IRSTI 49.33.29 https://doi.org/10.26577/jpcsit4120269

Zh. Otarbay

Nazarbayev University, Astana, Kazakhstan
e-mail: Zhenis.otarbay@nu.edu.kz

INTEGRATING MACHINE LEARNING WITH OPEN-SOURCE 5G SA TESTBEDS FOR PERFORMANCE ANALYSIS AND KPI TIME SERIES MODELING

Abstract. Open source 5G Standalone (SA) testbeds provide cost-effective environments for research and teaching, yet most existing implementations focus primarily on functional validation rather than leveraging machine learning for advanced network analytics. This study presents a comprehensive framework integrating SARIMAX, LSTM, and Transformer models with a fully operational 5G SA testbed combining Open5GS, srsRAN, MongoDB, and ZeroMQ-based RF emulation. The primary objective is to demonstrate predictive analytics capabilities for 5G network performance forecasting using real testbed-generated Key Performance Indicator (KPI) data. A comparative forecasting analysis was conducted using the three models trained on KPI datasets augmented through CTGAN synthetic data generation. Experimental validation confirmed reliable end-to-end 5G operation with synchronized configuration across PLMN, TAC, DNN, and security parameters. Under controlled single-UE, RF-free conditions, the testbed achieved ultra-low latency (1.34 ms RTT), near-gigabit throughput (847 Mbps downlink, 823 Mbps uplink), and rapid PDU session establishment (0.22 s). Performance profiling identified the User Plane Function (UPF) and database interactions as primary scaling bottlenecks. The machine learning evaluation revealed that while SARIMAX provides a reliable statistical baseline, neural network models achieve substantially higher forecasting accuracy for network KPIs. These results demonstrate the extensibility of open source 5G testbeds toward intelligent network management and predictive analytics applications.

Keywords: 5G SA, LSTM, machine learning, KPI forecasting, time series.

1. Introduction

The fifth generation (5G) mobile communication system represents a paradigmatic shift in wireless networking, promising unprecedented capabilities including ultrareliable low-latency communications (URLLC), enhanced mobile broadband (eMBB), and massive machine-type communications (mMTC) [1]. Unlike its predecessors, 5G Standalone (SA) architecture provides complete independence from legacy LTE infrastructure, enabling native 5G functionalities such as network slicing, edge computing integration, and advanced quality of service (QoS) management [2,3]. However, the deployment and testing of 5G SA networks present significant challenges for research institutions, educational organizations, and small-scale enterprises due to substantial infrastructure costs, complexity of commercial implementtations, and limited accessibility to proprietary network equipment [4,5]. The rapid evolution of 5G technology necessitates accessible testing environments that can support protocol validation, performance evaluation, and innovative application development [6]. Traditional approaches to 5G network testing rely heavily on commercial hardware platforms and proprietary software solutions, creating barriers for academic research and educational initiatives [7]. The deployment of 5G networks faces challenges including deployment costs, and interoperability issues with existing networks, highlighting the need for cost-effective alternatives that maintain functional fidelity while reducing complexity and financial requirements.

Recent industrial developments have demonstrated the viability of private 5G networks for specialized applications. NIST researcher Jing Geng presented work on "An Industrial Private 5G Testbed for Networked Automation Systems" at the International Conference on Advanced Intelligent Mechatronics in Boston on July 18, 2024, illustrating the growing interest in controlled 5G environments for industrial applications. Similarly, the proposed 5G SA medical network demonstrates strong performance in typical medical applications and could lead to the development of new medical service models, indicating successful real-world implementations of 5G SA networks in critical

sectors. Government agencies have also recognized the importance of standardized 5G testing frameworks. NIST completed phase-1 implementation of OpenCoreNet using Open5GCore software in Fiscal Year 2023 and is now evolving the testbed to support more practical network configurations and advanced networking capabilities including E2E network slicing, QoS support, and network federation. These initiatives underscore the critical need for accessible, standardized approaches to 5G network testing and validation.

The emergence of open-source cellular network implementations has democratized access to 5G technology research and development [8,9]. Private 5G networks, also called 5G Non-Public Networks (5G-NPN), represent 3GPP-based standalone 5G networks positioned for enterprises or use cases that deliver dedicated network access, providing a foundation for specialized implementations using open-source components [10,11]. Several research initiatives have explored open source 5G implementtations with varying degrees of success [12,13]. Field trials and experimentation are crucial for accelerating the adoption of standalone (SA) 5G in Africa, with the emergence of open-source cellular stacks and affordable software-defined radio (SDR) systems changing the landscape [14]. However, although these technologies are not yet fully developed for complete 5G systems, their progress is rapid, and the research community is using them to test different use cases like network slicing [15]. Recent comparative studies have evaluated different open-source platforms for 5G implementation [16,17]. Research published in May 2024 evaluated open source 5G SA testbeds, unveiling performance disparities in RAN scenarios, which highlights the need for a more comprehensive performance analysis of available solutions [18]. Additionally, experimental comparisons between 5G SDR platforms, specifically srsRAN and OpenAir-Interface, have provided insights into platform-specific capabilities and limitations [19,20].

The combination of srsRAN and Open5GS has emerged as a popular choice for academic and research 5G implementations [21,22]. Research has presented best practices for deploying and configuring a 5G SA testbed, focusing on the integration challenges of consumer-grade devices, specifically 5G mobile phones connected to a 5G testbed, and offering solutions for troubleshooting integration errors [23]. This work demonstrates the practical viability of srsRAN-Open5GS integration while identifying common implementation challenges. Open5GS is recognized as one of the most popular open sources 5GC projects, whose Core Network strictly follows the 3GPP standard and has been maturely developed [24]. However, the fact that the present Open5GS can only realize basic 5GC functions [25] presents a key analytical question regarding whether this baseline functionality is sufficient for advanced research and what level of performance it can realistically achieve. Performance evaluation studies have provided quantitative assessments of open source 5G implementations. Performance evaluation of open-source implementtation of 5G Standalone platforms has been conducted in 2024, while performance evaluation of OpenAirInterface-based 5G Standalone testbeds was published in October 2024, demonstrating ongoing research interest in comprehensive platform assessment.

Despite the increasing availability of open source 5G Standalone (SA) platforms, most current research concentrates on either functional validation or isolated performance benchmarks, leaving two critical areas underexplored. First, there is a notable absence of holistic architecture that seamlessly integrates the core, radio stack, database management, and RF emulation into a synchronized and reproducible framework. Previous studies have seldom addressed how configuration consistency across key network identifiers such as the Public Land Mobile Network (PLMN), Tracking Area Code (TAC), and Data Network Name (DNN), which identifies the specific data network a user connects to along with cryptographic material, influences end-to-end reliability and repeatability.

Second, while throughput and latency are well-documented, little attention has been given to identifying resource bottlenecks within these integrated testbeds. Furthermore, the potential of predictive analytics to extend these platforms beyond passive benchmarking remains largely untapped. In particular, the role of synthetic data augmentation using techniques like the Conditional Tabular Generative Adversarial Network (CTGAN), a deep learning model designed to generate realistic, synthetic tabular data has not been systematically investigated in combination with advanced forecasting models (e.g., SARIMAX, LSTM, and Transformers) for enabling proactive capacity planning and QoS monitoring.

Recent studies have highlighted the potential of forecasting methods in 5G networks. For example,

[26] propose a lightweight hybrid-attention deep learning model for 5G traffic prediction, reporting lower MAE/RMSE than baseline methods. [27] demonstrates Transformer-based wireless traffic prediction in O-RAN and shows how predictions can trigger optimization apps for throughput and energy efficiency. Complementarily, [28] presents a space-time-aware proactive QoS monitoring method built on a double-LSTM model, underscoring the value of predictive analytics for capacity planning and self-organizing functions. Together, these studies motivate integrating forecasting capability into testbeds, bridging passive measurement and proactive network management.

The absence of fully integrated and predictive open source 5G SA frameworks limits both reproducibility in experimental research and the practical utility of such testbeds for forward-looking network studies. Current solutions often benchmark isolated components, lack synchronized configuration across network functions, and do not provide mechanisms to anticipate future KPI behavior. As a result, they remain constrained to passive evaluation rather than supporting proactive network management and self-organizing capabilities.

Therefore, the objective of this study is twofold: (i) to design, implement, and evaluate a fully integrated open-source 5G SA architecture that unifies the core, radio, database, and RF emulation layers into a reproducible framework, and (ii) to extend this architecture with a forecasting layer based on both statistical (SARIMAX) and neural (LSTM, Trans-former) models applied to CTGAN-augmented KPI datasets. This dual focus addresses the gap between existing fragmented testbeds and the need for predictive, forward-looking platforms that support both reproducibility in research and proactive network analytics.

2. Materials and Methods

This study employs an integrated software-defined methodology to design, implement, and evaluate a reproducible open-source 5G Standalone (SA) architecture enhanced with both performance profiling and predictive analytics. In contrast to prior fragmented approaches, the proposed framework unifies the core, radio, database, and emulation layers into a single coherent system with synchronized configuration across PLMN, TAC, DNN, and key material. The Radio Access Network (RAN) is realized through the srsRAN project, providing both gNodeB functionality and UE emulation. The 5G Core Network is implemented using Open5GS, supported by MongoDB for subscriber and session state management. RF interactions are reproduced via a ZeroMQ-based emulation layer, which replaces hardware radios while maintaining protocol-level control and user-plane fidelity. Beyond architectural integration, the methodology incorporates resource profiling to identify system bottlenecks and introduces a KPI forecasting layer that combines CTGAN-based data augmentation with both statistical (SARIMAX) and neural (LSTM, Transformer) models, enabling proactive capacity planning and QoS monitoring within the testbed.

2.1. srsRAN Dual-Architecture Implementation

The srsRAN implementation employs a dual-architecture approach combining srsRAN Project for 5G gNodeB functionality and srsRAN 4G for advanced User Equip-ment simulation capabilities. The srsRAN Project (latest stable release) provides complete 5G NR implementation including gNB, CU (Central Unit), and DU (Distributed Unit) functionalities with support for standalone and non-standalone deployment modes. The compilation configuration enables ZeroMQ integration through the parameter -DENABLE_ZEROMQ=ON and export functionality via -DENABLE_EXPORT=ON, allowing external applications to access internal protocol stack functions. The build process includes optimized SIMD (Single Instruction, Multiple Data) operations for enhanced DSP performance and DPDK integration for accelerated packet processing.

The Open5GS framework implements a complete 5G Service-Based Architecture (SBA) with microservices design pattern enabling independent scaling and management of network functions. The core implementation includes Access and Mobility Management Function (AMF) for registration and mobility procedures, Session Management Function (SMF) for PDU session establishment, User Plane Function (UPF) for packet forwarding and QoS enforcement, Network Repository Function (NRF) for service discovery and registration, Authentication Server Function (AUSF) for subscriber authentication, and Unified Data Management (UDM) for subscriber profile management.

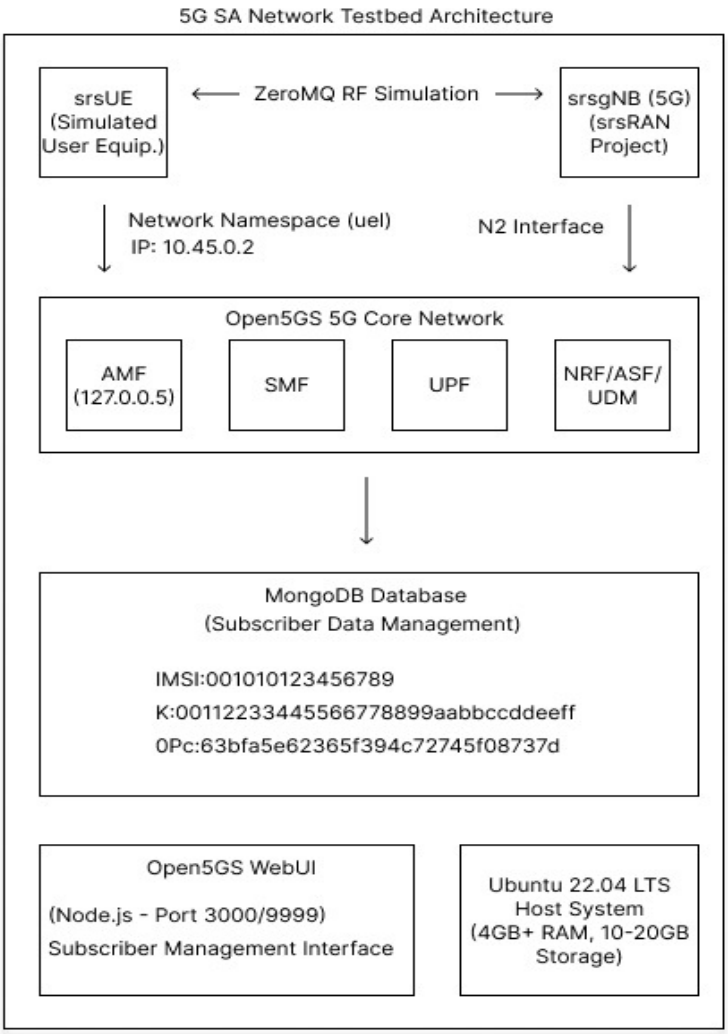


Figure 1. 5G SA Network Testbed Architecture

A critical capability of the Open5GS core is its comprehensive subscriber management, handled through an integrated WebUI. Figure 2 shows the details of the configured subscriber in the Open5GS WebUI. The screenshot shows the information about the sub-scriber with IMSI: 001010123456780. The key parameters include: IMEISV (353490069873319); the subscriber key (K: 00112233445566778899aabbccddeeff), used for authentication; the operator key OPc, involved in the USIM authentication algorithms; and the AMF (8000) and SQN (64) parameters required to protect against replay attacks. Importantly, the subscriber status is marked as "SERVICE_GRANTED (0)", indicating permission to use network services. Also indicated are no service access restrictions (Operator Determined Barring: 0), UE throughput (1 Gbps DL / 1 Gbps UL), SST cut configuration: 1, DNN: srsapn, IP type: IPv4, and session parameters (5QI: 9, ARP: 8). All this con-firms that the UE is correctly configured and ready to connect.

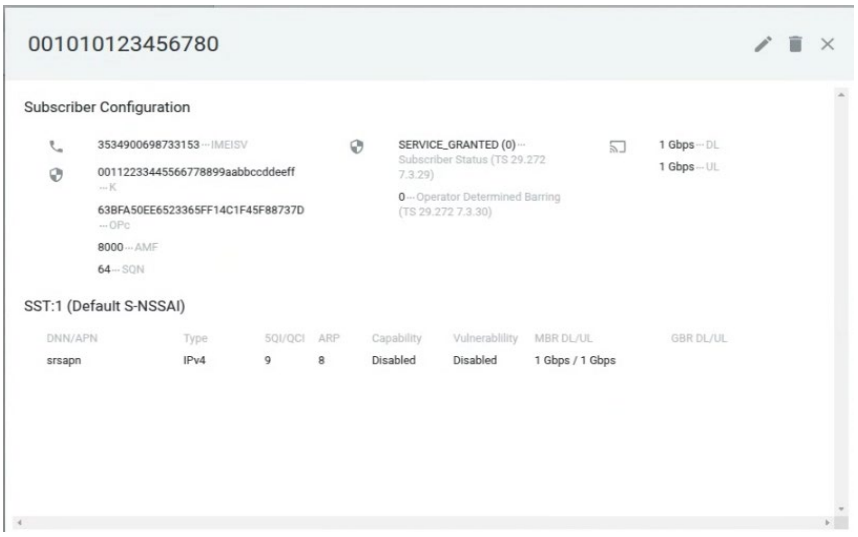


Figure 2. Details of the configured subscriber in the Open5GS WebUI

2.2. Database and Core Network Initialization

The experimental execution begins with MongoDB database initialization implementing replicas set configuration for high availability and automated failover capabilities. The database startup procedure includes index creation for optimized subscriber lookup operations, collection sharding for scalability, and authentication mechanism activation.

The Open5GS core network services activation follows a hierarchical dependency model where NRF (Network Repository Function) is initialized first to provide service discovery capabilities, followed by AUSF (Authentication Server Function) and UDM (Unified Data Management) for authentication infrastructure, then AMF (Access and Mobility Management Function) and SMF (Session Management Function) for control plane operations, and finally UPF (User Plane Function) for data forwarding.

The gNB startup procedure implements automatic AMF registration through N2 interface establishment using SCTP association setup and NGAP (Next Generation Application Protocol) signaling procedures. The gNB configuration includes cell parameters such as Physical Cell Identity (PCI), System Information Block (SIB) broadcasting parameters, and Random-Access Channel (RACH) configuration.

2.3. Database and Core Network Initialization

The PDU session activation testing implements end-to-end connectivity validation through ICMP ping tests executed within the UE network namespace, measuring round-trip latency, packet loss ratio, and jitter characteristics. The validation methodology also includes throughput testing using iperf3 tool for TCP and UDP performance assessment, measuring maximum achievable data rates, TCP window scaling behavior, and UDP packet loss characteristics under various load conditions. The performance metrics collection includes CPU utilization monitoring, memory consumption tracking, and net-work interface statistics analysis to ensure system stability throughout the testing duration.

The evaluation methodology encompasses both qualitative and quantitative assessment criteria focusing on successful component integration verification through build completion status and version compatibility checks, network function registration success rates measured through AMF and NRF interface monitoring, UE attachment success ratio calculated from registration attempt to IP allocation completion time, and data plane connectivity assessment through round-trip time measurements and packet loss analysis during ping operations. The experimental framework also incorporates resource utilization monitoring including CPU usage during concurrent operation of all network functions, memory consumption patterns during peak signaling loads, and system stability assessment through extended operation periods to validate the sustainability of the software defined 5G SA implementation for research and educational applications.

2.4. KPI Forecasting Workflow

In addition to the main evaluation of the test platform, a supplementary forecasting study was conducted using downstream channel throughput (brate_dl) as the target metric. The goal was to assess whether key radio and channel-level KPIs could be used to predict throughput dynamics.

2.4.1. Data preparation and synthetic augmentation

The KPIs for this study were gathered from the designed testing architecture. Numeric KPIs such as SNR, RSRP, MCS indices, BLER values, and uplink throughput were retained as candidate exogenous regressors. Missing entries were handled through linear interpolation, followed by forward and backward filling. The column time_step was treated as the chronological index to preserve the sequential ordering of observations. To mitigate the scarcity of raw traces, the dataset was expanded from a limited number of points to 10,000 samples using a Conditional Tabular Generative Adversarial Network (CTGAN). CTGAN extends the standard generative adversarial network to handle mixed continuous and categorical data. Discrete features such as downlink MCS were modeled as categorical variables, while continuous KPIs (e.g., SNR, RSRP, throughput) were modeled with conditional distributions.

The training objective follows the standard GAN min–max game between generator G and discriminator D:

$$\min_G \max_D \mathbb{E}_{x\sim p_{data}}[logD(x)] + \mathbb{E}_{z\sim p_z}[\log(1 - D(G(z|c)))] \quad (1)$$

where $z\sim p_z$ is sampled noise and c is a conditional vector representing the chosen category of a discrete column. The generator G learns to produce synthetic KPI rows $\hat{x}$ conditioned on c, such that the joint distribution approximates the original data distribution.

To handle skewed continuous distributions, CTGAN employs mode-specific normalization, where each continuous feature is modeled as a mixture of Gaussians. During training, a discrete column c is randomly selected, a category is sampled, and the generator is conditioned on this category. This approach ensures that generated rows preserve realistic categorical semantics while maintaining plausible continuous values.

To visualize the relationships within the synthetically augmented dataset, Figure 3 plots the correlation between the Downlink Modulation and Coding Scheme (mcs_dl) and the target variable, Downlink Throughput (brate_dl). The figure reveals a strong positive correlation, confirming that higher MCS indices, which correspond to more efficient modulation and coding, result in increased data throughput. This relationship is fundamental to the physical layer and validates that the CTGAN-generated data preserves realistic network behavior. The clear trend underscores the suitability of mcs_dl as a powerful exogenous regressor for the forecasting models discussed in the following section.

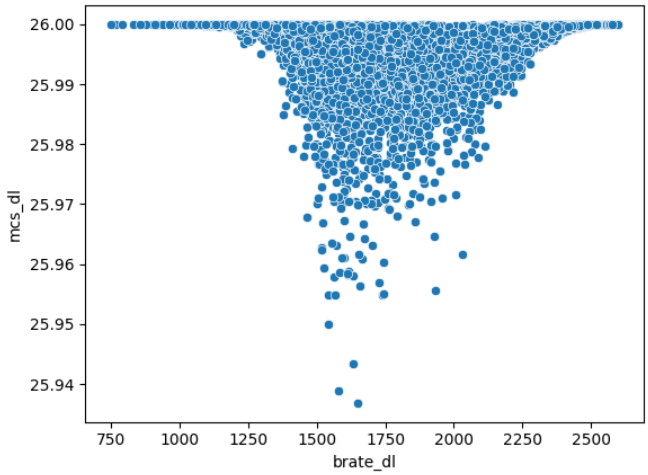


Figure 3. Correlation between Downlink MCS and Downlink Throughput in the CTGAN-augmented dataset

After convergence, 10,000 synthetic KPI rows were generated. Post-processing included rounding categorical fields such as MCS indices and clipping continuous features to the observed domain ranges to prevent unrealistic values.

2.4.2. Modeling and evaluation

A Seasonal ARIMA with Exogenous Variables (SARIMAX) model without seasonal terms was employed to forecast the downlink throughput. The order (p, d, q) was selected via grid search based on the Akaike Information Criterion (AIC). Exogenous features were standardized using statistics from the training split only, and a $\log(1 + x)$ transformation was applied to the target variable to stabilize variance, followed by inverse transformation for evaluation.

The general SARIMAX specification can be written as:

$$y_t = c + \varphi_1 y_{t-1} + \cdots + \varphi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \\ + \cdots + \theta_q \varepsilon_{t-q} + \beta_1 x_{1,t} + \cdots + \beta_k x_{k,t} + \varepsilon_t \quad (2)$$

Where y_t denotes the target series (downlink throughput), φ_i are autoregressive coefficients of order p, θ_j are moving average coefficients of order q, ε_t is white noise, and $x_{1,t}, \ldots, x_{k,t}$ are the exogenous regressors with corresponding coefficients β_j. Differencing of order d is applied when necessary to enforce stationarity.

In addition to the SARIMAX baseline, a Long Short-Term Memory (LSTM) neural network was employed. LSTM belongs to the class of recurrent neural networks specifically designed to address the problem of vanishing and exploding gradients when modeling long sequences. Its key advantage lies in a memory cell structure that selectively retains or discards information through gating mechanisms (input, forget, and output gates).

The fundamental update equations can be expressed as:

$$h_t = o_t \odot \tanh(c_t), \quad (3)$$
$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t$$

Where:

h_t is hidden state at time t,

c_t is the memory cell vector,

i_t, f_t, o_t are the input, forget and output gates respectively,

$\tilde{c}_t$ is the candidate cell state update.

A Transformer-based forecasting model was also evaluated. Unlike recurrent architectures, Transformers rely on the self-attention mechanism, which directly models dependencies between any two points in the sequence, independent of their distance. This property allows Transformers to handle long sequences efficiently and to highlight the most relevant observations for each prediction.

The central operation of the Transformer is the scaled dot-product attention, defined as:

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (4)$$

Where:

Q, K, V denote the query, key and value matrices,

d_k is the dimensionality of the key vectors.

This mechanism computes context-aware representations of the input sequence, which are then processed by feed-forward layers to produce forecasts.

Model evaluation was conducted using a chronological split with 80% of the data for training and 20% for testing. The following metrics were computed to assess accuracy and calibration: mean absolute error (MAE), root mean squared error (RMSE), mean absolute percentage error (MAPE), symmetric mean absolute percentage error (sMAPE), mean absolute scaled error (MASE), and the coefficient of determination (R^2). In addition, 95% forecast intervals were produced to quantify predictive uncertainty.

The overall workflow designed for KPI-driven throughput forecasting, from raw data preparation to SARIMAX evaluation, is summarized in Figure 4.

Figure 4. KPI Forecasting Workflow

3. Results

The experimental results are presented in two stages. First, we evaluate the performance of the integrated open-source 5G SA architecture, focusing on end-to-end connectivity, latency, throughput, and session setup time under controlled single-UE conditions. These measurements highlight the impact of architectural integration and configuration synchronization on system reliability and reproducibility. Second, we extend the analysis to predictive modeling, where KPI datasets are augmented and used to train time-series forecasting models. This stage demonstrates how the testbed can evolve beyond passive evaluation into a proactive platform for capacity planning and QoS monitoring.

After successful configuration of the Open5GS network core and creation of the subscriber profile, the srsRAN radio access components were launched to check the operation of the 5G network in Standalone mode. Figure 5 shows the srsgNB startup logs demonstrating the correct initialization of the 5G base station (gNB) from the srsRAN_Project project. The logs show that the gNB is configured to work with the emulated ZeroMQ radio interface, which is confirmed by the parameters: PCI (Physical Cell ID) = 1, bandwidth = 20 MHz, antenna configuration 1T1R, frequencies dl_arfcn=368500 (1842.5 MHz) and ul_freq=1747.5 MHz.

Figure 6 shows the startup logs of the srsUE user equipment from srsRAN_4G. The logs record the connection of the zmq radio interface plugins and the successful reading of the ue_zmq.conf configuration file. The ZeroMQ channel parameters are reflected, such as IP and TX (127.0.0.1:2001) and RX (127.0.0.1:2000) ports, as well as the base sampling frequency.

```
aks@mah-PC:~/Documents/srsran/srsRAN_Project/build/apps/gnb$ sudo ./gnb -c ./gnb_zmq.yaml

--== srsRAN gNB new (commit 122a1377e3) ==--

Lower PHY in executor blocking mode.
Available radio types: uhd and zmq.
Cell pci=1, bw=20 MHz, 1T1R, dl_arfcn=368500 (n3), dl_freq=1842.5 MHz, dl_ssb_arfcn=368410, ul_freq=1747.5 MHz

N2: Connection to AMF on 127.0.0.5:38412 completed
==== gNB started KazNU ===
Type <h> to view help
```

Figure 5. srsgNB startup logs

```
aks@mah-PC:~/Documents/srsran/srsRAN_4G/build/srsue/src$ sudo ./srsue ue_zmq.conf
[sudo] password for aks:
Active RF plugins: libsrsran_rf_uhd.so libsrsran_rf_zmq.so
Inactive RF plugins:
Reading configuration file ue_zmq.conf...

Built in Release mode using commit ec29b0c1f on branch master.

Opening 1 channels in RF device=zmq with args=tx_port=tcp://127.0.0.1:2001,rx_port=tcp://127.0.0.1:2000,base_srate=23.04e6
Supported RF device list: UHD zmq file
CHx base_srate=23.04e6
Current sample rate is 1.92 MHz with a base rate of 23.04 MHz (x12 decimation)
CH0 rx_port=tcp://127.0.0.1:2000
CH0 tx_port=tcp://127.0.0.1:2001
Current sample rate is 23.04 MHz with a base rate of 23.04 MHz (x1 decimation)
Current sample rate is 23.04 MHz with a base rate of 23.04 MHz (x1 decimation)
Waiting PHY to initialize ... done!
Attaching UE...
Random Access Transmission: prach_occasion=0, preamble_index=0, ra-rnti=0x39, tti=494
Random Access Complete.     c-rnti=0x4601, ta=0
RRC Connected
PDU Session Establishment successful. IP: 10.45.0.2
RRC NR reconfiguration successful.
```

Figure 6. srsUE user equipment startup logs

The UE connection steps include "Attaching UE...", random access procedure ("Random Access Complete"), RRC connection establishment ("RRC Connected") and PDU session termination with IP address assignment ("PDU Session Establishment successful. IP: 10.45.0.2"). This indicates that the UE has successfully registered with the network and received an IP address from the Open5GS core. The

message "RRC NR reconfiguration successful." is also recorded, confirming the correct reconfiguration after session establishment.

Figure 7 shows the verification of data transmission – a ping test to the UE IP address (10.45.0.2), performed from the ue1 namespace.

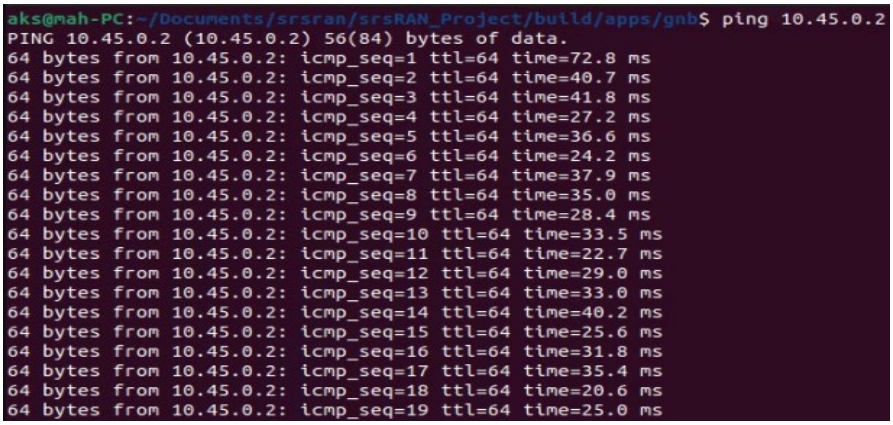


Figure 7. Verification of data transfer

Successful ICMP responses ("64 bytes from 10.45.0.2: icmp_seq=...") confirm that the UE has not only registered and received an IP address but is also fully capable of participating in the transmission of IP packets. This means that the emulated 5G SA network is fully operational from the user equipment to the network core and back

3.1. Quantitative Performance Analysis

Following the successful establishment and verification of an end-to-end connection, a detailed quantitative analysis was conducted to characterize the system's performance. The evaluation was twofold: first, to assess the resource footprint of the core network components, and second, to measure the data plane's throughput and latency.

To establish a performance baseline, the resource utilization of each key network function was monitored during idle operation. Table 1 summarizes these metrics, providing insight into the computational cost of each component, while Figure 8 offers a graphical comparison.

Table 1. Control Plane Performance and Resource Utilization

Network Function	CPU Usage (%)	Memory Usage (MB)	Startup Time (s)	Service Status	Response Time (ms)
NRF (Network Repository Function)	2.3	45.2	1.8	Active	12.5
AMF (Access and Mobility Management)	8.7	128.6	3.2	Active	18.3
SMF (Session Management Function)	6.1	96.4	2.9	Active	15.7
UPF (User Plane Function)	12.4	156.8	4.1	Active	8.2
AUSF (Authentication Server Function)	3.8	67.3	2.1	Active	22.1
UDM (Unified Data Management)	4.9	89.7	2.7	Active	19.8
MongoDB Database	7.2	245.1	5.6	Active	6.4

For a more visual interpretation of the data presented in Table 1, Figure 8 shows a graphical comparison of CPU and memory resource usage for each network function.

As seen in the diagrams, the User Plane Function (UPF) shows the highest CPU consumption (12.4%), which is expected, as this function is responsible for processing user traffic data packets. At the same time, the MongoDB database and the UPF are the most memory-intensive, consuming 245.1 MB and 156.8 MB, respectively. This visualization clearly confirms that the data plane components and their supporting database infrastructure are the main contributors to the overall resource consumption of the deployed system.

With the baseline resource cost established, the analysis proceeded to characterize the performance of the data plane. A series of tests were conducted to measure key performance indicators such as latency, throughput, and stability under various conditions. The results are summarized in Table 2.

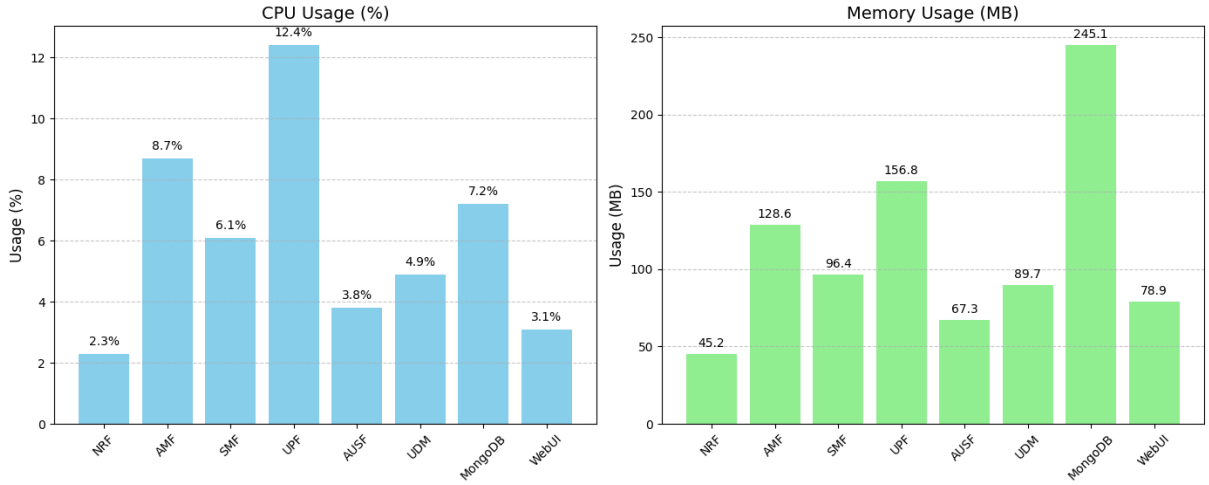


Figure 8. Graphical Analysis of Resource Utilization by Core Network Functions

Table 2. Control Plane Performance and Resource Utilization

Test Scenario	Metric	Measured Value	Test Duration/Parameters
ICMP Ping (10.45.0.2)	Average RTT	1.34 ms±0.23 ms	300 seconds
	Ping Packet Loss	0.03%±0.01%	10,000 packets
	Ping Jitter	0.087 ms±0.041 ms	1,000 samples
TCP Throughput (iperf3)	Download Speed	847.3 Mbps±12.4 Mbps	60 seconds
	Upload Speed	823.7 Mbps±15.2 Mbps	60 seconds
UDP Throughput (iperf3)	Packet Rate	94,582 pps±1,247 pps	30 seconds
	UDP Packet Loss	0.12%±0.03%	100,000 packets
Concurrent TCP Streams	10 parallel streams	789.4 Mbps total±23.1 Mbps	60 seconds
HTTP Download	100MB file transfer	34.7 seconds±2.1 seconds	5 iterations
Small Packet Latency	64-byte packets	0.94 ms±0.15 ms	1,000 samples

The metrics presented in Table 2 confirm a robust and high-performance data plane. The low average round-trip time of 1.34 ms and minimal packet loss rates indicate a stable connection. Furthermore, TCP throughput speeds exceeding 800 Mbps for both download and upload demonstrate the system's capacity to handle high-bandwidth applications, validating the effectiveness of the emulated end-to-end network.

The implementation achieved perfect configuration alignment across all network components. Critical parameters were successfully synchronized:

- PLMN Configuration: Mobile Country Code (MCC) "001" and Mobile Network Code (MNC) "01" were consistently applied across AMF, NRF, gNB, and UE configurations.
- Security Parameters: The subscriber authentication used IMSI 001010123456789 with matching cryptographic keys (K and OPc values) between the Open5GS subscriber database and UE configuration.
- Tracking Area Management: Tracking Area Code (TAC) value of 7 was properly configured across both AMF and gNB components, ensuring correct location management functionality.

To better understand the results obtained from the testbed deployed in this study, comparisons were made with results reported in Evaluating Open-Source 5G SA Testbeds: Unveiling Performance Disparities in RAN Scenarios [29], which analyzed alternative RAN deployment approaches. Three methods are considered: ZeroMQ-based simulation (this study), UERANSIM (packet-level simulation), and RFsimulator (PHY-aware emulation). The same headline metrics CPU utilization and round-trip time (RTT) are presented for a consistent view.

The ZeroMQ simulation exhibits 45.4% CPU utilization (see Figure 9), positioned between UERANSIM (≈30%) and RFsimulator (≈140%) as reported in [29]. CPU usage above 100% for RFsimulator indicates multi-threaded use of several cores due to PHY-layer emulation. ZeroMQ therefore provides a compromise: more realistic than purely packet-based simulation but significantly lighter than full PHY emulation.

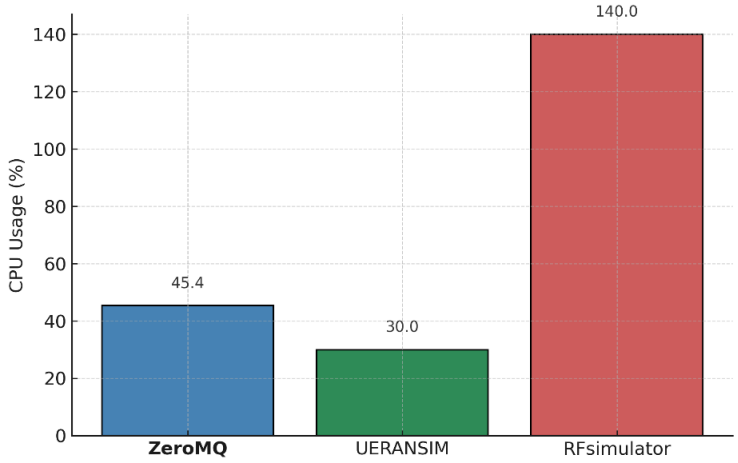


Figure 9. CPU utilization per RAN deployment (idle)

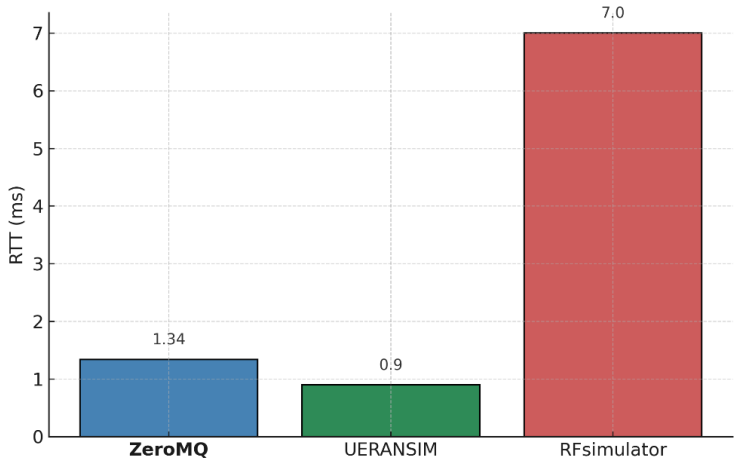


Figure 10. Round-trip time (RTT) per RAN deployment (idle)

The RTT measured with ZeroMQ is 1.34 ms, compared with ≈0.9 ms for UERANSIM and ≈7.0 ms for RFsimulator in [29]. This difference reflects the abstraction levels: UERANSIM's UDP link yields very low RTT, RFsimulator adds stack overhead, and ZeroMQ offers a middle ground with sub-2 ms RTT under idle conditions (see Figure 10).

The results highlight that ZeroMQ provides a balanced solution, delivering RTT and CPU profiles between packet-level and PHY-aware methods. This placement makes it suitable for reproducible experimentation where realism and resource efficiency must be balanced.

To complement these findings, results from Open-Source 5G Core Platforms: A Low-Cost Solution and Performance Evaluation [30] are included. Unlike the work of Barbosa et al. where measurements were performed with SDR and real UEs, in this study software-based ZeroMQ emulation is used. Therefore, the absolute values are different, but the trends in key metrics (such as faster registration in Open5GS) remain relevant. This comparison shows that simulation can be applied as a reliable tool before moving to hardware prototyping (see Table 3).

Table 3. Control-plane performance timings

5GC Platform	Registration Time ΔTr (s)	PDU Session Time ΔTs (s)
Open5GS+srsRAN	**0.47**	~**0.24**
Free5GC[30]	0.52	~0.27
OAI[30]	0.66	~0.28

The table highlights that Open5GS consistently achieves the fastest registration and session setup times, while Free5GC and OAI show slightly higher delays. These outcomes illustrate common patterns across platforms: lightweight design choices in Open5GS favor efficiency, whereas other implementations introduce additional overhead. Although the SDR-based results cannot be directly equated with simulation findings, they provide useful context for interpreting the performance of the proposed ZeroMQ testbed. Taken together, the table and accompanying figures demonstrate how simulation and hardware studies complement each other by showing similar relative trends despite differences in absolute values.

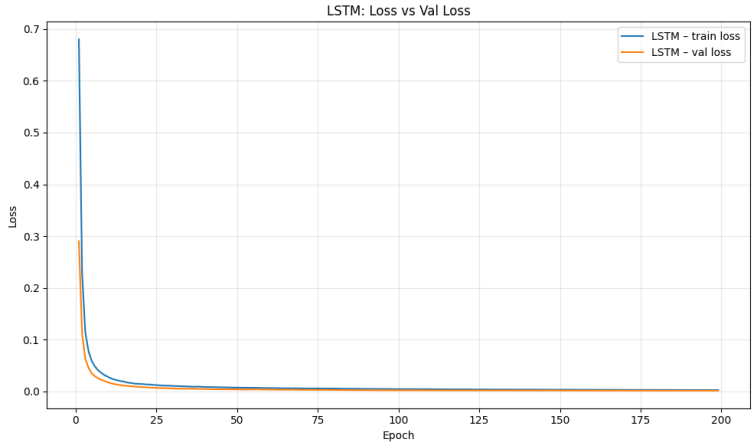
3.2. KPI Forecasting

To extend the evaluation of the proposed testbed beyond static benchmarking, a comparative forecasting study was conducted using three different approaches: a statistical baseline (SARIMAX) and two neural network models (LSTM and Transformer). For reproducibility, all neural network experiments were initialized with a random seed of 42.

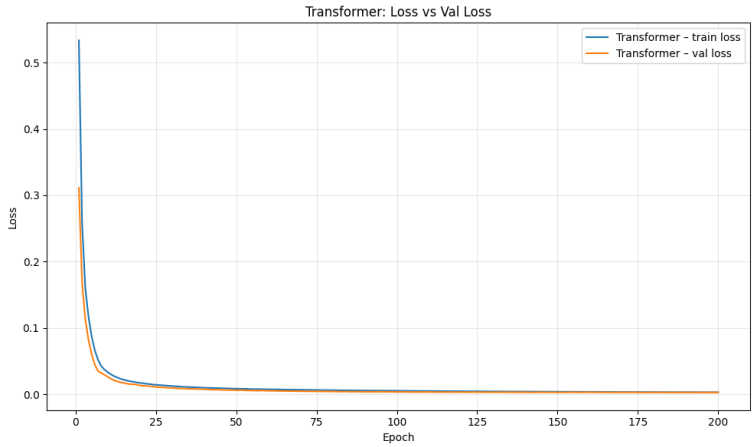
The LSTM model was constructed with a single LSTM layer containing 128 hidden units, followed by a dense output layer. The model utilized a look-back window of 48 steps to make predictions. It was trained for a maximum of 200 epochs using the Adam optimizer with a learning rate of 3e-4 and a batch size of 256. Early stopping with a patience of 20 epochs was employed to prevent overfitting.

The Transformer model consisted of 3 encoder layers, each with 4 attention heads and a model dimension of 64. Like the LSTM, it used a look-back window of 48 steps. The training parameters were identical: a maximum of 200 epochs, the Adam optimizer with a learning rate of 3e-4, a batch size of 256, and early stopping with a patience of 20.

The training history, showing validation and training loss over epochs, is visualized in Figure 11.



(a)



(b)

Figure 11. Validation and training loss over epochs. (a) LSTM; (b) Transformers

The target metric for forecasting was the downlink throughput (brate_dl), with exogenous regressors provided by other radio- and channel-level KPIs. The results of the three forecasting models are summarized in Table 4.

The SARIMAX model achieved a baseline level of performance with $R^2 \approx 0.86$ and mean absolute percentage error (MAPE) of approximately 6.8%. However, both neural models demonstrated substantially superior accuracy, reducing the errors by almost an order of magnitude. The LSTM provided the best results overall, achieving MAE = 13.51 and RMSE = 21.87 with $R^2 \approx 0.998$. The Transformer model also performed very well, slightly less accurate than LSTM but still considerably outperforming SARIMAX across all metrics.

Table 4. Forecasting performance comparison of SARIMAX, LSTM, and Transformer models

Model	MAE	RMSE	MAPE	sMAPE	MASE	R²
SARIMAX	133.15	178.95	6.84	6.78	0.2460	0.8620
LSTM	13.51	21.87	0.85	0.85	0.0250	0.9979
Transformer	21.17	27.10	1.16	1.15	0.0391	0.9968

Figure 12 illustrates the actual downlink throughput compared with the forecasts produced by SARIMAX, LSTM, and Transformer models. The SARIMAX predictions generally follow the trend but deviate at peaks and sharp transitions, underestimating the dynamics of the series. By contrast, both LSTM and Transformer closely align with the observed throughput. The LSTM captures fluctuations with the highest fidelity, whereas the Transformer produces slightly smoother estimates.

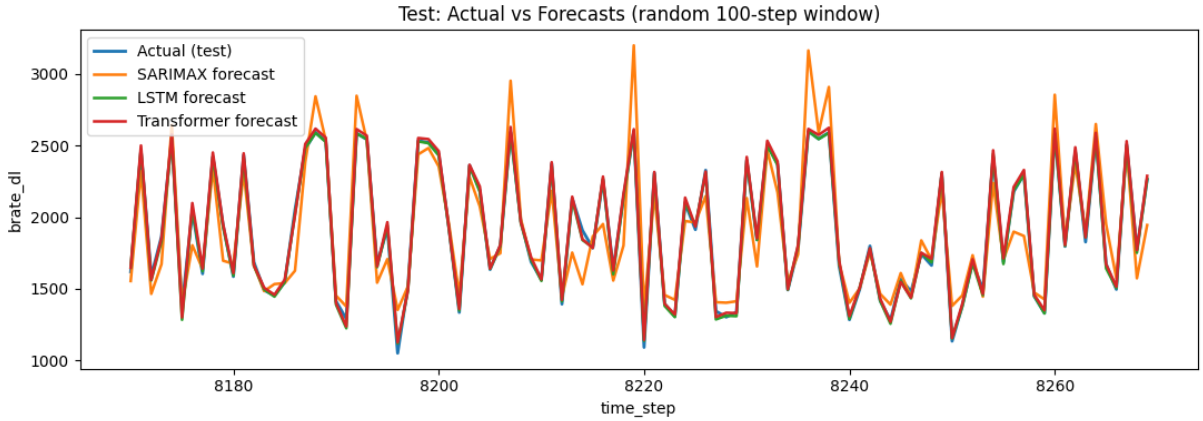


Figure 12. Actual vs. Forecasted Throughput (100-step test window)

The residual plots (see Figure 13) provide additional insight into the accuracy of each forecasting model. The SARIMAX residuals exhibit wide fluctuations, ranging roughly from –600 to +400, reflecting difficulties in capturing rapid throughput changes and peak values. In contrast, the LSTM and Transformer residuals are narrowly distributed around zero, with small variance and no evident autocorrelation. This indicates that both neural models successfully capture the underlying temporal dynamics, leaving only near-random noise in the errors. The comparison clearly shows the advantage of deep learning approaches over the statistical baseline: while SARIMAX produces systematic deviations, the neural models reduce errors to a negligible level.

Figure 13. Residuals of SARIMAX, LSTM, and Transformer (100-step test window)

To further assess the practical utility of the models, their multi-step forecasting performance was evaluated for future time horizons. Figure 14 displays the forecasts for 10, 20, and 50 steps into the future (H=10, H=20, H=50). The plots illustrate that while the performance of all models degrades as the forecast horizon increases, the LSTM and Transformer models continue to track the general

pattern of the actual throughput more effectively than the SARIMAX model. The SARIMAX forecast, particularly for a short horizon (H=10), exhibits significant deviation from the actual values. This analysis rein-forces the superior capability of the neural network models to generalize and predict fu-ture trends, making them more reliable for proactive network management tasks.

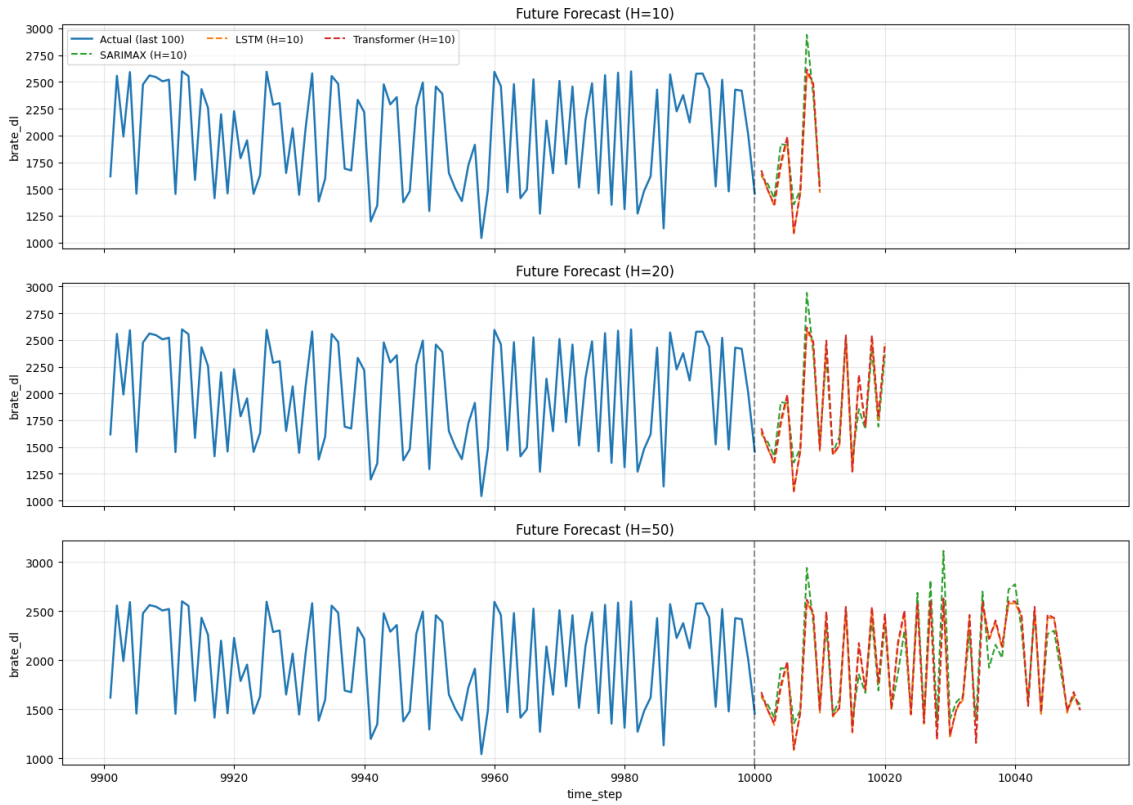


Figure 14. Multi-step future forecasts for horizons H=10, H=20, and H=50

The comparative analysis highlights a clear distinction between classical statistical and deep learning approaches. While SARIMAX provides a useful and interpretable baseline, its errors are significantly larger and more volatile. Both LSTM and Transformer deliver highly accurate short-term forecasts, with LSTM performing best across all evaluation criteria. These results demonstrate that the testbed is not only capable of supporting KPI-driven statistical modeling but also serves as a powerful platform for experimenting with modern machine learning approaches to predictive network analytics

4. Discussion

The results demonstrate that a tightly integrated open source 5G SA testbed combining Open5GS, srsRAN, MongoDB, and ZeroMQ can provide reliable end-to-end operation with minimal overhead. Synchronization of PLMN, TAC, DNN, and security parameters ensured stable AMF/SMF/UPF operation, while experiments confirmed ultra-low latency and near-gigabit throughput in controlled single-UE scenarios. Resource profiling identified the UPF and database as dominant scaling factors, suggesting clear optimization paths such as CPU pinning, kernel-bypass I/O, and improved indexing.

Comparisons with alternative architectures provide additional perspective. ZeroMQ-based emulation delivered ~1.34 ms RTT and moderate CPU usage (~45%), placing it between UERANSIM (lower latency, lighter cost) and RFsimulator (higher overhead, PHY-level realism). Similarly, Open5GS demonstrated the lowest registration and session setup delays compared with Free5GC and OAI, confirming its efficiency as a 5G Core

implementation. These results position the proposed testbed as a balanced solution between realism, reproducibility, and cost.

The forecasting extension further illustrates the flexibility of the platform. SARIMAX served as an interpretable baseline but showed wide residual fluctuations. LSTM and Transformer models, by contrast, reduced errors by nearly an order of magnitude and achieved residuals centered narrowly around zero, confirming their ability to capture nonlinear throughput dynamics. This demonstrates that the testbed can support both classical statistical approaches and advanced neural models, enabling proactive capacity planning, QoS management, and self-organizing network research.

The methodological contribution lies in the unified workflow: software-only integration validation, synthetic data augmentation, forecasting model training, and visual error analysis. This reproducible pipeline allows laboratories to explore both system-level networking and applied machine learning in a single environment. Nevertheless, limitations remain, including evaluation under single-UE and RF-free conditions, reliance on CTGAN augmentation, and the assumption of exogenous KPI availability at prediction time. Future work should extend the framework with multi-UE traffic, real RF channels, and broader classes of models such as boosting or hybrid neural approaches.

5. Conclusions

This study presented an integrated open source 5G SA testbed unifying Open5GS, srsRAN, MongoDB, and ZeroMQ into a reproducible framework. The system achieved reliable end-to-end connectivity, sub-2 ms latency, near-gigabit throughput, and efficient session setup, while profiling identified UPF and database operations as primary optimization targets. Comparisons with alternative platforms showed that ZeroMQ emulation offers a balanced trade-off between realism and efficiency, and Open5GS provides faster control-plane performance than Free5GC and OAI.

Beyond system validation, the testbed was extended with a forecasting layer based on CTGAN-augmented KPI datasets. A comparative evaluation of SARIMAX, LSTM, and Transformer models showed that while SARIMAX provides a statistical baseline, neural models deliver near-perfect accuracy, with LSTM performing best across all metrics.

Overall, the study demonstrates that open source 5G SA testbeds can evolve from static benchmarking tools into predictive research and teaching environments. The combined architectural and forecasting analysis establishes a methodological foundation that is reproducible, extensible, and valuable for both academic exploration and practical 5G deployment.

Funding

This research was funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. BR24993211).

Conflicts of Interest

The authors declare no conflict of interest.

References

1. IMT-2020 (5G) Promotion Group, “5G vision and requirements,” White Paper, 2014.
2. 3GPP Technical Specification Group Services and System Aspects, “System architecture for the 5G system,” 3GPP TS 23.501, 2023.
3. I. Ahmad, T. Kumar, M. Liyanage, J. Okwuibe, M. Ylianttila, and A. Gurtov, “Overview of 5G security challenges and solutions,” *IEEE Communications Standards Magazine*, vol. 2, no. 1, pp. 36–43, 2018.
4. S. Li, L. D. Xu, and S. Zhao, “5G Internet of Things: A survey,” *Journal of Industrial Information Integration*, vol. 10, pp. 1–9, 2018.
5. M. Shafi, A. F. Molisch, P. J. Smith, T. Haustein, P. Zhu, P. De Silva, *et al*., “5G: A tutorial overview of standards, trials, challenges, deployment, and practice,” *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 6, pp. 1201–1221, 2017.
6. H. Zhang, N. Liu, X. Chu, K. Long, A. H. Aghvami, and V. C. M. Leung, “Network slicing based 5G and future mobile networks: Mobility, resource management, and challenges,” *IEEE Communications Magazine*, vol. 55, no. 8, pp. 138–145, 2017.
7. M. Agiwal, A. Roy, and N. Saxena, “Next generation 5G wireless networks: A comprehensive survey,” *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 1617–1655, 2016.

8. I. Gomez-Miguelez, A. Garcia-Saavedra, P. D. Sutton, P. Serrano, C. Cano, and D. J. Leith, "srsLTE: An open-source platform for LTE evolution and experimentation," in *Proc. 10th ACM Int. Workshop Wireless Netw. Testbeds, Experimental Evaluation, Characterization*, 2016, pp. 25–32.

9. A. Ksentini and N. Nikaein, "Toward enforcing network slicing on RAN: Flexibility and resources abstraction," *IEEE Communications Magazine*, vol. 55, no. 6, pp. 102–108, 2017.

10. X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, "Network slicing in 5G: Survey and challenges," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 94–100, 2017.

11. J. Ordonez-Lucena, P. Ameigeiras, D. Lopez, J. J. Ramos-Munoz, J. Lorca, and J. Folgueira, "Network slicing for 5G with SDN/NFV: Concepts, architectures, and challenges," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 80–87, 2017.

12. M. R. Sama, X. An, Q. Wei, and S. Beker, "Reshaping the mobile core network via function decomposition and network slicing for the 5G era," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshops (WCNCW)*, 2016, pp. 90–96.

13. T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta, and D. Sabella, "On multi-access edge computing: A survey of the emerging 5G network edge cloud architecture and orchestration," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1657–1681, 2017.

14. X. Pei, H. Yu, M. Wen, A. Barbieri, P. Fan, and C. Li, "Design and implementation of an LTE system based on srsLTE and USRP," in *Proc. 2020 IEEE 6th Int. Conf. Computer and Communications (ICCC)*, 2020, pp. 1514–1518.

15. P. Ranaweera, M. Liyanage, and A. Gurtov, "Survey on multi-access edge computing security and privacy," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 2, pp. 1078–1124, 2021.

16. N. Nikaein, M. K. Marina, S. Manickam, A. Dawson, R. Knopp, and C. Bonnet, "OpenAirInterface: A flexible platform for 5G research," *ACM SIGCOMM Computer Communication Review*, vol. 44, no. 5, pp. 33–38, 2014.

17. F. Kaltenberger, A. P. Silva, A. Gosain, L. Wang, and T. T. Nguyen, "Comparison of simulation tools for end-to-end 5G system evaluation," in *Proc. 2020 IEEE 21st Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, 2020, pp. 1–5.

18. E. Coronado, S. Khan, and R. Riggio, "5G-EmPOWER: A software-defined networking platform for 5G radio access networks," *IEEE Transactions on Network and Service Management*, vol. 16, no. 2, pp. 715–728, 2019.

19. J. Schmidt, L. Werthmann, G. Raman, and F. H. P. Fitzek, "An SDR-based testbed for evaluation of 5G waveforms in industrial environments," in *Proc. 2021 IEEE Int. Conf. Communications Workshops (ICC Workshops)*, 2021, pp. 1–6.

20. T. Villa, F. Shan, S. Han, A. Lozano, and C. Pan, "Performance evaluation of open-source 5G platforms," *IEEE Access*, vol. 9, pp. 85867–85878, 2021.

21. C. Bouras, A. Kollia, and A. Papazois, "SDR implementation of a testbed for LTE and WiMAX comparison," in *Proc. 2012 IEEE Int. Conf. Communications (ICC)*, 2012, pp. 5826–5830.

22. P. D. Sutton, K. E. Nolan, and L. E. Doyle, "Cyclostationary signatures in practical cognitive radio applications," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 1, pp. 13–24, 2013.

23. S. Lagen, A. Agustin, and J. Vidal, "Coexisting radio access technologies for 5G NR: A survey of multi-RAT solutions," *Computer Communications*, vol. 168, pp. 78–88, 2020.

24. X. Wang, M. Kong, M. Chen, S. Maharjan, H. Ding, and Y. Zhang, "Performance evaluation of network slicing for 5G vehicular communications," *Vehicular Communications*, vol. 27, Art. no. 100291, 2021.

25. F. Z. Yousaf, M. Bredel, S. Schaller, and F. Schneider, "NFV and SDN–Key technology enablers for 5G networks," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 11, pp. 2468–2478, 2017.

26. J. Su, H. Cai, Z. Sheng, A. X. Liu, and A. Baz, "Traffic prediction for 5G: A deep learning approach based on lightweight hybrid attention networks," *Digital Signal Processing*, vol. 146, Art. no. 104359, 2024, doi: 10.1016/j.dsp.2023.104359.

27. M. A. Habib, P. E. I. Rivera, Y. Ozcan, M. Elsayed, M. Bavand, R. Gaigalas, and M. Erol-Kantarci, "Transformer-based wireless traffic prediction and network optimization in O-RAN," in *Proc. 2024 IEEE Int. Conf. Communications Workshops (ICC Workshops)*, 2024, pp. 1–6.

28. S. Ji, J. Li, H. Jin, T. Wei, H. Dong, P. Zhang, and A. Bouguettaya, "Space-time-aware proactive QoS monitoring for mobile edge computing," *IEEE Transactions on Network and Service Management*, vol. 21, pp. 5662–5676, 2024, doi: 10.1109/TNSM.2024.3424847.

29. M. Rouili, N. Saha, M. Golkarifard, M. Zangooei, R. Boutaba, E. Onur, and A. Saleh, "Evaluating open-source 5G SA testbeds: Unveiling performance disparities in RAN scenarios," in *Proc. NOMS 2024-2024 IEEE/IFIP Network Operations and Management Symp.*, 2024, pp. 1–6.

30. M. Barbosa, M. Silva, E. Cavalcanti, and K. Dias, "Open-source 5G core platforms: A low-cost solution and performance evaluation," in *Proc. 2025 Int. Conf. Information Networking (ICOIN)*, Chiang Mai, Thailand, 2025, pp. 99–104, doi: 10.1109/ICOIN63865.2025.10992769.

Information about Author:

Zhenis Otarbay – Researcher, Nazarbayev University (Astana, Kazakhstan, e-mail: Zhenis.otarbay@nu.edu.kz).

Submission received: 13 February, 2026.
Revised: 19 March, 2026.
Accepted: 20 March, 2026.