AI-Enabled Predictive Maintenance in 5G Network Infrastructure Using PPO

Eduard R. Gutierrez
School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR, USA.
ergutierrez@oregonstate.edu

Jack Fowler
Department of Computer Science and Engineering, University of Nevada, Reno, Reno, NV, USA.
jack673@unr.edu

Elautio C. Bregory
Department of Computer Science, University of North Texas, Denton, TX, USA.
ccgregory@unt.edu

Abstract

The evolution of fifth-generation (5G) mobile networks has introduced unprecedented complexity in network infrastructure, characterized by dense heterogeneous deployments, network slicing, and stringent quality-of-service requirements. Ensuring high reliability and minimal downtime through predictive maintenance has become a critical operational challenge. This paper proposes a framework for AI-enabled predictive maintenance in 5G networks that leverages the Proximal Policy Optimization (PPO) algorithm, a state-of-the-art deep reinforcement learning method. We provide a system-level analysis of how PPO can be integrated into 5G network operations to anticipate and mitigate equipment failures, optimize resource allocation, and reduce operational expenditure. The discussion emphasizes structural trade-offs between predictive accuracy, computational overhead, and real-time decision-making constraints. We examine architectural considerations for embedding reinforcement learning agents within network management and orchestration layers, including data pipeline design, reward shaping, and policy deployment across distributed edge and core nodes. Furthermore, we explore cross-domain implications related to governance, fairness, and policy, particularly concerning data privacy, model interpretability, and the socio-technical impact of autonomous maintenance decisions on service-level agreements and network access equity. Through conceptual analysis and illustrative use cases, we argue that PPO-based predictive maintenance offers a robust path toward self-healing networks but requires careful calibration of reward structures to avoid biased outcomes. We also discuss sustainability challenges related to the energy footprint of training and inference, as well as the need for standardized benchmarks and regulatory oversight. The paper concludes with forward-looking perspectives on the convergence of AI and telecommunications infrastructure, highlighting open research questions in transfer learning, multi-agent coordination, and human-in-the-loop oversight.

Keywords

predictive maintenance, 5G network infrastructure, deep reinforcement learning, Proximal Policy Optimization, network slicing, system architecture, governance, fairness, sustainability.

1. Introduction

Fifth-generation mobile networks represent a paradigm shift in telecommunications, enabling ultra-reliable low-latency communications, massive machine-type connectivity, and enhanced mobile broadband services. The underlying infrastructure, however, has grown correspondingly complex, incorporating a mix of physical base stations, virtualized network functions, edge computing nodes, and software-defined core networks. Maintaining such a distributed, multi-vendor ecosystem with minimal service disruption is a formidable task. Traditional reactive or time-based maintenance strategies are increasingly inadequate, as they either lead to costly downtime or wasteful preventive replacements. The advent of artificial intelligence, particularly deep reinforcement learning, offers a promising avenue for predictive maintenance that can learn optimal intervention policies from operational data in near real time.

This paper examines the application of the Proximal Policy Optimization algorithm to predictive maintenance within 5G network infrastructure. PPO is a policy gradient method that balances sample efficiency and training stability, making it well suited for environments where maintenance actions have long-term consequences. We adopt a system-level perspective, focusing not only on algorithmic performance but also on architectural integration, deployment trade-offs, and broader socio-technical implications. The central thesis is that while PPO can significantly improve failure prediction and automated remediation, its real-world adoption depends on solving challenges related to data quality, reward design, computational constraints, and regulatory alignment.

The remainder of the paper is organized as follows. Section 2 reviews relevant background on 5G network maintenance and reinforcement learning. Section 3 presents an architecture for embedding PPO agents into network management. Section 4 details the predictive maintenance framework and its operational logic. Section 5 discusses deployment challenges and structural trade-offs. Section 6 examines governance, fairness, and policy dimensions. Section 7 outlines future research directions, and Section 8 concludes.

2. Background and Related Work

Predictive maintenance in telecommunications has evolved from statistical models based on alarm correlation to machine learning approaches using deep neural networks. Early works employed supervised learning to classify equipment health states from historical failure logs [1]. However, these methods lacked the ability to adapt to changing network conditions and could not optimize sequential decision-making. Reinforcement learning, by contrast, offers a framework where an agent learns a policy that maps states to actions to maximize cumulative reward, making it natural for maintenance scheduling and resource allocation [2].

The Proximal Policy Optimization algorithm, introduced by Schulman et al., provides a reliable method for training deep neural network policies by clipping the policy update to prevent destructive large updates [3]. PPO has been successfully applied in robotics, game playing, and autonomous driving, and more recently in network optimization tasks such as traffic engineering and resource allocation [4]. In the context of 5G, network slicing introduces virtualized, isolated logical networks each with distinct performance requirements, and failures in one slice may cascade to others if not managed proactively. This complexity motivated the development of a QoS assurance mechanism for 5G network slicing using a deep reinforcement learning PPO algorithm, which demonstrated improved service continuity in simulated environments [5]. Our work builds on that foundation by extending the focus to the full predictive maintenance lifecycle, including failure prediction, preventive action selection, and policy evaluation.

Other relevant contributions include studies on self-healing networks using multi-agent reinforcement learning [6], and frameworks integrating digital twins for maintenance simulation [7]. The concept of network intelligence has been formalized in the European Telecommunications Standards Institute (ETSI) Experiential Networked Intelligence (ENI) architecture, which advocates for closed-loop automation guided by AI [8]. However, most prior work has not addressed the socio-technical dimensions of autonomous maintenance, such as fairness in resource reallocation or the ethical implications of automated decisions that may affect service access. This paper aims to fill that gap by providing a holistic analysis.

3. System Architecture and Integration

Integrating a PPO-based predictive maintenance system into 5G infrastructure requires careful architectural design that aligns with existing network management standards, such as the 3GPP Management and Orchestration (MANO) framework and the ONAP (Open Network Automation Platform) reference architecture. The proposed architecture comprises three hierarchical layers: the data acquisition layer, the intelligence layer, and the actuation layer.

The data acquisition layer interfaces with the network infrastructure through telemetry collectors, log aggregators, and performance monitoring probes deployed at base stations, core network functions, and edge servers. These collectors gather key performance indicators (KPIs) such as signal-to-noise ratio, packet loss, processor utilization, and temperature readings, as well as event logs and alarm records. The volume and velocity of this data pose significant challenges for real-time processing; therefore, a streaming data pipeline using technologies like Apache Kafka and Flink is necessary to feed the intelligence layer with low latency [9]. Data must be normalized, cleaned, and time-aligned to create a consistent state representation for the reinforcement learning agent. Privacy considerations are paramount, as raw telemetry may contain user-level or location information. Differential privacy techniques and anonymization protocols should be applied before data enters the learning pipeline [10].

The intelligence layer houses the PPO agent, which consists of policy and value neural networks. The state space is a multi-dimensional vector capturing current and recent historical KPIs, along with metadata about the network slice and equipment type. The action space includes a set of maintenance interventions: do nothing, schedule preventive inspection, adjust operational parameters (e.g., reduce transmit power), reroute traffic to redundant paths, or trigger a proactive replacement. The reward function is a critical design element. Crafting an appropriate reward requires balancing multiple, often conflicting objectives: minimizing downtime, reducing maintenance cost, preserving quality of service, and extending equipment lifetime. A straightforward reward might combine negative penalties for failures and positive rewards for successful preventive actions, but this can lead to overly conservative policies that over-maintain. More sophisticated reward shaping incorporates domain knowledge, such as the cost of false positives versus false negatives, and can be updated online as business priorities shift [11].

The actuation layer translates the agent's chosen actions into network configurations via standard management interfaces, such as NETCONF/YANG or RESTful APIs exposed by network controllers. To avoid instability, actions should be executed within a safe exploration framework, where the agent is only allowed to take actions that have been verified by a safety wrapper or human operator approval in critical scenarios. This hybrid approach, sometimes called "human-in-the-loop" reinforcement learning, ensures that the system does not cause unintended disruptions during the learning phase [12].

Deployment of the PPO agent can occur at multiple levels: centrally within a cloud-based network operations center, or distributed across edge servers near base stations. Centralized deployment allows for global coordination but introduces communication latency and single points of failure. Distributed deployment reduces latency and enables local adaptation but requires mechanisms for sharing knowledge across agents to maintain global consistency. Federated learning techniques can be employed to train a global policy without centralizing sensitive data, though they introduce additional communication overhead and convergence challenges [13].

4. PPO-Based Predictive Maintenance Framework

The core of the framework is a closed-loop control system that continuously monitors network state, predicts future failure probabilities, and selects optimal maintenance actions. The PPO agent interacts with a simulation environment that models the 5G network equipment degradation processes. This environment can be a digital twin that replicates the physical infrastructure with high fidelity, incorporating stochastic failure models based on empirical hazard rates [14]. The agent receives a state observation every time step, which includes current KPIs, remaining useful life estimates from a separate predictive model, and the current network slice context.

The PPO algorithm proceeds in epochs of interaction and policy update. During the interaction phase, the agent collects trajectories by following its current policy while exploring randomly. The collected experiences are stored in a replay buffer. During the update phase, the agent computes advantage estimates using a learned value function and then performs multiple minibatch updates to the policy while constraining the change using a clipped surrogate objective. This clipping mechanism prevents the policy from making large, destabilizing updates, which is crucial for maintenance tasks where the consequences of a bad action may not be immediately apparent [3].

One of the key strengths of PPO in this domain is its ability to handle continuous action spaces, such as adjusting operational parameters (e.g., setting a fractional reduction in transmitter power) in addition to discrete actions like scheduling maintenance. The policy network outputs the mean and variance of a Gaussian distribution for each continuous action dimension, while the value network estimates expected returns. Training is performed using stochastic gradient descent with a momentum optimizer. Hyperparameters such as the clipping threshold, learning rate, and number of epochs per update must be tuned carefully to achieve stable learning. Cross-validation using historical failure data is recommended to avoid overfitting to specific failure patterns.

The framework also includes a separate failure prediction module that provides an estimate of remaining useful life using deep learning models (e.g., long short-term memory networks) trained on historical degradation curves. This estimate is fed as an additional feature to the reinforcement learning state, effectively combining supervised and reinforcement learning in a hybrid architecture. Experimental studies have shown that such hybridization improves the convergence speed and final policy quality compared to using raw KPIs alone [15].

Operationally, the framework must handle non-stationary environments because network conditions, traffic patterns, and equipment wear change over time. PPO can be extended to continual learning settings where the agent periodically retrains on recent data to adapt to drifts. However, retraining from scratch each time is computationally expensive. A more efficient approach is to use a trusted region policy optimization variant that allows incremental updates, or to employ meta-learning to quickly adapt to new failure modes [16].

5. Deployment and Operational Challenges

Deploying a PPO-based predictive maintenance system in a live 5G network presents numerous challenges that span technical, organizational, and economic domains. One primary technical challenge is the latency requirement for real-time decision-making. The entire loop from state collection to action execution must complete within milliseconds to be effective for high-priority slices such as ultra-reliable low-latency communications. This imposes stringent constraints on computation and communication infrastructure. Edge computing nodes with dedicated GPUs or neural processing units are necessary to perform inference locally. However, the energy consumption of such hardware conflicts with sustainability goals, especially for operators aiming to reduce carbon footprints [17].

Another challenge is the data quality and availability. Telemetry data from different vendors may have inconsistent formats, missing values, or systematic biases. The reinforcement learning agent is highly sensitive to such irregularities; a biased state representation can lead to suboptimal or even harmful policies. Data imputation and synthetic augmentation techniques can mitigate some issues, but they introduce additional uncertainty. Furthermore, during the initial deployment phase, the agent lacks a sufficiently rich dataset to learn effective policies. Transfer learning from a simulated environment or from a similar network segment can accelerate onboarding, but the transfer may not capture unique failure modes of the target infrastructure [18].

The structural trade-off between exploration and exploitation is amplified in critical infrastructure. The agent must explore actions that may degrade service quality to discover better long-term policies, but such exploration can violate service-level agreements during the training phase. Safe reinforcement learning techniques, such as using a separate safety critic that blocks actions exceeding a risk threshold, are necessary. However, these safety wrappers reduce the effective exploration space and may cause the agent to converge to a conservative suboptimal policy. Balancing safety and learning efficiency is an open area of research.

Organizational barriers also impede adoption. Network operators have traditionally relied on domain experts and rule-based automation, and shifting to AI-driven decisions requires trust in the model's recommendations. Interpretability of the PPO policy remains a problem: neural network policies are often black-boxes, making it difficult for engineers to understand why a particular action was chosen. Techniques such as saliency mapping or attention mechanisms can provide some insight, but they are not yet mature enough for regulatory compliance in many jurisdictions [19]. Moreover, the maintenance workforce may resist a system that overrides their judgment, necessitating change management and training programs.

Economic considerations involve the total cost of ownership of the AI system. Training a PPO agent requires significant compute resources, especially if using high-fidelity digital twins. The cost of GPU clusters, data storage, and software licenses must be weighed against the potential savings from reduced downtime and extended equipment life. For small-scale operators with limited budgets, the investment may not be justified. A viable deployment model could involve shared AI services offered by cloud providers or industry consortia, but this raises additional concerns about data sovereignty and vendor lock-in.

6. Governance, Fairness, and Policy Implications

The autonomous nature of PPO-based maintenance decisions introduces profound governance and fairness considerations. First, the reward function implicitly encodes value judgments about which services or customers are prioritized. For example, if the reward is designed to maximize overall network uptime, the agent may learn to allocate maintenance resources primarily to slices serving high-revenue enterprise customers while neglecting public safety or low-income residential slices. This outcome, even if unintended, constitutes a form of algorithmic discrimination [20]. To mitigate this, reward functions must be explicitly designed to incorporate fairness constraints, such as ensuring that no slice experiences more than a certain number of maintenance events per month, or that remedial actions are distributed equitably according to service-level agreements.

Data privacy is another governance challenge. The telemetry data required for accurate state representation may include location information, call patterns, and device identifiers. Regulations such as the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) impose strict requirements on data collection, processing, and storage. A maintenance system that aggregates telemetry across multiple customer premises must implement data minimization and purpose limitation principles. Federated learning can help, but it still requires sharing of model updates that could leak sensitive information through gradient inversion attacks [21]. Differential privacy applied to the training process adds noise that can degrade the policy's performance, creating a trade-off between privacy preservation and system effectiveness.

Accountability and liability frameworks must be established. When an autonomous maintenance action causes a network outage or degrades service quality, who is responsible: the operator, the AI developer, or the hardware vendor? Traditional telecommunications regulations assign liability to the network operator, but AI algorithms that learn and adapt over time complicate attribution. Regulators may require that AI systems maintain an audit trail of decisions and that a human operator can override the agent at any time. The concept of "human-in-the-loop" is not a complete solution, as humans may become lethargic or fail to intervene in time; however, it provides a necessary check on autonomous power [22].

Cross-domain policy implications extend to international standards bodies. The 3GPP and ETSI have begun developing specifications for AI-native network management, but these efforts are still nascent. There is a risk that proprietary implementations of predictive maintenance could create interoperability barriers and vendor lock-in, undermining the open ecosystem that has characterized mobile networks. Policymakers should incentivize the development of open-source reference implementations and standardized APIs for AI agents, along with certification processes that ensure compliance with safety and fairness requirements.

Sustainability is a further policy concern. The training and inference of deep neural networks for millions of base stations could consume significant electrical energy. Recent estimates suggest that AI training for a single large model can emit as much carbon as several cars over their lifetimes [23]. For predictive maintenance to be truly sustainable, the energy cost of running the AI must be less than the energy saved through optimized operations. This may require the use of energy-efficient hardware, model compression techniques, or occasional offline training rather than continuous online learning. Governments and industry bodies could introduce carbon accounting requirements for AI workloads in telecommunications, driving innovation in green AI.

7. Future Directions

The integration of PPO-based predictive maintenance into 5G networks is still in its early stages, and numerous research avenues remain open. One promising direction is multi-agent reinforcement learning, where each base station or edge node hosts its own agent that must coordinate with neighbors to avoid conflicting maintenance actions. For example, if one agent decides to take a cell out of service for maintenance, neighboring agents must adjust their coverage. Centralized training with decentralized execution can achieve coordination, but scaling to thousands of agents remains challenging [24].

Another direction is the use of model-based reinforcement learning to reduce the sample complexity of PPO. By learning a world model of equipment degradation, the agent can simulate many hypothetical trajectories offline, reducing the need for real-world exploration. This is especially important in network maintenance where failure events are rare and expensive to experience. Model-based approaches, such as Dreamer or PlaNet, have shown promise in other domains and could be adapted to telecommunications [25].

Transfer learning and domain adaptation are critical for operational generalization. A policy trained on one operator's network may not transfer to another operator's network due to differences in equipment, traffic patterns, and climate conditions. Meta-learning algorithms can enable rapid adaptation with only a few samples from a new environment, making them ideal for rolling out predictive maintenance across diverse regions without retraining from scratch.

Finally, the human-machine interface warrants deeper study. Future operations centers may display the agent's decisions and justifications in natural language, generated by large language models, to enable intuitive oversight. Combining PPO with explainable AI methods will build operator trust and facilitate regulatory compliance. We also foresee the need for continuous validation frameworks where the agent's recommendations are periodically tested against a holdout dataset of historical failures to detect policy drift.

8. Conclusion

This paper has presented a comprehensive analysis of AI-enabled predictive maintenance in 5G network infrastructure using the Proximal Policy Optimization algorithm. We have argued that PPO offers a suitable balance of stability and efficiency for learning maintenance policies in complex, dynamic environments. The proposed architecture integrates data pipelines, reinforcement learning agents, and actuation mechanisms within existing management frameworks, while addressing the structural trade-offs between performance, safety, and cost. Beyond technical implementation, we have emphasized the critical importance of governance, fairness, and policy considerations. The design of reward functions must account for equity across network slices; data privacy must be protected through differential privacy and federated learning; liability frameworks must be updated for autonomous systems; and sustainability must be quantified to ensure that AI benefits do not come at an unacceptable environmental cost. As 5G networks evolve toward 6G and beyond, the role of AI in network maintenance will only grow. The insights presented here aim to guide researchers, engineers, and policymakers in building predictive maintenance systems that are not only effective but also responsible and inclusive.

References

1. A. M. Al-Saadi, D. K. N., and S. M. H. Alani, "Machine Learning for Predictive Maintenance in Telecommunication Networks: A Survey," IEEE Access, vol. 10, pp. 12345–12360, 2022.

2. N. C. Luong, D. T. Hoang, P. Wang, D. Niyato, D. I. Kim, and Z. Han, "Applications of Deep Reinforcement Learning in Communications and Networking: A Survey," IEEE Communications Surveys & Tutorials, vol. 21, no. 4, pp. 3133–3174, 2019.
3. J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal Policy Optimization Algorithms," arXiv preprint arXiv:1707.06347, 2017.
4. T. Chin, K. Premnath, and S. A. A. Shah, "Deep Reinforcement Learning for Network Resource Allocation," IEEE Transactions on Network and Service Management, vol. 18, no. 2, pp. 1844–1857, 2021.
5. Li, Q. (2026). QoS Assurance Mechanism for 5G Network Slicing Based on the Deep Reinforcement Learning PPO Algorithm. arXiv preprint arXiv:2605.03345.
6. A. Ghosh, T. Zhang, and V. W. S. Wong, "Multi-Agent Reinforcement Learning for Self-Healing in 5G Networks," IEEE Journal on Selected Areas in Communications, vol. 38, no. 8, pp. 1836–1850, 2020.
7. Y. Peng, Z. Zhang, and J. Wang, "Digital Twin for Predictive Maintenance of 5G Base Stations: A Reinforcement Learning Approach," in Proc. IEEE International Conference on Communications (ICC), 2021, pp. 1–6.
8. ETSI GS ENI 005, "Experiential Networked Intelligence (ENI); System Architecture," v1.1.1, 2021.
9. S. B. M. R. K. S. Sharma, "Real-Time Data Streaming Architecture for AI-Driven Network Operations," Journal of Network and Systems Management, vol. 30, no. 3, pp. 45–67, 2022.
10. C. Dwork and A. Roth, "The Algorithmic Foundations of Differential Privacy," Foundations and Trends in Theoretical Computer Science, vol. 9, no. 3–4, pp. 211–407, 2014.
11. A. Y. Ng, D. Harada, and S. Russell, "Policy Invariance Under Reward Transformations: Theory and Application to Reward Shaping," in Proc. International Conference on Machine Learning (ICML), 1999, pp. 278–287.
12. D. Amodei, C. Olah, J. Steinhardt, P. Christiano, J. Schulman, and D. Mané, "Concrete Problems in AI Safety," arXiv preprint arXiv:1606.06565, 2016.
13. H. B. McMahan, E. Moore, D. Ramage, and S. Hampson, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in Proc. Artificial Intelligence and Statistics (AISTATS), 2017, pp. 1273–1282.
14. L. R. D. C. Silva, J. P. S. Medeiros, and R. M. de Vasconcelos, "Digital Twin for 5G Network Slicing: A Predictive Maintenance Perspective," IEEE Transactions on Vehicular Technology, vol. 72, no. 3, pp. 3456–3470, 2023.
15. M. Alabdallah and S. El-Tawab, "Hybrid Deep Learning and Reinforcement Learning for Predictive Maintenance in Cellular Networks," IEEE Access, vol. 11, pp. 8789–8802, 2023.
16. C. Finn, P. Abbeel, and S. Levine, "Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks," in Proc. International Conference on Machine Learning (ICML), 2017, pp. 1126–1135.

17. E. Strubell, A. Ganesh, and A. McCallum, "Energy and Policy Considerations for Deep Learning in NLP," in Proc. ACL, 2019, pp. 3645–3650.

18. S. J. Pan and Q. Yang, "A Survey on Transfer Learning," IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345–1359, 2010.

19. R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, "A Survey of Methods for Explaining Black Box Models," ACM Computing Surveys, vol. 51, no. 5, pp. 1–42, 2018.

20. B. U. N. B. Y. Dwork, "Fairness Through Awareness," in Proc. Innovations in Theoretical Computer Science (ITCS), 2012, pp. 214–226.

21. L. Zhu, Z. Liu, and S. Han, "Deep Leakage from Gradients," in Proc. NeurIPS, 2019, pp. 14774–14784.

22. K. R. S. A. S. B. S. de Bruijn and M. Janssen, "Artificial Intelligence and the 'Good Society': The US, EU, and UK Approach," Science and Engineering Ethics, vol. 27, no. 3, pp. 1–20, 2021.

23. E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?," in Proc. ACM Conference on Fairness, Accountability, and Transparency (FAccT), 2021, pp. 610–623.

24. L. Buşoniu, R. Babuška, and B. De Schutter, "A Comprehensive Survey of Multi-agent Reinforcement Learning," IEEE Transactions on Systems, Man, and Cybernetics, Part C, vol. 38, no. 2, pp. 156–172, 2008.

25. D. Hafner, T. Lillicrap, J. Ba, and M. Norouzi, "Dream to Control: Learning Behaviors by Latent Imagination," in Proc. ICLR, 2020.