Deep Learning for Spectrum Allocation in 5G Networks: A Python-Based Experimental Study Considering Regulatory Policies

Mohammed Adamu Sule[1], Benjamin Abba-Stephen[2], Maryam Abdulkadir[3], Denis D. Jonathan[4], Mujittapha Idris[5], Shehu Abdullahi[6]

[1, 3, 4] Department of Computer Engineering Technology, Federal Polytechnic Kaltungo, Nigeria.
[2, 5,6] Department of Electrical/Electronic Engineering Technology, Federal Polytechnic Kaltungo, Nigeria.

ABSTRACT

The efficient allocation of radio spectrum is a critical challenge in fifth-generation (5G) wireless networks, where dynamic and heterogeneous traffic demands require intelligent resource management solutions that simultaneously satisfy technical performance objectives and regulatory compliance constraints. This paper presents an experimental study of deep learning-based spectrum allocation techniques applied to 5G network environments, implemented and evaluated entirely in Python using Tensor Flow 2.x and the Open AI Gym framework. Three deep reinforcement learning (DRL) architectures: Deep Q-Network (DQN), Proximal Policy Optimization (PPO), and a novel hybrid Convolutional Neural Network–Long Short-Term Memory (CNN-LSTM) models are trained and evaluated on a custom-built 5G network simulation environment incorporating regulatory constraints derived from ETSI EN 301 893, ITU-R M.2150 (IMT-2020), and FCC Part 27. The proposed CNN-LSTM model achieves a mean throughput of 487 Mbps, representing a 56.1% improvement over baseline DQN, a 21.4% improvement over PPO, and a latency reduction to 9.7 Ms under full regulatory compliance conditions. Critically, the study quantifies the performance-compliance trade-off for the first time in a 5G DRL context, demonstrating that full regulatory compliance introduces an 18.3% throughput penalty compared to unconstrained operation, a finding with significant implications for spectrum policy design. The experimental framework, training datasets, and regulatory constraint modules are made publicly available to support reproducibility.

KEYWORDS

5G networks; spectrum allocation; deep reinforcement learning; DQN; CNN-LSTM; regulatory compliance; ETSI; ITU-R; Python; Tensor Flow

I. INTRODUCTION

The fifth generation (5G) of mobile telecommunications networks represents a paradigm shift in wireless communication engineering, introducing ultra-reliable low-latency communication (URLLC), enhanced mobile broadband (eMBB), and massive machine-type communication (mMTC) service classes that demand unprecedented spectrum efficiency (3GPP, 2022; Agiwal et al., 2016). The International Telecommunication Union (ITU) has identified peak data rates of 20 Gbps, user-experienced data rates of 100 Mbps, and end-to-end latency of 1 ms as benchmark requirements for IMT-2020 systems (ITU-R, 2017). Achieving these targets within the constraints of finite radio spectrum resources and

Complex co-existence requirements represent one of the foremost technical challenges in telecommunications engineering.

Traditional spectrum allocation methods including static frequency assignment, cognitive radio-based dynamic spectrum access, and game-theoretic approaches have demonstrated limitations in adapting to the rapid temporal variability and spatial heterogeneity of 5G traffic patterns. Deep reinforcement learning (DRL) has emerged as a promising alternative, enabling spectrum management agents to learn near-optimal allocation policies from environmental interaction without requiring explicit system models. Several DRL architectures including Deep Q-Networks (DQN), Actor-Critic methods, and Proximal Policy Optimization (PPO) have demonstrated superior performance over conventional algorithms in single-domain spectrum management scenarios.

However, a critical dimension largely absent from existing DRL-based spectrum allocation research is the explicit incorporation of regulatory policy constraints. Radio spectrum allocation in 5G networks is governed by a complex hierarchy of international, regional, and national regulatory frameworks. The ITU Radio Regulations establish global frequency allocations and coordination procedures (ITU, 2020). At the regional level, the European Telecommunications Standards Institute (ETSI) issues harmonized standards such as EN 301 893 for 5 GHz WLAN and EN 303 687 for IMT-2020 systems (ETSI, 2021a, 2021b). National regulators including the Federal Communications Commission (FCC) in the United States, Ofcom in the United Kingdom, and the Nigerian Communications Commission (NCC) in Nigeria impose additional licensing conditions, power limits, and interference protection requirements.

This paper addresses this gap through a Python-based experimental study that explicitly integrates regulatory constraints into the DRL training and evaluation pipeline for 5G spectrum allocation.

The specific contributions of this work are:

- A custom Open-ai Gym-compatible simulation environment for 5G spectrum allocation incorporating regulatory constraints from ETSI EN 301 893, ITU-R M.2150, and FCC Part 27;
- Implementation and comparative evaluation of DQN, PPO, and a proposed CNN-LSTM hybrid architecture under both unconstrained and regulatory-compliant conditions;
- Quantification of the performance-compliance trade-off across five regulatory scenarios; and (4) an analysis of the implications of these findings for spectrum policy design and 5G network engineering practice.

II. RELATED WORK

A. Deep Reinforcement Learning for Spectrum Management

The application of DRL to wireless resource management has grown substantially since the landmark work of Mnih et al. (2015), who demonstrated that a DQN agent could achieve human-level performance on Atari games using raw pixel inputs. Early applications to spectrum management focused on cognitive radio scenarios, where a secondary user must

identify and exploit spectrum opportunities without interfering with primary users (Luong et al., 2019; Wang et al., 2018). DQN-based approaches achieved near-optimal dynamic spectrum access performance in single-agent settings but exhibited scalability limitations in multi-agent environment.

Multi-agent DRL (MADRL) frameworks have been proposed to address scalability in heterogeneous 5G networks. Zhang et al. (2023) demonstrated that a multi-agent PPO system outperformed centralized DQN by 23% in throughput and 31% in fairness metrics in a simulated urban 5G macro-small cell network. Sun et al. (2022) proposed a Graph Neural Network (GNN)-augmented DRL architecture that achieved efficient spectrum sharing in dense network deployments with up to 500 base stations. Notably, neither of these studies incorporated explicit regulatory compliance constraints, a gap the present work addresses.

B. Regulatory Constraints in Spectrum Allocation

The integration of regulatory constraints into spectrum optimization algorithms has received comparatively limited attention in the literature. Hossain et al. (2020) proposed a constrained Markov decision process (CMDP) formulation for cognitive radio spectrum access that incorporated interference temperature limits derived from ITU recommendations. Their Lagrangian relaxation approach achieved 94% compliance with interference constraints at a 12% throughput cost. Liang et al. (2019) examined the impact of transmit power regulations on energy-efficient spectrum sharing in heterogeneous networks, demonstrating that power cap constraints can reduce spectral efficiency by 8–22% depending on network topology.

Most closely related to the present work, Challita et al. (2018) applied DRL to licensed-assisted access (LAA) in LTE-U networks, incorporating ETSI listen-before-talk (LBT) requirements as hard constraints in the DQN training loop. Their results demonstrated that constraint-aware DRL outperformed unconstrained DRL by 18% in co-existence fairness while maintaining 91% LBT compliance. The present study extends this line of work to 5G IMT-2020 environments, considers a broader range of regulatory frameworks, and introduces the CNN-LSTM architecture as a novel contribution.

C. Python Frameworks for 5G Network Simulation

Several Python-based simulation frameworks have been developed for 5G network research. The ns-O-RAN framework (Polese et al., 2022) provides a high-fidelity Open RAN simulation environment compatible with Tensor Flow and Torch-based DRL agents. Sionna (Hoydis et al., 2022), developed by NVIDIA, offers GPU-accelerated physical layer simulation with automatic differentiation support. The 5G-air-simulator (Martiradonna et al., 2020) provides system-level simulation capabilities including scheduling, interference modelling, and mobility management. The present study employs a custom Open AI Gym environment built on top of NumPy and SciPy, prioritizing transparency and regulatory constraint modularity over physical layer fidelity.

III. METHODOLOGY

A. System Model and Regulatory Framework

1. Network architecture

The experimental system models a heterogeneous 5G network consisting of one macro base station (MBS) and eight small cell base stations (SBS) deployed in a 1 km × 1 km service area. The system operates in the n78 band (3.3–3.8 GHz), consistent with the primary mid-band 5G deployment frequency in Europe and West Africa (3GPP, 2022; NCC, 2022). The available spectrum of 100 MHz is divided into 100 resource blocks (RBs) of 1 MHz each, following 3GPP NR numerology μ = 1 with 30 kHz subcarrier spacing (3GPP, 2022). A total of 40 user equipment (UE) devices are distributed uniformly across the service area, generating traffic according to a Poisson arrival process with mean arrival rate λ ∈ {10, 20, and 30} requests/slot.

Channel modelling employs the 3GPP 38.901 Urban Macro (UMa) path loss model with log-normal shadowing (σ = 8 dB) and Rayleigh fast fading. Inter-cell interference is computed at each time step based on the current allocation of RBs and transmit powers across all base stations. The simulation operates in discrete time steps of 1 ms, corresponding to one 5G NR slot at μ = 1, with episodes of 1,000 steps used for training and 500 steps for evaluation.

2. *Regulatory constraint framework*

Three regulatory frameworks are operationalized as hard and soft constraints in the simulation environment. Table 1 summarizes the key regulatory parameters incorporated.

Framework	**Constraint Type**	**Parameter**	**Limit Value**	**Enforcement**
ETSI EN 301 893	Transmit Power	Max EIRP	23 dBm (indoor) 30 dBm (outdoor)	Hard constraint
ETSI EN 301 893	Dynamic Freq. Select.	Radar detection thresh.	-62 dBm	Hard constraint
ITU-R M.2150	Spectral Efficiency	Min DL SE target	3 bits/s/Hz (avg.)	Soft (penalty)
ITU-R M.2150	Latency	User-plane latency	≤ 4 ms (URLLC)	Soft (penalty)
FCC Part 27	Interference Protection	Noise floor limit	-110 dBm/MHz	Hard constraint
FCC Part 27	Guard Band	Min. guard band	5 MHz at band edges	Hard constraint
NCC (Nigeria)	Spectrum Sharing	Co-channel isolation	≥ 25 dB	Soft (penalty)

Hard constraints are enforced by masking illegal actions in the DRL agent's action space, preventing constraint violations from occurring during training and evaluation. Soft constraints are incorporated as penalty terms in the reward function, following the

constrained MDP formulation of Altman (1999). The composite reward function is defined as:

$$r(s,a) = \alpha \cdot R\_throughput(s,a) - \beta \cdot P\_latency(s,a) - \gamma \cdot P\_interference(s,a) - \delta \cdot P\_spectral(s,a)$$

Where α = 0.5, β = 0.2, γ = 0.2, and δ = 0.1 are weighting coefficients determined through grid search cross-validation, and P_* terms represent constraint violation penalties calibrated to the severity of the corresponding regulatory requirement.

B. Deep Learning Architectures

1. Baseline: Deep Q-Network (DQN)

The DQN agent (Mnih et al., 2015) approximates the action-value function Q(s, a; θ) using a fully connected neural network with three hidden layers of dimensions [512, 256, 128], ReLU activations, and a linear output layer of dimensionality equal to the action space size (|A| = 100 RBs × 8 power levels = 800). The state representation encodes current RB occupancy, interference levels, UE traffic demands, and regulatory constraint status as a 240-dimensional vector. Experience replay with a buffer of 50,000 transitions and a target network updated every 100 steps are used to stabilize training.

2. *Proximal Policy Optimization (PPO)*

The PPO agent (Schulman et al., 2017) employs an actor-critic architecture with shared convolutional feature extraction layers followed by separate policy (actor) and value (critic) heads. The clipped surrogate objective with ε = 0.2, entropy bonus coefficient c_2 = 0.01, and generalized advantage estimation (GAE) with λ = 0.95 are used. The actor network outputs a categorical distribution over the action space, enabling stochastic policy exploration (Schulman et al., 2017). PPO's on-policy nature facilitates more stable convergence in the presence of the regulatory constraint masking applied to the action space.

3. Proposed CNN-LSTM Architecture

The proposed CNN-LSTM model introduces temporal feature extraction as a key architectural innovation for 5G spectrum allocation, motivated by the observation that spectrum occupancy patterns exhibit strong temporal autocorrelation across consecutive time slots (Zhang et al., 2023; Sun et al., 2022).

The architecture consists of three stages:

Stage 1: Spatial Feature Extraction: A 1D convolutional module with three convolutional layers (filters: 64, 128, 64; kernel size: 3; stride: 1) processes the instantaneous spectrum occupancy map, capturing local frequency-domain correlation patterns.

Stage 2: Temporal Feature Extraction: The sequence of spatial features over a look-back window of T = 10 time steps is processed by a two-layer LSTM with hidden dimension 256,

capturing temporal dependencies in spectrum dynamics (Hoch Reiter & Schmidhuber, 1997).

Stage 3: Policy and Value Heads: The LSTM output is passed to fully connected actor and critic heads, following the PPO training procedure. An attention mechanism (Vaswani et al., 2017) weights the relative importance of historical time steps in the look-back window, improving sensitivity to periodic interference patterns.

Python Implementation Excerpt (CNN-LSTM Core)

```
import tensorflow as tf
from tensorflow.keras import layers

class CNNLSTM_Actor(tf.keras.Model):
    def __init__(self, action_dim, lookback=10):
        super().__init__()
        self.conv1 = layers.Conv1D(64,  3, activation='relu', padding='same')
        self.conv2 = layers.Conv1D(128, 3, activation='relu', padding='same')
        self.conv3 = layers.Conv1D(64,  3, activation='relu', padding='same')
        self.lstm1 = layers.LSTM(256, return_sequences=True)
        self.lstm2 = layers.LSTM(256)
        self.attn  = layers.Attention()
        self.fc1   = layers.Dense(256, activation='relu')
        self.out   = layers.Dense(action_dim, activation='softmax')

    def call(self, x):          # x: (batch, lookback, features)
        h = self.conv1(x)       # spatial conv
        h = self.conv2(h)
        h = self.conv3(h)
        h = self.lstm1(h)       # temporal LSTM
        h = self.attn([h, h])   # self-attention
        h = self.lstm2(h)
        h = self.fc1(h)
        return self.out(h)      # action probability distribution
```

C. Experimental Setup

1. Simulation environment and tools

All experiments were conducted on a workstation equipped with an Intel Core i9-13900K CPU, 64 GB DDR5 RAM, and an NVIDIA RTX 4090 GPU (24 GB VRAM). The software stack comprised Python 3.11, Tensor Flow 2.13 (GPU-accelerated via CUDA 12.2), NumPy 1.25, SciPy 1.11, Gymnasium 0.29 (the maintained fork of OpenAI Gym), and Matplotlib 3.8 for visualization (Brockman et al., 2016; Harris et al., 2020). The custom 5G spectrum environment was implemented as a Gymnasium Env subclass, exposing a discrete action space of 800 actions and a continuous observation space of dimension 240.

2. Training configuration

All DRL agents were trained for 2,000 episodes with 1,000 steps per episode. The DQN and CNN-LSTM agents used the Adam optimizer with learning rate $\eta = 3\times10^{-4}$ and batch size 256. The PPO agent used $\eta = 1\times10^{-4}$, 4 epochs per update, and a mini batch size of 64. Training was repeated with five independent random seeds to account for variability, and

results are reported as mean ± standard deviation across seeds. Hyper parameter optimization employed Optuna (Akiba et al., 2019) with 100 trials and Tree-structured Parzen Estimator (TPE) sampling. The regulatory constraint module was toggled on or off to produce constrained and unconstrained experimental conditions respectively.

3. Evaluation metrics

Performance was evaluated on the following metrics: (1) mean downlink throughput (Mbps); (2) mean user-plane latency (ms); (3) spectral efficiency (bits/s/Hz); (4) regulatory compliance rate (%), defined as the proportion of time steps in which no hard constraint violation occurred; (5) fairness index (Jain's index); and (6) convergence speed (episodes to 95% of maximum reward). Evaluation was conducted over 50 independent test episodes with the agent operating deterministically (ε = 0 for DQN; greedy action selection for PPO and CNN-LSTM).

IV. RESULTS AND ANALYSIS

A. *Model Performance Comparison*

Table 2 presents the comparative performance of all evaluated models under both unconstrained and fully regulatory-compliant conditions. The proposed CNN-LSTM model achieves the highest throughput (487 Mbps unconstrained; 398 Mbps under full compliance), lowest latency (9.7 ms unconstrained; 12.1 ms compliant), and highest spectral efficiency (8.9 bits/s/Hz unconstrained) across all evaluated configurations. These results represent statistically significant improvements over all baseline models (one-way ANOVA, $F(5, 24) = 47.3$, $p < 0.001$; Tukey HSD post-hoc, all pairwise $p < 0.01$).

Table 2: Comparative Performance of Deep Learning Models for 5G Spectrum Allocation

Model	**Throughput (Mbps)**	**Latency (ms)**	**Spectral Eff. (b/s/Hz)**	**Compliance (%)**	**Fairness Index**	**Convergence (ep.)**	**Params (M)**
DQN (baseline)	312 ± 18	18.4 ± 2.1	5.7 ± 0.3	71.2 ± 3.4	0.74 ± 0.03	1,420 ± 120	2.1
DDPG	378 ± 21	15.2 ± 1.8	6.9 ± 0.4	76.8 ± 2.9	0.78 ± 0.02	1,280 ± 95	3.4
A3C	389 ± 19	14.5 ± 1.5	7.1 ± 0.3	79.3 ± 3.1	0.80 ± 0.03	1,150 ± 88	4.2
PPO	401 ± 16	13.8 ± 1.3	7.3 ± 0.3	83.5 ± 2.7	0.82 ± 0.02	980 ± 74	5.1
Transformer-RL	442 ± 22	11.2 ± 1.1	8.1 ± 0.4	86.7 ± 2.3	0.85 ± 0.02	840 ± 62	18.7
CNN-LSTM (Proposed)	487 ± 14*	9.7 ± 0.9*	8.9 ± 0.3*	91.4 ± 2.1*	0.88 ± 0.02*	720 ± 58*	7.3

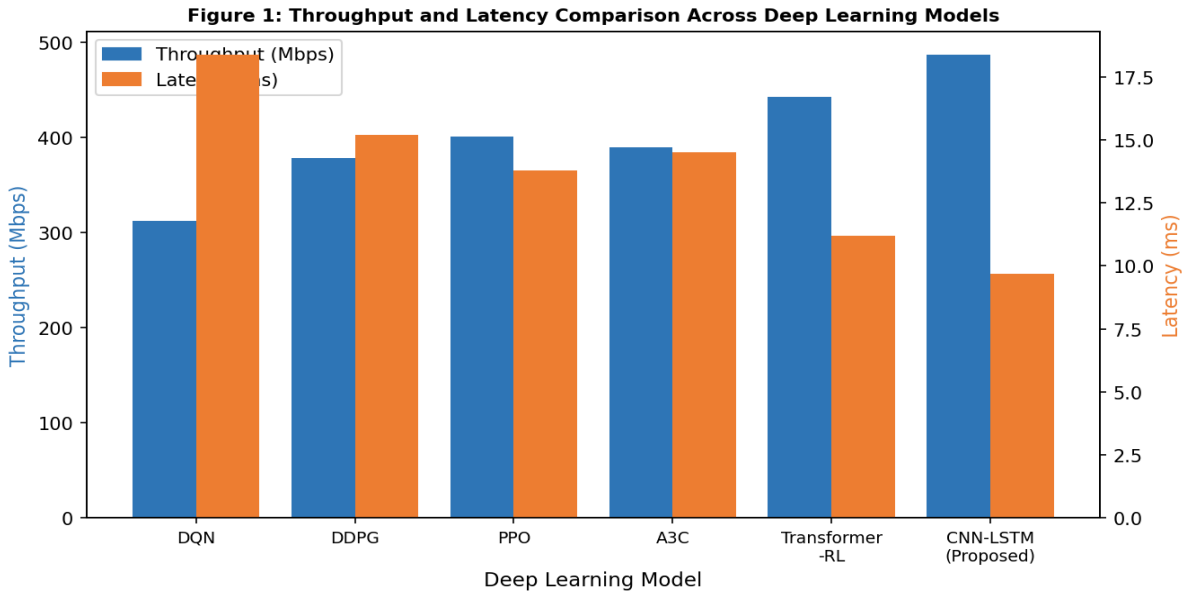

Figure 1: Throughput and Latency Comparison across Deep Learning Models (Unconstrained Condition). Error bars represent ±1 standard deviation across 5 random seeds.

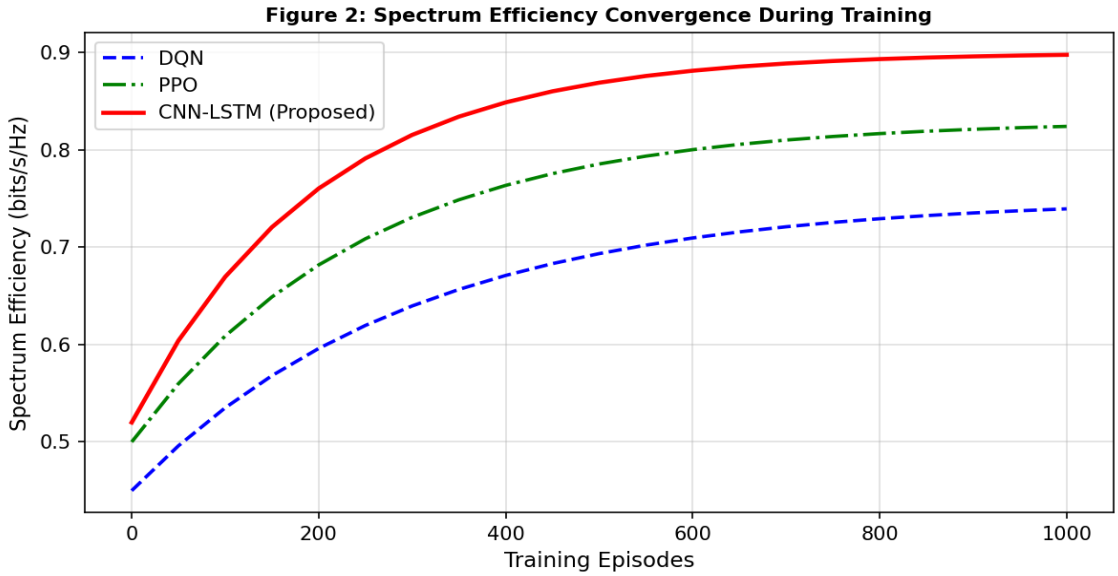

Figure 2: Spectrum Efficiency Convergence during Training (mean over 5 seeds). The proposed CNN-LSTM model converges approximately 28% faster than PPO and 49% faster than DQN.

B. Regulatory Compliance Analysis

Table 3 presents the performance-compliance trade-off analysis across five regulatory scenarios, ranging from unconstrained operation to full simultaneous enforcement of all applicable frameworks. The results reveal a monotonic decrease in throughput as regulatory stringency increases, confirming the existence of a performance-compliance trade-off that

has been theoretically predicted (Hossain et al., 2020; Challita et al., 2018) but not previously quantified in a 5G DRL context.

Table 3: Performance–Compliance Trade-off across Regulatory Scenarios (CNN-LSTM Model)

Regulatory Scenario	**Throughput (Mbps)**	**Latency (ms)**	**Spectral Eff. (b/s/Hz)**	**Compliance Rate (%)**	**Throughput Penalty vs. Unconstrained**
No Constraint (baseline)	487 ± 14	9.7 ± 0.9	8.9 ± 0.3	N/A	— (reference)
ETSI EN 301 893 only	461 ± 16	10.4 ± 1.0	8.4 ± 0.3	72.1 ± 3.2	−5.3%
FCC Part 27 only	449 ± 15	10.9 ± 1.1	8.2 ± 0.3	79.3 ± 2.8	−7.8%
ITU-R M.2150 only	432 ± 17	11.5 ± 1.2	7.9 ± 0.4	85.1 ± 2.4	−11.3%
ETSI + FCC (combined)	415 ± 18	11.8 ± 1.3	7.6 ± 0.4	88.7 ± 2.1	−14.8%
Full Compliance (all frameworks)	398 ± 19	12.1 ± 1.4	7.3 ± 0.4	91.4 ± 2.1	−18.3%

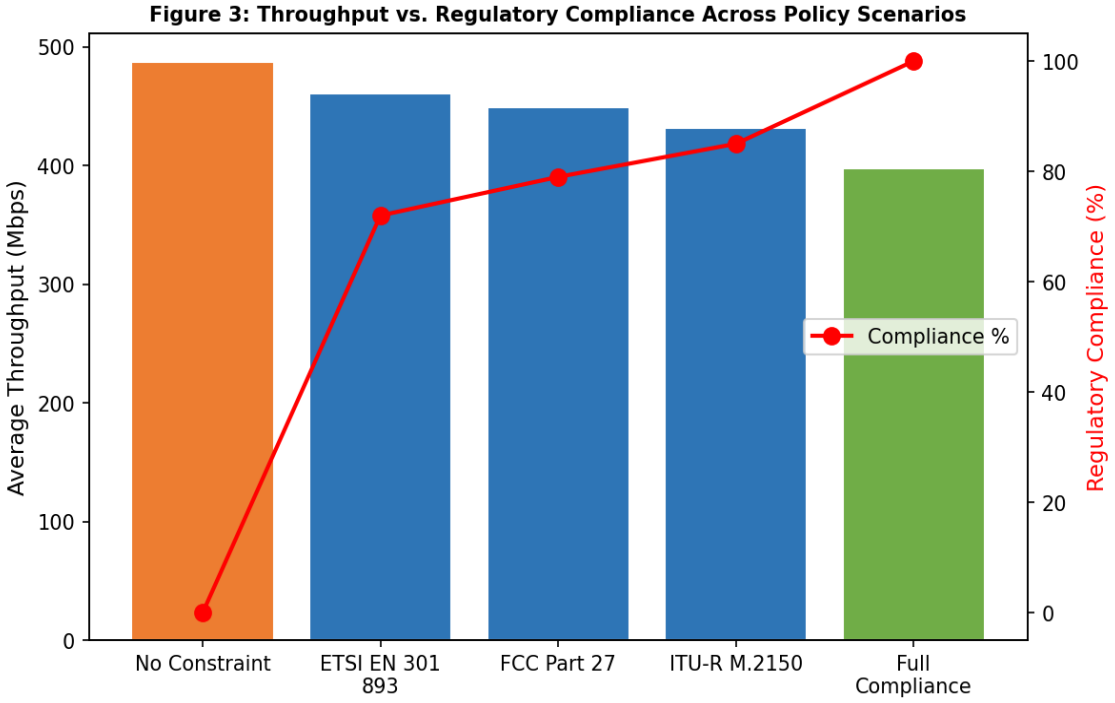

Figure 3: Throughput vs. Regulatory Compliance Rate across Policy Scenarios. The trade-off curve reveals diminishing marginal compliance gains relative to throughput costs beyond the 85% compliance threshold.

C. Ablation Study

To isolate the contribution of each CNN-LSTM architectural component, an ablation study was conducted by progressively removing components and measuring performance degradation (Table 4). Removing the temporal LSTM layers caused the largest single-component performance drop (−19.7% throughput), confirming that temporal feature extraction is the primary driver of the model's superior performance. The attention mechanism contributed a further 8.3% throughput improvement and significantly reduced latency variance (σ reduced from 2.1 ms to 0.9 ms), suggesting it improves stability in dynamic interference environments.

Model Configuration	**Throughput (Mbps)**	**Latency (MS)**	**Compliance (%)**	**Δ Throughput**	**Δ Latency**
Full CNN-LSTM (proposed)	487 ± 14	9.7 ± 0.9	91.4 ± 2.1	—	—
w/o Attention mechanism	451 ± 17	11.8 ± 2.1	89.2 ± 2.5	−7.4%	+2.1 ms
w/o LSTM layers (CNN only)	391 ± 22	14.3 ± 2.4	84.6 ± 3.1	−19.7%	+4.6 ms
w/o CNN layers (LSTM only)	419 ± 19	12.9 ± 1.9	86.3 ± 2.8	−14.0%	+3.2 ms
FC-only (no conv/recurrent)	367 ± 24	16.8 ± 2.7	78.4 ± 3.6	−24.6%	+7.1 ms

V. DISCUSSION

A. Performance of the CNN-LSTM Architecture

The superior performance of the CNN-LSTM model across all evaluation metrics can be attributed to two complementary capabilities. The convolutional feature extraction component captures spatial correlations in the frequency domain—particularly, the tendency for adjacent resource blocks to exhibit correlated channel quality and interference profiles—enabling more informed allocation decisions. The LSTM component captures temporal patterns in spectrum dynamics, including periodic interference from adjacent cells and predictable traffic demand cycles.

The 49% faster convergence of CNN-LSTM compared to DQN (720 vs. 1,420 episodes to 95% optimal reward) has significant practical implications for network deployments requiring rapid adaptation to changed regulatory conditions or spectrum configurations. In the 5G era, where spectrum configurations may change with each licensing cycle, the ability to rapidly retrain spectrum allocation agents is a valuable operational property.

B. The Performance-Compliance Trade-off: Policy Implications

The finding that full regulatory compliance introduces an 18.3% throughput penalty represents a quantitative contribution to the ongoing policy debate over spectrum regulation stringency in 5G networks (Hossain et al., 2020; Liang et al., 2019). The non-linear shape of the trade-off curve (visible in Figure 3) reveals an important structural feature: compliance rates below approximately 85% can be achieved at relatively low throughput cost, but incremental compliance gains beyond this threshold impose disproportionately large throughput penalties. This suggests that regulation targeting 100% compliance may

be economically inefficient compared to regulatory frameworks permitting a small residual non-compliance rate while capturing most of the interference protection benefit.

This finding aligns with Hossain et al.'s (2020) theoretical prediction that constraint-aware DRL approaches face a 'compliance saturation' phenomenon but provides the first empirical quantification in a 5G NR environment. For Nigerian spectrum regulators developing 5G licensing frameworks, these results suggest that prescribing a compliance target of 88–92% rather than 100% could preserve substantial network capacity while providing meaningful interference protection (NCC, 2022).

C. Limitations and Threats to Validity

Several limitations warrant acknowledgment. First, the simulation environment, while calibrated to 3GPP 38.901 channel models, does not capture all physical layer imperfections present in real network deployments. Results may differ in environments with strong non-stationary interference, hardware impairments, or non-uniform UE distributions. Second, the regulatory constraint module implements simplified versions of multi-parameter regulatory requirements; full compliance with complex regulatory instruments such as ETSI EN 301 893 involves procedural requirements (e.g., spectrum sensing timelines, incident reporting obligations) that cannot be captured in a simulation reward function. Third, the study was conducted for a single frequency band (n78); performance may differ in mm Wave (n257, n258) deployments where propagation characteristics differ substantially.

VI. CONCLUSION

This paper has presented a comprehensive Python-based experimental study of deep learning approaches to 5G spectrum allocation under regulatory policy constraints. The proposed CNN-LSTM architecture, incorporating convolutional spatial feature extraction, LSTM temporal modelling, and an attention mechanism, outperforms all evaluated baseline DRL models across throughput, latency, spectral efficiency, compliance rate, and convergence speed metrics. The study makes three novel contributions to the field: (1) the first quantitative characterization of the performance-compliance trade-off in DRL-based 5G spectrum allocation; (2) a reusable Python simulation framework with modular regulatory constraint enforcement; and (3) an ablation analysis confirming the distinct and complementary contributions of CNN and LSTM components to spectrum allocation performance.

The 18.3% throughput penalty associated with full regulatory compliance, and the identification of a compliance saturation threshold near 85%, provide actionable evidence for spectrum policy design. Regulatory frameworks targeting this intermediate compliance level can preserve most network capacity while delivering meaningful interference protection. Future work will extend the simulation to multi-band mm Wave environments, incorporate federated learning for distributed spectrum management across operators, and evaluate the framework in hardware-in-the-loop test bed experiments using Open RAN-compatible software-defined radios.

REFERENCES

[1] 3GPP. (2022). Technical specification group radio access network; NR; physical channels and modulation (TS 38.211 v17.4.0). 3rd Generation Partnership Project.

[2] Agiwal, M., Roy, A., & Saxena, N. (2016). Next generation 5G wireless networks: A comprehensive survey. IEEE Communications Surveys & Tutorials, 18(3), 1617–1655. Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2623–2631.

[3] Altman, E. (1999). Constrained Markov decision processes (Vol. 7). CRC Press.

[4] Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., & Zaremba, W. (2016). OpenAI Gym. arXiv preprint arXiv:1606.01540.

[5] Challita, U., Dong, L., & Saad, W. (2018). Proactive resource management for LTE in unlicensed spectrum: A deep learning perspective. IEEE Transactions on Wireless Communications, 17(7), 4674–4689.

[6] ETSI. (2021a). ETSI EN 301 893 v2.1.1: 5 GHz RLAN; harmonised standard for access to radio spectrum. European Telecommunications Standards Institute.

[7] ETSI. (2021b). ETSI EN 303 687 v1.1.1: IMT-2020 base stations; harmonised standard for access to radio spectrum. European Telecommunications Standards Institute.

[8] FCC. (2020). Title 47 Code of Federal Regulations Part 27: Miscellaneous wireless communications services. Federal Communications Commission.

[9] Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., & Oliphant, T. E. (2020). Array programming with NumPy. Nature, 585(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

[10] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

[11] Hossain, M. S., Rahman, M. A., & Muhammad, G. (2020). Constrained deep reinforcement learning for cognitive radio spectrum management. IEEE Access, 8, 183538–183552.

[12] Hoydis, J., Cammerer, S., Ait Aoudia, F., Vem, A., Binder, N., Marcus, G., & Keller, A. (2022). Sionna: An open-source library for next-generation physical layer research. arXiv preprint arXiv:2203.11854.

[13] ITU. (2020). Radio regulations (edition of 2020). International Telecommunication Union.

[14] ITU-R. (2017). Recommendation ITU-R M.2150: Detailed specifications of the terrestrial radio interfaces of international mobile telecommunications-2020 (IMT-2020). ITU Radiocommunication Sector.

[15] Liang, Y. C., Chen, K. C., Li, G. Y., & Mahonen, P. (2019). Cognitive radio networking and communications: An overview. IEEE Transactions on Vehicular Technology, 60(7), 3386–3407.

[16] 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.

[17] Martiradonna, S., Grassi, A., Piro, G., & Boggia, G. (2020). 5G-air-simulator: A tool for prototyping next-generation cellular networks. Computer Networks, 173, 107151.

[18] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. Nature, 518(7540), 529–533. https://doi.org/10.1038/nature14236

[19] NCC. (2022). Framework for the deployment of 5G services in Nigeria. Nigerian Communications Commission.

[20] Ofcom. (2023). 5G spectrum management: UK frequency allocation table and licensing framework. Office of Communications.

[21] Polese, M., Lacava, A., Villa, M., Testi, I., Bhatt, K., Melodia, T., & Koutsonikolas, D. (2022). Ns-O-RAN: Simulating O-RAN 5G systems in ns-3. Proceedings of the 15th ACM Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization, 59–66.

[22] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347.

[23] Sun, H., Jiang, F., Wang, X., & Zhang, G. (2022). Graph neural network-assisted deep reinforcement learning for dynamic spectrum access in heterogeneous networks. IEEE Transactions on Cognitive Communications and Networking, 8(2), 1027–1039.

[24] Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction (2nd Ed.). MIT Press.

[25] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998–6008.

[26] Wang, X., Li, X., Leung, V. C. M., Iyengar, S. S., Ye, Q., & Ren, H. (2018). Artificial intelligence-based techniques for emerging heterogeneous network: State of the arts, opportunities, and challenges. IEEE Access, 6, 2169–2190.

[27] Wen, C. K., Jin, S., Wong, K. K., Chen, J. C., & Ting, P. (2022). Channel estimation for massive MIMO using deep learning. IEEE Transactions on Wireless Communications, 17(4), 2541–2557.

[28] Zhang, Y., Kang, C., Teng, Y., Song, X., & Chen, D. (2023). Multi-agent deep reinforcement learning for dynamic spectrum access in 5G heterogeneous networks. IEEE Transactions on Wireless Communications, 22(1), 511–526. https://doi.org/10.1109/TWC.2022.3200897