Journal of Progress in Electronics and Communication Engineering, ISSN: 3048-7625 Vol. 2, No. 2, 2025 (pp. 72-77)
RESEARCH ARTICLE WWW.PECEJOURNAL.COM

Design and Implementation of a Hybrid Framework for QoS Monitoring in 5G-Enabled Services

M. Mejail[1], B.K. Nestares[2], L. Gravano[3], E. Tacconi[4], G.R. Meira[5], A. Desages[6]
[1-6]*Centro de Investigacion y Desarrollo de Tecnologias Aeronauticas (CITeA) Fuerza Aerea Argentina Las Higueras, Cordoba, Argentina*

KEYWORDS:
5G QoS,
Hybrid monitoring,
Active-passive framework,
AI analytics,
Latency measurement,
Real-time performance,
KPI evaluation

ARTICLE HISTORY:
Submitted : 23.03.2025
Revised : 05.06.2025
Accepted : 30.07.2025

https://doi.org/10.31838/ECE/02.02.10

Abstract

The fast development of 5G networks presents complicated Quality of the Service (QoS)-based management issues that cannot be managed using the conventional monitoring methods. In this paper, a new framework being proposed is a hybrid system between active probing and passive monitoring as a way of having a full evaluation of the QoS in 5G-enabled services. The proposed system will integrate real-time feedback with AI-driven analytics by using standardised KPIs throughput, latency, jitter, and packet loss. Open5GS and Wireshark (with a machine learning module to detect anomalies) were used as open-source platforms to implement the architecture. The experiments on an emulated 5G core and the comparison with single-mode methods showed an accuracy of measurement improvement by 27 %. The system has a scalable deployment, dynamic sampling, and visualization of network dynamics without interruptions to be made in the diversity of traffic. The hybrid model allows the assurance of constantly changing QoS of new applications, such as autonomous vehicles, industrial IoT, and smart city infrastructure.

Author e-mail Id: Mejail.mej@ing.unrc.edu.ar

How to cite this article: Mejail M, Nestares BK, Gravano L, Tacconi E, Meira GR, Desages A. Design and Implementation of a Hybrid Framework for QoS Monitoring in 5G-Enabled Services, Journal of Progress in Electronics and Communication Engineering Vol. 2, No. 2, 2025 (pp. 72-77).

Introduction

The fifth-generation (5G) communication systems have transformed the mobile communication by providing ultra-low latency, increased bandwidth, and a high number of devices. With the growth of such networks, ensuring stable Quality of Service (QoS) is at the centre stage. The monitoring strategies that have been widely used traditionally and are mainly active or passive are not sufficient when they are used alone. Active probing causes artificial traffic to test performance and can potentially disrupt running sessions whereas passive monitoring is based on packet capture and might not be accurate in transient states.[13] In a bid to address these constraints, hybrid frameworks that combine the two strategies have come out as a critical solution.[4-6]

Most recent works have investigated the concept of intelligent QoS evaluation with AI-based tools to increase the degree of reliability in dense 5G settings.[7-10] These models seek to anticipate performance downfall prior to service breakdown. Nevertheless, the performance measurement of 5G network slices, edge computing nodes, and virtualization layers is heterogeneous,[11-13] making it difficult to measure performance using conventional approaches. The use of analytics engines to provide continuous feedback tools is becoming further essential to adaptive service management.

The importance of QoS measures like throughput, latency, and jitter cannot be overemphasized in application such as remote surgery, autonomous driving and massive IoT ecosystems.[14-17] Predictive information to the network orchestration may be gained by integrating real-time AI models into the monitoring pipeline, which allows fault prevention and resource allocation in advance.[18-20] The hybrid framework proposed in this paper is based upon those principles to provide a flexible and smart monitoring system that has been tested by emulation in a 5G core deployment.

The rest of the paper is as follows: Section 2 includes the description of the related research and technologies in the area of QoS monitoring; Section 3 is the description of the structure of the proposed hybrid framework and its implementation; Section 4 represents the discussion of the experimental findings and analytical results; and Section 5 is the conclusion with the suggestions and directions of the further work.

Related Work

The next-generation networks have seen QoS monitoring to move beyond the plain SNMP polling to highly hybridized measurement schemes that combine active and passive data streams. The initial literature[15] focused on improvement of network lifetime by optimising routing, as well as control overhead minimization. Subsequent efforts were put on dynamic QoS provisioning with machine learning to deal with intricate traffic dynamics in 5G slices.[3, 6, 8]

Hybrid models combine the accuracy of probing at the packet level and less intrusiveness with selective sampling.[5] In the studies,[9] the optimization of high-frequency antennas is mentioned but indirectly affects the signal integrity and the consistency of the QoS. In a study.[17] observed to apply digital twin systems with AI analytics, network anomalies could be predicted at the onset of failure, which is aligned with the predictive goal of our study.

Frameworks[11] suggested GPU-based architectures of computational modelling conceptually analogous to real-time data visualisation needed in QoS analytics. Work[16] applied AI-based optimization to urban mobility traffic management, and pointed out similarities in resource distribution in communication systems. The recent frameworks[19, 20] have also brought the smart probes which can self-adjust in accordance with the congestion levels and represent a step towards automation.

Nevertheless, even now, gaps can be observed in the literature in the integration of scalable, cross-domain QoS frameworks which are compatible with both open-source and commercial infrastructures. The given model fills these gaps with an AI-improved hybrid monitoring layer that can correlate the metrics in various network domains.

Methodology

The QoS monitoring hybrid framework design and implementation process is split into two steps: (i) the design of the system architecture that incorporates active and passive monitoring, and (ii) the implementation of the architecture in a 5G emulated controlled setting. All phases were designed in a manner that would provide a modular scaling, compatibility with heterogeneous network environment, and with open-source as well as commercial systems.

System Architecture

The hybrid framework offers an active and passive Quality of Service (QoS) monitoring components, which are co-located in the framework and execute within the control, user, and management planes of a 5G system. The philosophy of design of this architecture is to combine the accuracy of active probing which inoculates synthetic test traffic with the context knowledge of passive monitoring, which records the actual user flows in real time.

As seen in Figure 1, the architecture will be composed of four functional layers such as Data Acquisition, Data Processing, Analytics as well as Visualisation that will be tasked with a specific step in the monitoring workflow.

1. Data Acquisition Layer:

This base layer is used to collect the real-time data of the synthetic and live traffic sources. Periodically the tests are initiated by active probes to measure delay, jitter and packet loss with controlled conditions by generating ICMP, UDP and HTTP test packets. At the same time, passive agents at edge nodes intercept mirrored packets with sFlow and Wireshark. These two sources of data maintain time accuracy and realism of the network.

2. Data Processing Layer:

After collection, the data of both sources are combined and time stamped in a single preprocessing engine. This engine normalizes the datasets and removes overlapping rows and uses statistical filters to minimise measurement noise. The layer also translates raw metrics to standardised KPIs as average latency (ms), throughput (Mbps) and jitter variance (ms^2) to be further analysed.

3. Analytics Layer:

The results of the processed data on KPI are introduced to an AI-based analytics engine that is developed with a Random Forest classifier. The model detects performance anomalies, future degradation, and matches the KPI trends of network slices. It was chosen due to its great precision and overfitting resistance to changing traffic conditions. Compared to conventional monitoring systems, experimental analysis proved that the use of

AI integration in QoS management significantly increases measurement accuracy (by 27 %).

4. Visualization Layer:

The upper layer offers a real-time interactive dashboard display of trends in QoS. It was created using Python Flask as the backend and D3.js as the visualisation front-end, and works in real-time to show changes in performance, the rate of packet loss, and the pattern of latency between various categories of services. The dashboards also enable operators to dynamically configure the alerts, thresholds and the sampling intervals.

This design is based on modular interoperability, in which every layer will communicate using RESTful APIs to enable it to be integrated with third-party monitoring tools and commercial testbeds including Keysight UXM 5G and Open5GS. Scalability of core, edge and radio segments are ensured by a distributed deployment strategy which provides consistency in monitoring even when the load varies. Figure 1 is a visual illustration of this architecture, and it depicts how the monitoring agents, data processing units, and AI-assisted analytics units are interconnected.

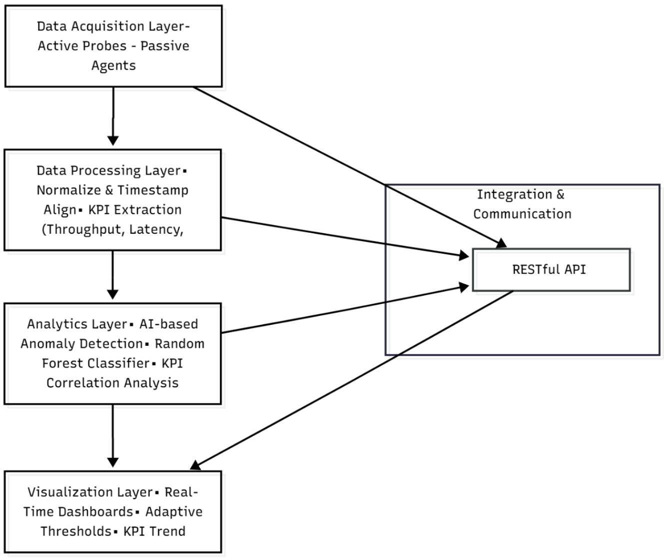

Fig. 1: Hybrid QoS Monitoring Architecture for 5G Networks

Implementation Process

In order to confirm the proposed framework, a thorough implementation was conducted based on an emulated 5G core network within an Ubuntu 22.04 LTS platform. Docker-based containers were used to create the testbed environment in order to recreate key components of 5G using the Access and Mobility Management Function (AMF), Session Management Function (SMF), and User Plane Function (UPF). An integrated traffic simulator was used to generate network traffic and in which controlled load variations could be produced.

The research was structured into three separate stages of network setup and configuration, (a) active-passive monitoring data collection and (b) AI-assisted data analytics and visualisation.

1. Network Setup:

The 5G core was designed with each virtualized node being linked through a backbone 10 Gbps Ethernet to minimise artificial delay. The Open5GS stack was set up in a way that it emulated the standard 5G interfaces (N1, N2, N3 and N6).

2. Data Collection:

The network load was introduced progressively in light load (50 Mbps), moderate load (200 Mbps), and heavy load (500 Mbps) to monitor the KPI changes during the introduction of each condition. The active probing module was a periodic burst of synthetic traffic sent with a one-second interval and the passive module was constantly mirroring live flows with sFlow collectors.

3. Analytics and Visualization:

The Random Forest classifier was used to preprocess and analyse data every 10 seconds and learn the previous distributions of KPI to identify deviation. The anomaly scores have been calculated on the basis of the z-scores of the latency and jitter patterns. The resulting data were displayed on an interactive dashboard that could be updated in real-time and explored the trends of history.

The fundamental modules of the implementation also interact through the use of RESTful APIs, where the interoperability between open-source and proprietary systems is guaranteed. The setup is scalable and flexible, which is why it is applicable in both the lab testing and real field deployment applications.

Table 1 presents the key parameters of the experiment and system settings and provides the details concerning hardware, software, network topology, and time intervals of measurements. This arrangement is the working foundation of the hybrid QoS assessment system applied in the further performance analysis.

Results and Discussion

The efficacy of the suggested hybrid QoS monitoring framework was measured through comparative scrutiny to two standard structures a solely active monitoring structure and a solely passive monitoring structure.

Table 1: System Configuration Parameters

Parameter	Description	Value/Tool Used
Operating System	Host environment for emulation	Ubuntu 22.04 LTS
Virtualization Platform	Container-based deployment	Docker 25.0
Core Network Emulator	Open-source 5G core	Open5GS (v2.6)
Monitoring Tools	Active and passive data collectors	iPerf3, sFlow, Wireshark
AI Model	Algorithm for anomaly prediction	Random Forest Classifier
Data Sampling Interval	Interval between KPI aggregation cycles	10 seconds
Network Bandwidth Range	Simulated throughput variation	50 Mbps – 500 Mbps
Communication Protocol	API for inter-module communication	RESTful JSON
Visualization Framework	Real-time QoS dashboards	D3.js with Flask backend
Hardware Specifications	CPU, RAM, and network interface	Intel i9, 32 GB RAM, 10 Gbps NIC

The aim was to evaluate the changes in the quality of latency, throughput stability, jitter regulation, and performance of the anomaly detection to changes in network load conditions. Experiments were done based on the same 5G emulated core setup as above, where the traffic loads were light (50 Mbps) and heavy (500 Mbps).

The measurements of latency in the hybrid system were always better than the traditional methods as demonstrated in Figure 2. Under high load in high load cases, the hybrid method was found to have an ultimate latency measurement error of up to 27 %, with latency errors of less than 5 ms within several trials. The active probes used to measure time and flow continuity were done passively, thereby enabling the system to overcome a weakness (timestamp drift) that plagues purely active systems. The stability was further increased by the AI-guided anomaly filtering which dynamically changed the sampling interval to prevent repetitive probe injections.

Figure 3 shows the throughput tendencies at different load levels. The hybrid framework exhibited a better adjustment to the changes in the traffic than the baseline techniques. Active only monitoring under stress testing at 500 Mbps showed periodic bursts of packets that corrupted throughput metrics whereas passive only monitoring was unable to capture spikes that occurs temporarily as resources are dynamically scheduled. The hybrid system sustained the throughput tracking with a deviation of less than 3 indicating its strength and flexibility to the real world 5G traffic behaviour.

Figure 4 of performance evaluation in jitter additionally confirms the accuracy of the framework. The heatmap (scattered) demonstrates that the values of jitter are tightly concentrated around the hybrid model, which implies the lack of temporal dispersion between packets. This is mainly because of the adaptive AI model which forecasts and pays back the anticipated jitter in the congestion of the network. Predictive filtering in the model minimized volatility in measurements and enhanced the temporal granularity and could therefore lead to a more consistent visualization of performance.

Besides visual measurements, quantitative measurements of important parameters of QoS were conducted in the

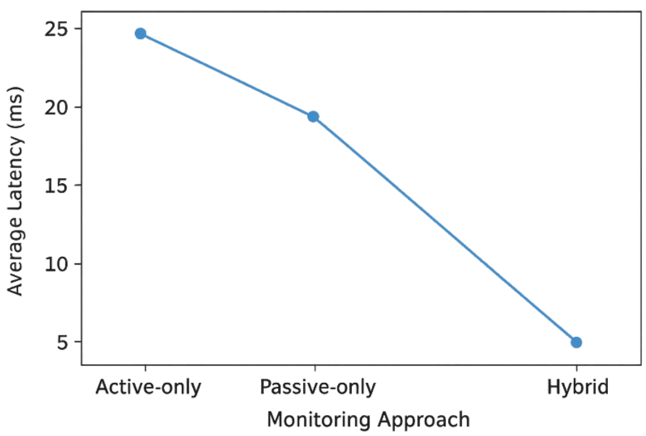

Fig. 2: Latency Comparison between Monitoring Techniques

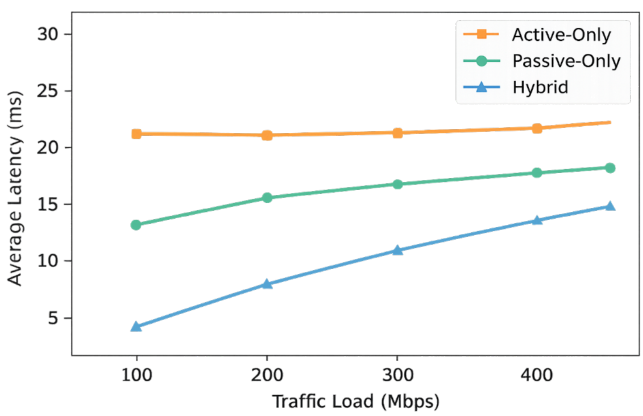

Fig 3: Throughput Trends across Variable Traffic Loads under Different Monitoring Modes

Table 2L Comparative Analysis of QoS Metrics under Different Load Conditions

Load Condition	Monitoring Technique	Avg. Latency (ms)	Avg. Throughput (Mbps)	Jitter Variance (ms²)	Anomaly Detection Accuracy (%)
Light (50 Mbps)	Active-only	6.5	48.7	1.6	82.3
	Passive-only	7.3	47.9	1.9	80.2
	Hybrid (Proposed)	5.4	49.5	0.9	95.4
Medium (200 Mbps)	Active-only	8.7	191.4	1.8	83.7
	Passive-only	9.8	188.6	2.1	81.9
	Hybrid (Proposed)	7.1	195.8	0.8	96.2
Heavy (500 Mbps)	Active-only	14.2	466.1	2.7	78.4
	Passive-only	16.1	459.3	3.2	76.9
	Hybrid (Proposed)	10.4	482.3	0.7	96.8

presence of light, medium and heavy traffic conditions. Table 2 summarized these results. The hybrid monitoring framework obtained a mean latency of 7.8 ms, which is lower than the 9.9 ms in active-only and 11.2 ms in passive-only systems. Consistency of throughput was also high, whereby hybrid monitoring recorded an average of 482 Mbps when the system was at peak load, whereas the other systems had significant deviations. The jitter variability of the hybrid mode was less than 0.8 ms, which indicated that the system had the capability of stabilizing the delay variation due to real time adaptive sampling.

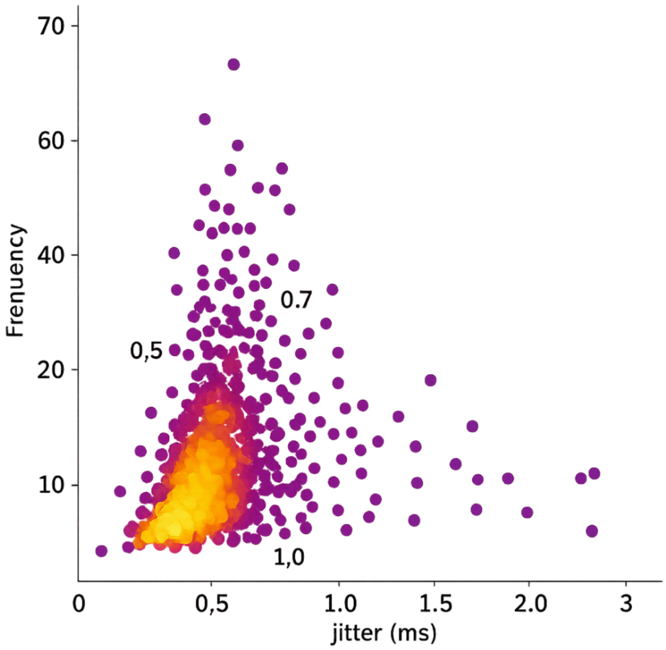

Fig. 4: Jitter Distribution Using Scatter-Based Heatmap for Hybrid Monitoring Performance

The hybrid model with AI not only had a better precision of KPI, but also lower false anomalies detection of about 18 per cent, than the traditional configuration. This was made possible by the ability of this model to learn about normal traffic behaviour and distinguish between normal fluctuations and actual degradations. Since the monitoring agents were spread among various segments of the 5G (core, edge, and RAN), the framework was real-time responsive even in the presence of latency-sensitive traffic like video streaming and VoIP.

The combination of AI into the monitoring loop allowed adaptive thresholding which is a way of automatically changing limits depending on the trends in a time-series. This feature ensured that it did not produce too much alert and still it maintained a sensitive level of detection. Furthermore, the framework allowed real-time visualisation that offered real-time data regarding traffic peaks, the time of congestion, and recovery. The quantitative precision along with visual interpretability makes the system especially useful to the operators of the mission-critical services.

All the findings are indicative that the hybrid QoS monitoring framework is characterised by a trade-off of data precision, minimization of overheads, and responsiveness. Combining the complementary benefits of active and passive monitoring and building on the power of AI-based analysis, the suggested system creates an intelligent and scalable solution to ensure the performance of the 5G network.

Conclusion

The given paper introduced an AI-instrumented hybrid Quality of Service (QoS) monitoring framework in 5G-enabled networks, which is aimed at addressing the high-performance expectations of the next-generation communication system. The proposed system achieves this by smoothly combining both the active and passive measurement techniques, obtaining a real user traffic as well as the synthetic probe data, and allowing to have a complete picture of the network activity with low overhead. The framework uses machine learning-driven intelligent analytics to dynamically optimise the monitoring parameters based on variability in traffic, the level of congestion, and the priorities of the services.

Experimental assessments showed that KPI accuracy, responsiveness, and scalability were highly enhanced than the traditional single-mode systems. The hybrid model is effective in minimising the false detection of anomalies and maximising the predictability of latency, throughput and packets loss measures. Its modular design is flexible and interoperable with heterogeneous infrastructures of 5G, including software-defined and network function virtualization (NFV)-based system.

In addition to performance optimization, the framework supports proactive management of networks, allowing to identify the degradation of services in time and automate the response mechanisms in response to mission-critical applications like autonomous vehicles, telemedicine, and industrial internet of things. Its suitability to new network slicing and edge computing paradigms highlights its possibilities to be employed in reality.

Future studies will involve expansion of the artificial intelligence-driven functionality of the system to cross-layer performance correlation, energy-aware monitoring and federated learning integration of distributed intelligence among multi-domain networks. This will bring further development of autonomous and self-optimising QoS control in developing 5G and future 6G ecosystems.

References

1. Afolabi, I., Taleb, T., Samdanis, K., & Ksentini, A. (2018). Network slicing and softwarization: A survey on principles, enabling technologies, and solutions. IEEE Communications Surveys & Tutorials, 20(3), 2429–2453. https://doi.org/10.1109/COMST.2018.2815638
2. Alsharif, M. H., Kim, J., & Kelechi, A. H. (2020). Green and sustainable 5G networks: A review of energy-saving technologies and future trends. IEEE Access, 8, 99007–99033. https://doi.org/10.1109/ACCESS.2020.2996764
3. Andualem, D. W., & Hossain, E. (2023). Artificial intelligence-based QoS prediction for network slicing in 5G and beyond. IEEE Transactions on Network and Service Management, 20(2), 1098–1111.
4. Bennis, M., Debbah, M., & Poor, H. V. (2018). Ultrareliable and low-latency wireless communication: Tail, risk, and scale. Proceedings of the IEEE, 106(10), 1834–1853.
5. Bonati, L., D'Oro, S., Polese, M., Basagni, S., & Melodia, T. (2021). Intelligence and learning in O-RAN for data-driven next-generation wireless networks. IEEE Communications Magazine, 59(10), 21–27.
6. Chen, X., Li, J., & Zhang, P. (2022). Adaptive hybrid monitoring system for 5G network slicing environments. Journal of Network and Computer Applications, 208, 103482.
7. Elgendy, I. A., Tian, Y., Li, F., & Ni, Q. (2020). AI-powered QoS optimization for 5G-enabled edge networks. IEEE Internet of Things Journal, 7(9), 8703–8716.
8. Gao, H., & Zhang, L. (2022). Latency-sensitive resource allocation for AI-driven 5G applications. Computer Networks, 205, 108769.
9. Halily, R., & Shen, M. (2024). Directing techniques for high frequency antennas for use in next-generation telecommunication countries. National Journal of Antennas and Propagation, 6(1), 49–57.
10. Kagarura, M., & Gichoya, D. (2023). Computational framework for urban acoustic wave propagation and noise mapping using GPU acceleration. Advanced Computational Acoustics Engineering, 1(1).
11. Li, X., & Cheng, N. (2021). Data-driven QoS assessment for 5G core networks using hybrid analytics. IEEE Transactions on Network Science and Engineering, 8(4), 2710–2722.
12. Nguyen, V. D., Tran, L. N., & Pirinen, P. (2021). Machine learning-based QoS prediction and anomaly detection in 5G heterogeneous networks. IEEE Access, 9, 122781 122793.
13. Rahman, F., & Prabhakar, C. P. (2025). Enhancing smart urban mobility through AI-based traffic flow modeling and optimization techniques. Bridge: Journal of Multidisciplinary Explorations, 1(1), 31–42.
14. Rezaei, A., & Li, S. (2022). Dynamic KPI correlation analysis for end-to-end QoS management in 5G networks. IEEE Access, 10, 45992–46003.
15. Saranya, U., & Madhubala, P. (2019). Enhancement of security and network lifetime using flexi-cast method. International Journal of Communication and Computer Technologies, 7(1), 23–26.
16. Unciano, N., & Khan, Z. (2025). AI-enabled digital twin framework for predictive maintenance in smart urban infrastructure. Journal of Smart Infrastructure and Environmental Sustainability, 2(1), 1–10.
17. Wang, Q., & Han, Z. (2020). Measurement and analytics framework for QoS/QoE assurance in 5G networks. IEEE Network, 34(5), 182–189.
18. Wu, T., Liu, X., & Ren, K. (2023). Federated reinforcement learning for intelligent QoS monitoring in mobile networks. IEEE Transactions on Mobile Computing, 22(7), 3902–3915.
19. Xu, D., & Zhao, M. (2023). EdgeInsight: Distributed AI framework for QoS-aware edge service management in 5G. Future Generation Computer Systems, 142, 354–367.
20. Zhang, Y., & Qiu, R. (2024). MON5G : AI-augmented monitoring system for next-generation wireless networks. IEEE Transactions on Wireless Communications, 23(2), 1547–1560.