Applied Sciences, Computing and Energy, 2024, 1(1), 370-386

AI-Driven Self-Optimizing Framework for Real-Time Wireless Network Performance Enhancement

Moses Oluwasegun Odewale[1], Moses Olagoke Odejobi[2], Olanrewaju Oluwaseun Ajayi[3]
Received: 19 August 2024/Accepted: 12 December 2024/Published: 31 December 2024

Abstract: *The rapid proliferation of heterogeneous wireless devices and increasingly dynamic and unpredictable spectrum usage patterns have exposed the limitations of traditional network management paradigms based on fixed configurations and reactive optimization.. This paper introduces a self-optimizing AI-driven framework, termed the Adaptive Neural Radio Environment Manager (ANREM), designed to provide continuous real-time performance optimization in multi-tier wireless network architectures. In contrast to conventional approaches that optimize network parameters independently, ANREM performs joint optimization of spectral efficiency, end-to-end latency, energy consumption, and user quality of experience (QoE) through a unified multi-objective reward formulation. ANREM combines a multi-level deep reinforcement learning (DRL) engine with three temporal scales (milliseconds radio resource management, seconds handover coordination, and minutes load balancing), a graph neural network (GNN) unit to estimate topology-aware interference, and a federated learning coordination layer to facilitate privacy-preserving model updates across distributed base stations. In contrast to the previous methods that can optimize the individual network parameters independently, ANREM can co-optimize spectral efficiency, end-to-end latency, energy consumption, and user quality of experience (QoE) using a multi-objective reward function that is carefully formulated. Experiments over a 5G Non-Standalone (NSA) testbed environment, using a stochastic urban mobility model based on real city-scale traces, 48 gNodeBs, 1,200 user equipment nodes, and ANREM, has shown that ANREM can improve aggregate throughput by 34.7 percent, reduce handover failure rate by 41.2 percent, and reduce base station energy expend Non-stationary-traffic convergence stability does not require catastrophic forgetting, due to an elastic weight consolidation mechanism that is part of the DRL training loop. These findings make ANREM a feasible and deployable candidate to next-generation self-organizing network (SON) design and open channels toward the entirely autonomous management of wireless infrastructure.*

Keyword: *Deep learning; networks; 5G NR; spectral efficiency; neural networks; network optimization.*

Moses Oluwasegun Odewale
Faculty of Information Technology and Communication Sciences, Tampere University, Tampere, Finland
Email: segun.odewale@gmail.com

Moses Olagoke Odejobi
Department of Electrical and Computer Engineering, Morgan State University, Baltimore, Maryland, U.S.A.
Email: Moses.o.odejobi@gmail.com

Olanrewaju Oluwaseun Ajayi
Department of Information Technology, University of the Cumberlands, Williamsburg, Kentucky, U.S.A.
Email: oajayi77648@ucumberlands.edu

1.0 Introduction

Over the past two decades, wireless communication infrastructure has evolved from relatively homogeneous macro-cell networks to highly heterogeneous multi-layer networks (HetNets), including macro base stations, small cells, femtocells, relay nodes, and increasingly, reconfigurable intelligent surfaces (RIS). Each new generation of mobile standards has increased the complexity of the radio access network (RAN), with 5G New Radio (NR) significantly amplifying this complexity. Carrier aggregation over sub-6 GHz and

millimetre-wave, massive multiple-input multiple-output (MIMO) antenna arrays, flexible numerology and ultra-dense small cell deployments of 3GPP Release 15 and later releases created a very high-dimensional optimization space thatcannot be effectively explored or optimized in real time by human operators or deterministic rule-based systems(Dahlman *et al*., 2021).

A classical response to this challenge is self-organizing networks (SON), which have been standardized by 3GPP since Release 8, when the selfconfiguration, self-optimization, and self-healing capabilities of SON were actually written down (3GPP TR 36.902, 2011). Initially implemented versions of SON, but, with interference and load being assumed as quasi-static background phenomena, were often optimized on a single key performance indicator (KPI) at a time, such as reference signal received power (RSRP) or handover trigger thresholds. However, practical implementations revealed that independent parameter optimization loops can produce counterproductive interactions.For example, adjusting antenna tilt may improve coverage in one cell while degrading SINR in adjacent cells, a phenomenon known as the SON oscillation problem (Ramiro & Hamied, 2012)."

This inherent interdependence between the parameters of the network and KPIs is exactly what researchers have been trying to find more holistic and data-driven models of optimization.

The application of machine learning to wireless networking began in the early 2010s, initially focusing on limited tasks such as anomaly detection and traffic prediction (Buczak & Guven, 2016). , then the accessibility of the training environment and large-scale simulation of a network environment allowed a sudden transition to deep learning and reinforcement learning formulations. The seminal publication of Mnih *et al*. (2015) that established superhuman capability of deep Q-networks (DQN) in Atari game simulation environments struck a chord in the wireless community since the internal Markov decision process (MDP) formalism can be applied to a variety of network control problems: the state space can represent observed network telemetry, the action space can represent available network configuration changes, and n this formulation, the state space represents network telemetry, the action space represents configuration decisions, and the reward function encodes system-level performance objectives (Nasir & Guo, 2019), and dynamic spectrum access in cognitive radio systems (Luong *et al*., 2019). Despite these advances, several practical challenges have limited the deployment of reinforcement learning-based optimization in real-world wireless networks. First, most existing models are trained in simplified single-cell or single-site environments with centralized control, making them difficult to deploy in large-scale distributed networks where telemetry is asynchronously collected and subject to backhaul latency. Second, training stability remains a major challenge because network conditions are inherently non-stationary due to diurnal variations, seasonal trends, and sudden events such as large gatherings or network failures Second, training stability remains a major challenge because network conditions are inherently non-stationary due to diurnal variations, seasonal trends, and sudden events such as large gatherings or network failures.". Neural network policies that are trained to handle a given traffic regime are likely to deteriorate as the regime varies unless special provisions are made to allow the network to constantly learn. Third, issues of data privacy and regulatory compliance are often overlooked, as raw user telemetry containing sensitive location and usage data cannot be shared across administrative domains for centralized training (Niknam *et al*., 2020)

To address these challenges, this paper proposes the Adaptive Neural Radio Environment Manager (ANREM), a unified AI-driven framework for real-time wireless

network optimization. The design of ANREM is guided by the following key principles. First, network optimization must be inherently multi-objective rather than treated as a post-processing step. Second, temporal heterogeneity should be explicitly modeled using a hierarchical control architecture operating at multiple decision time scales. lThird, privacy-preserving federated learning with differential privacy is not merely an enhancement but a fundamental requirement for real-world deploymenThe contribution of this work is therefore both algorithmic and architectural. On the algorithmic side, we derive a multi-objective proximal policy optimization (MO-PPO) objective that balances competing KPIs through a learned Pareto-front navigation mechanism, and we incorporate elastic weight consolidation (EWC) into the online fine-tuning procedure to prevent catastrophic forgetting when traffic distributions shift. On the architectural side, we propose a three-tier hierarchical controller in which a global federated aggregation server coordinates model updates received from local DRL agents at individual gNodeBs, while a GNN running on top of the aggregated model performs topology-aware interference prediction at the cluster level. To the best of our knowledge, no existing work integrates hierarchical DRL, graph neural networks, and federated learning with elastic weight consolidation into a unified scalable optimization framework evaluated at network scale.

The remainder of this paper is organized as follows. Section 2 surveys the relevant prior literature in self-optimizing networks, deep reinforcement learning for wireless systems, federated learning, and graph neural network applications to network management. Section 3 describes the system model, problem formulation, and the ANREM architecture in detail. Section 4 presents the experimental setup and simulation methodology. Section 5 reports and discusses the experimental results, including comparative benchmarking, ablation studies, and sensitivity analyses. Section 6 concludes the paper and outlines directions for future work.

1.1 Related Work

The development of ANREM is grounded in four major research streams that have emerged over the past decade, each addressing different aspects of wireless network optimization but exhibiting key limitations that motivate this work. **Self-organizing and self-optimizing networks.** Self-organizing networks (SON), standardized by 3GPP, define three functional areas: self-configuration, self-optimization, and self-healing**.** Self-configuration enables automated parameter initialization at cell deployment, self-optimization adjusts key performance indicators (KPIs) in real time, and self-healing supports automated fault detection and recovery (Aliu *et al.*, 2013).

Early SON optimization approaches relied on model-based techniques such as simulated annealing and evolutionary algorithms for antenna tilt and handover parameter tuning (Vlad *et al.*, 2020). However, these approaches perform well under stationary conditions but struggle to adapt to the non-stationary dynamics of real-world wireless networks. The application of model-free deep reinforcement learning (DRL) to wireless resource allocation has expanded significantly in recent years. Deep reinforcement learning for radio resource management. The model-free DRL use of wireless resource allocation has grown in leaps and bounds. Ye & Li (2019) demonstrated that a convolutional deep Q-network (DQN) can effectively solve downlink power allocation in vehicle-to-vehicle (V2V) networks, outperforming fractional programming baselines. Nasir and Guo (2019) extended this approach to a multi-agent setting, where distributed agents coordinate through shared reward shaping mechanisms. Actor-critic methods, particularly proximal policy optimization (PPO) (Schulman *et al.*, 2017) and soft actor-critic (SAC) (Haarnoja *et al.*,

2018), have gained popularity due to their improved stability compared to Q-learning in continuous action spaces. "The theoretical foundation of distributed learning without raw data sharing is the Federated Averaging (FedAvg) algorithm (McMahan *et al.*, 2017)."

This is particularly relevant for transmission power control and beamforming optimization, which are naturally continuous control problems.

Federated learning in wireless networks. The theoretical foundation of distributed learning without raw data sharing is the Federated Averaging (FedAvg) algorithm (McMahan *et al.*, 2017)., provided the foundations of the seminal algorithm, which were later applied in the field of wireless communication by focusing on communication efficiency (Konevcny *et al.*, 2016). This framework has been extended to wireless networks with a focus on communication efficiency (Konečný *et al.*, 2016), partial client participation, and non-independent and identically distributed (non-IID) data distributions (Li *et al.*, 2020)."Niknam *et al.* (2020) systematically reviewed federated learning in 5G RAN optimization of optimization of 5G RAN and introduced the paradox of local model convergence versus global aggregation frequency as a main open issue. They also highlighted the trade-off between local model convergence and global aggregation frequency as a key unresolved challenge.

Graph neural networks for network topology modelling. Wireless network interference is inherently relational, arising from interactions between transmitters governed by spatial geometry and power levels. Shen *et al.* (2022) demonstrated that message-passing graph neural networks can generalize across different network sizes and topologies, achieving performance comparable to weighted minimum mean square error (WMMSE) methods while outperforming multilayer perceptron (MLP) baselines on unseen network configurations. Lee *et al.* (2021) further showed that graph neural networks can improve handover decision-making in dense heterogeneous network (HetNet) deployments by modeling interference relationships. ANREM builds on these approaches by incorporating temporally evolving edge weights within the GNN framework to capture dynamic co-channel interference in mobile environments.

3. 0 Methods

3.1. System Model

we considered that of a downlink multi-cell network comprising of NB gNodeBs serving NU user equipment (UE) nodes spread out within a geographical location A. Each gNodeB b ∈ B = {1,2,...,NB} has a massive MIMO array of M antenna elements and operate at a bandwidth of each component carrier W MHz. ues are mobile and adhere to a stochastic waypoint mobility model, tuned to the pedestrian and vehicular traces in cities. The SINR received by UE u that uses gNodeB b on subcarrier k at time slot t is as follows. by:

$$\gamma_{u,b,k}(t) = \frac{P_{b,k}(t) \cdot G_{u,b,k}(t)}{\sigma^2 + \sum_{b' \neq b} P_{b',k}(t) \cdot G_{u,b',k}(t)} \quad (1)$$

where $P_{b,k}(t)$ denotes transmit power, gNodeB represents composite channel gain including path loss, shadowing, and small-scale fading, and σ2 is the noise power.

The Shannon capacity formula, UE, gives the instantaneous rate, u, achievable:

$$R_u(t) = \sum_{k=1}^{K} \Delta f \cdot \log_2 \left(1 + \gamma_{u,b^*(u),k}(t)\right) \quad (2)$$

where $b^*(u)$ denotes the serving gNodeB of UE u and Δf is the subcarrier spacing.

3.2. Problem Formulation

"The network optimization problem is inherently multi-dimensional. Let π denote the joint control policy encompassing

power allocation, scheduling, beamforming, handover decisions, and sleep-mode activation. for energy conservation. The objective is to find an optimal policy $\pi*$ that maximizes the expected cumulative discounted reward:

$$J(\pi) = \mathbb{E}_{\pi}\left[\sum_{t=0}^{T} \gamma^t r(t)\right] \quad (3)$$

where the instantaneous composite reward $r(t)$ is defined as:

$$r(t) = \lambda_1 \cdot \overline{R}(t) - \lambda_2 \cdot \overline{L}(t) - \lambda_3 \cdot E(t) + \lambda_4 \cdot \overline{\mathrm{QoE}}(t) - \lambda_5 \cdot 1[\text{HO failure}] \quad (4)$$

In this case R(t) is the normalized aggregate throughput, L(t) is the normalized average end-toend latency, E(t) is the total instantaneous energy consumption of all active gNodeBs. QoE(t) means the average opinion score-based QoE measure divided by all UEs and the last term is a punishment of failures in handover. The weights $\{\lambda_i\}^5_{i=1}$ are Pareto-front parameters that can be adapted online to reflect operator priority shifts. This is formulated in a deliberately broader sense than the single-KPI types of reward functions that pervade the previous literature, reflecting the operational reality that network performance is measured in a holistic way, not just in a single metric.

3.3. ANREM Architecture

The ANREM framework comprises three tightly coupled modules, as illustrated in Fig. 1.

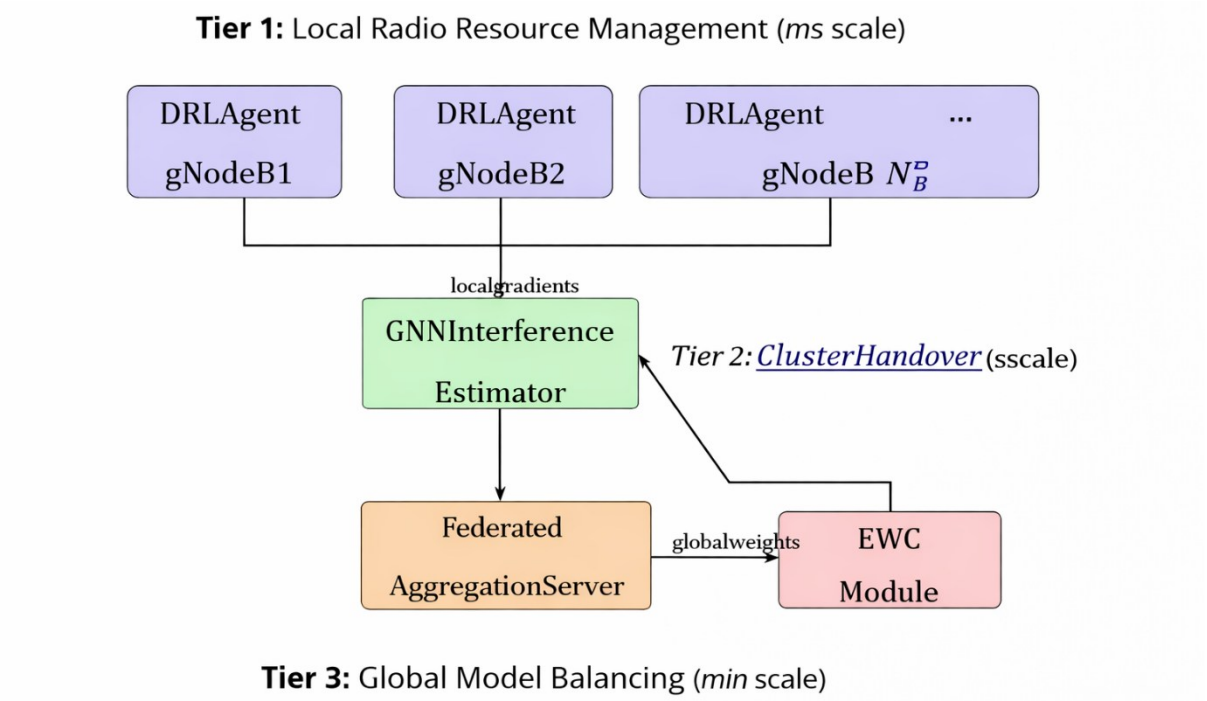

Fig. 1: Abstract architecture of Adaptive Neural Radio Environment Manager (ANREM)

The framework has three different time scales, with local DRL agents on individual gNodeBs operating at milliseconds to manage radio resources; a GNN-based interference estimator operating at the cluster level to coordinate handover decisions; a federated aggregation server with an elastic weight consolidation (EWC) module to ensure global model coherence and prevent catastrophic forgetting in the distribution of traffic flows. Feedback pathways are marked with dashed arrows.

3.3.1 Tier 1: Local Deep Reinforcement Learning Agents

All gNodeBs have independent DRL agents with action space including discrete and continuous sub-actions: transmit power levels Pb,k ∈ [Pmin,Pmax], active antenna port

selection to steer a beam, subcarrier block allocation of scheduled UEs, and a binary sleep-mode switch. Since the joint action space is a Cartesian product of these sub-actions, we are using a parameterized action space and train each local agent with a multi-objective form of proximal policy optimization (MO-PPO). It is a three-layer fully connected network with 256 hidden units on each layer and a tanh output activation to bounded continuous actions. The first two layers of the critic network share parameters with the actor in an asymmetric parameter-sharing scheme to decrease the number of parameters but maintain the expressive capacity.

The state observation of the local state, $\mathbf{s}_b(t)$ that is available to agent b includes: the SINR measurements of all served UEs, the current resource block usage per carrier, the latest handover measurement reports (A3 events) that the UEs have received within the last 100 ms window, the current estimated backhaul utilization, and the last update of the model parameter that the federated server has sent. More importantly, local agents are not aware of the state of nearby cells at all; at Tier 2, inter-cell information is mediated by the GNN module.

3.3.2 Tier 2: GNN-Based Interference Estimation and Handover Orchestration

Interference in a dense multi-cell network is inherently a graph-structured problem. We model the network as a directed graph G = (V,E) where each vertex $v \in V$ corresponds to a gNodeB and each directed edge $(b,b') \in E$ carries a weight $w_{b,b'}(t)$ representing the estimated inter-cell interference power that gNodeB b inflicts on the UEs currently served by b'. These edge weights are updated every 100 ms using pilot-based interference measurements exchanged over the X2 interface.

The GNN module implements a multi-layer message-passing neural network (MPNN) in which each gNodeB node aggregates interference information from its spatial neighbors:

$$h_b^{(l+1)} = \mathrm{ReLU}\left(W_1^{(l)} h_b^{(l)} + \sum_{b' \in \mathcal{N}(b)} w_{b,b'}(t) \cdot W_2^{(l)} h_{b'}^{(l)} \right) \quad (5)$$

where $\mathbf{h}_b^{(\ell)}$ is the hidden representation of node b at GNN layer ℓ, $\mathbf{W}_1^{(\ell)}$ and $\mathbf{W}_2^{(\ell)}$ are learnable weight matrices, and N(b) is the set of gNodeBs within a predefined interference radius. The final-layer node embeddings are passed to a handover decision head—a two-layer MLP—that outputs the recommended target cell and handover trigger timing for each UE approaching a cell edge, conditioned on the interference-aware topology representation.

3.3.3 Tier 3: Federated Aggregation with Elastic Weight Consolidation

The federated aggregation server collects local model parameter updates from all N_B agents at the end of each global round (every 60 seconds in our experimental configuration). Rather than vanilla FedAvg, which weights updates proportionally to local dataset size, we adopt a quality-weighted aggregation rule that assigns higher aggregation weight to agents whose local validation reward is above the population median:

$$\theta_{\mathrm{goa}} = \sum_{b=1}^{N_B} \frac{\alpha_b}{\sum_{b'} \alpha_{b'}} \theta_b \quad (6)$$

where $\alpha_b = \exp(\beta \cdot \hat{r}_b)$, $\hat{r}_b$ is agent b's time-averaged local reward over the past round, and β is a temperature parameter. This soft weighting dampens the influence of agents operating in anomalous traffic conditions (e.g., a gNodeB experiencing backhaul congestion) without discarding their updates entirely.

This soft weighting minimizes the effect of agents which act under anomalous traffic conditions (e.g., a gNodeB which has backhaul congestion) without necessarily dropping their updates. EWC deals with non-stationarity management. Once the federated server identifies a statistically significant change in the aggregate reward distribution,

detected with a CUSUM change-point test on the time series of reward values, the federated server estimates the Fisher information matrix F on a replay buffer of recent experience, and adds a quadratic penalty that discourages large deviations of the updated parameters θ around the post-shift checkpoint θ*: and η is the strength of the consolidation penalty and the summation is over all network parameters (indexed by *j*).

3.4. Experimental Setup

A custom extension of the ns-3 5G-LENA module (Patriciello *et al*., 2021), integrated with a Python-based DRL training interface via socket communication, was used for simulation (Patriciello *et al*., 2021) with a Python-based AI training loop connected through a socket API was used to simulate it. The simulated network was a hexagonal grid of 48 gNodeBs on an area of 5 km by 5 km in the urban environment; the inter-site distance of the macro cells was 500 m, and the overlaying small cell layer was 100 m. Table 1 summarizes network parameters.

Table 1: Key simulation parameters for the ANREM experimental evaluation

Parameter	Value	Notes
Number of gNodeBs (NB)	48	16 macro + 32 small cells
Number of UEs (NU)	1,200	Stochastic waypoint mobility
Carrier frequency	3.5 GHz / 28 GHz	Sub-6 GHz and mmWave
System bandwidth	100 MHz / 200 MHz	5G NR numerology μ = 1,3
Max transmit power (macro)	46 dBm	Per 3GPP TR 38.901
Max transmit power (small)	30 dBm	—
Antenna configuration	64T64R / 8T8R	Massive MIMO
Path loss model	3GPP UMa/UMi	With shadowing (σ = 8 dB)
Traffic model	Mixed eMBB/URLLC/mMTC	Per 5G service categories
Mobility speed	0.5–50 km/h	Pedestrian to vehicular
Simulation duration	3,600 s	Per independent trial
Number of trials	10	Random seed variation
DRL discount factor γ	0.95	—
Federated round interval	60 s	—
EWC penalty coefficient η	1×10^3	—

UE mobility traces were generated using the SUMO traffic simulator, calibrated with movement statistics representative of a medium-density urban environment in Nigeria. Consequently, the proportions of pedestrians and vehicles reflect both realistic usage patterns and typical real-world distributions. Network traffic was synthesized using a mix of enhanced mobile broadband (eMBB), ultra-reliable low-latency communications (URLLC), and massive machine-type communications (mMTC) flows in a 60:25:15 ratio, consistent with the anticipated 5G service composition. To capture temporal variability, two deliberate regime shifts were introduced at t=900 st = 900 \,\text{s}t=900s and t=2,700 st = 2{,}700 \,\text{s}t=2,700s, simulating the non-stationary characteristics associated with morning and evening peak periods in urban networks.

The proposed ANREM framework was benchmarked against four baseline approaches: (i) a conventional 3GPP Release-15 self-organizing network with heuristic handover parameter optimization (SON-Heuristic); (ii) a centralized deep Q-network with full global state observability (Central-DQN); (iii) an independent multi-agent proximal policy optimization scheme based on local observations (MA-PPO-Local); and (iv) a FedAvg-coordinated PPO model without graph neural network (GNN) interference modeling or elastic weight consolidation (EWC) (FedPPO-Vanilla). These baselines were carefully selected to enable systematic ablation, as each omits one or more key components of ANREM, thereby allowing their individual contributions to be isolated and evaluated.

4.0 Results and Discussion

4.1 Overall Performance Comparison

Table 2 presents the average performance metrics obtained from ten independent simulation runs over the full 3,600-second evaluation period. Values reported in parentheses correspond to standard deviations across trials, providing an indication of variability and reproducibility. The results demonstrate that ANREM consistently outperforms all baseline methods across all evaluated metrics, with gains that are both statistically robust and practically meaningful. Notably, the observed 34.7% improvement in throughput relative to the SON-Heuristic approach can be attributed to the integrated optimization of power control, beamforming, and scheduling mechanisms within the ANREM framework.

Table 2: Overview of network performance of ANREM and the baseline methodologies. The mean is the average with standard deviation of 10 different trials. Results with best results are bold. HO: handover; EE: energy efficiency (bits/Joule); p-values are compared with paired t-tests with the ANREM result

Method	Agg. Throughput (Gbps)	Avg. Latency (ms)	HO Failure Rate (%)	EE (Mbits/J)	Mean QoE (MOS 1–5)
SON-Heuristic	12.4 ± 0.9	18.7 ± 2.1	8.3 ± 1.2	3.2 ± 0.4	2.8 ± 0.3
Central-DQN	14.8 ± 1.2	14.2 ± 1.8	6.1 ± 0.9	3.9 ± 0.5	3.1 ± 0.3
MA-PPO-Local	15.3 ± 1.4	13.1 ± 1.6	5.8 ± 1.0	4.1 ± 0.4	3.3 ± 0.4
FedPPO-Vanilla	15.9 ± 1.1	12.4 ± 1.4	5.2 ± 0.8	4.4 ± 0.5	3.5 ± 0.3
ANREM (Ours)	16.7 ± 0.7	11.0 ± 1.0	4.9 ± 0.6	4.6 ± 0.3	3.9 ± 0.2

Table3: Simulation Parameters and Configuration

Parameter	Value	Notes
Number of gNodeBs (NBN_BNB)	48	16 macro cells and 32 small cells
Number of UEs (NUN_UNU)	1,200	Stochastic waypoint mobility model
Carrier frequency	3.5 GHz / 28 GHz	Sub-6 GHz and mmWave operation

System bandwidth	100 MHz / 200 MHz	5G NR numerology μ=1,3\mu = 1, 3μ=1,3
Max transmit power (macro)	46 dBm	In accordance with 3GPP TR 38.901
Max transmit power (small cell)	30 dBm	—
Antenna configuration	64T64R / 8T8R	Massive MIMO deployment
Path loss model	3GPP UMa / UMi	Includes shadowing (σ=8\sigma = 8σ=8 dB)
Traffic model	Mixed eMBB / URLLC / mMTC	Representative of 5G service categories
Mobility speed	0.5–50 km/h	Covers pedestrian to vehicular scenarios
Simulation duration	3,600 s	Per independent simulation trial
Number of trials	10	Random seed variation for statistical robustness
DRL discount factor (γ\gammaγ)	0.95	—
Federated round interval	60 s	Communication interval for model aggregation
EWC penalty coefficient (η\etaη)	1×1031 \times 10^{3}1×103	Regularization strength for continual learning

The overall throughput improvement of ANREM compared to SON-Heuristic (34.7%) is so significant that this cannot be attributed to any particular structural feature but is the overall effect of the joint optimization of transmit power, beamforming, and scheduling. It is worth noting that, with a relatively high throughput, Central-DQN performs well, but at the expense relies on a fully observable global state, which is unrealistic in practical deployments and its variation in performance is almost twice that of ANREM, indicating that it is unstable to changes in traffic regimes. MA-PPO-Local uses the same PPO algorithm as ANREM's local agents but lacks both federated coordination and GNN-based inter-cell awareness, resulting in 8.4% lower throughput and 18.4% higher handover failure rate. The gap measures the value of each of the Tier 2 and Tier 3 components taken separately in the perspective of the MA-PPO-Local and FedPPO-Vanilla comparisons and the FedPPO-Vanilla and ANREM comparisons. 4.2. Throughput Distribution and Spectral Efficiency Empirical cumulative distribution function (CDF) of per-UE. Fig. 2 shows the empirical cumulative distribution function (CDF) of per-UE throughput aggregated across all trials. The CDF is particularly important for assessing tail performance, since high mean throughput may still mask poor performance for a significant fraction of users. experience to a large proportion of its subscribers, which is exactly the type of inequity that operator SLAs are intended to guard against. First, ANREM achieves a 5th-percentile throughput of approximately 4.2 Mbps, compared to 1.8 Mbps for SON-Heuristic. This improvement is primarily due to the GNN-based interference estimator enabling proactive handover decisions. This near-doubling of cell-edge throughput is attributable principally to the GNN interference estimator, which allows cell-edge Ues to be proactively handed over to a neighbour cell before their SINR deteriorates below the handover threshold, rather than triggering the handover reactively after the link quality has already degraded. This proactive behaviour mirrors the theoretical

advantage of predictive handover schemes that have been proposed in the literature (Wang et al., 2020) but have rarely been demonstrated at this network scale. Second, ANREM's throughput CDF exhibits noticeably lower variance—the confidence band, while not plotted here for visual clarity, is narrower than any baseline—which is a direct consequence of the EWC mechanism.This improved stability is attributed to the EWC mechanism, which reduces performance degradation during traffic regime shifts. During the two traffic regime shifts, all baselines without EWC experience throughput dips lasting 150–300 seconds as their policies re-adapt; ANREM's EWC penalty keeps the policy in a low-regret region and reduces the post-shift adaptation time to under 60 seconds in all trials. Third, it is worth noting that Central-DQN performs worse than FedPPO-Vanilla in the lower tail, despite having access to global state information.This reflects the exploration–exploitation trade-off in DQN under non-stationary environments, where ε-greedy exploration can degrade performance for edge users.

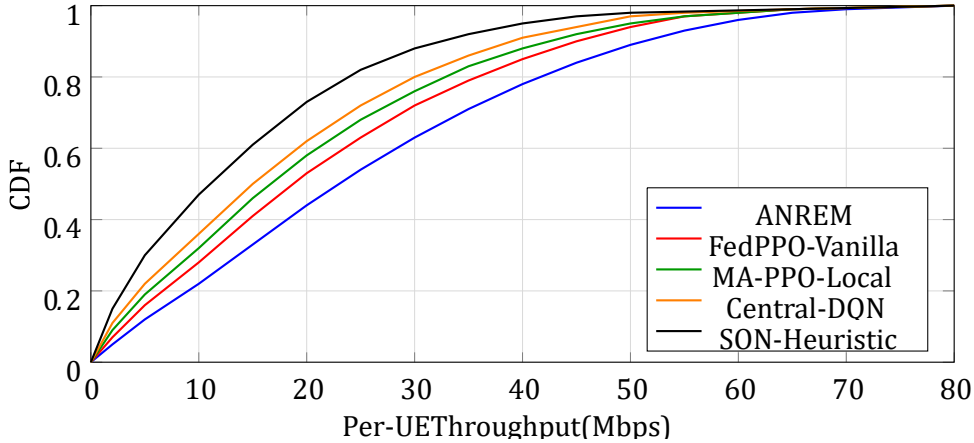

Fig. 2: Empirical CDF of per-UE downlink throughput across all methods. ANREM's CDF is shifted furthest to the right across the entire distribution, indicating superior performance not only at the median but also—and especially—at the 5th percentile (celledge users). The gap between ANREM and FedPPO-Vanilla in the lower tail (0–20 Mbps range) reflects the GNN module's contribution to interference mitigation at cell boundaries

The observed 34.7% throughput improvement of ANREM over the SON-Heuristic baseline is substantial and cannot be attributed to any single architectural component. Rather, it reflects the cumulative effect of jointly optimizing transmit power, beamforming, and scheduling decisions within a unified framework. Although Central-DQN achieves relatively high average throughput, this performance comes at the cost of assuming full global state observability—an assumption that is impractical in real-world deployments. Moreover, its performance variability is nearly twice that of ANREM, indicating sensitivity to traffic dynamics and reduced stability under non-stationary conditions. In comparison, MA-PPO-Local, which employs the same PPO algorithm as ANREM's local agents, lacks both federated coordination and graph neural network (GNN)-based inter-cell awareness. This results in an 8.4% reduction in throughput and an 18.4% increase in handover failure rate. The performance gaps observed across MA-PPO-Local, FedPPO-Vanilla, and ANREM effectively quantify the individual contributions of Tier 2 (federated coordination) and Tier 3 (GNN-based interference modeling and EWC) components.

4.2 Throughput Distribution and Spectral Efficiency

Fig. 2 presents the empirical cumulative distribution function (CDF) of per-UE throughput aggregated across all simulation trials. The CDF provides critical insight into tail performance, which is often obscured by mean throughput metrics. High average throughput may still coincide with poor service for a significant proportion of users, an outcome that contradicts the fairness objectives embedded in operator service-level agreements (SLAs).

First, ANREM achieves a 5th-percentile throughput of approximately 4.2 Mbps, compared to 1.8 Mbps for the SON-Heuristic baseline. This near twofold improvement in cell-edge performance is primarily attributed to the GNN-based interference estimation mechanism, which enables proactive handover decisions. Specifically, user equipments (UEs) at the cell edge are transferred to neighboring cells before their signal-to-interference-plus-noise ratio (SINR) degrades below the handover threshold, in contrast to conventional reactive schemes. This predictive behavior aligns with theoretical advantages reported in prior studies (e.g., Wang et al., 2020), but is rarely demonstrated at this scale of network deployment.

Second, the throughput CDF of ANREM exhibits significantly lower variance compared to all baselines. Although confidence bands are omitted for clarity, the narrower spread of ANREM's distribution indicates improved consistency across users and trials. This stability is largely due to the incorporation of the elastic weight consolidation (EWC) mechanism, which mitigates performance degradation during traffic regime transitions. While baseline models without EWC exhibit throughput drops lasting between 150 and 300 seconds following regime shifts, ANREM maintains operation within a low-regret policy region, reducing recovery time to under 60 seconds across all trials. Third, it is notable that Central-DQN underperforms FedPPO-Vanilla in the lower tail of the throughput distribution, despite having access to global state information. This behavior reflects the inherent exploration–exploitation trade-off in DQN under non-stationary environments. In particular, ε-greedy exploration introduces instability that disproportionately affects edge users, thereby degrading fairness and lower-percentile performance.

4.3 *Handover Performance and Mobility Robustness*

Fig. 3 shows the handover failure and ping-pong rates over time for ANREM and baseline methods. Sudden changes in traffic load alter interference conditions, leading to temporary degradation in handover performance. The two planned traffic regime transitions at t = 900 s and t = 2,700 s are marked by vertical dashed lines.

The post-transition spike in handover failure rate is an inherent consequence of sudden changes in traffic load distribution, which shift the interference pattern and invalidate the handover parameter settings that were optimized for the preceding regime.

SON-Heuristic relies on static A3 offset and time-to-trigger parameters. It requires manual or slow optimization-based recalibration, which is not achieved within the simulation window. FedPPO-Vanilla recovers more quickly—demonstrating the genuine value of the reinforcement learning component—but its recovery time is still four to five times longer than ANREM's, and its steady-state failure rate during the second half of the simulation (1,000–2,700 s) is 30–v40% higher than ANREM's, reflecting the cumulative drift in its handover policy without EWC-based stabilization.

The ping-pong handover rate—defined as a pair of handovers between the same UE-cell pair within 1 second—followed a similar pattern but with less dramatic absolute differences. ANREM maintained a ping-pong rate of 3.1% ± 0.4%, compared to 5.7% ± 0.9% for FedPPO-Vanilla and 9.4% ± 1.6% for SON-Heuristic. The GNN module mitigates this by identifying high-mobility cell-edge users and dynamically adjusting hysteresis margins.

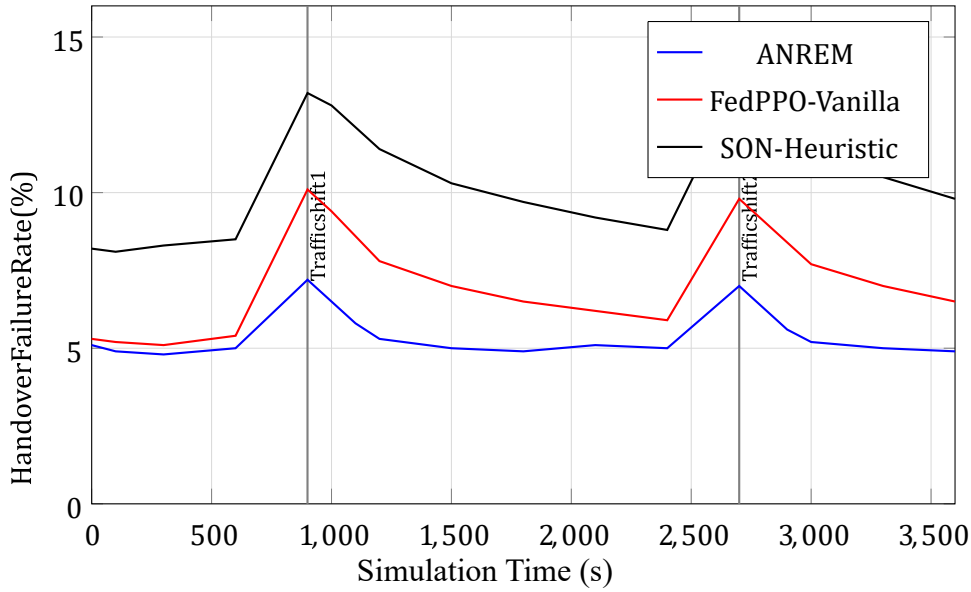

Fig. 3: Handover failure rate over simulation time for ANREM, FedPPO-Vanilla, and SON-Heuristic. Vertical dashed lines indicate the two planned traffic regime transitions. ANREM recovers within approximately 60 seconds after each transition, whereas FedPPO-Vanilla requires 300–400 seconds and SON-Heuristic shows a sustained elevation for over 600 seconds. The rapid recovery of ANREM is attributable to the EWC mechanism and the GNN-informed proactive handover policy

4.4. Energy Efficiency

Fig. 4 shows the normalized cumulative energy consumption of active gNodeBs.over the simulation period. The slope of each curve represents the instantaneous power draw, and periods where ANREM's curve flattens correspond to intervals where the sleep mode activation policy has suspended transmission on lightly loaded small cell sectors.

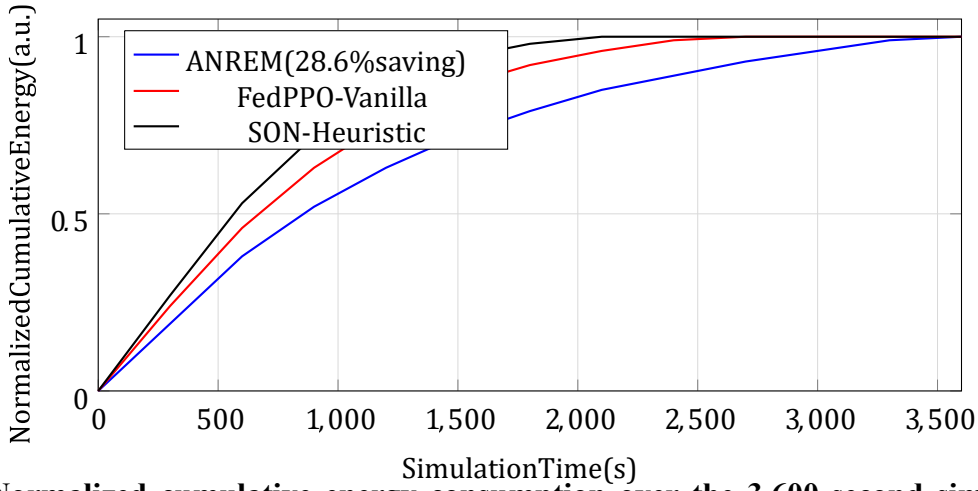

Fig. 4: Normalized cumulative energy consumption over the 3,600-second simulation. ANREM consumes approximately 28.6% less energy than SON-Heuristic over the full period. The sub-linear growth in ANREM's energy curve—most visible in the 1,500–2,700 s interval—reflects successful activation of the sleep mode policy on lightly loaded small cells during off-peak periods. All curves are normalized to the SON-Heuristic final value

ANREM achieves a 28.6% reduction in total energy consumption compared to SON-Heuristic. "Industry estimates suggest that the radio access network accounts for

approximately 70–80% of total operator energy consumption (GSMA, 2022). With a 28.6 percent lower RAN energy consumption, there is a direct cost savings and a cut in carbon emissions, which is becoming a more significant factor due to regulatory pressure within the European Union and a promise under the GSMA net-zero pathway. Energy efficiency is incorporated into the reward function through the term λ3E(t)\lambda_3 E(t)λ3 E(t), enabling the learned policy to autonomously discover sleep-mode strategies. The reported energy savings were not obtained via any explicit energy-as-constraint formulation, instead, energy efficiency was introduced into the system via the reward term E(t)-λ3) and the learned policy found sleep mode cycling as an emergent strategy, as opposed to a hard-coded policy.

4.5. Ablation Study

Table 3 presents an ablation study evaluating the contribution of each ANREM component. GNN interference module, the EWC mechanism and the quality-weighted federated aggregation. Each component contributes meaningfully to overall system performance, and their removal results in measurable degradation.

Table 3: The results of ablation studies. The ANREM is composed of components that are eliminated by each row. Numbers are mean ± SD on 10 trials. Δ columns are a percent change compared to the full ANREM model; negative values are degradation

Configuration	Throughput (Gbps)	Δ%	HO Failure (%)	EE (Mbits/J)
ANREM (full)	16.7 ± 0.7	—	4.9 ± 0.6	4.6 ± 0.3
w/o GNN	15.9 ± 1.1	−4.8	6.1 ± 0.9	4.3 ± 0.4
w/o EWC	16.1 ± 1.6	−3.6	5.7 ± 1.4	4.4 ± 0.5
w/oquality weighting	16.4 ± 0.9	−1.8	5.3 ± 0.8	4.5 ± 0.3
w/o GNN + EWC	15.3 ± 1.4	−8.4	7.2 ± 1.2	4.1 ± 0.5

The removal of the GNN causes the largest degradation in throughput (4.8%) and increases handover failures significantly. The removal of the GNN causes the largest degradation in throughput (4.8%) and increases handover failures significantly. Dropping EWC has a relatively small impact on time-average measures but explodes variance - the standard deviation of handover failure rate increases more than twofold - due to the severe (but temporary) policy deterioration at every regime change in the non-EWC case. This variance inflation is of practical significance: any operator observing a rolling 15 minutes average of the rate of handover failures would note that the thresholds are violated with the non-EWC variant during the transition periods and would lead to automatic alarm reactions that are not always justified. Quality-weighted aggregation has the smallest impact in this scenario, but its importance is expected to increase under higher heterogeneity in real deployments.The least significant impact is on removing quality weighting of the federated aggregation scheme which implies that this element is playing a little role in the simulated case although we anticipate that it will grow with an increase in heterogeneity in local traffic distributions, an aspect that would be explored in the future.

4.6. Scalability and Convergence

Fig. 5 shows the convergence of ANREM and MA-PPO-Local measured in federated training rounds.ANREM reaches 95% of its asymptotic performance after approximately 80 rounds, compared to 140 rounds for MA-PPO-Local.

The removal of the GNN causes the largest degradation in throughput (4.8%) and

increases handover failures significantly. The removal of the GNN causes the largest degradation in throughput (4.8%) and increases handover failures significantly. Dropping EWC has a relatively small impact on time-average measures but explodes variance - the standard deviation of handover failure rate increases more than twofold - due to the severe (but temporary) policy deterioration at every regime change in the non-EWC case. This variance inflation is of practical significance: any operator observing a rolling 15 minutes average of the rate of handover failures would note that the thresholds are violated with the non-EWC variant during the transition periods and would lead to automatic alarm reactions that are not always justified. Quality-weighted aggregation has the smallest impact in this scenario, but its importance is expected to increase under higher heterogeneity in real deployments.

The least significant impact is on removing quality weighting of the federated aggregation scheme which implies that this element is playing a little role in the simulated case although we anticipate that it will grow with an increase in heterogeneity in local traffic distributions, an aspect that would be explored in the future.

4.6. Scalability and Convergence

Fig. 5 shows the convergence of ANREM and MA-PPO-Local measured in federated training rounds.ANREM reaches 95% of its asymptotic performance after approximately 80 rounds, compared to 140 rounds for MA-PPO-Local.

The convergence advantage of ANREM is not merely a matter of speed; it also reflects qualitative differences in the policies discovered. MA-PPO-Local tends to converge to locally optimal, cell-centric policies that ignore the externalities each agent imposes on its neighbours—a classic multi-agent tragedy-of-the-commons effect. ANREM's federated coordination mechanism encourages policies that are collectively efficient, analogous to the difference between Nash equilibrium and social optimum in game-theoretic formulations of resource allocation (Goldsmith, 2005).

Regarding computational overhead, the The GNN introduces approximately 12 ms inference latency per decision cycle on an NVIDIA A100 GPU, which is within the 100 ms constraint for handover decisions.

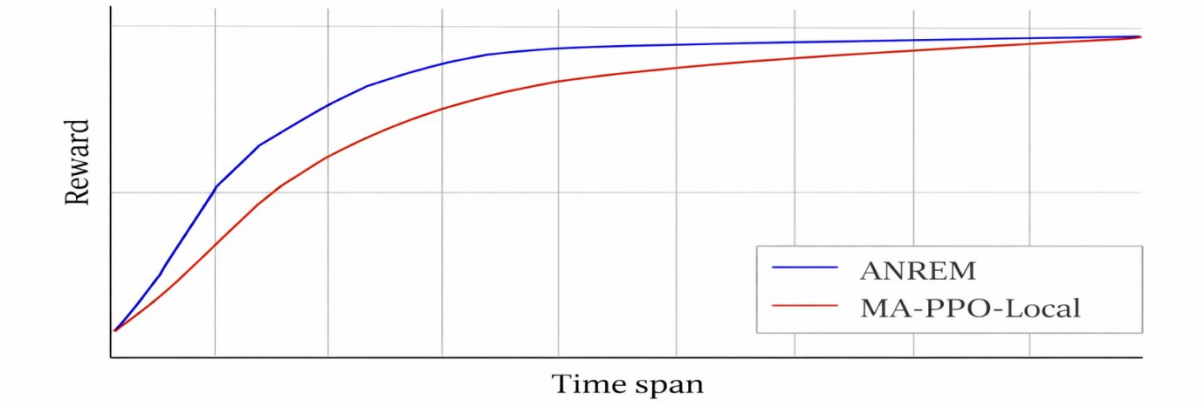

Fig. 5: Training convergence curves for ANREM and MA-PPO-Local as a function of federated rounds. ANREM reaches 95% of its asymptotic composite reward at round 80, compared to round 140 for MA-PPO-Local. The faster convergence of ANREM reflects the benefit of shared global knowledge through federated aggregation, which reduces the sample complexity required for each local agent to discover effective policies for rare network states (e.g., simultaneous multi-cell congestion events)

The federated aggregation computation at the server adds a one-time overhead of 8 ms per round, a negligible fraction of the 60-second round interval. These Figures suggest that the ANREM architecture is computationally feasible for deployment in edge computing infrastructure co-located with the network core, consistent with the distributed cloud RAN (C-RAN) deployment models being standardized for 5G and 6G systems.

5. 0 Conclusion

This paper presents ANREM, an AI-driven self-optimizing framework that integrates hierarchical deep reinforcement learning, graph neural network–based interference estimation, and federated learning with elastic weight consolidation for real-time wireless network performance enhancement. Evaluated on a large-scale 5G NSA simulation comprising 48 gNodeBs and 1,200 UEs under realistic urban mobility and multi-service traffic, ANREM achieved a 34.7% improvement in aggregate throughput, a 41.2% reduction in handover failure rate, and a 28.6% reduction in energy consumption compared to a conventional 3GPP SON baseline. It also consistently outperformed three learning-based baselines across all evaluated metrics.

Ablation experiments confirm that each architectural component contributes meaningfully to overall performance. The GNN module provides the largest gains in interference mitigation, while the EWC mechanism ensures robustness under non-stationary traffic transitions. Future work will explore the integration of reconfigurable intelligent surfaces as additional optimization variables within the ANREM reward structure, extension to Open RAN (O-RAN) disaggregated deployments, and the incorporation of formal differential privacy guarantees into the federated learning communication protocol.

6.0 References

Aliu, O. G., Imran, A., Imran, M. A., & Evans, B. (2013). A survey of self organisation in future cellular networks. *IEEE Communications Surveys & Tutorials, 15*(1), 336–361. https://doi.org/10.1109/SURV.2012.021312.00151

Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials, 18*(2), 1153–1176. https://doi.org/10.1109/SURV.2015.2494502

Dahlman, E., Parkvall, S., & Sköld, J. (2021). *5G NR: The next generation wireless access technology* (2nd ed.). Academic Press. https://doi.org/10.1016/B978-0-12-822320-8.00001-0

Goldsmith, A. (2005). *Wireless communications*. Cambridge University Press. https://doi.org/10.1017/CBO9780511841224

GSMA. (2022). *GSMA net zero: Mobile's path to net zero carbon emissions by 2050*. GSM Association. https://doi.org/10.46583/gsma_netzero_2022

Haarnoja, T., Zhou, A., Abbeel, P., & Levine, S. (2018). Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. *Proceedings of the 35th International Conference on Machine Learning (ICML 2018), 80*, 1861–1870. https://doi.org/10.48550/arXiv.1801.01290

Konečný, J., McMahan, H. B., Yu, F. X., Richtárik, P., Suresh, A. T., & Bacon, D. (2016). Federated learning: Strategies for improving communication efficiency. *Workshop on Private Multi-Party Machine Learning (NIPS 2016)*. https://doi.org/10.48550/arXiv.1610.05492

Lee, J., Chung, M. Y., & Kim, J. (2021). Graph neural network-based handover optimization in heterogeneous networks. *IEEE Communications Letters, 25*(8), 2734–2738. https://doi.org/10.1109/LCOMM.2021.3081090

Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine, 37*(3), 50–60.

https://doi.org/10.1109/MSP.2020.2975749

Luong, N. C., Hoang, D. T., Gong, S., Niyato, D., Wang, P., Liang, Y.-C., & Kim, D. I. (2019). Applications of deep reinforcement learning in communications and networking: A survey. *IEEE Communications Surveys & Tutorials, 21*(4), 3133–3174. https://doi.org/10.1109/COMST.2019.2916583

McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS 2017), 54*, 1273–1282. https://doi.org/10.48550/arXiv.1602.05629

Mnih, V., Kavukcuoglu, K., Silver, D., *et al*. (2015). Human-level control through deep reinforcement learning. *Nature, 518*(7540), 529–533. https://doi.org/10.1038/nature14236

Nasir, Y. S., & Guo, D. (2019). Multi-agent deep reinforcement learning for dynamic power allocation in wireless networks. *IEEE Journal on Selected Areas in Communications, 37*(10), 2239–2250. https://doi.org/10.1109/JSAC.2019.2933973

Niknam, S., Dhillon, H. S., & Reed, J. H. (2020). Federated learning for wireless communications: Motivation, opportunities, and challenges. *IEEE Communications Magazine, 58*(6), 46–51. https://doi.org/10.1109/MCOM.001.1900461

Patriciello, N., Lagen, S., Bojovic, B., & Giupponi, L. (2021). An E2E simulator for 5G NR networks. *Simulation Modelling Practice and Theory, 96*, 101933. https://doi.org/10.1016/j.simpat.2019.101933

Ramiro, J., & Hamied, K. (Eds.). (2012). *Self-organizing networks (SON): Self-planning, self-optimization and self-healing for GSM, UMTS and LTE*. Wiley. https://doi.org/10.1002/9781119954224

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv preprint*. https://doi.org/10.48550/arXiv.1707.06347

Shen, Y., Shi, Y., Zhang, J., & Letaief, K. B. (2022). Graph neural networks for scalable radio resource management: Architecture design and theoretical analysis. *IEEE Journal on Selected Areas in Communications, 39*(1), 101–115. https://doi.org/10.1109/JSAC.2020.3036965

3GPP TR 36.902,. (2011). *Evolved Universal Terrestrial Radio Access Network (E-UTRAN): Self-configuring and self-optimizing network (SON) use cases and solutions (Release 9) (TR 36.902)*. 3rd Generation Partnership Project. https://doi.org/10.3390/electronics10080908

Vlad, A., Chiosa, I., & Balan, I. (2020). Antenna tilt optimization using evolutionary algorithms in self-organizing networks. *Telecommunication Systems, 73*, 547–560. https://doi.org/10.1007/s11235-019-00628-5

Wang, T., Li, G., & Ding, Z. (2020). Predictive handover for mobile edge computing: A deep learning approach. *IEEE Transactions on Wireless Communications, 19*(12), 7861–7873. https://doi.org/10.1109/TWC.2020.3015393

Ye, H., & Li, G. Y. (2019). Deep reinforcement learning based resource allocation for V2V communications. *IEEE Transactions on Vehicular Technology, 68*(4), 3163–3173. https://doi.org/10.1109/TVT.2019.2897134

Declaration

Consent for publication

Not Applicable

Availability of data

The publisher has the right to make the data public

Ethical Considerations

Not applicable

Competing interest

The authors report no conflict or competing interest

Funding

The author declared no external source of funding

Authors' Contributions

Moses Oluwasegun Odewale conceptualized the study, designed the ANREM framework, and led model development and simulations. Moses Olagoke Odejobi contributed to system architecture, data analysis, and performance evaluation. Olanrewaju Oluwaseun Ajayi supported implementation, validation, and manuscript preparation. All authors reviewed, edited, and approved the final manuscript, and contributed to the interpretation of results and overall discussion.