU, ensuring greater efficiency.

2. Related Work

In 6G THz communication, predicting beam directions early is essential for reducing the time and resources required for beam training, enabling faster communication. The authors in [8] proposed a deep learning approach using gated recurrent units (GRU) to predict which base station or RIS a drone should connect to and which data beam to use.

To address challenges in THz hybrid beamforming systems, such as precise 3D angle estimation and reducing beam tracking overhead, the authors in [9] developed an off-grid ultra-high-resolution direction-of-arrival (DoA) estimation algorithm. This method works with the array-of-subarrays (AoSA) architecture and achieves millidegree precision.

Beam training often becomes expensive with large antenna arrays due to increased overhead. To tackle this, the authors in [10] proposed a multi-modal machine learning approach that combines location and visual (camera) data from the communication environment to predict beams quickly. They validated their approach using a real-world vehicular dataset containing position, camera vision, and mmWave beam data.

Transformer-based architectures [11] have recently gained attention in solving wireless communication problems. For instance, to address antenna spacing limitations in massive MIMO systems, the authors in [12] developed a transformer-based framework for holographic MIMO. Their method optimizes channel capacity by allocating power effectively for beamforming.

In [13], the authors introduced a masked autoencoder and deep reinforcement learning-based RIS-enabled system to improve pilot signal allocation policies. Their method reduces the required pilot signals and enhances channel estimation performance.

3. System Model and Problem Formulation

A scenario with multiple drones (users) communicating through a Base Station (BS) and a Reconfigurable Intelligent Surface (RIS) was considered. The BS acts as the central hub, managing signals, control information, bandwidth allocation, and processing AI-assisted algorithms. The RIS is stationary, placed above the ground with a clear Line of Sight (LoS) to the BS. The RIS acts as an Aerial Base Station (AEB), linking the backhaul (connected to the BS) and the access network (connected to drones).

The RIS ensures reliable communication for drones, even when there is no direct LoS between the drones and the BS, by reflecting and adapting signals to maintain connectivity. In this study, drones or flying objects are referred to as users. These users connect to the BS directly via LoS or indirectly via the RIS in non-LoS conditions.

3.1. System and Channel Model

The system model is shown in Figure 1. The Base Station (BS) is equipped with N_B antenna array elements, represented by the beamforming vector $\mathbf{b} \in \mathbb{C}^{N_B \times 1}$. A predefined beam codebook $\mathcal{P}$ of size N_{CB} is used to select beamforming vectors. If the direct LoS connection between the BS and the drone is disrupted, the RIS with N_{RIS} antennas assists the drone to ensure continued communication.
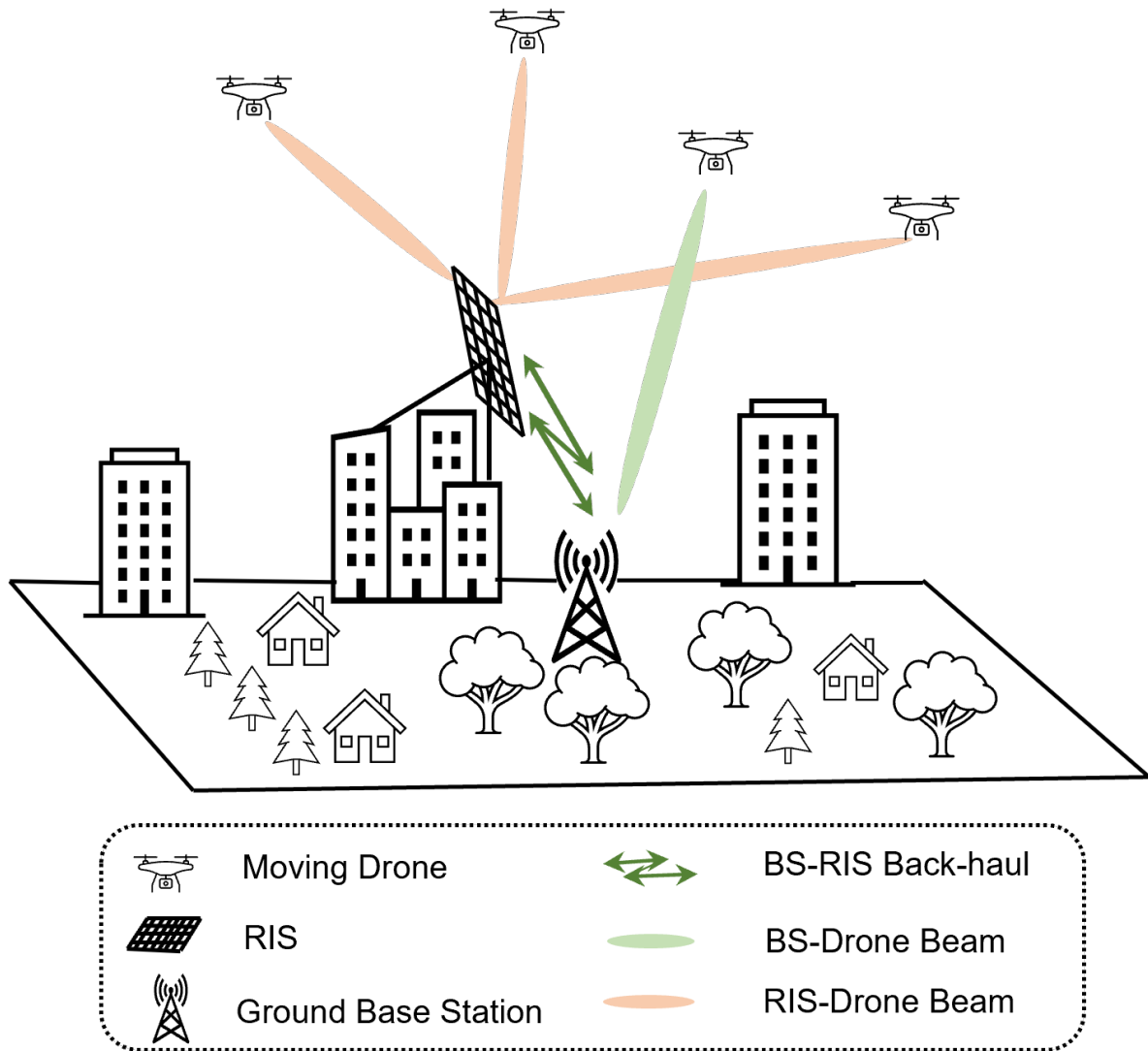
Figure 1. System Model for Outside THz Drones with RIS.

Let $\mathbf{ff}_{T,k}$ and $\mathbf{ff}_{T,k}^T$ represent the uplink and downlink transmitter channel coefficients to the RIS. Similarly, $\mathbf{ff}_{R,k}$ and $\mathbf{ff}_{R,k}^T$ represent the uplink and downlink receiver channel coefficients of the RIS. The received signal strength is given by:

$$r_k = \mathbf{ff}_{R,k}^T \mathbf{\Phi} \mathbf{ff}_{T,k} d + u,$$

where $d \in \mathbb{C}$ is the data symbol ($\mathbb{E}\big[|d|^2\big] = P_T$), P_T is the transmit power, and $u \sim \mathcal{M}_{\mathbb{C}}\big(0, \omega^2\big)$ is noise. The interaction matrix $\mathbf{\Phi} = \mathrm{diag}(\phi)$ represents the interaction of the incident signal with the RIS. Each RIS element modifies the signal phase by an interaction factor $[\phi]_e = e^{j\psi_e}$, where $e = 1, 2, \dots, E$. The beamforming vector ϕ is selected from a predefined reflection codebook $\mathcal{B}$.

A clustered wideband geometric THz channel model with W_c clusters was used, where each cluster contributes a single ray. Each ray has a time delay τ_w, azimuth-elevation angles of arrival (AoA) (θ_w, ϕ_w), and a complex path gain α_w, which includes path loss. The channel delay δ is modeled as:

$$\mathbf{h}_\delta = \sum_{w=1}^{W_c} \alpha_w p_{\mathrm{rc}}(\delta T_s - \tau_w) A_{\mathrm{rv}}(\theta_w, \phi_w),$$

where $p_{\mathrm{rc}}(\tau)$ is a pulse-shaping function, and $A_{\mathrm{rv}}(\theta_w, \phi_w)$ is the BS array response vector at the AoA.

In the frequency domain, the channel coefficient vector at subcarrier k is:

$$\mathbf{h}_k = \sum_{t_p=0}^{T_p-1} \mathbf{h}_\delta e^{-j\frac{2\pi k}{K}d}.$$

Here, T_p is the total number of taps, K is the total number of subcarriers, and the channel coefficients $\{\mathbf{h}_{n,k}\}_{k=1}^{K}$ remain constant during the coherence period T_C [14].

3.2. Problem Formulation

The main focus of the study is to predict RIS beamforming trajectories. To achieve this, an attention based DL model was created to effectively direct the future RIS Beam. The base station adjusts its beam $\mathbf{b}$ to match the user's movement and ensure a stable connection. The duration of beam coherence can change based on variables such as the user's velocity and the base station antenna number. The beam used to communicate with the drone at a specific moment is denoted as $\mathbf{bs}^{(t)}$, where t represents the beam coherence time. A sequence of beams is established based on this t-step.

$$\mathcal{B}_t = \left\{ \mathbf{bs}^{(1)}, \mathbf{bs}^{(2)}, \dots, \mathbf{bs}^{(t)} \right\}. \tag{1}$$

In a similar vein, when considering the RIS, the beamforming vector at time t as $\phi^{(t)}$, the sequence of steps can be expressed as:

$$\mathcal{F}_t = \left\{ \Phi^{(1)}, \Phi^{(2)}, \dots, \Phi^{(t)} \right\}. \tag{2}$$

When beam $\mathbf{b}^{(t)}$ was selected, let $\mathbf{l}^{(t)}$ denote the drone location at time t (i.e.,). Then, the location sequence of t time was defined as:

$$\mathcal{L}_t = \left\{ \mathbf{l}^{(1)}, \mathbf{l}^{(2)}, \dots, \mathbf{l}^{(t)} \right\}. \tag{3}$$

$d^{(t)}$ is used to indicate whether there is a Line of Sight (LoS) connection between the base station and the drone user at time step t:

$$\mathcal{D}_t = \left\{ d^{(1)}, d^{(2)}, \dots, d^{(t)} \right\}. \tag{4}$$

The problem of directing the future beam at time $t+1$ based on the given beam sequence B_t is addressed. A multivariate attention-based deep learning system is utilized to map the beam sequence from B_t to B_{t+1}. The system accurately predicts the appropriate communication channel and beam to use, and the process of training a deep neural network to achieve this high level of accuracy is explained.

4. Solution Approach

In this study, a multivariate time-series prediction model was developed using a transformer network with attention mechanisms to predict RIS beam trajectories. The Transformer [11] uses attention to extract features. The architecture of the attention-based (transformer) beam prediction model is shown in Figure 2. Specifically, the model computes correlations among input data in the attention layer (Figure 2 right) and applies multiplication to create a weighted input matrix. The attention layer uses three matrices: Query $\mathbf{Q}$, Key $\mathbf{K}$, and Value $\mathbf{V}$. The Query $\mathbf{Q}$ and Key $\mathbf{K}$ matrices hold the input sequence features. The self-attention score is calculated in the final step of the attention layer. Multi-head self-attention improves the model by processing multiple input elements at once. This helps the Transformer capture both local and global features by looking at correlations across the entire input sequence. The Transformer is highly effective at analyzing sequential and time-series data due to its ability to detect long-range correlations, process data in parallel, and retain positional information without needing to analyze order explicitly.
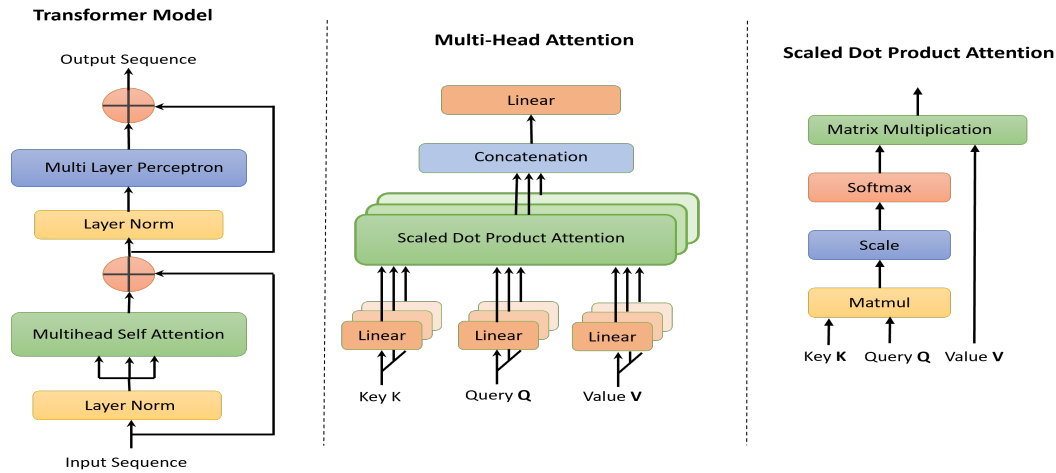
Figure 2. Attention Based Beam Prediction.

A new deep learning model based on the transformer's self-attention architecture was proposed to predict regression time-series data. The model is designed to process sequences using self-attention and convolutional layers to capture temporal patterns. The model architecture includes encoder blocks with normalization for stable learning, multi-head self-attention, and dropout for regularization. A feed-forward network with convolutional layers extracts local patterns. Global average pooling was used to extract important features for analysis. Then, a multi-layer perceptron (MLP) with ReLU activation is used to capture high-order relationships in the sequence. Finally, a dense output layer provides the model's prediction for continuous output values.

5. Experimental Analysis and Discussion

In this section, the proposed Deep Learning model for 6G UAV is evaluated and compared with traditional LSTM and GRU models.

5.1. Experimental Settings

To evaluate the DL-Attention-based model, DeepMIMO framework [15] was used. A drone scenario was set up in a downtown area, with a ground base station placed 6 meters above the ground and a hovering RIS 80 meters above. The system includes a 3D drone grid consisting of four parallel 2D grids at different heights, with the base positioned 40 meters above the ground. The grid points are spaced 0.81 meters apart horizontally and 0.8 meters vertically. The intersection of these axes is surrounded by buildings with varying heights but uniform base dimensions. The drones follow random waypoint trajectories as defined in the random waypoint model [15].

The DeepMIMO script generates a dataset for the 6G drone scenario, including channel and beam information between the base station, RIS, and drones. An experimental dataset was created by generating drone trajectories based on the parameters in Table 1.

Table 1. Dataset Generation Input Parameters.

Parameter	Value
First, Last User Row	1, 496 [8]
BS Antenna Quantity	64, 1, 1 [8]
RIS Antenna Quantity	256, 1, 1 [8]
Space Between Antennas	0.5 [8]
Center Frequency [GHz]	200 [8]
Bandwidth [GHz]	1 [8]
OFDM Subcarriers	512 [8]
Sampling Factor (OFDM)	1 [8]
OFDM Range	1 [8]
Path Count	1 [8]

The dataset contains beam vector information such as arrival angles, phase, time of arrival, and power. Beam data from the RIS, base station, and drone sequence numbers are used. If there was no communication path or blockage with the base station, the beam data for that drone position was set to zero. A 10-step time sequence dataset for drone trajectories was created, including base station and RIS beamforming vectors, sequence number, number of paths, 3D drone positions, and LoS from the BS and RIS. The dataset contains 10,000 rows, with 21 feature vectors as input and one predicted row per 10-step sequence. The dataset was split into training, validation, and testing subsets, with 20% for testing and 20% for validation.

The features in the dataset have different scales, so scaled them to a range of 0 to 1 using the formula:

$$XScaled = \frac{X - XMin}{XMax - XMin}$$

Feature scaling was applied to improve the performance of the machine learning model. The scaling parameters were learned from the training data and then applied to the training, validation, and test datasets to ensure consistency.

The proposed model's parameters are listed in Table 2.

Table 2. Proposed Model's Parameters.

Parameter	Value
Head Size	4
Number of Heads	4
Feed Forward Dimension	16
Number of Transformer Blocks	2
MLP Units	32
MLP Dropout	0.3
Dropout	0.3
Batch Size	32
Optimizer	Adam (LR=0.001)
Loss Function	Mean Squared Error
Evaluation Metric	Root Mean Squared Error
Epochs	200

The proposed model was compared with LSTM and GRU models. Both models are designed for sequential tasks, using 10-length input sequences with 21 features per time step. The LSTM/GRU models consist of 32 units, with dropout layers (rate 0.2) to reduce overfitting, followed by a fully

connected dense layer with 16 neurons and a *ReLU* activation function. A second dropout layer was added for regularization.

The final output layer is a dense layer with 21 neurons and a linear activation function for regression tasks. For a fair comparison, the same number of parameters, optimizer, learning rate, and training epochs (200) are used for the LSTM/GRU models.

5.2. Result Analysis

The LSTM model has 7,797 parameters, GRU has 6,165, and the proposed model has 5,457. As shown in Figure 3, despite having fewer parameters, the model learns faster, as evidenced by the Root Mean Squared Error (RMSE) vs. Epoch plot.
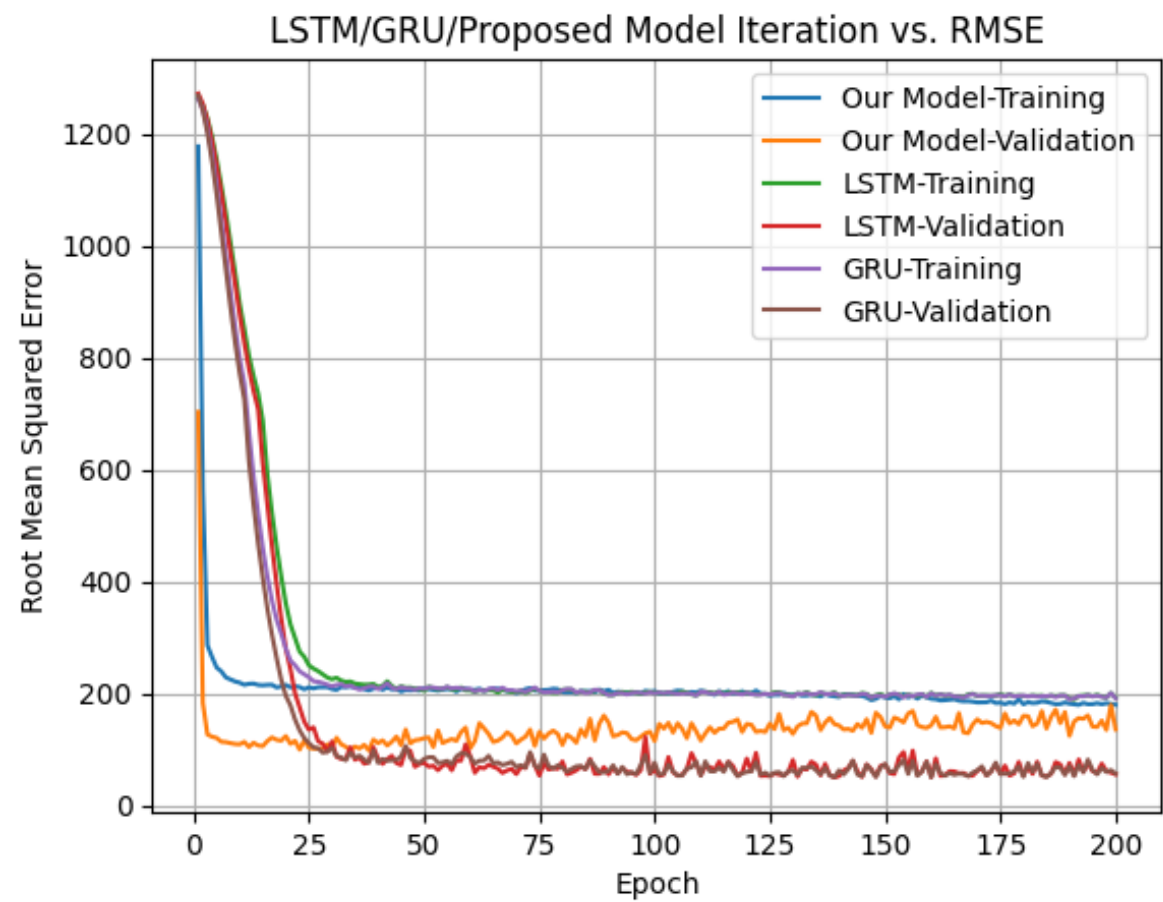

Figure 3. Iteration vs RMSE plot for RIS Beam.

By the 10th epoch, the training RMSE stabilizes, while LSTM/GRU models stabilize after the 27th epoch. LSTM/GRU models perform similarly, but the validation curve of the proposed model at epoch 200 merges with the training curve, indicating better performance.

The model's performance is evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) values, as shown in Table 3. Despite the higher MAE and RMSE values, the proposed model performs well compared to LSTM and GRU.

Table 3. Model Performance on RIS's Beam.

Model	MAE	MSE	RMSE
LSTM	22.94199	1887.49869	43.44535
GRU	22.95527	1887.95526	43.45061
Proposed Model	32.21674	2609.64263	51.08466

Possible reasons for the Transformer model's higher MAE and RMSE include:

- The Transformer might not be the best fit for data with strong temporal patterns, where LSTM/GRU models excel.
- For data with short-term patterns, LSTM/GRU models perform better, while the Transformer is more suited for long-term dependencies.
- Transformer models are larger and more complex, requiring more data to perform effectively. Only 10,000 data points were used in this study.

6. Conclusion

In this study, the future beam of the RIS is predicted in a 6G drone communication scenario for different drone users. The goal is to improve the reliability of communication between the drone user and the base station by predicting the future beam. The proposed model is based on the transformer architecture and is compared with traditional LSTM and GRU models.

The dataset used in this study includes drone trajectories, beamforming vectors from both the base station and RIS, 3D positions of the drones, Line-of-Sight (LoS) information from both the base station and RIS, along with sequence numbers and the number of paths. This large and comprehensive dataset represents the spatial movement of drones and communication data, which is crucial for the 6G drone scenario.

The transformer-based model can capture complex temporal dependencies in sequential data. The model's performance was evaluated using Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) metrics, and it consistently outperformed the LSTM and GRU models.

References

1. Chen, Z.; Han, C.; Wu, Y.; Li, L.; Huang, C.; Zhang, Z.; Wang, G.; Tong, W. Terahertz wireless communications for 2030 and beyond: A cutting-edge frontier. *IEEE Communications Magazine* **2021**, *59*, 66–72.
2. Sarieddeen, H.; Saeed, N.; Al-Naffouri, T.Y.; Alouini, M.S. Next generation terahertz communications: A rendezvous of sensing, imaging, and localization. *IEEE Communications Magazine* **2020**, *58*, 69–75.
3. Akyildiz, I.F.; Jornet, J.M.; Han, C. Terahertz band: Next frontier for wireless communications. *Physical communication* **2014**, *12*, 16–32.
4. Huq, K.M.S.; Busari, S.A.; Rodriguez, J.; Frascolla, V.; Bazzi, W.; Sicker, D.C. Terahertz-enabled wireless system for beyond-5G ultra-fast networks: A brief survey. *IEEE network* **2019**, *33*, 89–95.
5. Castro, C.; Elschner, R.; Merkle, T.; Schubert, C.; Freund, R. Experimental demonstrations of high-capacity THz-wireless transmission systems for beyond 5G. *IEEE Communications Magazine* **2020**, *58*, 41–47.
6. Kokkoniemi, J.; Lehtomäki, J.; Juntti, M. A discussion on molecular absorption noise in the terahertz band. *Nano communication networks* **2016**, *8*, 35–45.
7. Basar, E.; Di Renzo, M.; De Rosny, J.; Debbah, M.; Alouini, M.S.; Zhang, R. Wireless communications through reconfigurable intelligent surfaces. *IEEE access* **2019**, *7*, 116753–116773.
8. Abuzainab, N.; Alrabeiah, M.; Alkhateeb, A.; Sagduyu, Y.E. Deep learning for THz drones with flying intelligent surfaces: Beam and handoff prediction. In Proceedings of the 2021 IEEE International Conference on Communications Workshops (ICC Workshops). IEEE, 2021, pp. 1–6.
9. Chen, Y.; Yan, L.; Han, C. Millidegree-level direction-of-arrival (DoA) estimation and tracking for terahertz wireless communications. In Proceedings of the 2020 17th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON). IEEE, 2020, pp. 1–9.
10. Charan, G.; Hredzak, A.; Stoddard, C.; Berrey, B.; Seth, M.; Nunez, H.; Alkhateeb, A. Towards Real-World 6G Drone Communication: Position and Camera Aided Beam Prediction. In Proceedings of the GLOBECOM 2022 - 2022 IEEE Global Communications Conference, 2022, pp. 2951–2956. https://doi.org/10.1109/GLOBECOM48099.2022.10000718.
11. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.u.; Polosukhin, I. Attention is all you need. *Advances in Neural Information Processing Systems* **2017**, *30*.
12. Adhikary, A.; Raha, A.D.; Qiao, Y.; Munir, M.S.; Kim, K.T.; Hong, C.S. Transformer-based Communication Resource Allocation for Holographic Beamforming: A Distributed Artificial Intelligence Framework. In Proceedings of the 2023 24st Asia-Pacific Network Operations and Management Symposium (APNOMS), 2023, pp. 243–246.

13. Kim, K.; Tun, Y.K.; Munir, M.S.; Saad, W.; Hong, C.S. Deep Reinforcement Learning for Channel Estimation in RIS-Aided Wireless Networks. *IEEE Communications Letters* **2023**, *27*, 2053–2057. https://doi.org/10.1109/LCOMM.2023.3280821.

14. Va, V.; Choi, J.; Shimizu, T.; Bansal, G.; Heath, R.W. Inverse multipath fingerprinting for millimeter wave V2I beam alignment. *IEEE Transactions on Vehicular Technology* **2017**, *67*, 4042–4058.

15. Alkhateeb, A. DeepMIMO: A Generic Deep Learning Dataset for Millimeter Wave and Massive MIMO Applications. In Proceedings of the Proc. of Information Theory and Applications Workshop (ITA), San Diego, CA, Feb 2019; pp. 1–8.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.