Computer Networks 279 (2026) 112152

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

OTT-MNO Collaboration for a network-layer ML-based QoE prediction for video streaming over 5G O-RAN

Claudia Carballo González [a], Ernesto Fontes Pupo [a], Alessandro Floris [b,c,*], Simone Porcu [b,c], Maurizio Murroni [b,c], Luigi Atzori [b,c]

[a] i2CAT Foundation, 08034, Barcelona, Spain
[b] DIEE, University of Cagliari, 09123, Cagliari, Italy
[c] CNIT, University of Cagliari, 09123, Cagliari, Italy

A R T I C L E I N F O

Keywords:
5G, O-RAN
Quality of experience
Machine learning
Video streaming
Over the top
Mobile network operator

A B S T R A C T

It is well-known that, without access to application-layer parameters controlled by Over-The-Top (OTT) providers, Mobile Network Operators (MNOs) struggle to accurately predict customers' Quality of Experience (QoE). While some previous proposals have suggested interaction between OTTs and MNOs, they have faced challenges in terms of practical implementation and limited application scenarios. This work aims to advance these solutions with two key contributions. First, following the Open Radio Access Network (O-RAN) architecture, we propose adding components that integrate a machine learning (ML)-based QoE prediction model, deployed by the MNO, into the O-RAN system. By establishing specific data-sharing interfaces between OTTs and MNOs, our approach helps MNOs overcome the limitations in updating their quality prediction modules. Second, we present a network-aware, ML-driven QoE prediction model that captures the relationship between the resulting QoE and various network parameters, such as signal-to-interference-noise ratio (SINR), channel quality indicator (CQI), network resource blocks (RBs), throughput, and device mobility. Among seven considered ML regressors, the Gradient Boosting (GB) achieved the highest QoE prediction performance in terms of R^2 (0.906) and RMSE (0.259).

1. Introduction

The rapid advancement of mobile network technology has led to an exponential increase in video streaming performance, significantly enhancing the network Quality of Service (QoS) and the user Quality of Experience (QoE). Consequently, end users have become increasingly quality-demanding, making video service management more complicated for both Mobile Network Operators (MNOs) and Over-the-Top (OTT) providers. Additional complexity is because the QoE depends upon many influencing factors, including system, human, and context factors, which are not in the same hands [1]. Indeed, while the MNO can only measure and control network parameters (e.g., throughput, delay), the OTT provider can only measure application-level information, such as bitrate and resolution of the played video.

A prompt solution for MNOs is estimating the QoE using QoS-to-QoE models [2]. However, these models typically consider only the impact of the overall network performance on the user's QoE, neglecting the type of application running on the end device (ED), which is due to the widespread use of encryption in OTT traffic. Instead, QoE models based on application parameters utilized by OTTs can provide QoE estimations closer to the actual user's perceived QoE [3,4]. However, as mentioned before, the MNO does not know the application data and cannot use these models for QoE estimation. Since communication between MNO and OTT is hindered by the lack of standardized interfaces, these two actors estimate the QoE based on the information they gather using independent network-aware [5] and application-aware [6] QoE estimation models, respectively.

In recent years, machine learning (ML) has opened new possibilities for MNOs by providing the means to train predictive algorithms on a huge amount of network-related data and the corresponding QoE predicted with state-of-the-art application-layer QoE models [7–11]. Therefore, these ML-based models can predict the QoE solely based on network data. Still, the obstacle is the lack of application data for the MNO to continuously train the network-layer predictor, making use of more accurate application-layer estimated QoE to update these models with novel data or for diverse services. To overcome this limitation, the study in [9] exploited the 5G standardized interfaces for sharing information between a third-party Network Function (NF) implemented by the OTT

* Corresponding author.
E-mail address: alessandro.floris84@unica.it (A. Floris).

https://doi.org/10.1016/j.comnet.2026.112152
Received 8 May 2025; Received in revised form 23 December 2025; Accepted 22 February 2026
Available online 26 February 2026

and the 5G control plane NF implemented by the MNO. However, their evaluations were based on simulated 4G data, which did not allow them to fully exploit the capabilities of the 5G technology. Indeed, the considered video streaming flows were restricted to 0.5-3 Mbps and supported up to full high-definition (FHD) resolution with a maximum frame rate of 30 fps.

Although it is true that 5G significantly extends the bandwidth boundaries and overall performance range, it introduces technological features such as millimeter-wave bands and highly directive communications, which result in a system more sensitive to the user's network conditions and mobility behavior [12]. Additionally, 5G network slicing allows for the dynamic creation of multiple virtual networks with distinct service-level requirements and configurations, leading to heterogeneous QoS provisioning across users and applications. These characteristics significantly exacerbate the QoE prediction challenge from the MNO side, increasing the variability and context sensitivity of user experience and thereby limiting the viability of prior 4G-based solutions [8]. Indeed, 4G systems are generally optimized for video services with low to medium resolutions and standard frame rates, which are insufficient to meet the demands of modern 5G use cases involving higher bitrates, ultra-low latency, and more dynamic conditions, such as 4K video, 60 fps streaming, and immersive applications.

In this paper, we aim to advance the state-of-the-art in two directions. First, we leverage the Open Radio Access Network (O-RAN)[1] to integrate additional components aimed at supporting network management operations in 5G and future 6G networks. In particular, the proposed solution concerns integrating an ML-based QoE prediction model deployed by the MNO into the O-RAN framework and implementing RESTful APIs for data sharing between the OTT and the MNO. This allows for overcoming the key limitation of MNO to update the modules for quality prediction. Second, we propose a network-aware ML-based QoE prediction model that captures the relationship between several network parameters, such as signal-to-interference-noise-ratio (SINR), channel quality indicator (CQI), network resource blocks (RBs), throughput, and the device mobility behavior, with the corresponding QoE. To this aim, we have trained and compared the performance and computational complexity (CC) of seven supervised ML regression models. In addition, to compute the ground-truth QoE, we have utilized the state-of-the-art ITU-T P.1204 algorithm, which covers diverse video codecs (H.264, H.265, and VP9) and, especially, higher resolutions (up to 4K) and frame rates (up to 60 fps) [13] to fully exploit the 5G network capabilities. Extensive analyses have been performed to identify the most important features for the prediction task, the accuracy of the different prediction models, and when the monitored channel quality conditions vary continuously over time. The Gradient Boosting (GB) regressor achieved the highest QoE prediction performance, as indicated by R^2 (0.906) and RMSE (0.259) metrics.

The paper is structured as follows. Section 2 presents the related works on QoE assessment and prediction for video streaming in mobile networks. Section 3 provides an overview of the proposed solution and its integration into the O-RAN framework. Section 4 presents the proposed ML-based QoE prediction model, whereas the collected dataset and preliminary analysis are discussed in Section 5. The achieved prediction performance of the trained models is analyzed in Section 6, and we applied the proposed model for temporal QoE estimation in Section 7. Section 8 offers insights into the challenges and scope related to OTT-MNO collaboration and the synthetic dataset. Finally, Section 9 concludes the paper.

[1] https://www.o-ran.org/

2. Related work

This section discusses literature studies on QoE assessment for video streaming, particularly on ML-based QoE prediction models for mobile networks.

2.1. QoE assessment of video streaming

QoE assessment and modeling approaches for video streaming services can be divided into two categories: subjective and objective [14]. Subjective QoE assessment methods involve organizing controlled experiments that require a preset group of users to evaluate the quality of video content using the Mean Opinion Score (MOS) metric. These experiments are conducted in a controlled environment and must follow strict guidelines defined by the International Telecommunication Union (ITU) to achieve reliable QoE ground-truth data that can be used to build and validate objective QoE models [15]. While subjective assessments leverage personal user characteristics (e.g., expectation, previous experience) to assess the QoE, these methodologies are time-consuming, costly, and unsuitable for real-time QoE measurements.

Objective quality models overcome these drawbacks by providing an estimated measure of the QoE based on monitored network and/or service parameters. These models are classified into full-reference (FR), reduced-reference (RR), and no-reference (NR), depending on the need for the original video content to measure the quality of the received content. For instance, the Video Multi-method Assessment Fusion (VMAF) is a state-of-the-art FR video quality assessment algorithm developed by Netflix. While the VMAF provides QoE estimates highly correlated with the MOS, the downside is the requirement of the original video content [3]. As another example, the ITU-T P.1203 module algorithms are NR bitstream-based models that assess the QoE of adaptive video streaming services based on the measurement of content and application-related parameters, such as video quality and resolution [4]. Advanced versions of these models, the ITU-T P.1204 algorithms, cover diverse video codecs (H.264, H.265, and VP9) and, especially, higher resolutions (up to 4K) and frame rates (up to 60 fps) [13]. The ITU-T P.1204 has proven higher QoE prediction performance than FR metrics, such as the VMAF [16,17].

However, traditional QoE objective approaches have evidenced limitations due to the accelerated growth of diverse EDs, the combination of wired and wireless network infrastructures, and the increased interactivity of multimedia streaming services. To handle this, recent QoE management research efforts have been conducted based on ML to provide a dynamic and real-time optimization loop for adaptive streaming applications. ML allows the modeling of complex problems with high accuracy, such as quantifying the correlation between QoS and QoE. Nevertheless, selecting the optimal ML model for each multimedia application is still an open research issue [18].

2.2. QoE prediction models in mobile networks

Accurate assessment and prediction of QoE in mobile networks is not straightforward. Indeed, most of the network resource management approaches presented in the literature are QoS-based, i.e., different values of the CQI are mapped into diverse classes of QoS requirements (in terms of, for instance, packet delay and error rate) to satisfy the specific service running on the ED [21,22]. In [23], a framework based on a neural network (NN) is proposed to predict the QoS, considering network parameter fluctuation over time and multiple services mapped into different slices. This method overcomes the issue of the service status (e.g., workload) and network environment (e.g., congestion) changing over time. A QoS-aware virtualization resource management mechanism is proposed in [24] to solve the problem of resource management difficulties of multiple MNOs sharing the same 5G backhaul network. The proposed resource scheduling algorithm based on QoS priority can maintain high resource utilization, reduce wavelength tuning overhead, and improve traffic delay. An xApp to enhance the functionality of a Near-Real Time RAN Intelligent Controller (Near-RT RIC) into the 5G O-RAN framework is proposed in [25]. This xApp runs an Adaptive Genetic Algorithm to associate each user equipment with the base station (BS) that offers the highest throughput (i.e., high channel quality index, good channel conditions). Experimental tests demonstrated the potential to provide the users with an overall higher data rate and quality level.

However, these purely network-centric QoS-aware approaches do not fully capture the quality perceived by the end user, i.e., the QoE, which is not only affected by the network performance but also by further aspects, such as the used device, the context, and the user's preferences and expectations. For this reason, to the same QoS, different QoE can be perceived on the user's side. This is even more true in 5G and beyond networks, which open the doors to immersive and interactive services where the subjective QoE perception is of foremost importance. Thus, QoE objective models have started to be used to predict the end user's QoE based on the measured QoS.

A model for predicting the QoE of gaming video streaming applications is proposed in [7], which is built using a convolutional NN (CNN). Well-known QoS-to-QoE models were used to compute ground-truth QoE values to feed the CNN together with the corresponding network data, namely the logistic mapping function, IQX hypothesis, Weber-Fechner law, and Stevens' power law. Although the authors discussed the integration of their QoE prediction model into the O-RAN framework, they did not delve into the main O-RAN elements, such as the RICs, and how they must interact to collect data, train, and deploy the ML model. In [26], subjective tests were conducted to collect data from users under diverse network conditions. Then, the mapping between QoS and QoE in video and web services using an ML approach was evaluated. In [8], a QoE model based on an NN is implemented to correlate several QoS parameters collected for diverse video streaming applications, such as YouTube, in a Long-term evolution (LTE) network. The QoE was measured by conducting a subjective test and computing the MOS. Even if this model exhibited good QoE estimation, the subjective tests are time-consuming and expensive, limiting what the model learns to the criteria of a small group of young people with similar interests. Therefore, the generalization of this model would require retraining it with a larger and more diverse group of people.

In [9], five supervised ML regression models were trained to estimate the QoE of video-on-demand (VoD) and voice-over IP (VoIP) services based on several network Key Performance Indicators (KPI). The network environment was simulated using the Network Data Analytics Function (NWDAF), a 5G core element. Two models were used for ground-truth QoE estimation: the P.1203.1 mode 0 (bitstream-based NR model) and the cumulative quality model (CQM). Similarly, in [10], the round-trip time (RTT), jitter, and packet loss rate between the VoD server and the clients were measured and collected to train a decision tree model to predict QoE. In this case, the P.1203.1 mode 3 (full bitstream) model was used for the ground-truth QoE estimation. In [11], the authors proposed a Random Forest (RF) model to capture the relationship between network parameters (e.g., delay, jitter, packet loss) and QoE. They extracted segment files from video streaming sessions and calculated MOS using the ITU P.1203 model. The approaches in [9,10], and [11] are focused on the prediction of QoE only using network parameters. However, the considered P.1203 has only been validated for QoE estimation of H.264 encoded videos with resolutions up to FHD and frame rates up to 30 fps. Particularly, the video bitrates considered in [9,10] were restricted to 0.5-3 Mbps and 0.2–12 Mbps, respectively, which limits the 5G network environment, as they failed to capture the full potential of 5G-enabled services that demand significantly higher bitrates. On the other hand, none of these previous works, except [7], investigated the use of O-RAN for deploying ML-based real-time QoE prediction mechanisms in 5G environments.

The approaches in [19] and [20] are hybrid because they leverage both network and application data to predict the QoE. A Multi-layer Perceptron (MLP) model is proposed in [19] to estimate the QoE for live video streaming. The input data comprises video resolution, screen size, performance type, network bandwidth, and RTT. A non-standardized QoE objective model is used to compute the ground-truth QoE, including the quality of the video source, interactive quality, and watching quality. Although the proposed MLP model achieves high QoE prediction accuracy, it was trained on application parameters (video resolution and screen size) that are generally not available to the MNO (and often also to the OTT, such as the screen size), and how these could be provided by the OTT has not been discussed. Similarly, in [20], the authors proposed an ML-based QoE prediction model for video streaming based on network (packet loss, latency, jitter) and application (rebuffering, resolution switching) data by relying on VMAF as a ground truth quality for training. In addition, they generated synthetic data to capture mobility and wireless fluctuations in 4G/5G mobile networks for autonomous driving systems. However, as we previously mentioned, VMAF requires access to the original video content, which is generally inaccessible to the MNOs. While hybrid approaches may achieve accurate QoE prediction performance thanks to detailed knowledge of application-related data (e.g., video resolution, rebuffering and resolution switching events), the measurement of these metrics is not straightforward and may require high-frequency measurements corresponding to a relevant amount of data to be shared between OTT and MNO. Moreover, application-related data affecting the QoE, such as those mentioned, are a direct consequence of some network impairments; for instance, a reduced throughput may cause a rebuffering event and, if persistent, a subsequent resolution switching event aimed at adapting the video quality to the new available network throughput. Thus, QoE prediction models based only on network measurements are preferable because they describe a direct relationship between network statistics and the user's QoE regardless of the used application. The only required information is the perceived user's quality, which can be measured by the specific OTT application and shared with the MNO.

Compared with previous works, we provide an ML model that allows the MNOs to estimate the QoE for video streaming only based on the network and the user's mobility data. The proposal is a unified tool for fixed and mobile devices playing videos encoded with different codecs at up to 4K resolution and up to 60 fps, exploiting the capabilities of the 5G network. Finally, we propose integrating the model into the 5G O-RAN framework, taking advantage of its disaggregated and flexible elements and open interfaces. Table 1 summarizes the distinguished features of our proposal in comparison to a selected set of state-of-the-art works.

3. Overview of the proposed solution

In this work, we propose a novel architecture which is aimed at integrating an ML-based QoE prediction model into the O-RAN framework managed by MNOs, as shown in Fig. 1 (details of this figure will be clarified in Section 3.1). In particular, the MNO leverages the O-RAN architecture to exchange data with collaborative OTTs that decide to share QoE estimations of their services provided through the MNO's network infrastructure. This kind of collaboration overcomes the MNO limitation in accurately predicting the user's QoE when using video streaming services due to the lack of knowledge of the application's parameters (such as video encoding, resolution, and bitrate information for video streaming services).

According to the proposed framework, the OTT monitors and utilizes application-level parameters to predict the perceived QoE in terms of MOS through QoE objective models. The QoE results are then conveyed to the MNO. The rationale for OTT providers to collaborate with

Table 1
Comparison of the proposed approach with the state-of-the-art. CDN: Content Delivery Network; MEC: Mobile Edge Computing.

Reference	Our proposal	[7]	[8]	[9]	[10]	[19]	[20]
QoE estimation	Seven supervised ML models	CNN	NN	Five supervised ML models	Decision tree	MLP	NN
Input data	Network and mobility data	Network data	Network data and device position	Network data	Network data	Network and application data	Network and application data
Ground-truth QoE	ITU P.1204.3	Non-linear QoS/QoE correlation	Subjective test	ITU P.1203.1	ITU P.1203.1	Non-standardized QoE model	VMAF
Video specifications	Up to 4K (0.15–45 Mbps) and 60 fps, H.264/H.265/VP9	2K, 60 fps	Not specified	Up to FHD (0.5-3 Mbps), H.264, up to 30 fps	Up to FHD (0.2–12 Mbps), H.264, up to 30 fps	Up to 8K and 60 fps, H.264	Up to 50 Mbps, H.264/H.265
Network implementation	4G/5G	5G	4G	5G	Small scale CDN-MNO or MEC	4G/5G	4G/5G
O-RAN framework	✓	✓	×	×	×	×	×
MNO-OTT collaboration	✓	×	×	✓	×	×	×

MNOs by sharing predicted QoE data is primarily driven by user retention strategies. When end-users experience degraded application performance, they are liable to churn from the application rather than the network operator, as they typically lack visibility into the underlying network causes of the malfunction. Furthermore, this OTT-MNO data exchange establishes a potential revenue stream, as OTTs may require financial compensation for providing actionable QoE insights. Such cross-layer, multi-stakeholder collaboration is widely supported in the literature, particularly within the context of network softwarization paradigms, such as Software-Defined Networking [27].

The MNO can then perform correlation analysis and train predictive models between the monitored network status (i.e., available physical resources, channel state information, throughput, energy consumption, mobility behaviors) and the QoE values shared by the OTT. Collecting a dataset, including correlated network parameters and QoE scores, allows the MNO to define an accurate ML-based QoE model that, once deployed, can predict the QoE of the connected users solely based on the device's mobility and network-related data. Whereas this solution has been tailored to video streaming services, the proposed approach can be extended to other application scenarios.

3.1. Leveraging the O-RAN framework

The main components of the O-RAN framework shown in Fig. 1 are defined by the O-RAN ALLIANCE, whose mission is to reshape the RAN industry towards more intelligent, open, virtualized, and fully interoperable mobile networks to improve user experience [28]. We propose integrating the ML-based QoE prediction model into the O-RAN framework, using its disaggregated elements and open interfaces. Specifically, the BS functionalities are disaggregated into the O-RAN Central Unit Control Plane (O-CU-CP) and User Plane (O-CU-UP), Distributed Unit (O-DU), and Radio Unit (O-RU) [29]. Such a logical split allows these functional units to be flexibly deployed at different network locations and hardware platforms.

One of the main modules in the O-RAN framework is the Non-Real Time RIC (Non-RT RIC), which is a Service Management and Orchestration (SMO) element. It implements optimization actions through microservices termed rApps on a timescale of more than 1 second. Moreover, it trains and updates ML models that will be executed nearer to the end-user, such as in the Near-RT RIC. Furthermore, the Non-RT RIC realizes long-term monitoring via the *O1* interface and guarantees policy-based guidance of applications/features in the Near-RT RIC via the *A1* interface [30]. The Near-RT RIC executes ML tasks through microservices termed xApps in control loops between 10 ms and 1 s, collecting data via the *E2* interface [31]. In this context, the network metrics used in our study (e.g., throughput and channel state information) correspond to the data types typically exposed by the E2 Service Model for Key Performance Measurement (E2SM-KPM). The isolation of xApps is crucial for the independent operation of O-RAN services and accurate decision-making at the Near-RT RIC. In this context, the Security, Conflict Mitigation, and Subscription Management components play a critical role [32].

While our simulations focus on validating the ML logic, the O-RAN framework provides the standardized functional elements that make the integration of such third-party QoE intelligence technically feasible. The O-RAN architecture facilitates QoE prediction and real-time optimization with proactive closed-loop network management. Leveraging this framework, the proposed contributions to the O-RAN framework concern the “Training/testing ML-based QoE Prediction" and “QoE Prediction" modules highlighted with dashed lines in Fig. 1, which perform the offline training of the ML model in the Non-RT RIC and the model inference in the Near-RT RIC at the network edge, respectively. As the first step, the Non-RT RIC collects the devices' mobility information and the network data from the O-CU/O-DU components via the *O1* interface and the QoE value via the Application server through RESTful APIs defined between the MNO and the OTT. In particular, the shared data concerns a JSON-formatted file including timestamp, application type, application utilization time, and predicted QoE in terms of MOS. Authentication mechanisms can be used to guarantee secure communications, whereas no privacy issues are identified because no user information is involved. Indeed, the predicted MOS is computed from application-related data, such as video resolution and bitrate in the case of a video application. The collected corresponding network-related data is prepared (e.g., data partitioning and feature selection) and used to train offline diverse ML models to estimate the QoE on the continuous MOS scale during an application session [15].

Once the ML models are trained and tested, the one achieving the best compromise between QoE prediction performance and CC is stored in the ML Catalog of the Non-RT RIC, which deploys the QoE policy model. The Near-RT RIC receives the ML-based QoE prediction model via the *O1* interface and stores it in the ML inference entity as an xApp to estimate the QoE only based on mobility and network-related informa-

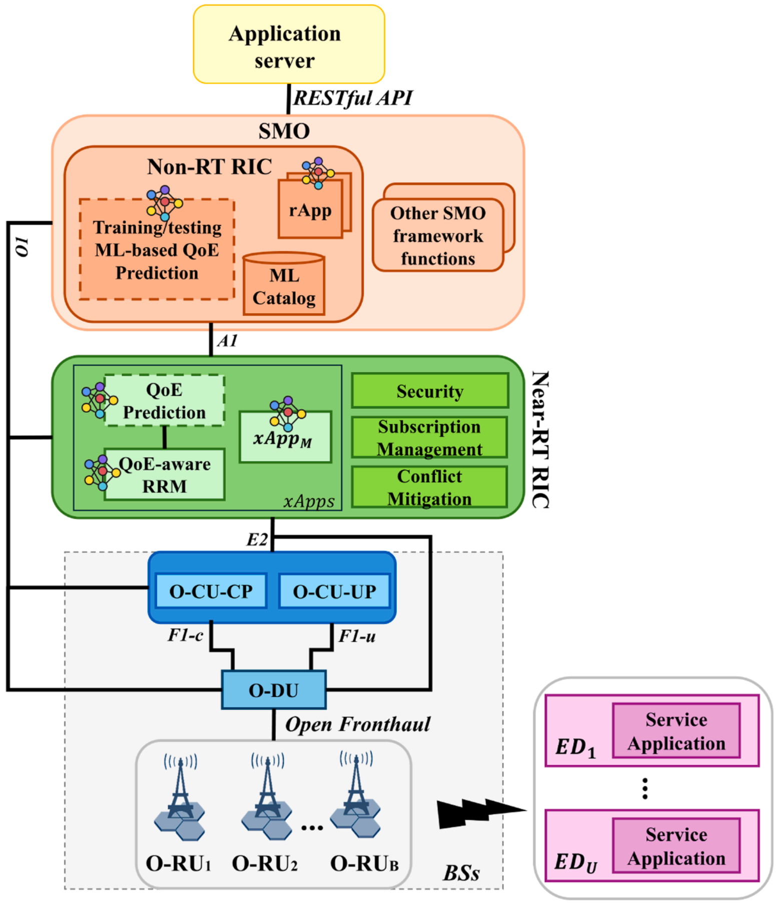

Fig. 1. The ML-based QoE prediction model inserted into the O-RAN framework.

tion (i.e., obtained via the *E2* interface). At this stage (during inference), the Near-RT RIC does not receive any application information from the OTT via RESTful APIs. Additionally, the Non-RT RIC assesses the performance of the deployed ML-based QoE model, retrains it if necessary, and updates it in the Near-RT RIC.

The predicted QoE is essential to drive the QoE-aware radio resource management (RRM) module, implementing dynamic radio resource optimization to provide high-quality application sessions to the connected users. In particular, resource management may involve the increase/decrease of physical RBs to specific users or a handover process to improve user experience. The advantage of managing these resources based on the predicted user's QoE (instead of based on QoS metrics) lies in the fact that the QoE considers further factors than network performance metrics, such as the used device or the user's preferences. This allows for the allocation of radio resources for those users who need them to achieve a satisfactory experience. On the other hand, QoS-based optimization approaches typically consider each user in the same way. The use of the Near-RT RIC is essential to satisfy the sub-second latency requirements of QoE-aware radio management, which would be unfeasible in traditional, non-disaggregated RAN deployments.

4. The ML-based QoE prediction model

This section presents our methodology to define an ML-based QoE prediction model for a video streaming service provided over an MNO 5G network, leveraging the O-RAN architecture. Fig. 2 illustrates the workflow concerning the data collection, data preparation, training/testing, deployment, and update of the proposed ML-based QoE prediction model in the O-RAN framework. The ML workflow steps will be further detailed in the following sections.

As described in the previous section, the assumption is that the OTT collaborates with the MNO by computing and sharing the estimated QoE of the application served through the MNO network via RESTful APIs. The QoE estimated values will be considered as the *ground-truth data* for the MNO to train a QoE prediction model solely based on network-related data. We assume that the application layer mechanisms devoted to the management of the video streaming session work adaptively with respect to the network layer conditions in a way that influences the final quality. For instance, video streaming services based on HTTP Adaptive Streaming solutions have become the most prevalent technologies utilized by OTT providers for live video streaming applications because they allow for adaptive download, according to the monitored network fluctuations, appropriate video representations from media servers using adaptive bitrate algorithms, consequently improving users' QoE [6,33]. Thus, we believe that there is a correlation between the network status and the overall provided quality, which we intend to exploit with the proposed approach.

First, we describe the considered mobile network environment with particular attention to the resources allocated to the users and the parameters used to describe their status. Then, we explain the data preparation and the ML training and testing phases of the QoE models.

4.1. Mobile network environment

We assume a 5G deployment, including B New Radio BSs indexed by $b \in \{1, 2, \ldots, B\}$. The smallest frequency resource the BSs can allocate is the RB, whose bandwidth corresponds to 12 consecutive equally spaced subcarriers, $W = 12 \times \Delta f$. The subcarrier spacing value, expressed in kHz, is defined as $\Delta f = 15 \times 2^{\mu}$, where μ is the considered numerology according to the 5G New Radio standard [12].

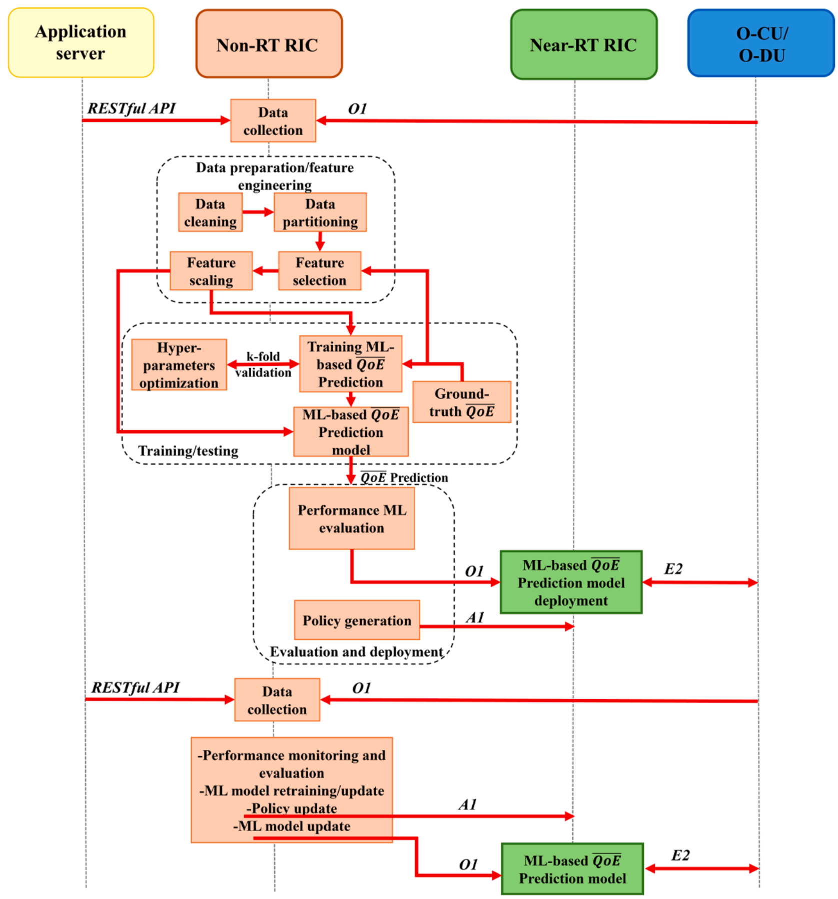

Fig. 2. Workflow concerning the proposed ML-based QoE prediction model in the O-RAN framework.

We assume that the network serves a total of U EDs indexed by $u \in \{1, 2, \dots, U\}$. We consider four types of EDs randomly distributed in the service area: mobile (MO), tablet (TA), personal computer (PC), and television (TV). The mobility status of the ED is defined by the parameter $v \in \{0, 1\}$, where $v_u = 0$ represents a static position and $v_u = 1$ a random directional mobility. The TVs and PCs are static, while the TAs and MOs can be static or follow random directional mobility behavior [34,35].

We assume the U EDs are receiving a certain throughput (Th_u), expressed in bits per second (bps), which depends on the network conditions, the type of ED, and the video resolution. To deliver the requested video service effectively, the set of available RBs at the BSs is managed by executing the fast link adaptation procedures and selecting the most appropriate modulation coding scheme based on the users' channel state information. The transmitters carry out such selection every transmission time interval t', collecting from the U EDs their experienced CQI (CQI_u) and SINR ($SINR_u$, expressed in dBm), as detailed in [12]. The reported CQI_u is directly related to the $SINR_u$ experienced by the ED_u. Such feedback procedures allow the BSs to execute the fast link adaptation and corresponding RRM.

We denote as eff_u the efficiency value (in bps/Hz) associated with the channel reception conditions (based on the CQI_u) of each ED_u [12]. Then, the Th_u is computed as

$$Th_u = n_u^{used} \times W \times eff_u, \tag{1}$$

where n_u^{used} is the number of RBs assigned to the ED_u for proper reception of the video service encoded in a specific resolution. n_u^{used} is less than or equal to the number of available RBs (n_b^{av}) in the BS_b.

Ec_u, expressed in joules (J), is the energy consumption at the ED when accessing the video service via the BS_b

$$Ec_u = P_u * D_u, \tag{2}$$

where P_u, expressed in watts (W), is the power consumed by the ED_u for the specific video reception. D_u, expressed in seconds (s), is the delay experienced by the ED_u to access the video service through the BS_b [32]. D_u is inversely proportional to Th_u. Consequently, a higher Th_u means a lower value of Ec_u because the user will be able to download the video faster, wasting less energy.

Table 2 summarizes the definitions of the considered network elements and mathematical notations.

4.2. Prediction models parameters

The MNO monitors and manages the network resources assigned to each ED, but it is not capable of measuring the user's perceived QoE. This is the role of the OTT, which applies state-of-the-art QoE objective models to estimate the user's perceived QoE based on the monitored application parameters ("Server application" module in Fig. 1). Thus, the data collection concerns the creation of a dataset including network-based resource parameters of the EDs reached by the network environment, with the corresponding user's perceived QoE shared by the OTT. This dataset will be the source for training QoE predictive models ("Training/testing ML-based QoE prediction" module in Fig. 1) that aim to estimate the QoE based solely on the MNO's network parameters.

Specifically, aiming to execute the ML-model offline training, the MNO collects the following information: the number U of EDs, their status ED_u, the mobility information v_u and the network status parameters per device, which are represented by $SINR_u$, CQI_u, n_b^{av}, n_u^{used}, Ec_u, and Th_u.

At the same time, the OTT utilizes four application-related parameters, i.e., video codec, frame rate, video encoding resolution, and display resolution, to compute the corresponding QoE_u (label of the ML

Table 2
Definitions of the considered network elements and mathematical notations.

Symbol	Definition
B	Number of new radio BSs
b	Index for the BSs
RB	Physical resource block
W	RB's bandwidth
μ	New Radio numerology
U	Number of EDs served by the network
u	Index for the EDs
MO	Mobile end device
TA	Tablet end device
PC	Personal computer end device
TV	Television end device
Th	Throughput
v	ED's mobility parameter
t'	Transmission time interval
CQI	Channel Quality Indicator
SINR	Signal-to-interference-noise-ratio
eff	Channel efficiency
n^{used}	Number of assigned RBs
n^{av}	Number of available RBs
Ec	Energy consumed at the ED
P	Power consumed by the ED
D	Transmission delay

model during the training process) perceived for each one of the ED_u. To this aim, it was considered the state-of-the-art QoE objective model described in the ITU-T Recommendation P.1204.3 concerning the video quality assessment of streaming services up to 4K resolution [13]. The model takes the four video parameters mentioned as input and returns the QoE value on the five-point MOS scale representing the per-second video quality estimation. Then, the data collection at the MNO and OTT sides occurs with a frequency of 1 s.

In particular, we have used the mode 0 bitstream model, which settles on the degradation-based modeling principle: the higher the degradation, the lower the video quality. This model considers three different degradations: quantization, upscaling, and frame rate degradation. The quantization degradation accounts for coding-related degradations that can be perceived by the user as blockiness and other artifacts. This degradation is codec-dependent. The upscaling degradation is due to the encoded video being upscaled to a higher display resolution during playback, thereby resulting in blurring artifacts. Finally, frame rate degradation relates to the degradation introduced by playing out the distorted video at a reduced frame rate compared to the display's native frame rate, resulting in jerkiness.

4.3. MNO prediction models

According to the works of the state of the art [9,36], to estimate the QoE on the continuous MOS scale, we considered seven ML regression techniques: Least Absolute Shrinkage and Selection Operator (LASSO), Linear Regression (LR), Support Vector Regression (SVR), k-Nearest Neighbors (kNN), MLP, RF and Gradient Boosting (GB). The considered models are structurally different from each other, meaning different hyperparameters and computational complexity (CC). Moreover, these models were chosen to balance prediction accuracy with CC. This trade-off is important because the ML models are required to run in the Near-RT RIC, which has limited computing resources. For this reason, we have not considered training deep neural networks, which would create significant overhead in terms of latency, memory use, and retraining time. As an example, the study in [37] compared an SVM and a Convolutional Neural Network (CNN), which have reached similar prediction accuracy, but the SVM was about 84 times faster to train. These results support our choice that simple supervised models are better suited for fast and efficient QoE estimation at the network edge.

Table 3
ML models' comparison in terms of CC.

ML model	CC (training)	CC (testing)
LASSO	$O(f^3 + f^2 \times r)$	$O(f')$
LR	$O(f^3 + f^2 \times r)$	$O(f)$
SVR	$O(r^3 + r^2 \times f)$	$O(f \times f'')$
MLP	$O(e \times r \sum_{l=1}^{L-1} n_{l-1} n_l)$	$O(\sum_{l=1}^{L-1} n_{l-1} n_l)$
kNN	$O(1)$	$O(k \times f \times r)$
RF	$O(f \times r^2 \times t)$	$O(f \times t)$
GB	$O(f \times r^2 \times t)$	$O(f \times t)$

Table 3 summarizes the CC of the seven considered supervised ML regression algorithms [9,38]. The training process of LASSO and LR depends on the number of features (f) and training samples (r). For each test sample, the CC is $O(f')$ and $O(f)$, where f' is the number of non-zero weights after training, and $f' \ll f$. The SVR CC is $O(f'' \times f)$, where f'' is the number of support vectors. The MLP CC depends on the number of epochs (e) required for convergence, r, the number of layers (L), and the number of neurons (n_l) in each layer l, where $n_0 = f$. In the case of kNN, since it does not have a separate training phase, the CC only regards the testing phase and increases linearly with the number of neighbors (k), f, and r.

The CC of RF and GB is linearly affected by f and the number of trees (t). Additionally, during training, r has a quadratic influence. Thus, to adequately compare the different structures of the selected models, the training phase required a hyperparameter grid-search step and a 5-fold cross-validation to avoid overfitting issues. The results of this process are evaluated in terms of estimation performance and CC.

5. Dataset and preliminary analysis

To obtain a dataset that could be used to analyze the performance of our proposed solution, we resorted to a simulation environment from which we extracted the data of interest, as explained in the first subsection. In the second subsection, we provide an analysis of the importance of the identified features in terms of correlation with the estimated QoE.

5.1. Synthetic dataset

Based on the network environment described in the previous section, we have generated the network data through a link-level simulator (LLS) on a per-second scale granularity to simulate realistic data collection. We have used our ad-hoc developed Python-based tool [35] to model the link communication between the BSs and the EDs, according to [39]. To recreate a wide range of ED reception conditions, we executed multiple simulation runs of a New Radio urban micro station (UMi) service area, collecting the information regarding a total of 10,000 EDs randomly distributed in the grid (250×250 m). We chose the UMi Street Canyon scenario for its wide adoption in 3GPP and 5G evaluation studies [39,40], recreating an urban microcell environment, capturing key propagation characteristics, such as line-of-sight obstructions, reflections, and scattering typical of dense city streets. Parameters, such as frequency, bandwidth, antenna gain, and height, were chosen according to standardized UMi deployment guidelines [39,40], ensuring realistic path loss, fading, and interference conditions. In general, these settings allow the simulation to capture a diverse range of ED reception conditions and realistic user experiences, making the results applicable to urban 5G network planning and QoE analysis. We set up the simulations to ensure that 50% of the EDs are MOs or TAs, and 50% TVs or PCs. In each run, we collected the network- and application-level pa-

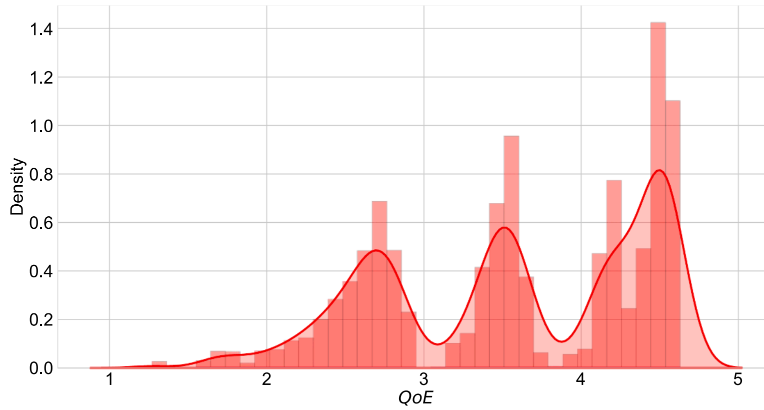

Fig. 3. Density distribution of the obtained QoE during the data collection.

Table 4
Simulation parameters.

Parameter	Value
Scenario type	UMi Street Canyon [39]
Grid size	250 × 250 m
BS operating frequency	28 GHz [39,40]
New Radio numerology μ	3
BS Bandwidth	400 MHz
RB's bandwidth, W	1.44 MHz
Subcarrier spacing	120 kHz
BS Transmission power	26 dBm
BS Antenna gain	10 dBi
Power spectral density of noise	-174 dBm/Hz
BS Height	10 m
Small-scale fading models	Jakes [35]
Large-scale fading models	[35]
Dynamic line of sight	Yes
Type of end devices	MO, TA, PC, TV
Mobility model MO, TA / PC, TV	Static, random directional/ static [35]
MO/TA speed	0–2 m/s
Reception mode MO, TA, PC / TV	Indoor-outdoor/Indoor
Display resolution MO, TA / PC, TV	2560 × 1440 / 3840 × 2160 [13]

Table 5
The considered video resolutions [13].

Video resolution (MO/TA)	Bitrate (Mbps)	Video resolution (PC/TV)	Bitrate (Mbps)
480 × 270	[0.09,1)	720 × 540	[0.15,4)
720 × 540	[1,4)	1280 × 720	[4,10)
1280 × 720	[4,10)	1920 × 1080	[10,15)
1920 × 1080	[10,15)	2560 × 1440	[15,20)
2560 × 1440	[15,20]	3840 × 2160	[20,45]

rameters for each simulated ED without temporal evolution. Table 4 summarizes the simulation parameters.

As summarized in Table 5, for each kind of device (MO/TA or PC/TV), we have considered minimum video resolution (and corresponding encoding bitrate), as suggested in [13], to simulate realistic scenarios. For instance, we have considered the display of 4K videos only on PC/TV and the display of low-resolution videos (i.e., 480 × 270) only on MO/TA. Moreover, we have considered videos encoded with three different codecs (H.264, H.265, and VP9), and we have assumed a fixed display resolution for MO/TA (2560 × 1440) and PC/TV (3840 × 2160) and a fixed frame rate of up to 60 fps.

Fig. 3 shows the resulting QoE density distribution, distinguishing EDs perceiving poor (MOS < 3), average ($3 \leq$ MOS < 4) and good QoE value (MOS ≥ 4). The QoE values were estimated using the ITU P.1204.3 model as discussed in the previous section.

5.2. Features analysis

This subsection describes the procedures followed to preprocess the collected MNO features. We did not apply data cleaning during the data preparation because there were no missing values or outliers. First, the data was randomly divided into 80% for training and 20% for testing, ensuring each new subset is a representative part of the entire dataset to avoid a possible bias. The feature selection process was only performed on the training data to investigate the relationship between the features and the QoE scores without being affected by the test data.

Figs. 4 and 5 show the relationship of $SINR$ and CQI with the QoE, respectively, along the overall training dataset. Even if the reception conditions are not good (i.e., small values of $SINR$ and CQI), the QoE is not necessarily low. The reason is that the user perception not only depends on these variables, but it also depends on n_b^{av} to guarantee at least the minimum Th required by the video service. For high values of $SINR$ and CQI, fewer n_u^{used} are required, and, consequently, it is more likely to achieve high QoE scores. The above explanation is evident in Fig. 3, where there are considerably more average and good QoE samples than low QoE samples.

Fig. 6 complements the previous explanation, showing that even for $n_u^{used} = 1$, the resulting QoE could be any possible value in the MOS scale because the QoE mostly depends on the reception conditions and the available resources at the BSs, n_b^{av}. Fig. 7 describes the relationship between Th and QoE, distinguishing among the considered different types of EDs. This figure evidences that devices with smaller screen sizes, like MOs and TAs, achieve higher QoE with the same Th than devices with larger screens, such as PCs and TVs. This result, which aligns with the findings achieved by the literature [41,42], highlights that users may be unable to distinguish between two different video qualities on a small screen.

We used the Pearson correlation coefficient (PCC) and Spearman's rank correlation coefficient (SRCC) to quantify the relationship between the MNO features and QoE scores. The PCC captures the linear relationship among two variables, where $+1$ represents a positive correlation, zero means the variables are not correlated, and -1 denotes a negative correlation. The SRCC can capture non-linear relationships between two variables and ranges from -1 to $+1$, where 0 means no correlation and -1 or $+1$ highlights a negative or positive correlation, respectively.

According to the results in Table 6, only the first four features highlight a strong correlation ($> |0.7|$ for both coefficients) with the QoE. However, upon analyzing the relationships between the features and the QoE scores, we observed that although v does not achieve a high correlation value with the QoE, it can still be discriminative for QoE prediction. This is illustrated in Fig. 7, which shows that the device

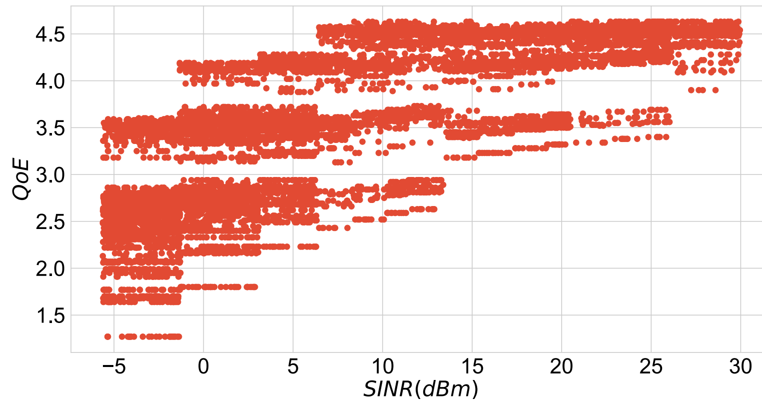

Fig. 4. $SINR$ and QoE relationship.

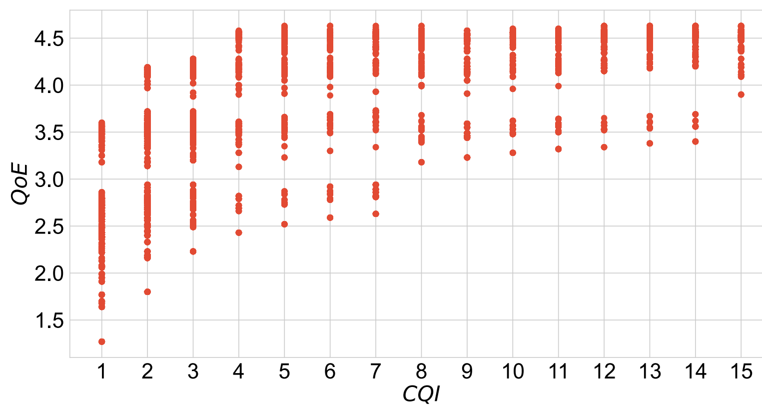

Fig. 5. CQI and QoE relationship.

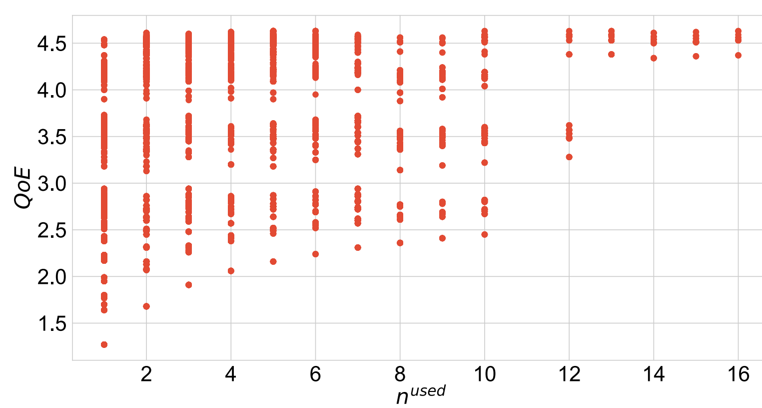

Fig. 6. n^{used} and QoE relationship.

type (mobile or non-mobile) impacts how the users perceive the quality. Thus, even if v alone is not strongly correlated with the QoE, its contribution could be informative when combined with the other features.

While a deeper analysis of feature interactions (e.g., between v and Th) using techniques such as partial dependence plots or interaction metrics would indeed offer further insights, our current focus is limited to evaluating individual correlations with respect to QoE prediction. To partially address this, we include all features in the ML training process (with their importance analyzed in detail later) and rely on the model's internal mechanisms to implicitly capture potential nonlinear interactions, without introducing bias or overfitting.

To conclude the data preparation step, the $MinMaxScaler$ feature scaling module has been used to normalize the input data, leading to more stable and efficient training [36].

Table 6
Feature ranking using PCC and SRCC.

Ranking	PCC	SRCC
1	$SINR$ (0.76)	Th (0.91)
2	CQI (0.73)	Ec (-0.91)
3	Th (0.72)	CQI (0.79)
4	Ec (-0.70)	$SINR$ (0.79)
5	n^{av} (0.27)	n^{av} (0.29)
6	n^{used} (0.22)	n^{used} (0.26)
7	v (0.15)	v (0.13)

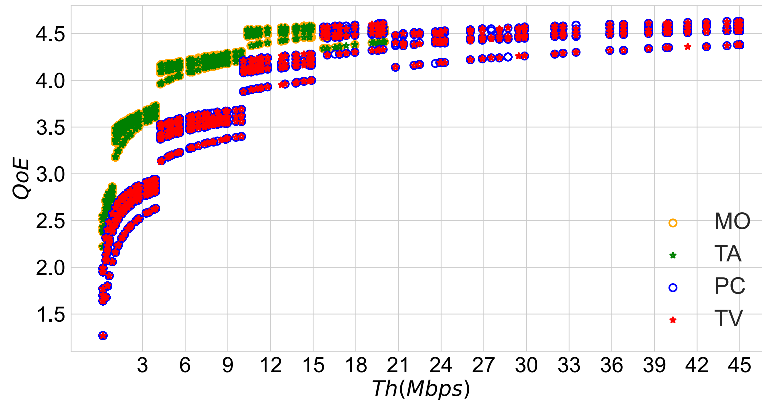

Fig. 7. Th and QoE relationship.

Table 7
Best hyperparameters for each ML model.

ML model	Hyperparameters (Best values)
LASSO	$alpha = 0.001$, $fit\_intercept : True$
LR	$fit\_intercept : True$
SVR	$C = 100$, $epsilon = 0.2$, $gamma : scale$, $kernel : rbf$
MLP	$hidden\_layer\_sizes = (72, 72, 72)$, $activation : relu$, $solver : sgd$, $learning\_rate = 0.05$, $batch\_size = 32$, $max\_iter = 3000$
kNN	$n\_neighbors = 10$, $weights : uniform$, $algorithm : auto$, $leaf\_size = 10$, $metric : manhattan$
RF	$n\_estimators = 150$, $max\_depth = 10$, $bootstrap : True$, $max\_features : None$, $min\_samples\_split = 10$, $min\_samples\_leaf = 1$, $max\_leaf\_nodes = 100$
GB	$n\_estimators = 100$, $learning\_rate = 0.05$, $max\_depth = 5$, $min\_samples\_leaf = 5$, $min\_samples\_split = 2$, $max\_features : None$, $subsample = 0.5$

6. Analysis of QoE estimation performance

To evaluate the ML performance, we used the following metrics: the coefficient of determination (R^2), mean absolute error (MAE), median absolute error ($MedAE$), mean absolute percentage error ($MAPE$), mean squared error (MSE), root mean squared error ($RMSE$), and PCC. Even if R^2 and $RMSE$ are generally the preferred performance evaluation metrics for regression tasks [36,38], we have also considered the other metrics to complement our analysis. The obtained performance has been maximized using the hyperparameters grid-search optimization. Thus, to better understand and compare the obtained results with their CCs, we have summarized the optimal configuration of the hyperparameters for each model in Table 7.

Table 8
ML training and testing results.

Metric	LASSO	LR	SVR	MLP	kNN	RF	GB
R^2 (training)	0.814	0.815	0.889	0.900	0.904	0.906	0.906
R^2 (testing)	0.811	0.811	0.886	0.899	0.894	0.900	0.902
MAE (testing)	0.297	0.296	0.206	0.191	0.193	0.193	0.185
$MedAE$ (testing)	0.269	0.268	0.146	0.132	0.121	0.124	0.112
$MAPE(\%)$ (testing)	9.120	9.100	6.510	6.260	6.290	6.330	6.080
MSE (testing)	0.128	0.128	0.077	0.070	0.072	0.069	0.068
$RMSE$ (testing)	0.357	0.357	0.277	0.264	0.270	0.262	0.259
PCC (testing)	0.900	0.901	0.942	0.948	0.945	0.949	0.950

6.1. Performance analysis

The analysis of PCC and SRCC showed a strong correlation between the Th, Ec, CQI, and $SINR$ features and the MOS scores. However, the other features can also provide valuable information in addition to these four features for estimating the quality, as suggested by the feature-QoE trend shown in Fig. 7. As a result, we decided to train the base models using the first four features initially. Then, based on the feature ranking obtained from PCC and SRCC, we incrementally trained the models by adding the remaining features (i.e., from 4 to 7).

Fig. 8 shows the resulting outcomes of R^2 during testing, which highlights the contribution provided by each feature to the performance enhancement of the different ML models. In particular, the figure underlines the importance of the v feature, which identifies the ED as a mobile or fixed device, intrinsically including information concerning the device's screen size. Indeed, mobile devices have smaller screens than fixed ones, affecting the perceived end user's QoE. Then, even if the feature v is weakly correlated with the QoE scores, the analysis in Fig. 8 demonstrates that adopting the whole set of features can improve model performance for QoE prediction.

Table 8 shows the ML training results in terms of R^2 using the whole set of features. Moreover, the table comprises the corresponding ML testing results in terms of R^2, MAE, $MedAE$, $MAPE$, MSE, $RMSE$, and PCC. The GB model achieved the best performance with an $R^2 = 0.902$ and $RMSE = 0.259$ during the ML testing and exhibited a strong correlation between the predicted QoE and the ground-truth QoE ($PCC = 0.950$). Furthermore, as reported in Table 7, the best results were obtained with the following hyperparameters: $n\_estimators = 100$, $learning\_rate = 0.05$, $max\_depth = 5$, $min\_samples\_leaf = 5$, $min\_samples\_split = 2$, $max\_features = None$, and $subsample = 0.5$. For completeness, although slightly lower than GB, the RF, kNN, MLP, and SVR models have also achieved good QoE prediction results.

The weaker performance achieved by LASSO and LR can be explained by their structural limitations. These models assume a linear relationship between inputs and the target variable, which restricts their capacity to model complex interactions. In QoE estimation, the influence of features such as bitrate, delay, and resolution is often interdependent and nonlinear. Therefore, linear models may fail to capture key aspects of the data distribution. Models based on decision trees or nonlinear kernels (e.g., GB and SVR) are more flexible and better equipped

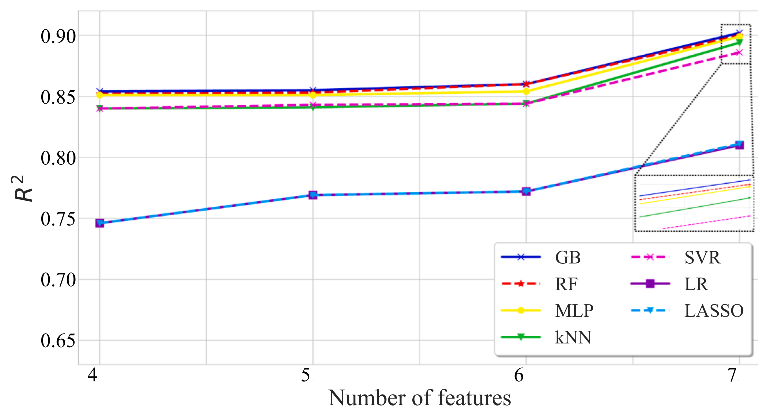

Fig. 8. R^2 results for an incremental number of features (ML testing).

to approximate these relationships, which accounts for their stronger performance in our evaluation.

6.2. CC analysis and performance trade-Off

We have excluded the LASSO and LR models based on the performance analysis results. Among the remaining models, it is important to note that their performance is similar regarding the metrics in Table 8. However, they differ in terms of complexity. Then, the complexity analysis focuses on identifying the lightest-weight estimation model. As shown in Table 3, models such as GB and RF involve higher training complexity due to the iterative construction of multiple decision trees. SVR and MLP also require substantial training effort, especially as the number of support vectors or hidden layers increases. On the other hand, kNN requires virtually no training, but its inference cost grows with the size of the dataset. These variations in training complexity are acceptable in our architecture because model training is performed offline within the Non-RT RIC, where resource constraints are relaxed. This design allows us to focus on inference efficiency at the Near-RT RIC, ensuring low-latency operation while still benefiting from expressive and accurate models.

In the context of the Near-RT RIC, CC is a key factor influencing model deployability. Control loops in this component typically operate within a latency window ranging from 10 ms to 1 s [43], which requires models to deliver fast and consistent inference. As shown in Table 3, the GB algorithm provides a good balance between accuracy and CC, with testing complexity of $O(f \times t)$, where f is the number of features and t the number of trees. This makes it a strong candidate for real-time QoE-aware Radio Resource Management. By contrast, models such as SVR and MLP can be less suitable for this layer due to their higher memory consumption and evaluation time.

As highlighted in recent literature, a key challenge for deploying AI/ML models within the O-RAN architecture lies in balancing prediction accuracy with computational efficiency [28]. While more complex models, such as MLP, may yield higher accuracy, they often incur higher inference costs, which can hinder their suitability for latency-sensitive tasks at the Near-RT RIC. In contrast, models like GB or even simpler ensemble methods (e.g., shallow RFs) offer an advantageous trade-off; they maintain competitive prediction performance while meeting tight latency constraints (e.g., sub-100 ms) expected in near-real-time control loops.

Given that model training is handled offline in the Non-RT RIC, the real-time burden shifts entirely to the inference phase. This further motivates the selection of models with lightweight runtime behavior. Ultimately, the model choice must consider both statistical performance and architectural feasibility in O-RAN, and our results aim to evaluate such decisions under the covered system constraints.

7. Application of the ML model for temporal QoE estimation

The previous sections demonstrated the capabilities of the proposed model in estimating the users' perceived QoE. As described in Section 5.1, the training process was based on a dataset of uncorrelated samples (each run provided the simulated network and application parameters of each device without any temporal evolution), recreating a New Radio UMi service area and collecting the information regarding a total of 10,000 EDs randomly distributed in the grid. An important distinction between the static and temporal datasets lies in the presence of network dynamics. While the static dataset assumes fixed conditions for each sample, the temporal dataset captures the evolution of QoE over time under varying channel quality and user mobility. This temporal modeling is essential, as users may experience fluctuations in perceived quality during a session due to handovers, interference, or congestion. These changes, although momentary, can significantly influence the overall perception of quality, especially in delay- or bitrate-sensitive applications. Therefore, analyzing temporal patterns enables a more realistic and robust assessment of user experience in dynamic network environments.

Thus, in this section, we aim to test the performance of the proposed ML-based QoE model for temporal QoE estimations of specific EDs over a predefined time frame. The goal is to investigate whether the estimated QoE remains accurate when the monitored channel quality conditions vary continuously over time. Therefore, this section evaluates the performance and generalization capacity of the proposed ML QoE model under a scenario where a set of users move in the simulation area. The used model was the GB, which performed better than the others, as presented in Section 6.

7.1. Simulation environment

As general simulation settings, we consider a 5G environment delivering a video service to 200 EDs, where 50% are randomly distributed between MOs and TAs, and the other 50% between PCs and TVs. For each simulation run, we registered 50 s with a resolution of 1 s. The simulations were conducted using the same ad-hoc 5G LLS used before [35].

In the defined scenario, we considered a New Radio UMi BS in the center of a grid (250×250 m), where 200 EDs are randomly distributed, ensuring a full range of CQI values from 1 to 15 among the samples. We executed 20 simulation runs, ensuring a 95% confidence value (CV) in the simulation outcomes. The goal of this scenario is to assess how the model performs for temporal QoE estimations. As presented in Section 4, the proposed model gives as output a QoE estimation per chunk, defined as the QoE_u perceived by the user ED_u in terms of the five-point scale of the MOS [13]. Then, we can obtain the overall QoE by averaging the per-second scores over time for each ED_u [44]. The equation for the

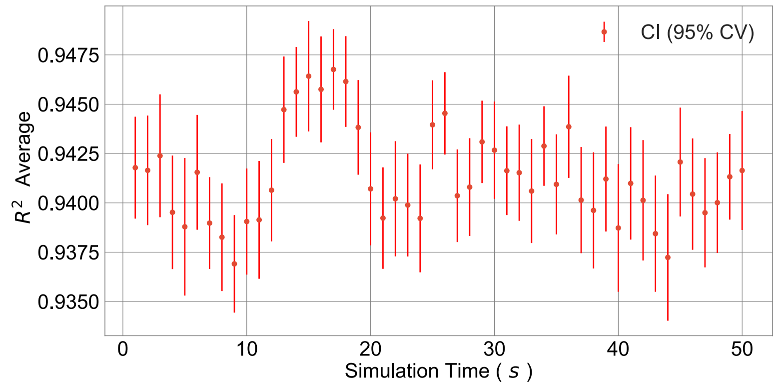

Fig. 9. R^2 average among the 200 EDs over the 50 s of simulation and for 20 simulation runs.

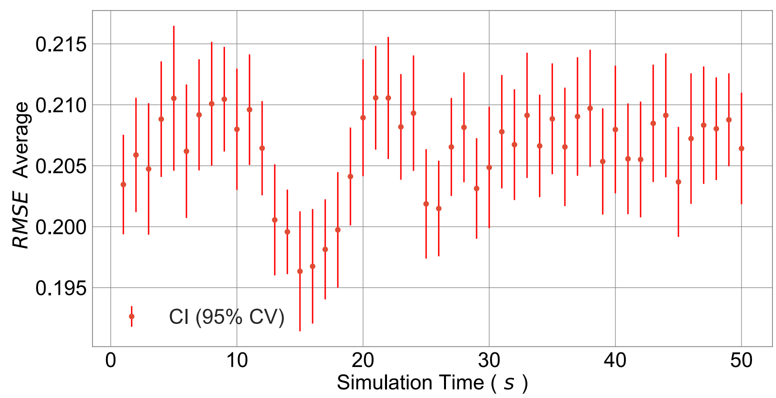

Fig. 10. $RMSE$ average among the 200 EDs over the 50 s of simulation and for 20 simulation runs.

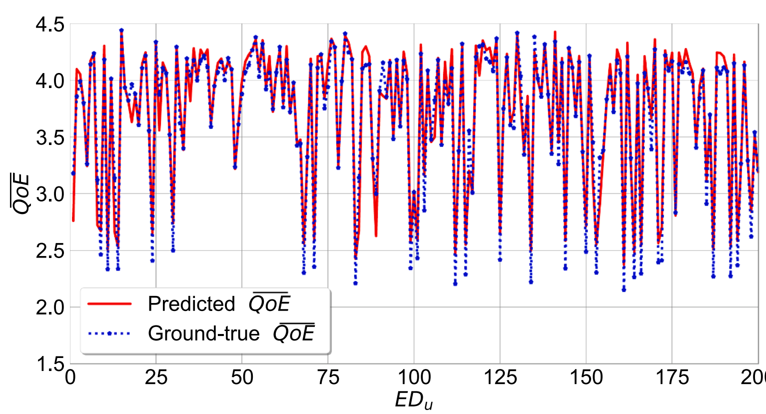

Fig. 11. $\overline{QoE}$ for 200 EDs in the network.

overall QoE experienced by ED_u is defined as

$$\overline{QoE_u} = \frac{\sum_{t'=1}^{T'} QoE_u^{t'}}{T'}, \tag{3}$$

where T' is the number of registered samples along the video session with a resolution of 1 s.

7.2. Results and analysis

This subsection provides the results for the analyzed scenario. Evaluating the proposed GB-ML model over the new unseen dataset (i.e., $200 \times 50 = 10000$ samples for each simulation run), we obtain a per-second QoE estimation with $R^2 = 0.941$, $MAE = 0.138$, $MSE = 0.043$, $MedAE = 0.079$, $MAPE = 4.66$ %, and $RMSE = 0.206$. This result proves the generalization capacity of the proposed solution facing a completely new dataset.

To assess the performance of the GB-ML model estimating the QoE when tracking the EDs over time, Figs. 9 and 10 display the R^2 and $RMSE$ average of estimating the QoE of 200 EDs during 50 s and 20 simulation runs. In every case, we can see how the CV is less than 1 % regarding the R^2 and $RMSE$ average.

On the other hand, Fig. 11 shows the overall QoE for each ED_u in the network ($\overline{QoE_u}$), including the ground-truth $\overline{QoE_u}$ and the predicted $\overline{QoE_u}$ resulting from averaging the QoE during 50 s, as in Eq. (3). In Fig. 11, the MAE of the predicted $\overline{QoE}$ is 0.1 with a maximum error always lower than 0.48. These results prove the high performance of the proposed solution, estimating the overall QoE ($\overline{QoE}$) when tracking the EDs along the simulation. Also, from Fig. 11, we can highlight how the higher absolute errors in the prediction occur mostly for $\overline{QoE} \leq 2.5$.

Table 9
Estimation's results per different levels of quality.

Metric	High QoE	Low QoE
R^2	0.906	0.895
MAE	0.070	0.172
$MedAE$	0.050	0.105
$MAPE(\%)$	1.695	6.689
MSE	0.011	0.0603
$RMSE$	0.106	0.246

This is also shown in Table 9, where the results for the cases of low and high quality are shown separately. This happens because, in the training dataset, only 10% of the samples have a QoE value lower than or equal to 2.5, as can be extrapolated from Fig. 3. This imbalance in the dataset between samples of EDs with QoE value lower than or equal to 2.5 and QoE higher than 2.5 produces this lower performance regarding the EDs with low QoE values. From a practical perspective, this impacts the system's ability to promptly detect and respond to service degradation. In real-time applications, such as adaptive streaming or QoE-driven RRM, failing to accurately identify poor-quality conditions may delay corrective actions, leading to a suboptimal user experience. This limitation suggests the need for future improvements through either real-world dataset enrichment or the integration of cost-sensitive learning strategies.

7.3. Data balancing

The original dataset of 10,000 samples exhibited a skewed QoE distribution (Fig. 3) with approximately 12% of samples having MOS ≤ 2.5, 18% in the range (2.5–3.0], 30% in (3.0–4.0], and 40% in (4.0–5.0]. To address this imbalance, the dataset was split into four MOS bins and balanced through a hybrid resampling procedure. Random undersampling was applied to the majority MOS bins, while Synthetic Minority Over-sampling Technique for Regression (SMOTER) [45] was used for the minority MOS bins. Each bin was then equalized to 2,500 samples, maintaining the total dataset size at 10,000 samples and ensuring a uniform MOS distribution.

Before balancing, the GB model achieved an overall performance of $R^2 = 0.900$ and $RMSE = 0.262$, while for low-QoE samples (MOS ≤ 2.5) the MAE and RMSE were 0.172 and 0.246, respectively. After retraining on the balanced dataset, the model achieved improved performance for low-QoE conditions, with $MAE = 0.131$ and $RMSE = 0.198$, and slightly higher global accuracy ($R^2 = 0.921$, $RMSE = 0.242$). These results confirm that the balancing strategy effectively enhanced the model's sensitivity to lower-quality scenarios without compromising global performance.

8. Limitations of the study

8.1. OTT-MNO Collaboration

The proposed solution assumes OTTs are willing to collaborate with MNO and share QoE data. However, there could be diverse barriers to collaboration, related to technological, economic, and data confidentiality aspects. Regarding technology, it is essential that MNO and OTT rely on IT systems, ensuring reliable and secure communication. The proposed solution considers open and standardized technological solutions (O-RAN, RESTful APIs, JSON data), which can easily overcome technological barriers. Concerning information exchange, the use of CAMARA[2] APIs represents a realistic and standardized alternative for obtaining network information from MNOs. However, such integration would require access to a real operational network exposure function (NEF) configuration, which is beyond the scope of the current simulation-based study. Nonetheless, the inclusion of CAMARA APIs in a future validation setup based on real field data will be the next step of this study. Moreover, the scalability of the solution needs to be considered to ensure the system achieves the same performance even in the case of a much larger number of users and different types of video streaming services.

The economic aspect is more complex, and it is likely to be the driving factor towards collaboration between the two actors. While a QoE-aware network management is beneficial to both MNOs and OTTs, as it likely enhances services and reduces users' churn, the OTT may demand compensation for the shared data. Thus, an economic agreement is required, which should consider the potential benefits received by the MNO from the shared QoE information. These benefits include the possibility to train an ML model on the network-related data, which gives the MNO the potential to manage the network resources in a QoE-aware manner, even if the OTT decides to stop collaboration. Finally, data confidentiality limitations may regard sharing QoE information of OTT applications with the MNOs. QoE is not the user's sensitive information, but it results from the mathematical computation of objective models on application-related data. However, some OTTs may consider this information confidential and neglect to share it.

8.2. Synthetic dataset

In the proposed solution, we have generated a synthetic dataset for the network data using a link-level simulator on a per-second scale granularity to simulate realistic data collection. Although synthetic data offers a flexible and controlled environment for model evaluation, we are aware that it cannot fully capture the complexity of real 5G networks. Indeed, particular factors, such as user mobility, interference, and diverse traffic profiles, introduce dynamics that may be difficult to reproduce synthetically. Additional factors may fail to be synthetically reproduced in the generation of the dataset for temporal QoE estimation, such as handovers, fading, and network congestion.

Moreover, we acknowledge the class imbalance present in our dataset, where low-value MOS samples are underrepresented. To solve this issue, we leveraged a hybrid resampling procedure by applying random undersampling to the majority MOS bins and SMOTER-based oversampling to the minority MOS bins. The QoE prediction performance of the ML model improved on the balanced dataset, which confirms this balancing strategy to effectively enhance the model's sensitivity to lower- quality scenarios without compromising global performance. However, we are aware that the application of undersampling and oversampling techniques on synthetic data can introduce additional bias. Thus, future work will focus on validating the models with real network traces and exploring techniques like cost-sensitive learning or weighted loss functions to improve performance across the full QoE spectrum. These steps will help ensure both robustness and generalizability in practical deployments. In particular, robustness to noise and anomalies in the input data is required to be investigated by, for example, injecting controlled noise and outlier perturbations into the collected real data to evaluate the model's stability and generalization under distorted input conditions, as typically observed in operational O-RAN systems.

9. Conclusion

This paper proposes an ML-based QoE prediction model for a video streaming service over an MNO 5G network, leveraging the O-RAN architecture. The proposed solution addresses the challenge of predicting QoE solely based on network data, overcoming the limitations of traditional QoS-based models and the lack of application data available to MNOs. By leveraging collaborative OTT and MNO interaction, our approach enables accurate QoE predictions through data sharing and an advanced ML model. The performance evaluation on per-second video chunks demonstrated that the Gradient Boosting model outperformed

[2] https://camaraproject.org/api-overview/

other ML models, achieving an RMSE score of 0.259, highlighting accurate QoE estimations with minimal computational complexity. Furthermore, we investigated the GB performance for temporal QoE estimations of specific EDs over a predefined timeframe, which confirmed the robustness of the proposed model in a real-time scenario (an RMSE score of 0.206 was achieved in this case).

Our findings suggest that the proposed ML-based QoE prediction model can significantly enhance user experience in 5G networks by improving resource management and dynamic optimization of network resource allocation. Future work will focus on validating the models with real network traces obtained with an operational 5G network testbed and by relying on open standardized APIs, such as CAMARA APIs, to collect real network information from MNOs. The implementation of the solution in a real O-RAN infrastructure will provide the means for testing the actual performance of the QoE prediction model in terms of accuracy and latency, as well as the robustness of the solution to noise in the input data and scalability issues. Additionally, whereas this solution has been tailored to video streaming services, the proposed approach can be extended to other application scenarios.

CRediT authorship contribution statement

Claudia Carballo González: Writing – original draft, Software, Methodology, Investigation, Formal analysis, Conceptualization; **Ernesto Fontes Pupo:** Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization; **Alessandro Floris:** Writing – review & editing, Supervision, Project administration, Methodology, Conceptualization; **Simone Porcu:** Writing – review & editing, Supervision, Methodology, Formal analysis, Conceptualization; **Maurizio Murroni:** Writing – review & editing, Resources, Project administration, Conceptualization; **Luigi Atzori:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

Data availability

Data will be made available on request.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Maurizio Murroni reports financial support was provided by EU Horizon Europe Framework Programme (HORIZON). Luigi Atzori reports financial support was provided by European Union - Next Generation EU under the Italian National Recovery and Resilience Plan (NRRP). If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was carried out within the "HEAT - Hybrid Extended reAliTy" Project GA 101135637 funded by the EU Horizon Europe Framework Programme (HORIZON). This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them. This work has also been partially supported by the European Union - Next Generation EU under the Italian National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.3, CUP C29J24000300004, partnership on "Telecommunications of the Future" (PE00000001 - program "RESTART"), and by the European Union under the Italian NRRP of NextGenerationEU, "Sustainable Mobility Center" Centro Nazionale per la Mobilità Sostenibile, CNMS, CN_00000023.

References

[1] P. Le Callet, S. Möller, A. Perkis, Qualinet White Paper on Definitions of Quality of Experience (2012), 2012, . European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), Lausanne, Switzerland, Version 1.2, March 2013.
[2] D. Tsolkas, E. Liotou, N. Passas, L. Merakos, A survey on parametric QoE estimation for popular services, J. Netw. Comput. Appl. 77 (1) (2017) 1–17. https://doi.org/10.1016/j.jnca.2016.10.016
[3] Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy, M. Manohara, Toward a practical perceptual video quality metric, Online: https://netflixtechblog.com/toward-a-practical-perceptual-video-quality-metric-653f208b9652 (2016).
[4] ITU, ITU-T Rec. p.1203.1 - Parametric bitstream-Based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport–video quality estimation module (2019).
[5] B. Palit, S. Argha, A. Mondal, A. Zunaid, J. Jayatheerthan, S. Chakraborty, Improving UE energy efficiency through network-Aware video streaming over 5G, IEEE Trans. Netw. Serv. Manage. 20 (3) (2023) 3487–3500. https://doi.org/10.1109/TNSM.2023.3250520
[6] R. Farahani, E. Çetinkaya, C. Timmerer, M. Shojafar, M. Ghanbari, H. Hellwagner, ALIVE: A latency- and cost-Aware hybrid P2P-CDN framework for live video streaming, IEEE Trans. Netw. Serv. Manage. 21 (2) (2024) 1561–1580. https://doi.org/10.1109/TNSM.2023.3335190
[7] G. Kougioumtzidis, A. Vlahov, V.K. Poulkov, P.I. Lazaridis, Z.D. Zaharis, QoE Prediction for gaming video streaming in O-RAN using convolutional neural networks, IEEE Open J. Commun. Soc. 5 (2024) 1167–1181. https://doi.org/10.1109/OJCOMS.2024.3362275
[8] P. Uthansakul, P. Anchuen, M. Uthansakul, A.A. Khan, Estimating and synthesizing QoE based on QoS measurement for improving multimedia services on cellular networks using ANN method, IEEE Trans. Netw. Serv. Manage. 17 (1) (2020) 389–402. https://doi.org/10.1109/TNSM.2019.2946091
[9] S. Schwarzmann, C.C. Marquezan, R. Trivisonno, S. Nakajima, V. Barriac, T. Zinner, ML-Based QoE Estimation in 5G networks using different regression techniques, IEEE Trans. Netw. Serv. Manage. 19 (3) (2022) 3516–3532. https://doi.org/10.1109/TNSM.2022.3179924
[10] G. Miranda, D.F. Macedo, J.M. Marquez-Barja, Estimating video on demand QoE from network QoS through ICMP probes, IEEE Trans. Netw. Serv. Manage. 19 (2) (2022) 1890–1902. https://doi.org/10.1109/TNSM.2021.3129610
[11] P.H.S. Panahi, A. Hossein Jalilvand, A. Diyanat, An efficient network-Based QoE assessment framework for multimedia networks using a machine learning approach, IEEE Open J. Commun. Soc. 6 (2025) 1653–1669. https://doi.org/10.1109/OJCOMS.2025.3543750
[12] E. Fontes Pupo, C. Carballo González, J. Montalban, P. Angueira, M. Murroni, E. Iradier, Artificial intelligence aided low complexity RRM algorithms for 5G-MBS, IEEE Trans. Broadcast. 70 (1) (2024) 110–122. https://doi.org/10.1109/TBC.2023.3311337
[13] ITU, ITU-T Rec. p.1204.3 - Video quality assessment of streaming services over reliable transport for resolutions up to 4K with access to full bitstream information, 2020.
[14] Y. Chen, K. Wu, Q. Zhang, From QoS to QoE: a tutorial on video quality assessment, IEEE Commun. Surv. Tutorial. 17 (2) (2015) 1126–1165. https://doi.org/10.1109/COMST.2014.2363139
[15] ITU, ITU-T Rec. p.910 - Subjective video quality assessment methods for multimedia applications (2023).
[16] R.R.R. Rao, S. Göring, P. List, W. Robitza, B. Feiten, U. Wüstenhagen, A. Raake, Bitstream-Based model standard for 4K/UHD: ITU-T p.1204.3 – model details, evaluation, analysis and open source implementation, in: 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX), 2020, pp. 1–6. https://doi.org/10.1109/QoMEX48832.2020.9123110
[17] N. Barman, Y. Reznik, M.G. Martini, A subjective dataset for multi-Screen video streaming applications, in: 2023 15Th International Conference on Quality of Multimedia Experience (QoMEX), 2023, pp. 270–275. https://doi.org/10.1109/QoMEX58391.2023.10178645
[18] G. Kougioumtzidis, V. Poulkov, Z.D. Zaharis, P.I. Lazaridis, A survey on multimedia services QoE assessment and machine learning-Based prediction, IEEE Access 10 (2022) 19507–19538. https://doi.org/10.1109/ACCESS.2022.3149592
[19] L. Yang, J. Liu, S. Li, D. Zhang, Z. Xia, Enhancing user experience in ultra HD cloud performing arts live streaming: a QoS-to-Qoe mapping approach, IEEE Trans. Broadcast. 70 (2) (2024) 413–428. https://doi.org/10.1109/TBC.2024.3358756
[20] D. Kafetzis, N. Fotiou, S. Argyropoulos, J. Nasreddine, I. Koutsopoulos, Video Quality Monitoring for Remote Autonomous Vehicle Control, arXiv:2506.03166 (2025).
[21] Y.L. Lee, T.C. Chuah, J. Loo, A. Vinel, Recent advances in radio resource management for heterogeneous LTE/LTE-A networks, IEEE Commun. Surv. Tutorial. 16 (4) (2014) 2142–2180. https://doi.org/10.1109/COMST.2014.2326303
[22] T.O. Olwal, K. Djouani, A.M. Kurien, A survey of resource management toward 5G radio access networks, IEEE Commun. Surv. Tutorial. 18 (3) (2016) 1656–1686. https://doi.org/10.1109/COMST.2016.2550765
[23] J. Zhou, D. Ding, Z. Wu, Y. Xiu, Spatial context-Aware time-Series forecasting for QoS prediction, IEEE Trans. Netw. Serv. Manage. 20 (2) (2023) 918–931. https://doi.org/10.1109/TNSM.2023.3250512
[24] H. Zhang, C. Huang, J. Zhou, L. Chen, QoS-Aware Virtualization resource management mechanism in 5G backhaul heterogeneous networks, IEEE Access 8 (2020) 19479–19489. https://doi.org/10.1109/ACCESS.2020.2967101
[25] B. Agarwal, M.A. Togou, M. Ruffini, G.-M. Muntean, QoE-Driven Optimization in 5G O-RAN-Enabled hetnets for enhanced video service quality, IEEE Commun. Mag. 61 (1) (2023) 56–62. https://doi.org/10.1109/MCOM.003.2200229

[26] M. García-Torres, D.P. Pinto-Roa, C. Núñez-Castillo, B. Quiñonez, G. Vázquez, M. Allegretti, M.E. García-Diaz, Feature selection applied to QoS/QoE modeling on video and web-based mobile data services: an ordinal approach, Comput. Commun. 217 (2024) 230–245. https://doi.org/10.1016/j.comcom.2024.02.004
[27] A.A. Barakabitze, R. Walshe, SDN And NFV for QoE-driven multimedia services delivery: the road towards 6G and beyond networks, Comput. Netw. 214 (2022) 109133. https://doi.org/https://doi.org/10.1016/j.comnet.2022.109133
[28] M. Polese, L. Bonati, S. D'Oro, S. Basagni, T. Melodia, Understanding O-RAN: architecture, interfaces, algorithms, security, and research challenges, IEEE Commun. Surv. Tutorial. 25 (2) (2023) 1376–1411. https://doi.org/10.1109/COMST.2023.3239220
[29] O-RAN Alliance, O-RAN Work group 1 (use cases and overall architecture) O-RAN architecture description (February 2024).
[30] O-RAN Alliance, O-RAN Work group 2 (non-RT RIC: architecture-R003 v05.00) (February 2024).
[31] O-RAN Alliance, O-RAN Work group 3 (near-Real-time RAN intelligent controller and E2 interface-R003 v05.00) (October 2023).
[32] C.C. González, E.F. Pupo, E. Iradier, P. Angueira, M. Murroni, J. Montalban, Network selection over 5G-Advanced heterogeneous networks based on federated learning and cooperative game theory, IEEE Trans. Veh. Technol. 73 (8) (2024) 11862–11877. https://doi.org/10.1109/TVT.2024.3373638
[33] M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hoßfeld, P. Tran-Gia, A survey on quality of experience of HTTP adaptive streaming, IEEE Commun. Surv. Tutorial. 17 (1) (2015) 469–492. https://doi.org/10.1109/COMST.2014.2360940
[34] P. Nain, D. Towsley, B. Liu, Z. Liu, Properties of random direction models, in: Proceedings IEEE 24Th Annual Joint Conference of the IEEE Computer and Communications Societies., 3, 2005, pp. 1897–1907 vol. 3. https://doi.org/10.1109/INFCOM.2005.1498468
[35] E.F. Pupo, C.C. González, E. Iradier, J. Montalban, M. Murroni, 5G Link-Level simulator for multicast/broadcast services, in: 2023 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), 2023, pp. 1–6. https://doi.org/10.1109/BMSB58369.2023.10211507
[36] C. Carballo González, E.F. Pupo, D.P. Ruisánchez, D. Plets, M. Murroni, From MFN to SFN: performance prediction through machine learning, IEEE Trans. Broadcast. 68 (1) (2022) 180–190. https://doi.org/10.1109/TBC.2021.3132804
[37] W. Fu, T. Menzies, Easy over hard: a case study on deep learning, in: Proceedings of the 2017 11Th Joint Meeting on Foundations of Software Engineering, ESEC/FSE 2017, Association for Computing Machinery, 2017, p. 49–60. https://doi.org/10.1145/3106237.3106256
[38] A. Géron, Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow, "O'Reilly Media, Inc.", 2022.
[39] 3GPP TR 38.901, 5G; Study on channel model for frequencies from 0.5 to 100 GHz (3GPP TR 38.901 version 16.1.0 release 16) (2020).
[40] N. Chukhno, O. Chukhno, S. Pizzi, A. Molinaro, A. Iera, G. Araniti, Efficient management of multicast traffic in directional mmwave networks, IEEE Trans. Broadcast. 67 (3) (2021) 593–605. https://doi.org/10.1109/TBC.2021.3061979
[41] G. Bingöl, A. Floris, S. Porcu, C. Timmerer, L. Atzori, Are quality and sustainability reconcilable? a subjective study on video QoE, luminance and resolution, in: 2023 15Th International Conference on Quality of Multimedia Experience (QoMEX), 2023, pp. 19–24. https://doi.org/10.1109/QoMEX58391.2023.10178513
[42] C.C. González, E.F. Pupo, G. Bingöl, A. Floris, S. Porcu, M. Murroni, L. Atzori, A QoE-based energy-aware resource allocation solution for 5G heterogeneous networks, in: 2024 16Th International Conference on Quality of Multimedia Experience (QoMEX), 2024, pp. 29–35. https://doi.org/10.1109/QoMEX61742.2024.10598282
[43] E. Municio, G. Garcia-Aviles, A. Garcia-Saavedra, X. Costa-Pérez, O-RAN: Analysis of latency-Critical interfaces and overview of time sensitive networking solutions, IEEE Commun. Stand. Magaz. 7 (3) (2023) 82–89. https://doi.org/10.1109/MCOMSTD.0001.2200041
[44] A. Raake, S. Borer, S.M. Satti, J. Gustafsson, R.R.R. Rao, S. Medagli, P. List, S. Göring, D. Lindero, W. Robitza, G. Heikkilä, S. Broom, C. Schmidmer, B. Feiten, U. Wüstenhagen, T. Wittmann, M. Obermann, R. Bitto, Multi-Model standard for bitstream-, pixel-Based and hybrid video quality assessment of UHD/4K: ITU-T p.1204, IEEE Access 8 (2020) 193020–193049. https://doi.org/10.1109/ACCESS.2020.3032080
[45] L. Torgo, R. Ribeiro, B. Pfahringer, P. Branco, SMOTE For regression, 8154, 2013, pp. 378–389. https://doi.org/10.1007/978-3-642-40669-0_33