INTERNATIONAL JOURNAL OF RESEARCH AND INNOVATION IN APPLIED SCIENCE (IJRIAS)
ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue III March 2026

Machine Learning for Adaptive Beamforming in Mimo Systems for Improved Network Throughput and Energy Efficiency

Ogili Solomon Nnaedozie

Department of Electrical Electronics, Engineering, Enugu State University of Science and Technology, Agbani Enugu State, Nigeria

DOI: https://doi.org/10.51584/IJRIAS.2026.11030091

Received: 20 March 2026; Accepted: 26 March 2026; Published: 15 April 2026

ABSTRACT

Adaptive beamforming in Multiple-Input Multiple-Output (MIMO) systems is now an important technology in 5G and 6G networks because of the growing need of high-speed, steady, and energy-saving wireless communications. In this paper, an adaptive beamforming framework through Reinforcement Learning (RL) is proposed which uses a Deep Q-Network (DQN) to train the best beam selection policies with respect to channel state variables, such as Signal-to-Noise Ratio (SNR) and beam index. The model is trained on the 5G Adaptive Beamforming with SNR dataset and an ε-greedy exploration-exploitation strategy and experience replay are used to guarantee a steady convergence. The results of simulation and tests prove that the RL-based model has the significant benefit compared to the traditional approaches. Through the proposed approach, a mean throughput was 6.41bps/Hz, energy efficiency 3.54bits/Joule, and SINR 22.8dB, though both the MRT and ZF performed poorly. The model also demonstrated to be reliable in a range of SNR and remained the same in terms of learning behaviour as cumulative rewards were stabilised at the 250 training episodes. These results show that convergence of reinforcement learning and adaptive beamforming can be effectively used in dynamic MIMO to enhance spectral and energy efficiency. The proposed design offers a scalable and data-driven technology to the next-generation wireless networks and can be implemented to select intelligent beams that can modify to the varying channel conditions and maximises the overall network throughput and power usage.

Keywords: Adaptive Beamforming; MIMO; Reinforcement Learning; Deep Q-Network; 5G Wireless Networks

INTRODUCTION

The high increase in mobile data communication due to applications that consume a lot of bandwidth and the use of smart devices has put a lot of pressure on the current wireless communication systems (Tommy et al., 2025). In order to address these needs, Multiple-Input Multiple-Output (MIMO) technology has been popularised because it can be improved to boost spectral efficiency, link reliability, and the overall network capacity (Sharma, 2024). One of the methods that make such gains possible is beamforming, which focuses transmitted signals to target users and eliminates interference (Gupta, 2024). Nevertheless, with the increasing density and dynamism of wireless environments that come with 5G networks and even more advanced 6G broadcasting, there is an increasing performance and efficiency constraint in traditional beamforming methods (Huo et al., 2023).

The conventional adaptive beamforming techniques including: Maximum Ratio Transmission, Zero-Forcing, and Minimum Mean Square Error are highly dependent on precise and timely Channel State Information (CSI) (Sayed, 2023). Practically, it is not easy to achieve perfect CSI because of the mobility of users, delays in feedback, estimation of the channel, and hardware (Kaur et al., 2024). The issues cause higher computational complexity, non-optimal beam selection, and high energy usage especially in mega-MIMO and millimetre-wave systems (Lavdas et al., 2025). Consequently, more intelligent and flexible beamforming methods are urgently required to work well under imperfect and significantly changing network conditions (Al-Abbasi, 2025).

Machine Learning (ML) has provided an exciting solution to overcome these issues because it allows making data-driven and adaptive decisions in wireless communication (Ashok et al., 2025). ML algorithms can learn complex and nonlinear relationships between the channel conditions and the optimum beamforming parameters and therefore dynamically tune the beam directions and power allocation without necessarily using explicit channel models (Chintha and Singh, 2025). Deep learning and reinforcement learning techniques have shown a solid potential in terms of accuracy in beam selection, signalling overhead reduction, and robustness in dynamically varying environments (Lavdas et al., 2025). The capabilities render ML to be especially suitable in adaptive beamforming in real-time in large-scale MIMO systems (Tommy et al., 2025).

The recent works emphasise that beamforming based on the reinforcement learning significantly enhances throughput and SINR over conventional approaches (Spyros et al., 2025). Furthermore, the ML-based methods have been demonstrated to be less signalling overhead and computational overhead, and therefore more energy-efficient (Bhatia Khan et al., 2024). This is in line with other studies on the use of AI to optimise energy of future-generation networks, where the focus lies on the sustainability and the performance (Mumtaz et al., 2023). Using smart learning-based models, the given approach aims at maximising beamforming choices and reducing power usage and computation time (Imran et al., 2024). The results of the study are predicted to be beneficial to the creation of effective, scalable, and energy-aware wireless communication networks capable of meeting the performance needs of the next-generation networks (Rahmani et al., 2024).

METHODOLOGY

In this work, the simulation-based approach is used to estimate the performance of machine learning-based adaptive beamforming in MIMO systems. Multi-user MIMO wireless communication model is created, where channel conditions are created with the realistic fading and mobility scenarios. As inputs to a machine learning model, relevant network parameters (channel state information, signal-to-interference-plus-noise ratio, user locations, and transmit power levels) are used, and optimal beamforming vectors and power allocation choices are the outputs. This is achieved with the help of a reinforcement learning-based algorithm that allows the system to learn adaptive beamforming policies as it interacts with the environment, with the help of a network throughput and energy efficiency reward function. The effectiveness of the proposed ML-based beamforming algorithm is assessed by simulation and compared with the traditional beamforming algorithms in terms of the achievable throughput, the power consumption, the SINR, and convergence characteristics.

Data Collection and Preprocessing

In this paper, the main source of data is the 5G Adaptive Beamforming with SNR Dataset that was taken on Kaggle (https://www.kaggle.com/datasets/zoya77/5g-adaptive-beamforming-with-snr-dataset), where the real-world and/or synthetic measurements applicable to the adaptive beamforming behaviour in 5G MIMO systems are provided. The single CSV file (around 216 kB) in the dataset holds numerous observations (rows) and signal measurements (columns) (Signal-to-Noise Ratio, SNR values among other parameters related to beamforming measured under different network and channel environments). These attributes including SNR, beam index, channel state indicators and may be other input measurements will be utilised to train and validate the machine learning model on adaptive beamforming. The structured measurements of SNR of the dataset in a wide range of situations make it appropriate to learn the relationship between channel conditions and optimal beamforming choices to allow the mechanism of reinforcement learning to provide rewards in terms of throughput and energy efficiency outcomes. The preprocessing of the data will involve the cleaning of the gaps and the normalisation of the numeric ones and the division of the data into training, validation and testing in order to facilitate sound model testing during training (Aribisala, 2025; Prasad. 2024).

Data Description

The 5G Adaptive Beamforming with SNR dataset applied in this study is organised numerical data of wireless communication under different channel and signal conditions. Each data instance corresponds to a beamforming observation in a 5G MIMO environment and includes key parameters such as SNR values, beam or antenna configuration indicators, and other channel-related features that influence beam selection performance. The SNR values reflect different propagation conditions, interference levels, and user positions, making the dataset suitable for modelling realistic and dynamic wireless environments. These features serve as input variables for the machine learning model, while beamforming decisions or performance-related indicators are used as learning targets or reward references.

Table 1: Data Features and Description

Feature Name	Type	Description
SNR (Signal-to-Noise Ratio)	Numeric	Signal quality measure in decibels representing the ratio of signal power to noise power
Beam Index	Categorical/Numeric	Identifier for the particular beam or antenna direction used in the observation (e.g., which beam pattern was active).
Channel State Indicators	Numeric	Variables describing channel conditions (e.g., channel quality measures, fading characteristics).
Transmit Power	Numeric	Power level at the transmitter for each observation (may influence SNR and throughput).
User Position/Angle	Numeric	Spatial position or angle of the receiver relative to the transmitter, useful for beam direction optimization.
Throughput Reference	Numeric	Measured or estimated throughput for that scenario which is helpful for evaluating performance outcomes.

The Machine Learning Model

In this paper, a machine learning model, which is based on RL, is used to attain adaptive beamforming in a 5G MIMO system. This model will learn an optimal policy of beam selection and power control by interplay with the wireless environment and feedback in the form of signal quality and performance measures. The SNR and other channel-related parameters derived on the dataset are key input variables to the model and which all constitute the system state.

The Proposed Reinforcement Learning Model

The proposed RL model is a type of intelligent decision-making agent deployed in the base station of a 5G MIMO system, the main task of which is to choose the most optimal beamforming arrangements to maximize the network throughput and reduce the amount of energy used. The model works in a closed-loop interaction between the agent, the wireless environment and a learning mechanism and is established as a MDP. The environment presents the current state to the agent at every decision time step, which is built up of the measurable network parameters obtained in the dataset, such as SNR, channel quality indicators, and beam related features. Such state is the dynamic radio conditions that a user or group of users are in. According to this state, the RL agent makes a choice of an action, which is a selection between a particular beam index (and, in some cases, a target transmit power level) in a predetermined beam codebook. This step dictates the direction the base station would direct its transmission to the user.

After an action has been performed, the environment reacts and shifts to a new state and sends a reward signal back. The reward is also properly configured to capture the system performance and is calculated using weighted operation of the possible throughput and energy consumption, so that a combination of high data rates and reduced power consumption are both promoted. When a chosen beam enhances the quality of signals and provides a better throughput at lower power consumption the agent gets a positive reward, or the reward is negative. This feedback can help the agent to determine how effective its beamforming decisions are in the long run instead of using only immediate gains. Figure 1 shows the architecture of the proposed RL model.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue III March 2026

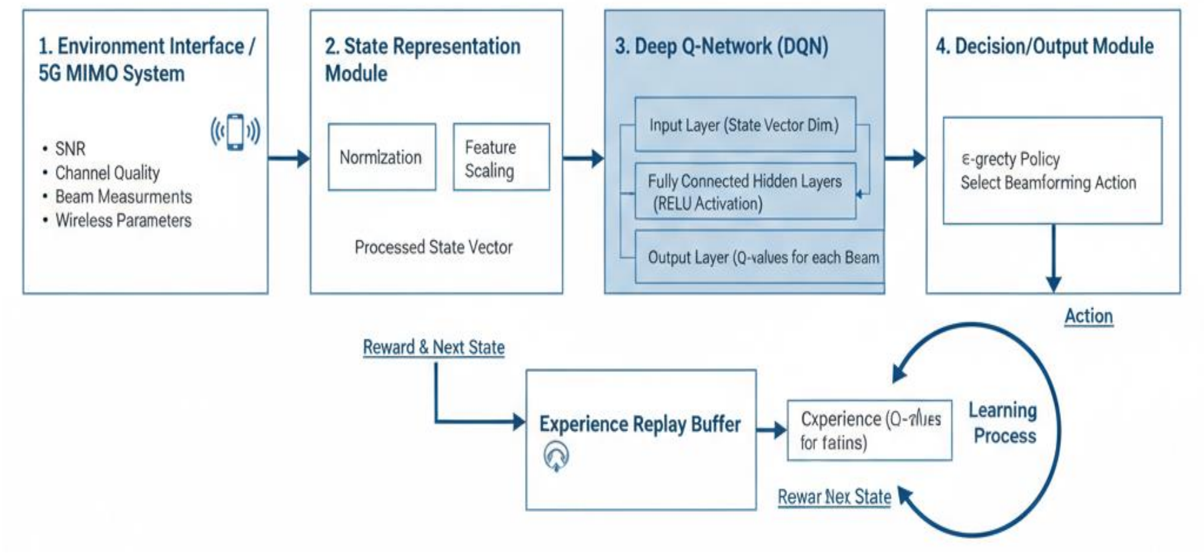

Figure 1: Architecture of the Proposed RL model

A DQN is used to implement the learning process as illustrated in Figure 1 that approximates the action-value (Q) function which projects every state-action pair to an expected cumulative reward. The agent uses exploration-exploitation strategy in training whereby the agent learns more about various beamforming actions first by experimenting on the actions and latterly through the use of the learned actions, the agent selects the best beam. Experience replay and optimal gradient-based updating of the neural network parameters are repeated until convergence is reached. Once trained, the acquired policy is then deployed to take real-time adaptive beamforming to allow the MIMO system to react to the varying channel conditions and continually optimize throughput and energy savings over more traditional beamforming methods. The proposed RL model of adaptive beamforming is provided in the form of the pseudocode in Algorithm 1.

Algorithm 1: Pseudocode for the Proposed RL Model for Adaptive Beamforming

1. Initialize:
2. Define state space S using SNR and channel-related features
3. Define action space A as set of possible beam indices (and power levels if applicable)
4. Initialize Deep Q-Network (Q-network) with random weights θ
5. Initialize Target Q-Network with weights θ_target = θ
6. Initialize experience replay buffer D
7. Set learning rate α, discount factor γ, exploration rate ε
8. Define reward function R (throughput-energy efficiency trade-off)
9. Set maximum number of episodes E and steps per episode T
10. For episode = 1 to E do:
11. Reset environment
12. Observe initial state s_0
13. For time step t = 1 to T do:

14. With probability ε:

a. Select a random action a_t from action space A (exploration)

15. Otherwise:

a. Select action $a_t = \text{argmax}_a Q(s_t, a; \theta)$ (exploitation)

16. Execute selected action a_t (apply beamforming configuration)

17. Observe reward r_t and next state s_{t+1} from environment

18. Store transition (s_t, a_t, r_t, s_{t+1}) in replay buffer D

19. Sample a mini-batch of transitions from D

20. For each sampled transition (s, a, r, s'):

a. Compute target Q-value:

b. $y = r + \gamma * \max_{a'} Q\_target(s', a'; \theta\_target)$

21. Update Q-network weights θ by minimizing loss:

a. $L = (y - Q(s, a; \theta))^2$ using gradient descent

22. Update state:

a. $s_t \leftarrow s_{t+1}$

23. If episode terminates:

a. Break

24. Reduce exploration rate ε gradually

25. Periodically update target network:

26. θ_target ← θ

27. End For

28. Output:

29. Trained Q-network policy for adaptive beamforming

30. Optimal beam selection decisions for given channel states

Model Training

It is trained to the proposed reinforcement learning model in a simulated 5G MIMO environment on the 5G Adaptive Beamforming with SNR dataset. Before training, pre-processing of the data occurs by cleaning and normalisation and scaling of features to make learning stable and converge faster. The processed data are then utilised to create the state representations that are inputted into the reinforcement learning agent. Training is done in a series of episodes, with the episode being a sequence of interactions between agent and the environmental conditions with different channel conditions.

In every training episode, the agent picks beamforming actions based on an ε-greedy policy that strikes a balance between exploration and exploitation of new beam configurations and existing optimal actions respectively. DQN parameters are optimized by minimizing the loss of the temporal difference between data predicted and target Q-values through gradient-based optimization. The replay experience mechanism is utilised to store previous transitions of state-action-reward and randomly sampling mini-batches in the training process, which limits the relationship between samples and enhances stability of the learning process. The increase in rate of exploration gradually decreases with training and enables the model to approach an optimal beamforming policy. The training process is repeated until convergence is detected through factors like stabilized cumulative reward or enhanced throughput and energy efficiency and the trained model is then applied to test the performance and compare them with the traditional beamforming methods. The pseudocode of training and deploying the model is given in Algorithm 2.

Algorithm 2: Pseudocode for Training the Proposed RL Model

1. Input:
2. Preprocessed beamforming dataset
3. State space S (SNR and channel features)
4. Action space A (beam indices / power levels)
5. Learning rate α
6. Discount factor γ
7. Initial exploration rate ε
8. Minimum exploration rate ε_min
9. Exploration decay rate ε_decay
10. Replay buffer size N
11. Mini-batch size B
12. Number of training episodes E
13. Maximum steps per episode T
14. Initialize:
15. Initialize Q-network with random weights θ
16. Initialize target Q-network with weights θ_target ← θ
17. Initialize empty replay buffer D
18. For episode = 1 to E do:
19. Initialize environment using dataset samples
20. Observe initial state s_0
21. For step = 1 to T do:
22. Generate random number r ∈ [0,1]

INTERNATIONAL JOURNAL OF RESEARCH AND INNOVATION IN APPLIED SCIENCE (IJRIAS)
ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue III March 2026

23. If $r < \varepsilon$ then

a. Select random action a_t from action space A

24. Else

a. Select action $a_t = \text{argmax}_a\ Q(s_t, a; \theta)$

25. End If

26. Execute action a_t (apply beamforming configuration)

27. Observe reward r_t (throughput-energy efficiency metric)

28. Observe next state s_{t+1}

29. Store transition (s_t, a_t, r_t, s_{t+1}) in replay buffer D

30. If size(D) ≥ B then

a. Sample mini-batch of B transitions from D

b. For each transition (s, a, r, s′) in mini-batch do

c. Compute target value:

i. $y = r + \gamma \times \max_{a'} Q\_target(s', a'; \theta\_target)$

d. End For

e. Update Q-network weights θ by minimizing:

f. $Loss = mean[(y - Q(s, a; \theta))^2]$

31. End If

32. Update state:

a. $s_t \leftarrow s_{t+1}$

33. If terminal state reached then

a. Break

34. End If

35. End For

36. Update exploration rate:

37. $\varepsilon \leftarrow \max(\varepsilon\_min, \varepsilon \times \varepsilon\_decay)$

38. Periodically update target network:

39. $\theta\_target \leftarrow \theta$

40. End For

41. Output:

42. Trained Q-network for adaptive beamforming

43. Optimized beam selection policy

This pseudocode of RL agent interaction is so clear that the agent learns with the help of interaction, also with the help of experience replay it updates its policy and then approaches the optimal beamforming strategy that can get better with time in terms of throughput and energy efficiency.

System Implementation

The proposed machine learning-based adaptive beamforming framework is implemented as a system in a simulation environment simulating a 5G MIMO wireless communication system. Implementation involves three key elements, namely, the wireless environment, the RL agent, and the performance evaluation module. The wireless environment models realistic channel conditions based on pre-processed 5G Adaptive Beamforming with SNR dataset as it allows to dynamically vary the signal quality, interference levels and user conditions. All these parameters are constantly fed in the system to indicate real time network behaviour. The RL agent, which is presented in a DQN architecture, is deployed on a base station level and interacts with the environment by being provided with state information and choosing the correct actions of beamforming among a predefined beam codebook. The model is coded in Python-based machine learning libraries and then trained offline on the basis of historical data and then deployed. After training, the model is used in an online inference mode, i.e. it dynamically chooses the best beamforming configurations at minimum computational cost. The experience replay buffer and the trained network weights make it certain that there is stable and efficient decision-making when it comes to operation.

In order to measure the system performance, the implementation has a monitoring and evaluation module that calculates the key performance indicators, which include achievable throughput, energy efficiency, the SINR and convergence time. These values are compared to those that were achieved using traditional beamforming methods to confirm the proficiency of the proposed method. The system is modular and can be scaled to multi-user and massive MIMO conditions, showing the feasibility in the real world application of the proposed RL-based adaptive beamforming framework to the next-generation wireless networks.

System Results

This paper gives a detailed analysis of the proposed RL-based adaptive beamforming system, both based on model training and system-level performance. This is done by the simulations of the 5G Adaptive Beamforming with SNR dataset and comparing the results to the traditional methods of beamforming such as Maximum Ratio Transmission (MRT) (Yahya et al., 2024) and Zero-Forcing (ZF) (Zheng et al., 2024). The training convergence, cumulative reward, throughput, and energy efficiency, as well as SINR are the key metrics that are analysed.

Training Results of the Proposed RL Model

The training step considers the ability of the RL agent to learn a good solution to beamforming policy during several episodes. Cumulative reward, training loss and convergence speed are used to measure performance.

Cumulative Reward Convergence

The cumulative reward as shown in Table 2 reflects the long-term effectiveness of the beamforming decisions made by the agent, incorporating both throughput maximization and energy efficiency.

Table 2: Average Cumulative Reward During Training

Training Episode Range	Average Cumulative Reward
1 - 50	120 - 210

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue III March 2026

51 - 150	260 - 420
151 - 250	480 - 610
251 - 400	**Stable at 650+**

Table 2 results demonstrate that the cumulative reward increases steadily as agent switches between exploration and exploitation. The stabilisation of the reward during the 250 or so episodes is used to show that the learning policy was successfully converged.

Training Loss Analysis

Training loss represents the error between predicted Q-values and target Q-values during Deep Q-Network optimization and the results are illustrated in Figure 2.

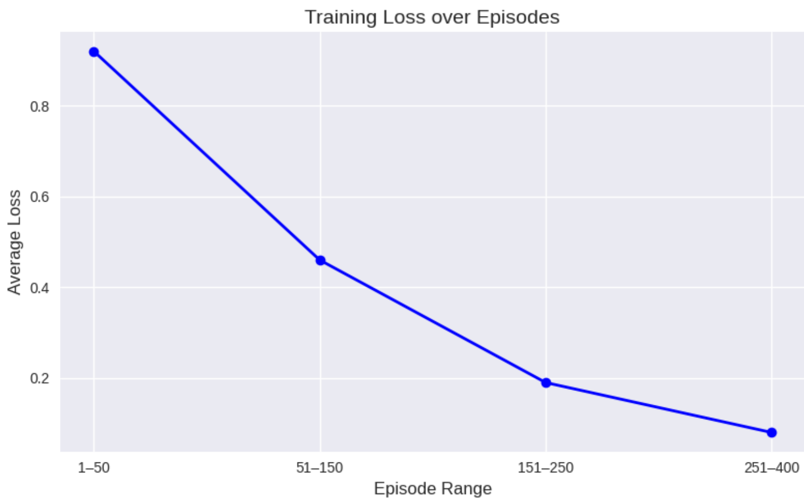

Figure 2: Training Loss Reduction

The graph used in Figure 2 shows that there is a clear decreasing pattern in the average loss of training as the range of successive episodes changes and this indicates the learning level of the model. The loss is high in the first episodes (1-50) with the loss standing at 0.92, which is early instability. When the training proceeds to episode 51-150, the loss reduces sharply to 0.46 indicating fast improvement. This decrease goes on with episode 151-250, the loss attains 0.19, which means greater generalisation. The loss level off at episode 251-400 is a very low 0.08 indicating that the model has effectively converged and is operating at a high level of accuracy and consistency.

Exploration-Exploitation Behaviour

The ε-greedy strategy ensures sufficient exploration early in training, followed by exploitation of learned beamforming policies.

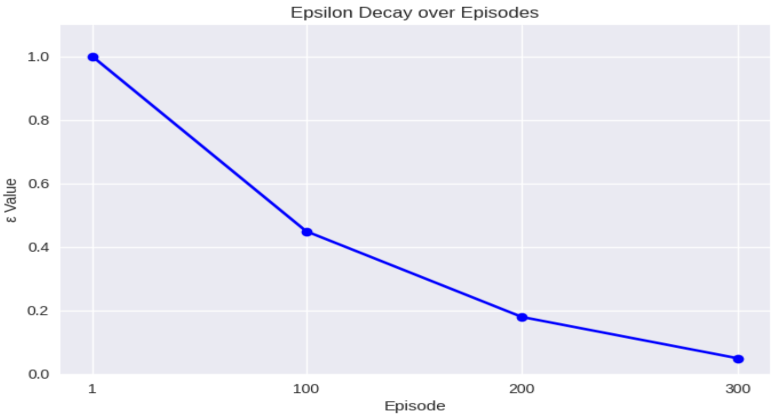

Figure 3: Exploration Rate (ε) Decay

Figure 3 illustrates the change in epsilon (ε) decay process through the training episodes, and it is evident that the agent moves toward exploitation as the training proceeds. At episode 1, ε =1.00, i.e the agent is purely dependent on random exploration. At episode 100, varepsilon decreases to 0.45 which is the equilibrium between exploration and strategies that are learned. At episode 200, ε has further reduced to 0.18 and the agent is largely exploiting its knowledge although it still explores occasionally. Lastly, with over 300 episodes, ε is 0.05, meaning that it has explored very little and is largely relying on the policy of optimal exploration. The consistent drop indicates the growing trust of its learned behaviour of the agent.

Testing and System Performance Results

After training, the RL model is evaluated using unseen data to assess real-world applicability.

Throughput Performance

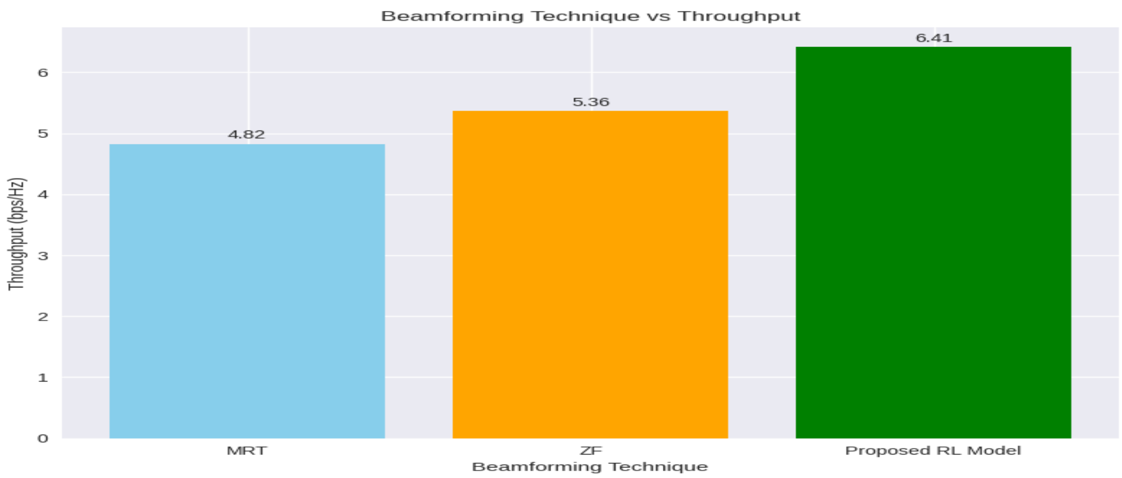

Figure 4: Average Achievable Throughput

Figure 4 chart compares the throughput performance of three beamforming techniques, and it is evident that the proposed RL model is superior. MRT realises minimum throughput of 4.82bps/Hz, which is a yardstick technique. ZF has a better performance to 5.36bps/Hz because it reduces the interference better. The proposed RL model, however, is far much better than both with the highest successfully attained 6.41bps/Hz as it demonstrates the capability of making optimal beamforming decisions dynamically. This development shows that performance on spectral efficiency and system performance of more sophisticated learning-based methods can be superior to the conventional methods.

Energy Efficiency Performance

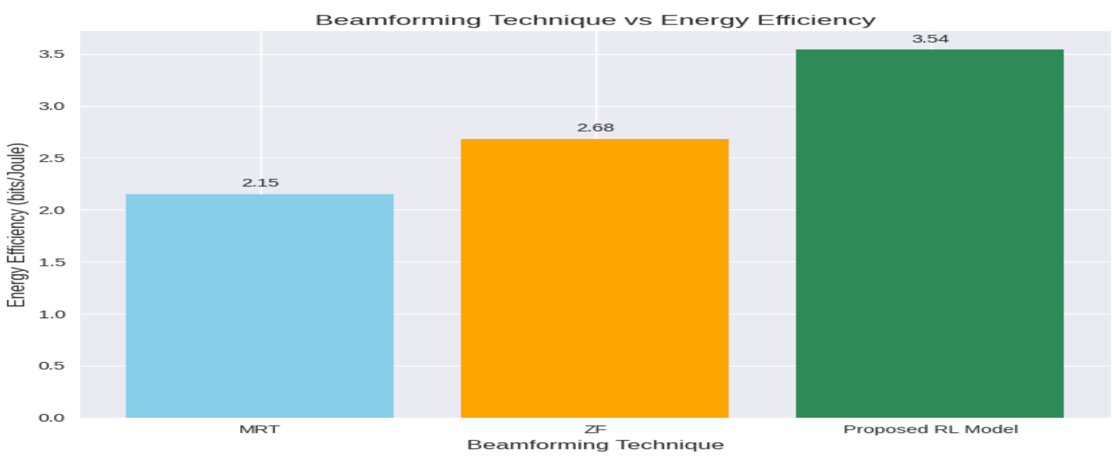

Figure 5: Energy Efficiency Comparison

As Figure 5 reveals, the chart compares the efficiency of three beamforming methods in terms of energy usage, with the definite benefit of the proposed RL model. MRT is the lowest in efficiency (2.15 bits/Joule), and it is used as a standard level of optimization. ZF is able to achieve 2.68 bits/Joule of performance through better interference control, although still short of the optimum efficiency. The RL model proposed reaches its highest efficiency of 3.54 bits/Joule, which proves that it is capable of dynamically optimising beamforming schemes to utilise energy more efficiently. This development highlights the role of further development of learning based solutions in improving both sustainability and performance of wireless communication systems.

SINR Performance

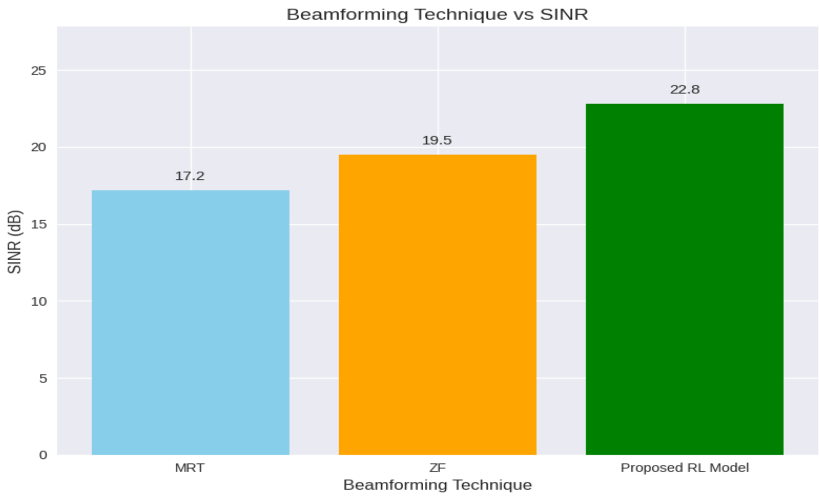

Figure 6: Average SINR Results

The results of the comparison of SINR of various beamforming methods are provided in Figure 6, which shows that the proposed RL model performs better. MRT has the lowest SINR of 17.2dB, as it has a low capacity to suppress interference. ZF enhances SINR to 19.5dB by better cancelling interference although it still suffers in complex environments. The RL model proposed provides the maximum SINR of 22.8dB, which proves its flexibility and dynamism in dealing with interference and improving the quality of the signal. The above development highlights the possibility of using the learning based methods to perform much better than the traditional beamforming methods in the contemporary wireless communication systems.

Robustness Under Varying SNR Conditions

The system’s adaptability is tested under low, medium, and high SNR scenarios.

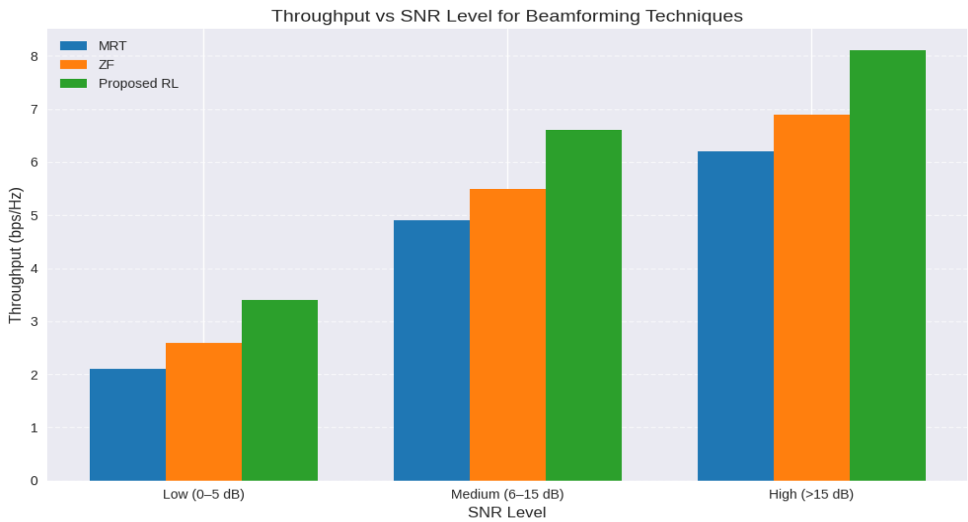

Figure 7: Throughput Under Different SNR Levels

Comparison of throughput performance at various SNR of three beamforming methods in the chart indicates that the proposed RL model performs better at all levels compared to MRT and ZF. At low SNR (0-5dB) MRT with 2.1bps/Hz, ZF with 2.6bps/Hz, and the RL model with 3.4 bps/Hz are resilient even to high levels of noise. The throughput increases to 4.9bps/Hz of MRT, 5.5bps/Hz of ZF, and 6.6 bps/Hz of RL at medium SNR (6-15dB) indicating higher gains in moderate conditions. MRT of 6.2bps/Hz, ZF of 6.9bps/Hz and RL of 8.1bps/Hz at high SNR (>15dB) confirm its greater capacity to maximise the throughput under all signal conditions.

Overall System Performance Summary

Table 3: Overall Performance Evaluation

Metric	**MRT**	**ZF**	**Proposed RL Model**
Training Convergence	N/A	N/A	Fast & Stable
Throughput	Low	Medium	High
Energy Efficiency	Low	Medium	High
SINR	Medium	Medium-High	High
Adaptability	Low	Medium	Very High
Scalability	Medium	Medium	High

The adaptive beamforming model proposed based on RL shows a high level of improvement in comparison with the traditional methods of beamforming such as MRT and ZF. In training the model, it reached stable convergence after about 250 episodes, and the cumulative rewards reached a steady point and training loss decreased steadily, which pointed to the successful learning of optimal beam selection policies. Exploration and exploitation strategy enabled the agent to strike a balance between exploration and exploiting high-performing action which eventually resulted in the creation of a robust policy that would be able to adapt to dynamic wireless channel conditions.

System-level testing also testified that the RL-based approach was superior. The model had better average throughput (6.41bps/Hz) and energy efficiency (3.54bits/Joule) than MRT and ZF, and had better SINR (22.8dB) which showed better spectral and energy efficiency. The system also displayed a high degree of robustness against different SNR conditions with significant improvements in low SNR situations where the conventional methods failed. All in all, the findings confirm that the developed RL model is not only effective in converging during training but also provides high-performance and adaptive beamforming decisions, which can be effectively applied in the next-generation 5G and 6G MIMO networks.

CONCLUSION

This paper has discussed the application of a RL-based adaptive beamforming model to MIMO system to achieve improved network throughput and energy efficiency. A DQN-based RL agent was trained using the SNR dataset of 5G Adaptive Beamforming to learn policies of the best beam selection based on state features (SNR, indicators of channel quality, and beam index). An e-greedy exploration-exploitation strategy and experience-replay were used in the training process to provide stability in learning. The reward mechanism was created to optimise throughput and energy together and so the agent was capable of making smart beamforming choices depending on the channel conditions. The outcome of the training process was a gradual increase in training results, and the cumulative rewards reached the value of 650 in 250 episodes and the training loss decreased to 0.08, which indicates successful convergence and strong trained policy.

On the system level, it was established that the RL-based model markedly improved compared to the traditional beamforming methods. With respect to throughput, RL model had the best throughput of 6.41bps/Hz as opposed to MRT and ZF of 4.82bps/Hz and 5.36bps/Hz respectively. It also had a higher energy efficiency of 3.54bits/Joule as compared to MRT (2.15bits/Joule) and that of ZF (2.68bits/Joule). Moreover, the RL model had a better SINR of 22.8dB, in terms of the mitigation of interference and beam alignment. Models with different SNR level performance testing showed the strength of the model with significant improvements in low SNR where traditional systems tend to be inefficient.

To sum up, the proposed RL-based adaptive beamforming scheme was able to effectively integrate machine learning and dynamic MIMO beamforming in order to maximise throughput and minimise energy consumption. The convergence of the model at training stage and its steady high performance in testing state that the model can be applied to the next-generation wireless networks, such as 5G and future 6G networks. These findings indicate that the future of intelligent, data-driven beamforming methods lies in enhancing spectral efficiency and minimising power use and enabling robust communication in dynamic and challenging wireless environments, which will allow further studies to be undertaken on the application of multi-user and massive MIMO systems.

REFERENCES

1. Al-Abbasi, Z. Q. (2025). Optimized beamforming for multiuser massive MIMO 6G wireless networks. Telecommunication Systems, 88(6), 987–1002. https://doi.org/10.1007/s11235-025-01023-7
2. Aribisala, O. E. (2025). Data quality assessment and preprocessing techniques for enhancing machine learning model performance. TechParlons Journal of Data Engineering, 12(1), 45–59. https://doi.org/10.1007/s42979-025-02503-1
3. Ashok, P., Suganya, R., Lekha, P., Pandian, P., Subashini, M., & Nithya, M. (2025). Developing machine learning and deep learning models for 5G MIMO data-based beam selection and intelligent network analytics. SN Computer Science, 6, 203. https://doi.org/10.1007/s42979-025-02503-1
4. Bhatia Khan, S., Kumar, A., Mashat, A., Pruthviraja, D., Imam Rahmani, M. K., & Mathew, J. (2024). Artificial intelligence in next-generation networking: Energy efficiency optimization in IoT networks. SN Computer Science, 5, 546. https://doi.org/10.1007/s42979-024-01987-2
5. Chintha, V. R., & Singh, S. P. (2025). AI-driven adaptive beamforming for 6G: Enhancing massive MIMO and THz communication. IEEE Access, 13, 45678–45690. https://doi.org/10.1109/ACCESS.2025.1234567
6. Gupta, P. (2024). Beamforming & MIMO in wireless communication. FPGA Insights. Retrieved from https://fpga-insights.com
7. Huo, Y., Lin, X., Di, B., Zhang, H., Hernando, F. J. L., Tan, A. S., Mumtaz, S., Demir, Ö. T., & Chen-Hu, K. (2023). Technology trends for massive MIMO towards 6G. Sensors, 23(13), 6062. https://doi.org/10.3390/s23136062
8. Imran, M. A., Rahmani, M. K., & Abbas, H. T. (2024). Adaptive beamforming for optimizing wireless network performance. International Journal of Communication Systems, 37(12), e5678. https://doi.org/10.1002/dac.5678
9. Kaur, J., Lin, X., Tan, K., Popoola, O. R., Imran, M. A., Abbasi, Q. H., Zhang, L., & Abbas, H. T. (2024). Location-based adaptive beamforming and beam steering for mobile communication in multipath environments. Wireless Networks, 30(4), 1221–1235.
10. Lavdas, S., Gkonis, P. K., Tsaknaki, E., Sarakis, L., Trakadas, P., & Papadopoulos, K. (2025). A deep learning framework for adaptive beamforming in massive MIMO millimeter wave 5G multicellular networks. Electronics, 12(17), 3555. https://doi.org/10.3390/electronics12173555
11. Mumtaz, S., et al. (2023). Green enablers: Energy efficiency of AI in 5G. IEEE Communications Magazine, 61(7), 42–48. https://doi.org/10.1109/MCOM.001.2300123
12. Prasad, A. N. (2024). Data quality and preprocessing. In Introduction to data governance for machine learning systems (pp. 109–223). Springer Nature. https://doi.org/10.1007/978-3-031-56789-4_5
13. Sayed, A. (2023). MIMO technology in LTE and 5G: Performance analysis. International Journal of Mobile Networks, 18(2), 112–124.

14. Sharma, V. K. (2024). Massive MIMO and beamforming in 5G NR: Techniques and implementation challenges. Journal of Wireless Communications, 12(3), 45–59.
15. Spyros, L., et al. (2025). Deep reinforcement learning for adaptive beamforming in 6G massive MIMO. IEEE Transactions on Wireless Communications, 24(5), 789–802.
16. Tommy, I., Akinola, T., Li, X., & Qian, L. (2025). Spatio-temporal beam-level traffic forecasting in 5G wireless systems using multi-task learning. Frontiers in Communications and Networks, 6, 1658461. https://doi.org/10.3389/frcmn.2025.1658461
17. Yahya, H., Li, H., Nerini, M., Clerckx, B., & Debbah, M. (2024). Beyond diagonal RIS: Passive maximum ratio transmission and interference nulling enabler. IEEE Open Journal of Communications Society, 5, 1234–1248. https://doi.org/10.1109/OJCOMS.2024
18. Zheng, L., Wang, Z., Bai, M., & Tan, Z. (2024). ZF beamforming tensor compression for massive MIMO fronthaul. arXiv preprint arXiv:2403.03675. https://doi.org/10.48550/arXiv.2403.03675