IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 17 November 2023, accepted 12 December 2023, date of publication 18 December 2023,
date of current version 22 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3343846

RESEARCH ARTICLE

Deep Learning-Aided QoE Prediction for Virtual Reality Applications Over Open Radio Access Networks

GEORGIOS KOUGIOUMTZIDIS[1], ATANAS VLAHOV[2], VLADIMIR K. POULKOV[1], (Senior Member, IEEE), PAVLOS I. LAZARIDIS[3], (Senior Member, IEEE), AND ZAHARIAS D. ZAHARIS[4], (Senior Member, IEEE)

[1]Department of Communication Networks, Technical University of Sofia, 1000 Sofia, Bulgaria
[2]Intelligent Communication Infrastructure Laboratory, Sofia Tech Park, 1784 Sofia, Bulgaria
[3]School of Computing and Engineering, University of Huddersfield, HD1 3DH Huddersfield, U.K.
[4]School of Electrical and Computer Engineering, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece

Corresponding author: Georgios Kougioumtzidis (gkougioumtzidis@tu-sofia.bg)

This research was supported by the European Union, through the Horizon 2020 Marie Skłodowska-Curie Innovative Training Networks Programme "Mobility and Training for beyond 5G Ecosystems (MOTOR5G)" under grant agreement no. 861219, and the "Intelligent Communication Infrastructure Laboratory" at Sofia Tech Park, Sofia, Bulgaria.

ABSTRACT Nowadays, innovative applications in the field of virtual reality (VR) are being developed, attracting the interest of both academia and industry. Wireless VR applications focus on various aspects of daily life, such as smart education, entertainment, healthcare, tourism, architecture, automotive, and industrial automation. All these inherently interactive applications that aim to create immersive experiences for users are closely related to the concept of quality of experience (QoE), which expresses the quality of a service as perceived by end-users. In this paper, we develop an objective QoE prediction model based on deep learning techniques. The prediction model examines the impact of wireless network operation on the quality of VR 360-degree video streaming. It is based on an encoder-decoder long short-term memory (LSTM) neural network and is able to predict in real-time the overall transmission-related QoE value using only measurable quality of service (QoS) parameters. The prediction model is tested and evaluated on an open radio access network testbed with interfaces based on the O-RAN specifications.

INDEX TERMS Deep learning, deep neural network (DNN), long short-term memory (LSTM), open radio access network (Open RAN), quality of experience (QoE), QoE prediction, recurrent neural network (RNN), virtual reality (VR).

I. INTRODUCTION

Alongside the remarkable advancement of Internet technologies, network data traffic driven by video streaming services has been constantly growing in recent years, bringing new challenges to current wireless communication networks. The popularity of streaming media is growing, creating new markets and business models, and spawning new video formats such as virtual reality (VR) 360-degree video. VR 360-degree video offers omnidirectional content that creates a much more immersive viewing experience compared to conventional two-dimensional (2D) videos. This has the consequence of attracting the attention of over-the-top and video-on-demand platforms, which are constantly increasing their content in this type of video. In addition, the head-mounted displays (HMDs) required to view VR 360-degree videos are becoming more affordable and available to the public, allowing a larger audience to access them.

VR 360-degree video and all other examples of VR applications are based on interactivity and immersion, making the users' quality of experience (QoE) a particularly important aspect for their effective deployment. Both in the

The associate editor coordinating the review of this manuscript and approving it for publication was Alessandro Floris.

design of new use cases and in the implementation of VR applications, it is critical to capture and understand the user experiences of VR services as defined and quantified by QoE. As a metric, QoE can be considered a multidisciplinary quantity influenced by a variety of parameters attributed to different scientific domains. In recent years, QoE has been established as an important tool for evaluating the quality of operation of wireless communication networks. This is because it allows a deeper understanding of how the quality of service (QoS) parameters of a network affect the quality of communication services as perceived by end-users.

Today's wireless communication networks cannot adequately meet the extremely high demands of VR applications for very high bandwidth, ultra-low latency, high energy efficiency, and high computational capabilities. Consequently, there is a need to develop improved wireless network architectures capable of providing programmable, flexible, scalable, and virtualized communication solutions to address these challenges. The technology that holds great promise for transforming the radio access network (RAN) architecture of wireless and mobile networks into tailored communication solutions is that of the open RAN (Open RAN), as proposed by the O-RAN Alliance [1]. Open RAN can be regarded as a disruptive and innovative technology within the wireless communications ecosystem based on two cardinal concepts: openness and intelligence. A major advantage of developing communication networks based on Open RAN technology lies in its architecture, where the implementation of specialized intelligent network controllers is foreseen.

The benefits of QoE prediction in a wireless communication system with respect to system performance arise from the ability to proactively analyze the patterns and dynamic characteristics of QoE, based on which a network optimization policy can be implemented. Consequently, accurate QoE prediction allows for efficient planning of the allocation of both radio and computational resources, improving network energy efficiency, limiting network congestion, as well as reducing capital expenditure (CapEx) and operating expenditure (OpEx) and improving profitability.

In this paper, we propose an objective QoE prediction model for VR 360-degree video streaming supported by deep learning methodologies. Specifically, the prediction model is based on an encoder-decoder long short-term memory (LSTM) deep neural network (DNN), which is applied to an Open RAN wireless network based on the O-RAN architecture and specifications. The purpose of the prediction model is to quantify in real-time the effect of wireless network operation on VR service quality as perceived by end-users. The QoE evaluation can be classified as an objective method, since the prediction model is able to predict the overall QoE value using only measurable QoS parameters. These QoS parameters, which are the key performance indicators (KPIs) of the network, are collected and recorded using a QoS monitoring system based on open-source monitoring tools. The prediction model maps the collected values of the network KPIs, namely bandwidth, delay, and packet loss, to mean opinion score (MOS) values, and the encoder-decoder LSTM network examines the interdependencies between them to determine the overall QoE value for the VR 360-degree video streaming service.

The structure of the paper is as follows: in Section II, we review state-of-the-art models for QoE management for wireless VR video streaming and VR 360-degree video QoE prediction. In Section III, we classify QoE assessment methods in relation to VR applications. Moreover, we examine the VR QoE influencing factors (IFs) and analyze the network/transmission-related IFs. In addition, we analyze the QoS/QoE mapping methodologies. In Section IV, we present the proposed QoE prediction model. First, we describe the Open RAN testbed. Next, we introduce the QoS monitoring system and data generation process. Then, we define the QoS/QoE mapping model, and finally, we analyze the deep learning-based QoE prediction model. In Section V, we present a comparative study of DNN techniques that belong to the RNN category and demonstrate the performance evaluation and measurement results for the proposed prediction model. Finally, Section VI includes final remarks and conclusions.

To the best of the authors' knowledge, this is the first work that quantifies and predicts the QoE of a wireless VR application over Open RAN. In addition, it presents for the first time a comparative study of DNN techniques belonging to the recurrent neural network (RNN) category for this type of application. Real-time QoE prediction should be considered a critical parameter for the design of next-generation wireless communication networks, especially for the transmission of services with inherent interactivity, such as VR applications.

II. RELATED WORK

In recent years, there has been significant interest in developing QoE management models for wireless streaming of ultra-high-definition (UHD) and VR video and in QoE prediction for VR 360-degree video.

In [2], an analysis of the impact of network parameters on the evaluation of QoE, taking into account both subjective and objective approaches, is offered. This analysis took into account video and receiver characteristics and wireless channel capacity. The model includes the use of the least absolute shrinkage and selection operator regression (LASSO) algorithm to predict QoE and adjust the video quality accordingly.

In [3], a generalized QoE model derived from experiments involving subjective quality evaluations is presented. Initially, the model uses VR 360-degree video processing and wireless transmission methods to conduct subjective experiments. Then, the effect of various factors on the user experience, such as user viewing angle, tiling, stall and resolution switch, is mathematically analyzed. Finally, a universal QoE model is proposed to manage wireless VR 360-degree video streaming.

In [4], a platform for reinforcement learning (RL) is introduced, showcasing its ability to dynamically adjust the feedback control loop within a wireless network. This study focuses on the influence of reconfigurable queuing on video streaming. The model encompasses both model-free and model-based RL strategies to address the challenge of assigning clients to specific queues during each decision period, especially in high-load scenarios.

In [5], a transmission strategy for VR video is proposed. This method constructs an integer nonlinear programming (INLP) problem with the goal of maximizing the average QoE in wireless VR 360-degree video transmission. By exploiting the correlation between the structure of viewport tiles and spectral efficiency expression, the model introduces a multi-stream clustering technique to enhance the overall VR video transmission performance.

In [6], an examination of a network structure tailored for mobile VR video streaming applications is undertaken. The study encompasses various modulation and coding schemes offering different rates, with a subsequent analysis of performance metrics including blocking probability, throughput, queueing delay, and average packet error rate, taking into account Nakagami-m fading on the wireless link. Numerical findings show that the proposed system ensures high QoE levels for diverse mobile VR video streaming applications.

In [7], a two-phase prediction method is employed to assess the perceived quality of adaptive VR video streaming in mobile networks. In the first phase, the prediction model utilizes a decision tree regression algorithm, providing an estimate of video playback performance by incorporating network QoS metrics as predictive factors. In the next phase, it models and predicts end-user perceived quality by considering the expected metrics for VR video playback performance.

In [8], a method that combines neural network-driven viewport estimation with a rate control system is introduced. The proposed model utilizes a three-dimensional convolutional neural network (3D CNN) and LSTM. It considers two algorithms for non-tiled adaptive video streaming, namely the buffer-based algorithm (BBA) and the buffer occupancy-based Lyapunov algorithm (BOLA), along with two variants of tiled adaptive video streaming algorithms, namely viewport only and viewport plus.

In [9], a model for predicting QoE in VR 360-degree videos is proposed, employing both the linear regression algorithm and a neural network utilizing the stochastic gradient descent (SGD) optimization algorithm. This model considers two dimensions of QoE: perceptual quality and cybersickness. Additionally, it introduces two supplementary parameters influencing QoE evaluation: the level of familiarity with VR and the degree of interest in 360-degree video.

In [10], a QoE prediction technique is created using a neural network optimized through SGD. This method is designed to estimate the extent of cybersickness induced by VR 360-degree videos, particularly during stalling events. The evaluation and prediction of cybersickness levels involve metrics such as the simulator sickness questionnaire (SSQ), virtual reality sickness predictor (VRSP), virtual reality sickness assessment (VRSA), and visual comfort assessment (VCA). Additionally, the user's sense of presence is assessed using the igroup presence questionnaire (IPQ).

In [11], a method for estimating QoE using a decision tree algorithm is introduced. This method involves a subjective exploration of factors influencing QoE in VR 360-degree videos, such as quantization parameter (QP), resolution, and initial delay. The model predicts four VR QoE factors based on subjective evaluation: immersion, acceptability, reality judgment, and attention captivation.

Our work is distinct from the above state-of-the-art approaches, as it focuses on the effect of the transmission channel on QoE. The vast majority of works on QoE assessment and prediction focus on human and system IFs, as analyzed in Section III-C, without examining the impact of network/transmission-related IFs. In the few cases where the impact of the wireless transmission channel is considered, this is achieved by using network simulators and emulators. Our work is the first to use a real 4G/5G wireless network based on Open RAN technology and examine the effect of network/transmission-related IFs in real-world conditions. In addition, our work focuses on developing a purely objective QoE prediction model based on the utilization of QoS/QoE mapping functions. In particular, we consider real-world network design scenarios where the use of subjective QoE evaluation and prediction models is not feasible due to their limitations, as described in Section III-A.

III. VIRTUAL REALITY QOE ASSESSMENT

The perceived quality of an application or service by end-users is defined by the QoE measure. Consequently, in the context of mobile and wireless networks, QoE expresses how network operation affects the perceived quality of communication services. The evaluation of QoE is based on the juxtaposition of the intended quality features that set the user's expectations with the perceived features resulting from a natural stimulus. Immersive applications present some peculiarities in QoE evaluation, which mainly focus on the experience itself, as well as the IFs [12]. In immersive applications, the following can have an important impact on QoE: 1) the "sense of being there", which refers to a sense of embodiment in a VR environment; 2) the "place of illusion", which permits users to perceive virtual objects as real; and 3) the interaction among the user and the VR environment [13]. Two methods for QoE evaluation are mentioned in the literature: subjective and objective evaluation. Subjective models use human evaluators who assess the quality of a service after being exposed to a series of tests or stimuli. In contrast, objective models capture end-users' perceived quality using only objective quality metrics [14].

A. SUBJECTIVE METHODOLOGIES

In subjective methods, the evaluators' perception of service quality is quantified using psychophysical and psychometric methodologies. In addition, qualitative approaches such as focus groups, interviews and profile surveys are used to find out which IFs affect QoE and to what extent [15]. Typically, evaluators assign a series of perceived quality attributes to a MOS scale, which is scaled from 1 to 5 (i.e., poor to excellent), indicating the level of satisfaction with a service [16]. Subjective evaluation gives precise results due to the immediate reception of data from end-users. Subsequently, these results can be used as a reference for training and validating QoE prediction models. The inherent disadvantages of subjective methods arise from the fact that they usually take place in a laboratory setting and, consequently, are expensive, time-consuming, incapable of real-time use, and unreproducible [17].

Regarding VR applications, evaluators should use HMDs to maintain immersive features and ensure accurate perception of panoramic video quality [18]. In addition, the spatial and temporal sensory content of the test video sequence is of particular importance, as it determines the acceptable degree of video compression and, hence, the level of quality degradation during the wireless transmission of the test video sequence [19]. Subjective VR QoE evaluation methodologies need to map evaluators' scores to a range of perceptual scales with respect to parameters such as audiovisual quality, simulator sickness symptoms, and exploration behavior [20].

B. OBJECTIVE METHODOLOGIES

Due to the restraints imposed by subjective methodologies, there has been intense interest in developing objective models that forecast the perceived quality of end-users based purely on the quality characteristics of communications networks. The principle of operation of the objective methodologies is based on achieving a prediction of the QoE values that approximate the evaluations of the subjective methodologies. The main advantage of objective methodologies is their simplicity of application and modification, since the evaluation process needs to consider only measurable QoS parameters and mathematical models that relate these measurements to QoE values. The disadvantage of objective methodologies is their inaccuracy, since the calculated QoE is an estimation and not precise representation of the perceived quality by end-users [21].

C. VIRTUAL REALITY QOE INFLUENCING FACTORS

The actual state and adjustment associated with any characteristic of a user, system, service, application or environment that may affect the perceived quality of the end-user can be considered a QoE IF [12]. IFs contain, among others, features that pertain to the characteristics of the application or service, the context of use, the accomplishment of user expectations, as well as the user's cultural identity, social and economic standing, and cognitive and emotional status [22]. VR QoE IFs are classified in three major categories: 1) human-related; 2) system-related; and 3) context-related [4], [23]:

1) *Human IFs* can be any changeable or immutable feature or characteristic of a user. Human IFs affect both the growth and overall quality of an experience. They are particularly complex due to their subjective nature and relationship with inner human processes [24]. Human IFs in VR-based applications contain the following: 1) physiological characteristics of the user; 2) anomalies in the human visual system (HVS) [25] and impairments in the human auditory system (HAS) [26]; 3) simulator sickness [27]; 4) immersion; and 5) the user's expectations and competence with VR applications.

2) Properties and characteristics that affect the functionality of a VR application in terms of technical parameters constitute *system IFs*. VR system IFs are subdivided into four classes, as content-related, media/encoder-related, network/broadcast-related, and hardware-related:

- Content characteristics significantly affect the overall system QoE and refer to attributes such as spatial audio, spatial depth, and spatiotemporal complexity.
- Media/codec IFs are properties of media modulation characteristics such as compression, video codecs, audio, storage and transport, bit rate, resolution, frame rate, audio sample rate, and encoding delay.
- Network/transmission IFs are affected by transmission channel deficiencies and pertain to network QoS parameters of delay, bandwidth, and packet loss.
- Hardware IFs refer to features of user devices, including HMDs, headphones, decoder performance, head tracking, field of view (FoV), screen resolution, and refresh rate.

3) *Context IFs* consider the user's environment and are divided into physical context factors, temporal context factors, social context factors, and task context factors.

D. NETWORK/TRANSMISSION INFLUENCING FACTORS

A wireless VR 360-degree video streaming service requires the transmission of a huge amount of omnidirectional video content with extremely low latency. Also, to create truly immersive VR experiences, the ability to smoothly downlink high-resolution VR video and haptic feedback is crucial. Therefore, the network/transmission parameters of latency, bandwidth, and packet loss play an important role in maintaining immersive experiences. Because of the limitations of human perception, a procedure for mapping these parameters to attributes that directly affect the perceived quality of VR applications by end-users can be implemented [28].

1) *Delay*: VR applications must meet stringent latency requirements to deliver a truly immersive experience. The most common forms of latency include queuing delay, over-the-air delay, buffer delay, and motion-to-photon delay. Delay effects attributes like initial loading time and stalling, causing poor presentation quality [29]. Because human perception requires accurate and seamless visual stimuli, VR should be treated as a latency-critical application, with latency being the

most important IF, as it leads to dissociated VR experiences and can worsen simulator sickness [30].

2) *Bandwidth*: The amount of necessary data to create immersive experiences in a VR 360-degree video streaming environment is significant. Consequently, if the necessary bandwidth is not provided during transmission, the quality of the VR setup will deteriorate. Therefore, overload can result in substantial packet loss and excessive latency, lowering the perceived quality of the VR service [23]. Network bandwidth is defined as the throughput necessary to transmit a VR 360-degree video and can be regarded as one of the most crucial aspects affecting a VR application [30].

3) *Packet loss*: The transmission scheme designates how packet loss affects the creation of immersive VR experiences. In reliable transmission modes, the necessary packet retransmissions due to packet loss increase the overall delay. In unreliable transmission modes, audiovisual quality is degraded due to packet loss, which causes video freezing and tiling artifacts [23].

E. QOS/QOE MAPPING

The idea behind QoS/QoE mapping is to derive QoE values based on measurable input parameters. To determine the connection between QoS parameters and their distinct and reciprocal influence on QoE, a correlation model must be created. The QoS parameters are related to the operational quality of a communication network and refer to the KPIs of the network. In a QoS/QoE mapping process, the QoE value can be calculated from these parameters using appropriate mathematical models [31].

1) MAPPING FUNCTIONS

To transform the measurable QoS parameters into MOS scores, a suitable mapping function is required. Mapping functions can be linear or non-linear. In case where objective and subjective ratings are arbitrarily scaled, any variation corresponds with a matching observed quality variation across the full range, allowing the linear mapping function to be defined as follows [32]:

$$QoE = a + b \cdot QoS. \tag{1}$$

In order to evaluate the objective measurement, the predicted QoE values need to be mathematically related to the observed outcomes. However, because objective quality metrics are rarely uniform, the linear mapping function may underestimate the overall outcome. In most cases, non-linear mapping functions are used to address this issue. Nonlinear mapping functions achieve higher correlations compared to the corresponding linear functions [17]. The most frequently utilized mapping functions are shown in the expressions below and include the logistic (2), cubic (3), exponential (4), logarithmic (5) and power function (6):

$$QoE = a/(1 + e^{-b(QoS-c)}), \tag{2}$$
$$QoE = a + bQoS + cQoS^2 + dQoS^3, \tag{3}$$
$$QoE = ae^{bQoS} + ce^{dQoS}, \tag{4}$$
$$QoE = a - b\log(QoS), \tag{5}$$
$$QoE = aQoS^b + c. \tag{6}$$

All these different types of mapping functions can be used to correlate different QoS metrics to QoE values [32].

2) IQX HYPOTHESIS

The IQX hypothesis is an exponential method that describes QoE as a parameterized negative exponential function of a QoS impairment attribute. QoE can be defined as a function of n influence factors I_j, $1 \leq j \leq n$ [33]:

$$QoE = \Phi(I_1, I_2, \ldots I_n). \tag{7}$$

IQX hypothesis focuses on a single influence component, I = QoS, to obtain the primary correlation QoE = f(QoS) [34]. In general, subjective sensitivity to QoE changes becomes more pronounced when there are high QoE values. This happens because if the QoE is too high, even a minor degradation in quality will be immediately noticeable, whereas if the QoE is already poor, additional degradation is not as noticeable. In this view, it is stated that changes in QoE values are dependent on the current QoE level, while changes in QoS values have the opposite algebraic sign. We fit the differential equation below, considering a linear connection at the QoE level:

$$\frac{\partial QoE}{\partial QoS} =\sim -(QoE - \gamma). \tag{8}$$

This equation's outcome is specified as an exponential function, reflecting the fundamental relationship of the IQX hypothesis:

$$QoE = \alpha e^{-\beta QoS} + \gamma. \tag{9}$$

3) WEBER-FECHNER LAW

The Weber-Fechner Law (WFL) is a logarithmic approximation that heralds the birth of psychophysics as a scientific discipline. WFL relates the perceptual capacities of the human sensory system to the awareness of scarcely perceptible changes in a salient stimulus. As far as the human senses are concerned, a barely perceptible variation turns out to be a constant ratio of the original magnitude of the stimulus. Experiments on the sensation of touch, for instance, have shown that individuals can detect a 3% growth in the weight of an object, independent of its exact magnitude [35]. The following differential equation explains it:

$$\frac{\partial Perception}{\partial Stimulus} \sim -\frac{1}{Stimulus}. \tag{10}$$

Therefore, the ensuing mathematical equation is logarithmic and can be utilized to express the coupling of stimulus and perception as follows:

$$QoE = a\log(QoS) + b. \tag{11}$$

4) STEVEN'S POWER LAW

Stevens' power law (SPL) is a psychophysics law that describes how the intensity of a physical stimulus affects human perception [36]. The following equation can be used to describe SPL:

$$P(S) = KS^b. \tag{12}$$

where P stands for human perception as a product of stimulus strength S, K is a constant that varies with the measurement setting, and the exponent b denotes the kind of stimulus and defines the curvature of the function power. This formula spans nearly every perceptual continuum based on stimulus magnitude over a given range and forms the basis of the mapping function (6).

IV. QOE PREDICTION MODEL

The proposed QoE prediction model is based on encoder-decoder LSTM DNN and applied to a small-scale experimental wireless network built on Open RAN technology. The QoS monitoring system is developed on the user's side using open-source monitoring tools. The captured QoS attributes of bandwidth, latency, and packet loss are fed into the QoS/QoE mapping model, which maps them to MOS values. Finally, the encoder-decoder LSTM prediction model examines the interdependencies between the mapped input values to determine the overall QoE value.

A. OPEN RAN TESTBED

Open RAN technology demonstrates strong potential to transform the mobile RAN domain and is expected to be an important component in the development of next-generation wireless networks. The Open RAN architecture, as suggested by the O-RAN Alliance, is anticipated to be able to provide tailored communication services to meet the demanding and differentiated service requirements of emerging innovative usage scenarios, including wireless VR applications. The key elements of the Open RAN technology are founded on the principles of openness and intelligence and can achieve flexible, programmable and scalable network implementations by making full use of the capabilities of open and standardized interfaces as well as the separation between hardware and software. In this way, RAN is compatible with interoperable application programming interfaces (APIs) and, thus, can be considered an enabling technology for building multi-vendor ecosystems. Open RAN leverages network softwarization technologies, such as software-defined networking (SDN) and network function virtualization (NFV), as well as cloud computing capabilities, to reshape RAN operations and enable the development of novel applications and services over RAN [37].

The Open RAN experimental network is built on a small-scale testbed supporting the 4G architecture as well as the 5G non-standalone (NSA) architecture, which uses a combination of the legacy 4G long-term evolution (LTE) infrastructure with a 5G RAN. It is based on OpenAirInterface (OAI), an open-source software suite that supports 3GPP architectures on multipurpose x86 computer equipment and commercial off-the-shelf (COTS) software-defined radio (SDR) systems such as universal software radio peripherals (USRPs) [38]. The testbed supports three types of evolved node base station (eNB) architectures (monolithic, split option 2 and split option 7.2), one type of next generation node base station (gNB) architecture (monolithic), and evolved packet core (EPC) [39]. EPC is containerized using Docker technology, while RAN is deployed as bare metal. However, all network elements are deployed on the same physical hardware, a personal computer (PC) equipped with 64 GB of RAM and an Intel i9 3.6 GHz processor. For eNB architecture, the transmission is configured in frequency division duplexing (FDD) mode with 20 MHz bandwidth in band 7, while for gNB architecture, the transmission is configured in time division duplexing (TDD) mode with 40 MHz bandwidth in band n78. The major attributes of the testbed are shown in Fig. 1 and described below [40]:

- The remote radio head (RRH) is implemented for all RAN architectures via a USRP B210 in single-input single-output (SISO) mode, connected to the PC with a universal serial bus (USB) 3.0 cable.
- For split option 7.2, the DU implements higher-layer physical layer (high-PHY) functions, including medium access control (MAC) and radio link control (RLC) layer, while the CU deploys radio resource control (RRC), packet protocol data convergence process (PDCP), and service data adaptation process (SDAP). The two units are connected to each other via an F1 interface.
- The virtual baseband unit (vBBU) is present in the cases of monolithic 4G and 5G architectures, performing the operations of the CU and DU units.
- EPC comprised of the 3GPP-based units of home subscriber server (HSS), mobility management entity (MME), service gateway (SGW) and packet gateway (PGW). However, service and packet gateways (SPGWs) have separate user (SPGW-U) and control (SPGW-C) planes.

FIGURE 1. Monolithic gNB (5G NSA) architecture [40].

B. QOS MONITORING AND DATASET GENERATION

The QoS monitoring system to capture the network/transmission parameters of bandwidth, latency, and packet loss is built on the monitoring tools of Prometheus, Telegraf, and Grafana. Prometheus is an open-source cloud-native monitoring system that captures real-time measurements into a time-series database using customizable queries and real-time alerts [41]. Telegraf is an open-source server agent for gathering, processing, aggregating and writing measurements from stacks, sensors and systems [42]. Grafana is an open-source, cross-platform analytics and interactive visualization tool for providing charts, graphs and web alerts [43]. The QoS monitoring system is user-centric, as the network/transmission parameters are collected on the user side using end-user device probes to provide application-level measurements [44].

TABLE 1. Network requirements for different VR phases [45].

	Bandwidth	Round-Trip Time	Packet Loss
Pre-VR	25 Mb/s	40 ms	$1.4 \cdot 10^{-4}$ %
Entry-Level VR	100 Mb/s	30 ms	$1.5 \cdot 10^{-5}$ %
Advanced VR	418 Mb/s	20 ms	$1.9 \cdot 10^{-6}$ %
Ultimate VR	1 Gb/s for smooth play 2.35 Gb/s for instant interactive	10 ms	$5.5 \cdot 10^{-8}$ %

The dataset for training the neural network has been generated by monitoring the QoS parameters of bandwidth, latency, and packet loss during wireless transmission of VR 360-degree YouTube videos. The QoS monitoring system collected data from November 2022 to January 2023 using a variety of video streaming loads. The monitoring period lasted a total of 8 weeks and the values of network bandwidth, delay and packet loss were collected with a 1-minute measurement interval, yielding a total of 80640 data samples for each QoS parameter.

VR applications have been categorized into four phases, each with its own set of network and video requirements, as well as technical equipment specifications, as shown in Table 1: pre-VR, entry-level VR, advanced VR, and ultimate VR [45]. YouTube's VR 360-degree videos, like the majority of today's commercial VR applications, belong to the pre-VR phase and have corresponding network requirements. In order to create a representative dataset, we used a mix of low, medium, and high motion VR 360-degree videos. The technical specifications of these videos are as follows: resolution 3840×2160 pixels (4K UHD), frame rate 30 FPS, video bitrate 45 Mbps, MP4 container and H.264 video encoding.

The collected samples of QoS parameters are depicted in Fig. 2, and the five-number summary of the dataset

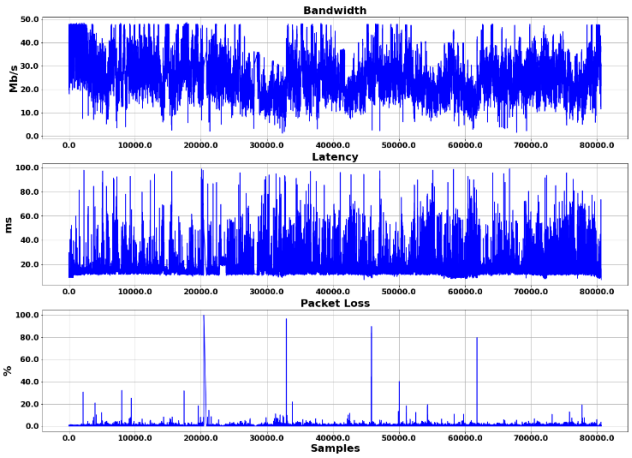

FIGURE 2. Samples of QoS parameters.

(i.e., minimum, first quartile, median, third quartile, and maximum) is depicted in Fig. 3.

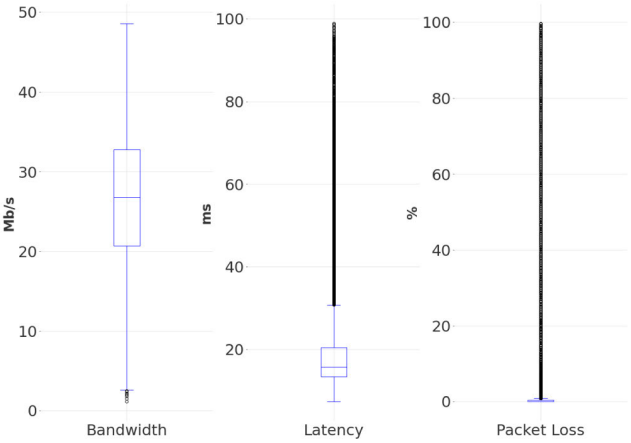

FIGURE 3. Five-number summary of QoS parameters.

C. QOS MONITORING AND DATASET GENERATION

To create the QoS/QoE mapping model, we leveraged the logistic, IQX, WFL, and SPL mapping methods, applying equations (2), (9), (11), and (12) to the measurements of network QoS parameters, respectively. To construct an approximate curve $y = f(x)$ that best fits the discrete set of measurement points (x_i, y_i), where $i = 1, 2, 3, \ldots, n$, we used the curve fitting methodology. Specifically, we used the least squares method, which is one of the most widely used methods for finding the curve of best fit to a given data set [46]. The coefficients a, b, and c of the mapping functions are adjustable parameters in the formula $f(x)$. The goal of the least squares method is to determine these parameters in such a way that the fitting error, i.e., the variance among the data values y_i and the y values $f(x_i)$ on the fitted curve, is minimized. The residuals are defined as the variances among the observed y values and those given by the fitted curve to the x values where the data were originally collected.

Let the data points be $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ where x is the independent variable and y is the dependent variable. The deviation error e_i of the fitted curve $f(x)$ from each data point is determined as follows:

$$\begin{aligned} e_1 &= y_1 - f(x_1), \\ e_2 &= y_2 - f(x_2), \\ &\vdots \\ e_n &= y_n - f(x_n). \end{aligned} \tag{13}$$

According to the principle of the least squares, the best fitting curve has the property that the sum of the squares of errors in formula (14) is minimum, and hence, the calculated value of the parameters a, b, and c minimizes the error e_i.

$$\sum_1^n e_i^2 = \sum_1^n [y_i - f(x_i)]^2. \tag{14}$$

To implement the least squares curve fitting, we used Python's SciPy, NumPy, and Pandas open-source libraries [47]. SciPy stands for scientific Python and is a scientific computation library that contains a collection of mathematical algorithms and functions, NumPy is the main library for scientific computing in Python, and Pandas is a widely used library for data analysis and machine learning applications.

To evaluate the curve fitting accuracy, we used R^2 and mean squared error (MSE) metrics, which are defined as follows:

$$R^2 = 1 - \frac{SSR}{SST}. \tag{15}$$

where SSR denotes the residual sum of squares and SST denotes the total sum of squares from the regression.

$$MSE = \frac{\sum_{i=1}^n (y_i - \lambda(x_i))^2}{n}. \tag{16}$$

where the error prediction for an instance is the variance among the actual and forecasted values.

TABLE 2. Curve fitting accuracy.

	R^2	MSE
IQX Bandwidth	0.99	$7.33 \cdot 10^{-15}$
IQX Latency	0.99	$1.92 \cdot 10^{-16}$
IQX Packet Loss	0.99	$5.46 \cdot 10^{-19}$
WFL Bandwidth	0.99	$1.12 \cdot 10^{-14}$
WFL Latency	1	0
WFL Packet Loss	1	0
SPL Bandwidth	0.99	$3.18 \cdot 10^{-14}$
SPL Latency	1	0
SPL Packet Loss	1	0
Logistic Bandwidth	0.99	$3.43 \cdot 10^{-15}$
Logistic Latency	0.99	$8.51 \cdot 10^{-15}$
Logistic Packet Loss	0.99	$1.06 \cdot 10^{-13}$

The curve fitting accuracy values for IQX, WFL, SPL, and logistic QoS/QoE mapping functions are contained in Table 2, while the corresponding curves are depicted in Fig. 4. As we can observe in Fig. 4, the QoS/QoE mapping process converts the measured values of bandwidth, latency, and packet loss to MOS values. It is worth noting that in the case of latency and packet loss as QoS attributes, their effect on the perceived service quality grows reversely with the impairment, i.e., the higher the value of QoS, the lower the objective quality is. On the contrary, in the case of bandwidth, the higher the value of QoS, the higher the objective quality. Therefore, the algebraic signs in equations (2), (9), (11), and (12) have been adjusted accordingly.

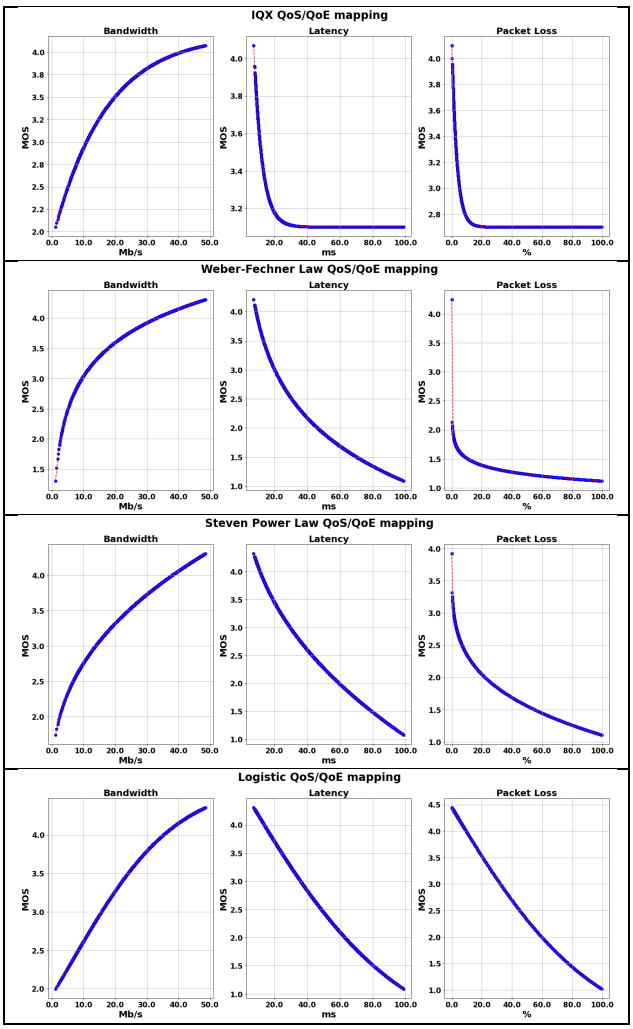

FIGURE 4. Curve fitting for QoS/QoE mapping.

D. DEEP LEARNING PREDICTION MODEL

The development of the QoE prediction model is based on deep learning techniques, specifically the LSTM DNN. LSTM is a subclass of the RNN and is used in many different real-world applications where the utilization of a conventional RNN is inefficient. The architectural design of LSTM is specifically aimed at addressing the primary drawback of RNNs, which is their inability to retain information from the previous state. This fact gives rise to the vanishing gradient problem. The basic design of an LSTM block incorporates the elements of cells as well as the input

and output gates. Combined with the use of constant error carousel (CEC) units, LSTM is capable of handling exploding and vanishing gradient problems [48].

The LSTM design was primarily developed to address the problem of long-term dependencies, as long-term retention of information is its main advantage. RNNs have repeating neural network chain units, consisting of a simple structure such as the *tanH* layer. The LSTM architecture, in contrast, rather than consisting of a single neural network layer, has a chain structure. As shown in Fig. 5, the LSTM includes four gates that constitute single-layer neural networks: 1) the forget gate, with sigmoid activation; 2) the candidate gate, with *tanH* activation; 3) the input gate, with sigmoid activation; and 4) the output gate, with sigmoid activation. The input cell comprises the input vector x_t, the preceding hidden state h_{t-1}, and the preceding memory state C_{t-1}. The output cell comprises the current hidden cell h_t and the current memory cell C_t. W and U are weight vectors.

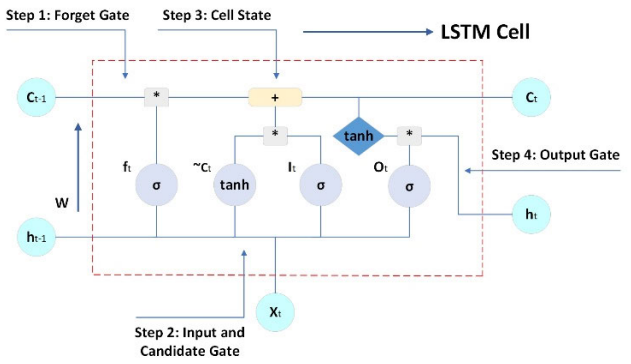

FIGURE 5. LSTM cell architecture [40].

The four steps of LSTM operation, as shown in Fig. 5, are as follows:

1) First, to forecast an impending sequence based on the set of subsequent timestamps, the cell state stores information about the current input. When new input data is received, it is associated with the preceding pattern in the time series sequential data as follows:

$$f_t = \sigma(W_f\,[h_{t-1}, x_t] + b_f). \tag{17}$$

2) The next stage involves determining what data should be saved. This is done as follows: initially, the input gate layer containing the sigmoid layer determines which values to update, and then a *tanH* layer procreates a vector of next potential values called the candidate gate ($\widetilde{C}_t$). These two gates can be correlated together to produce an updated state. The following are the input and candidate gate formulas:

$$i_t = \sigma(W_i\,[h_{t-1}, x_t] + b_i), \tag{18}$$
$$\widetilde{C}_t = \tanh(W_C\,[h_{t-1}, x_t] + b_C). \tag{19}$$

3) Third, the previous cell state C_{t-1} is multiplied by f_t to update it to the new cell state C_t. The input gate i_t is then multiplied by the candidate gate $\widehat{Ct}$, resulting in the new candidate value. The cell state equation is as follows:

$$C_t = f_t C_{t-1} + i_t \widetilde{C}_t. \tag{20}$$

4) Finally, the output state decision is determined using a filtered version of the cell state. The sigmoid layer determines which part of the cell state output to produce, and subsequently, the cell state is set to *tanH* and multiplied by the output of the sigmoid gate. The following are the formulas for the output state and hidden state:

$$O_t = \sigma(W_o\,[h_{t-1}, x_t] + b_o), \tag{21}$$
$$h_t = O_t \tanh(C_t). \tag{22}$$

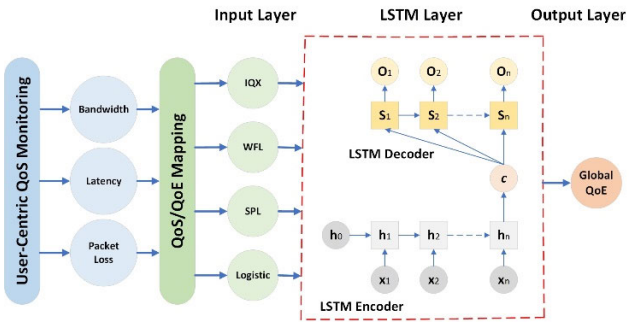

FIGURE 6. Deep learning QoE prediction model architecture.

To develop the deep learning prediction model, we used TensorFlow and Keras. TensorFlow is an open-source end-to-end platform that was initially developed by the Google Brain Team for large numerical computations but proved to be very effective for deep learning as well as conventional machine learning applications [49]. Keras is an open-source high-level library developed by Google for implementing neural networks that runs on top of TensorFlow [50]. The development of the proposed QoE prediction model uses a multistep encoder-decoder LSTM with multivariate input. As shown in Fig. 6, the multivariate input consists of the mapped values of bandwidth, latency, and packet loss. The hidden layer is built on an encoder-decoder LSTM, a subcategory of LSTM designed to solve sequence problems. Sequence prediction problems are demanding since the number of elements in input and output sequences can differ, and prediction often involves the provision of the next value in an actual sequence. This is often framed as a prediction task involving a sequence of one or multiple input timesteps to one output timestep. In our case, we use the QoS measurements of the previous 2 days to forecast the QoE of the next day. The encoder-decoder LSTM network examines the interdependencies between the input values to determine the global QoE value in the output layer. In Fig. 7, the diagram of the complete small-scale testbed architecture including the Open RAN network, the QoS monitoring system and the prediction model, is depicted.

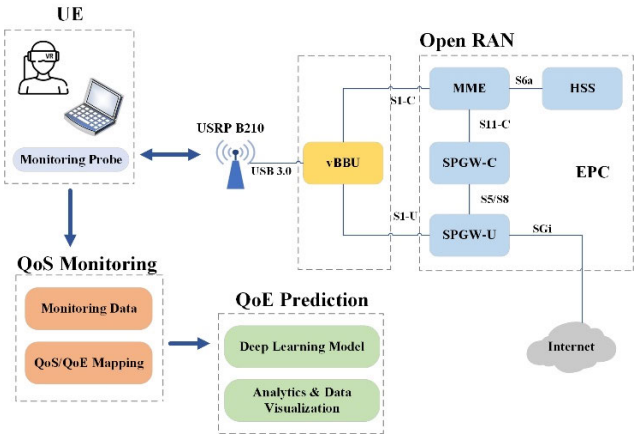

FIGURE 7. Complete testbed architecture.

The integration of the computationally demanding prediction model into a large-scale communication system architecture is possible in the form of an xApp according to the O-RAN specifications. The training of deep learning models in O-RAN systems is performed offline in a separate cloud cluster with the necessary computational resources. Once the model is trained, it is converted into a format compatible with the xApps runtime environment. In practice, integration into the system is done through the two components of the RAN intelligent controller (RIC): the non-real-time RIC (non-RT RIC) and the near-real-time RIC (near-RT RIC). Specifically, the non-RT RIC implements the non-real-time control loop to manage and train the deep learning model and interacts through standardized interfaces with the near-RT RIC, which is the xApps host platform during the inference operation, including model execution and online learning [39].

V. PREDICTION MODEL EVALUATION

The proposed prediction model of the encoder-decoder LSTM is compared to the most significant and widely used neural networks in the RNN category, including SimpleRNN, LSTM, autoencoder LSTM, bidirectional LSTM, and gated recurrent unit (GRU).

In data preparation, we resampled the dataset, as neural networks belonging to the LSTM class are most efficient when handling samples as single sequences of observations that do not exceed 400 timesteps. Therefore, we resampled the minute input data as hourly samples, yielding a total of 1344 samples. Our forecasting model is framed as a multivariate, multistep time series forecasting model that predicts the next day's QoE values based on the observations of the past 2 days. Specifically, we use 48 hourly samples (2days $\times$ 24h) as backward timesteps to predict the QoE of the following 24 timesteps (1day $\times$ 24h). The prediction model has 12 input sequences that correspond to the mapped QoS parameters, as shown in Fig. 4, and 1 output that yields the global QoE value of the wireless network.

A. COMPERATIVE EVALUATION OF RNN PREDICTION MODELS

Initially, we developed a *Naive* prediction model, which provides a basic performance based on which the more sophisticated prediction models of SimpleRNN, LSTM, autoencoder LSTM, bidirectional LSTM, and GRU are evaluated. Naive forecasting is an estimation approach that uses the known values of the previous period as prediction values for the current period without modifying them or determining a causality rule. It is only useful for comparing these predictions with the results of more advanced approaches.

SimpleRNN is a variant of conventional feedforward neural networks (FNN) that is suitable for handling sequential or time series data, as it can be trained to retain past or historical information to predict future values. Conventional FNN is effective for applications involving data values that are unrelated to one another. Nonetheless, if there is data in a sequence where each value is contingent on the preceding one, the neural network must be altered to reflect the dependencies among these values. SimpleRNN may save the information of preceding inputs in order to compute the sequence's subsequent output.

As mentioned in the previous section, *LSTM* constitutes a type of RNN designed to learn long-term dependencies, as it is particularly designed to overcome the exploding and vanishing gradient problems that characterize conventional RNN.

Autoencoder LSTM is a neural network implementation suitable for sequence data based on the LSTM encoder-decoder architecture. An autoencoder is a model designed to learn a compressed version of the input. During training, the encoder learns a subset of attributes from the input data, and the decoder is trained to reconstruct the data based on these features. Models of this kind are referred to as self-supervised and can be classified as unsupervised learning methods. Its most common use is the automatic feature extraction.

Bidirectional LSTM is an enhanced LSTM model capable of increasing efficiency in sequence data applications. To maintain upcoming and prior knowledge, the input fluxes in both directions, thus, during training, two models must be created. The first model learns the input sequence, and the second model learns the inverse of that sequence. These models significantly improve the available quantity of information in the network, consequently enriching the context provided to the algorithm.

The *encoder-decoder LSTM* model is very efficient with sequential data as well as with time series data, a type of sequential data produced by collecting data points in a sequence of time values. In this model architecture, the encoder receives the input sequence data at each time step, learns the information from the input and propagates it for further processing. The feature vector is an internal and intermediate state responsible for maintaining the sequential information of the input. Finally, the decoder decodes the output by converting it back to sequential form.

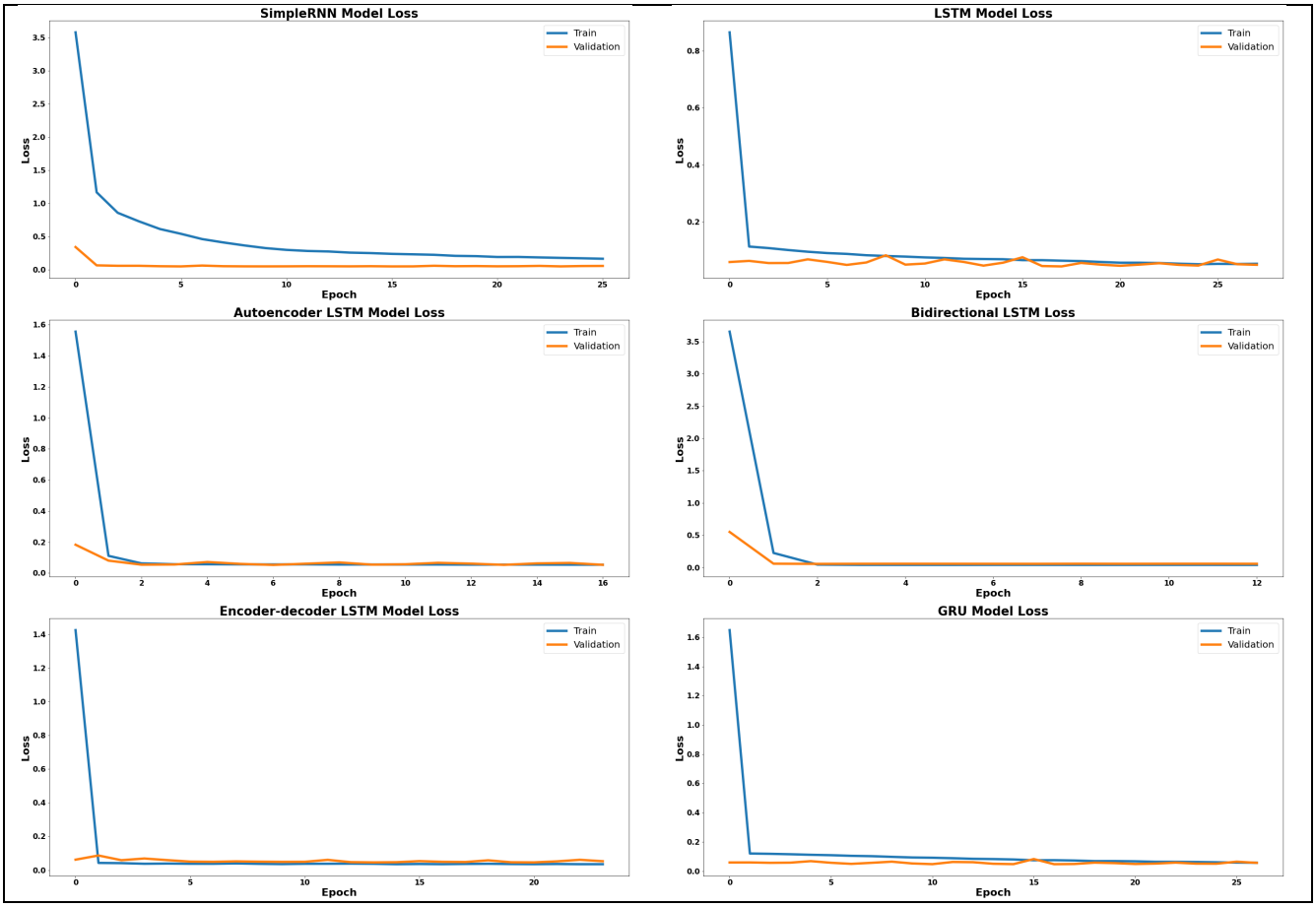

FIGURE 8. Learning curves of QoE prediction models.

GRU is an improved model of the RNN class, suitable for solving the vanishing gradient issue of the conventional RNN. GRU, likewise LSTM, employs two gates to control the flow of information: the update gate and the reset gate. Having a simpler architecture than LSTM, GRU can be considered an improved version since it does not have a separate cell state (C_t), but only a hidden state (H_t). Due to the simpler architecture, the training time in GRU is improved substantially.

The 8 weeks of QoS monitoring yielded a total of 56 days of observations. We split these 56 days as follows: 38 days were used as a training dataset, 9 days as a test dataset, and 9 days were reserved as a validation dataset. During compiling the models, we used Keras callbacks to improve training efficiency. In particular, we used ModelCheckpoint to save the model weights at specific time steps, EarlyStopping to halt the process when the observed evaluation metric stops improving, and ReduceLROnPlateau to minimize the learning rate once the observed metric is no longer improving.

To evaluate the accuracy of the prediction models, in addition to the MSE, we used the root-mean-square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and median absolute error (MedAE) metrics, which are defined as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum (y_i - \hat{y}_i)^2}, \tag{23}$$

which expresses the level of data dispersion.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \left|y_i - \hat{y}_i\right|, \tag{24}$$

which computes the average level of errors in a sequence of forecasts.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{y_i - \hat{y}_i}{y_i}\right|, \tag{25}$$

which defines the accuracy of a forecasting method as the average of the absolute percentage errors in comparison with the actual values.

$$MedAE(y, \hat{y}) = median(\left|y_1 - \hat{y}_1\right|, \ldots, \left|y_n - \hat{y}_n\right|). \tag{26}$$

where the loss is calculated by taking the median of all absolute differences between the actual values and the predictions. In the above expressions y signifies the real values and $\hat{y}$ signifies the forecasted values.

The results of the prediction accuracy metrics are presented in Table 3. We can observe that the proposed encoder-decoder LSTM modem outperforms the rest of the models by

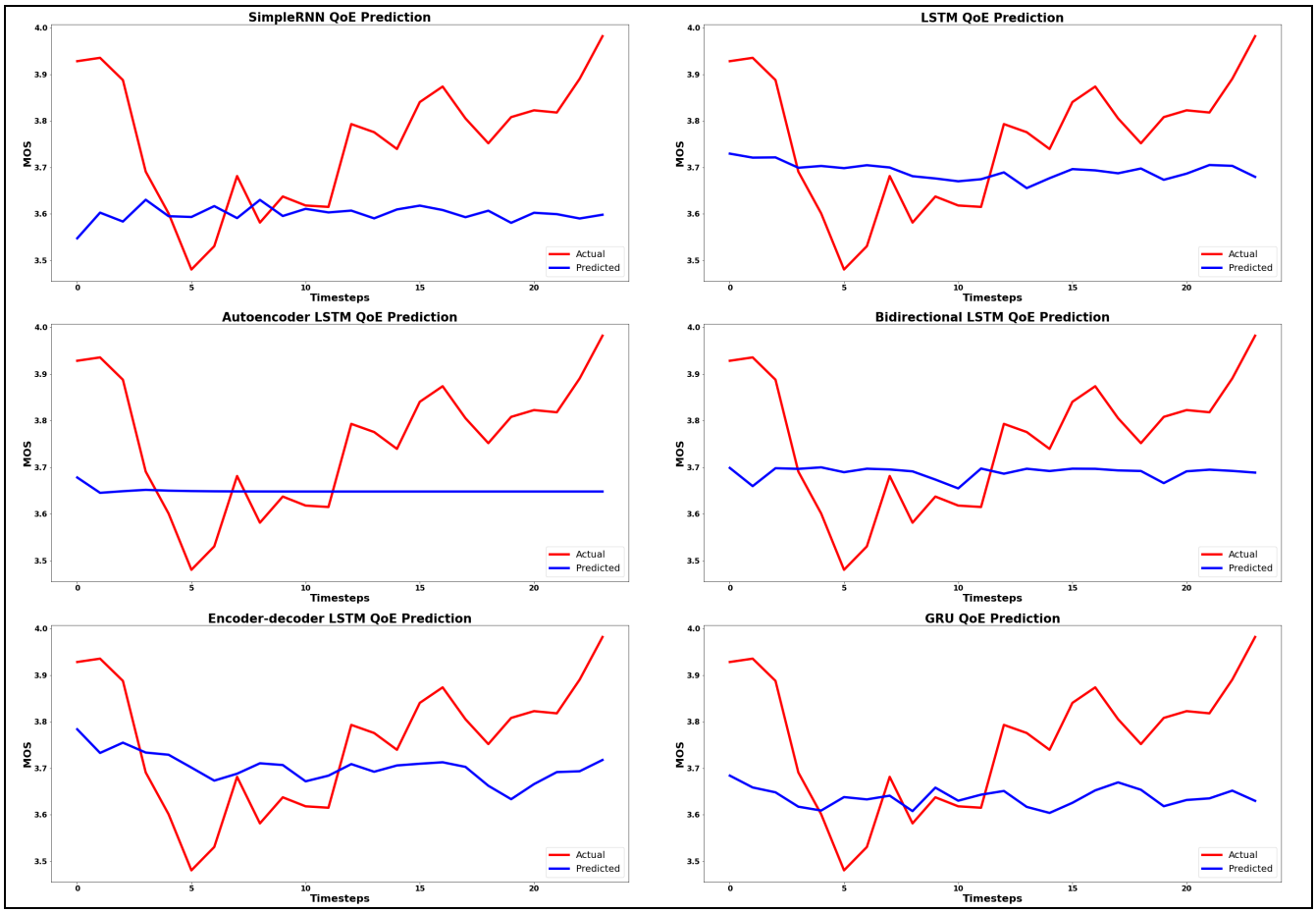

FIGURE 9. Representation of actual versus predicted QoE value.

achieving better scores in all accuracy metrics. It is worth noting that all prediction models perform considerably better than the naive model and that the differences between them are small, thus confirming the effectiveness of RNN class models in handling time series data applications. The effectiveness of the prediction models can also be examined by the learning curve of the models' learning performance, as shown in Fig. 8. The models have been evaluated on the training dataset utilized to fit each model and on the holdout validation dataset, which offers an unbiased evaluation of the model fitting. The learning curves can be used to diagnose problems during training, such as underfitting and overfitting.

In order to minimize overfitting and improve the performance of the models we used the dropout regularization technique. As we can see in Fig. 8, the learning curves show good fit, confirming the good results in the accuracy metrics presented in Table 3. A good fit is also an indication that the predictive models will be able to generalize well to new data. In the event that the learning curves showed overfitting, this would mean that the prediction models have learned the training data set too well, including statistical noise and random fluctuations, resulting in increased generalization error. Accordingly, in the event of underfitting, the prediction models would not be capable of apprehending the correlation among the input and output variables accurately, resulting in an increased error rate in both the training and validation datasets.

TABLE 3. QoE prediction model evaluation.

	MSE	RMSE	MAE	MAPE (%)	MedAE
Naive	0.29636	0.54439	0.49610	15.2277	0.46343
Simple RNN	0.04126	0.20312	0.17529	4.68265	0.17242
LSTM	0.03383	0.18395	0.15186	4.08651	0.13727
Autoencoder LSTM	0.03134	0.17704	0.14987	4.05443	0.14518
Bidirectional LSTM	0.029015	0.17033	0.13958	3.81615	0.12350
Encoder-decoder LSTM	**0.02541**	**0.15943**	**0.12491**	**3.42698**	**0.09981**
GRU	0.02767	0.16361	0.13198	3.63324	0.10320

It is also worth noting that all prediction models show lower validation loss than training loss. This is due to the use of the dropout, which probabilistically precludes inputs and recurrent connections from activation and weight updates during model training, reducing model capacity and resulting

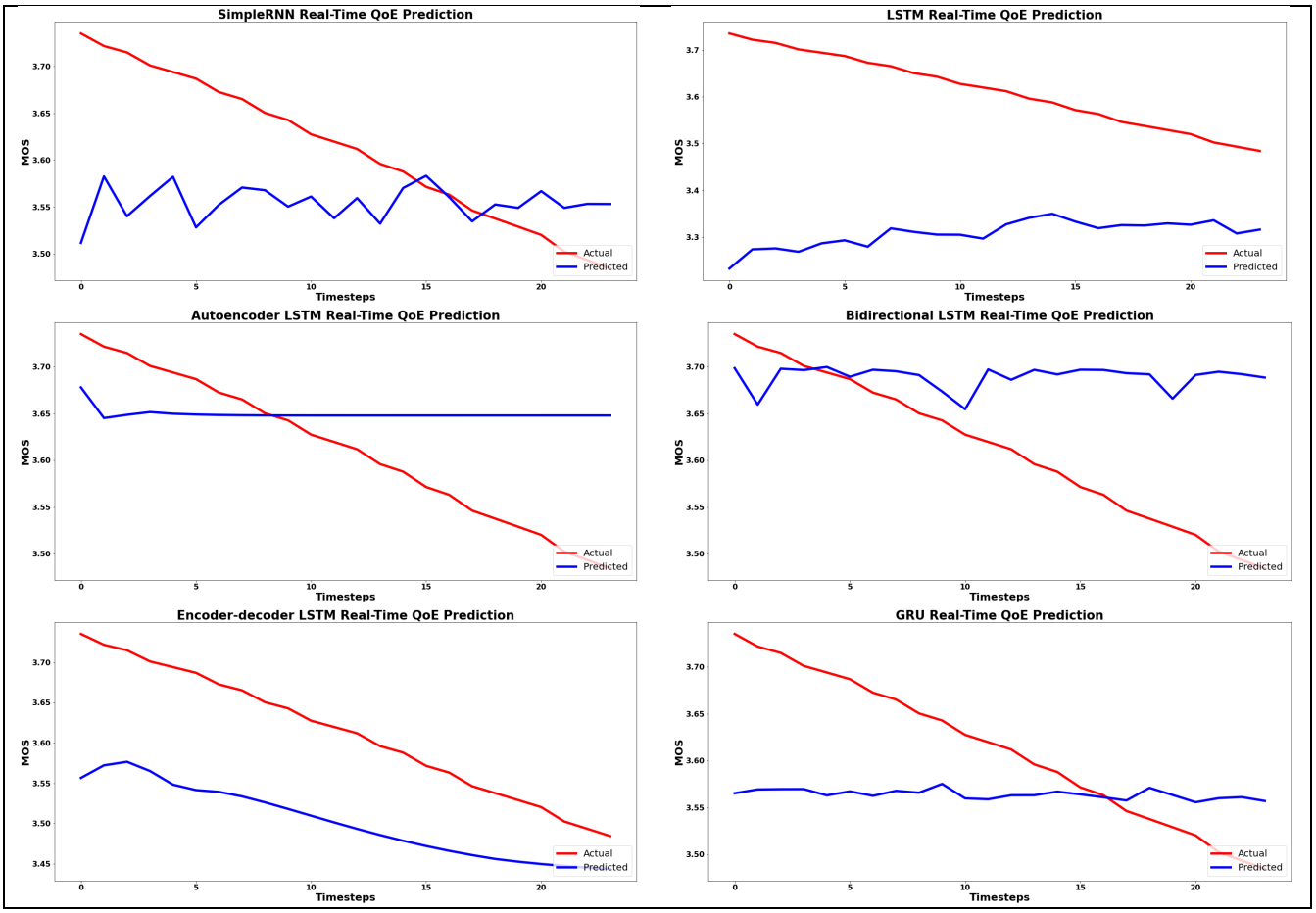

FIGURE 10. Representation of real-time QoE prediction versus actual values.

in lower validation loss. Moreover, in Keras, regularization techniques are disabled during validation, so their effect impacts only the training loss. Additionally, the training loss is computed as the average of the losses for every batch of training data for the current epoch. Since the models vary over time, the loss during the early batches of an epoch is typically greater than the last batches, resulting in a lower average value. Contrariwise, the validation loss for each epoch is calculated at the end of the epoch, leading to a lower value.

The results of the QoE prediction of the next day over the validation dataset values of the previous 2 days are depicted in Fig.9. The next-day prediction corresponds to 24 forward time steps, and the validation basis of the previous 2 days corresponds to 48 backward timesteps. Table 4 includes the mean QoE values as derived from the considered prediction models. We can observe that the proposed model approximates the actual QoE value with greater accuracy compared to the rest of the prediction models. Due to the small batch size that we employed and the stochastic behavior of the models, the same model will learn a subtly varying mapping of inputs to outputs on every training session. Consequently, when a model is assessed, the outcomes may differ. Therefore, we ran the models repetitively and averaged the model performance with regard to the mean QoE value.

TABLE 4. QoE prediction mean value.

	Mean QoE (MOS)
Actual	**3.70014**
Naive	3.60808
SimpleRNN	3.62492
LSTM	3.63622
Autoencoder LSTM	3.64957
Bidirectional LSTM	3.68915
Encoder-decoder LSTM	**3.70277**
GRU	3.69029

B. REAL-TIME QOE PREDICTION

The trained models are used for real-time QoE prediction, as depicted in Fig.10, where we can observe the prediction of QoE values for 24 forward timesteps. The predicted QoE values are depicted versus the actual QoE values, which are based on real-time QoS monitoring and QoS/QoE mapping.

Specifically, the experimental data collection spans a 4-hour time period of monitoring the QoS parameters of a VR 360-degree video stream over the Open RAN testbed, yielding a total of 240-minute samples. These samples are used as backward time steps to evaluate the accuracy of the real-time QoE prediction and are the dataset against which the ability of the model to generalize well to new data is validated.

Each of the 12 time series input variables is used by the prediction models to predict the global QoE value for the following 24 time steps. This is accomplished by feeding each one-dimensional time series into the models as a distinct input sequence. In turn, the prediction models generate an internal representation of each input sequence, which will be interpreted by the output. Incorporating multivariate inputs is necessary for our application, where the output sequence represents a function of earlier time step observations influenced by several independent attributes instead of predicting a single feature.

TABLE 5. Real-time QoE prediction model evaluation.

	MSE	RMSE	MAE	MAPE (%)	MedAE
Simple RNN	0.54320	0.73702	0.50636	21.9287	0.32311
LSTM	0.47626	0.69011	0.46084	20.1354	0.26658
Autoencoder LSTM	0.52514	0.72466	0.53395	22.3827	0.40065
Bidirectional LSTM	0.56478	0.75152	0.55616	23.2374	0.43710
Encoder-decoder LSTM	**0.43370**	**0.65856**	**0.42022**	**18.7329**	**0.24058**
GRU	0.51687	0.71894	0.48108	20.9888	0.26736

As shown in Table 5, the proposed encoder-decoder LSTM model outperforms the prediction models of SimpleRNN, LSTM, autoencoder LSTM, bidirectional LSTM, and GRU by achieving the best score in all accuracy metrics. We can observe that the real-time QoE prediction accuracy scores for all models are reduced compared to the scores during model evaluation. This is due to the small size of the experimental real-time dataset, which affects the intrinsic variance of the input data, resulting in a larger internal variance of the training dataset compared to the validation dataset. The LSTM encoder- decoder model outperforms the rest of the prediction models as it exhibits greater ability to handle small datasets. The prediction accuracy of the models improves as the size of the dataset increases, but we wanted to experimentally test the effectiveness of the models under the conditions of a real-world application.

In Table 6, we can see the real-time mean QoE values as derived from the prediction models under study. The real-time mean QoE value represents the overall QoE value of the Open RAN testbed, as derived from the synthesis of the interdependencies between the values of the 12 inputs. Similar to previous observations, we can observe that the encoder-decoder LSTM model approximates the actual QoE value more accurately compared to the state-of-the-art RNN class prediction models. This is due to the fact that the encoder-decoder LSTM model has proven to be particularly efficient for small sequence problems, making it suitable for real-time applications where the performance evaluation of a wireless network relies on feeding a limited amount of new data to the QoE prediction model due to time constraints.

TABLE 6. Real-time QoE predicted mean value.

	Mean QoE (MOS)
Actual	**3.17505**
SimpleRNN	3.63920
LSTM	3.50846
Autoencoder LSTM	3.64948
Bidirectional LSTM	3.68913
Encoder-decoder LSTM	**3.27873**
GRU	3.60822

VI. CONCLUSION

Real-time QoE prediction should be considered as critical parameter in the design of next generation wireless communication networks, specifically for the transmission of the highly demanding VR applications, which are characterized by intrinsic interactivity. In this paper, we develop a QoE prediction model based on deep learning techniques. In particular, we propose an objective QoE prediction model for VR 360-degree video streaming based on the encoder-decoder LSTM DNN. The prediction model is tested on an Open RAN testbed and is able to quantify the influence of the wireless network operation on VR video quality by predicting the QoE value with the use of the QoS features of bandwidth, delay, and packet loss.

The proposed prediction model outperforms state-of-the-art DNN models belonging to the RNN class, including SimpleRNN, LSTM, autoencoder LSTM, bidirectional LSTM, and GRU. The LSTM encoder-decoder model exhibits higher prediction accuracy and is able to generalize more effectively to new data. Furthermore, since the inputs and outputs of this model are uncoupled and their lengths can vary, this model is capable of mapping sequences of different lengths to each other and is hence appropriate for multistep time series forecasting problems with multivariate inputs. Also, the encoder-decoder LSTM model has been proven by our experiments to be suitable for real-time forecasting applications, as it performs efficiently with small sequences.

This paper is the first to offer a comparative study on DNN techniques belonging to the RNN class for QoE prediction of wireless VR applications. Moreover, this is the first time that a QoE prediction model for VR applications is applied and tested on a testbed based on the Open RAN technology.

REFERENCES

[1] *O-RAN: Towards an Open and Smart RAN*, O-RAN Alliance, Alfter, Germany, Oct. 2018.

[2] M. Taha, A. Canovas, J. Lloret, and A. Ali, "A QoE adaptive management system for high definition video streaming over wireless networks," *Telecommun. Syst.*, vol. 77, no. 1, pp. 63–81, May 2021.

[3] J. Li, R. Feng, Z. Liu, W. Sun, and Q. Li, "Modeling QoE of virtual reality video transmission over wireless networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 1–7.

[4] R. Bhattacharyya, A. Bura, D. Rengarajan, M. Rumuly, S. Shakkottai, D. Kalathil, R. K. P. Mok, and A. Dhamdhere, "QFlow: A reinforcement learning approach to high QoE video streaming over wireless networks," in *Proc. 20th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, Jul. 2019, pp. 251–260.

[5] L. Teng, G. Zhai, Y. Wu, X. Min, W. Zhang, Z. Ding, and C. Xiao, "QoE driven VR 360° video massive MIMO transmission," *IEEE Trans. Wireless Commun.*, vol. 21, no. 1, pp. 18–33, Jan. 2022, doi: 10.1109/TWC.2021.3093305.

[6] T. M. C. Chu and H.-J. Zepernick, "Performance analysis of an adaptive rate scheme for QoE-assured mobile VR video streaming," *Computers*, vol. 11, no. 5, p. 69, Apr. 2022, doi: 10.3390/computers11050069.

[7] R. I. T. da Costa Filho, M. C. Luizelli, M. T. Vega, J. van der Hooft, S. Petrangeli, T. Wauters, F. De Turck, and L. P. Gaspary, "Predicting the performance of virtual reality video streaming in mobile networks," in *Proc. 9th ACM Multimedia Syst. Conf.* New York, NY, USA: Association for Computing Machinery, Jun. 2018, pp. 270–283.

[8] S. Park, A. Bhattacharya, Z. Yang, M. Dasari, S. R. Das, and D. Samaras, "Advancing user quality of experience in 360-degree video streaming," in *Proc. IFIP Netw. Conf. (IFIP Netw.)*, May 2019, pp. 1–9.

[9] M. S. Anwar, J. Wang, W. Khan, A. Ullah, S. Ahmad, and Z. Fei, "Subjective QoE of 360-degree virtual reality videos and machine learning predictions," *IEEE Access*, vol. 8, pp. 148084–148099, 2020.

[10] M. S. Anwar, J. Wang, S. Ahmad, A. Ullah, W. Khan, and Z. Fei, "Evaluating the factors affecting QoE of 360-degree videos and cybersickness levels predictions in virtual reality," *Electronics*, vol. 9, no. 9, p. 1530, Sep. 2020.

[11] M. S. Anwar, J. Wang, S. Ahmad, W. Khan, A. Ullah, M. Shah, and Z. Fei, "Impact of the impairment in 360-degree videos on users VR involvement and machine learning-based QoE predictions," *IEEE Access*, vol. 8, pp. 204585–204596, 2020.

[12] K. Brunnström, *Qualinet White Paper on Definitions of Quality of Experience*. Novi Sad, Serbia: Output of the Fifth Qualinet Meeting, Mar. 2013.

[13] A. Perkis, "QUALINET white paper on definitions of immersive media experience (IMEx)," 2020, *arXiv:2007.07032*.

[14] G. Kougioumtzidis, V. Poulkov, Z. Zaharis, and P. Lazaridis, "QoE assessment aspects for virtual reality and holographic telepresence applications," in *Proc. 6th EAI Int. Conf. Future Access Enablers Ubiquitous Intell. Infrastructures*. Cham, Switzerland: Springer, 2022, pp. 171–180.

[15] R. Schatz, T. Hoßfeld, L. Janowski, and S. Egger, "Quality of experience as a new measurement challenge," in *Data Traffic Monitoring and Analysis*. Berlin, Germany: Springer, 2013, pp. 219–263.

[16] *Mean Opinion Score (MOS) Terminology*, document Rec. ITU-T P.800.1, ITU-T, 2006.

[17] M. A. Alreshoodi and J. C. Woods, "Survey on QoE/QoS correlation models for multimedia," *Int. J. Distrib. Parallel Syst.*, vol. 4, no. 3, pp. 53–72, 2013.

[18] Y. Zhang, Y. Wang, F. Liu, Z. Liu, Y. Li, D. Yang, and Z. Chen, "Subjective panoramic video quality assessment database for coding applications," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 461–473, Jun. 2018.

[19] *Subjective Video Quality Assessment Methods for Multimedia Applications*, document Rec. ITU-T P.910, ITU-T, 2008.

[20] *Subjective Test Methodologies for 360° Video on Head-Mounted Displays*, document ITU-T P.919, ITU-T, 2020.

[21] Y. Wang and P. Zhang, *QoE Management in Wireless Networks*. New York, NY, USA: Springer, 2016.

[22] *Vocabulary for Performance, Quality of Service and Quality of Experience*, document ITU-T, Rec. ITU-T G.1035, 2017.

[23] *Influencing Factors on Quality of Experience for Virtual Reality*, document ITU-T G.1035, ITU-T, May 2020.

[24] U. Reiter, "Factors influencing quality of experience," in *T-Labs Series in Telecommunication*. Cham, Switzerland: Springer, 2014, pp. 55–72.

[25] C. A. Burbeck and D. H. Kelly, "Spatiotemporal characteristics of visual mechanisms: Excitatory-inhibitory model," *J. Opt. Soc. Amer.*, vol. 70, no. 9, p. 1121, 1980.

[26] S. Greenberg and W. A. Ainsworth, "Speech processing in the auditory system: An overview. Speech processing in the auditory system," in *Springer Handbook of Auditory Research*, vol. 18. New York, NY, USA: Springer, 2004.

[27] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal, "Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness," *Int. J. Aviation Psychol.*, vol. 3, no. 3, pp. 203–220, Jul. 1993.

[28] G. Kougioumzidis, A. Vlahov, V. Poulkov, Z. Zaharis, and P. Lazaridis, "QoE-oriented open radio access networks for virtual reality applications," in *Proc. 25th Int. Symp. Wireless Pers. Multimedia Commun. (WPMC)*, Herning, Denmark, Oct. 2022, pp. 491–496.

[29] A. Singla, S. Fremerey, W. Robitza, and A. Raake, "Measuring and comparing QoE and simulator sickness of omnidirectional videos in different head mounted displays," in *Proc. 9th Int. Conf. Quality Multimedia Exper. (QoMEX)*, Erfurt, Germany, May 2017, pp. 1–6.

[30] S. Mangiante, G. Klas, A. Navon, Z. GuanHua, J. Ran, and M. D. Silva, "VR is on the edge: How to deliver 360° videos in mobile networks," in *Proc. Workshop Virtual Reality Augmented Reality Netw.*, Aug. 2017, pp. 30–35.

[31] G. Kougioumtzidis, V. Poulkov, Z. D. Zaharis, and P. I. Lazaridis, "A survey on multimedia services QoE assessment and machine learning-based prediction," *IEEE Access*, vol. 10, pp. 19507–19538, 2022.

[32] J. Korhonen, N. Burini, J. You, and E. Nadernejad, "How to evaluate objective video quality metrics reliably," in *Proc. 4th Int. Workshop Quality Multimedia Exper.*, Melbourne, VIC, Australia, Jul. 2012, pp. 57–62.

[33] R. Schatz, T. Hoßfeld, L. Janowski, and S. Egger, "From packets to people: Quality of experience as a new measurement challenge," in *Data Traffic Monitoring and Analysis*. Berlin, Germany: Springer, 2013, pp. 219–263.

[34] T. Hoßfeld, P. Tran-Gia, and M. Fiedler, "Quantification of quality of experience for edge-based applications," in *Proc. Int. Teletraffic Congr. (ITC)*, Ottawa, ON, Canada, Jan. 2007, pp. 361–373.

[35] P. Reichl, S. Egger, R. Schatz, and A. D'Alconzo, "The logarithmic nature of QoE and the role of the weber-fechner law in QoE assessment," in *Proc. IEEE Int. Conf. Commun.*, Cape Town, South Africa, May 2010, pp. 1–5.

[36] S. Khorsandroo, R. M. Noor, and S. Khorsandroo, "The role of psychophysics laws in quality of experience assessment: A video streaming case study," in *Proc. Int. Conf. Adv. Comput., Commun. Informat.*, Chennai, India, Aug. 2012, pp. 446–452.

[37] G. Kougioumtzidis, V. Poulkov, Z. D. Zaharis, and P. I. Lazaridis, "Intelligent and QoE-aware open radio access networks," in *Proc. 3rd URSI Atlantic Asia–Pacific Radio Sci. Meeting (AT-AP-RASC)*, Gran Canaria, Spain, May 2022, pp. 1–4.

[38] F. Kaltenberger, A. P. Silva, A. Gosain, L. Wang and T.-T. Nguyen, "OpenAirInterface: Democratizing innovation in the 5G Era," *Comput. Netw.*, vol. 176, Jul. 2020, Art. no. 107284.

[39] *O-RAN Use Cases and Deployment Scenarios*, O-RAN Alliance, Alfter, Germany, Feb. 2020.

[40] A. Vlahov, G. Kougioumtzidis, A. Mihovska, and V. Poulkov, "Performance analysis of evolved RAN architectures with open interfaces," *J. Mobile Multimedia*, vol. 19, no. 1, pp. 239–262, Sep. 2022.

[41] *Prometheus*. Accessed: Dec. 2022. [Online]. Available: https://prometheus.io/

[42] *InfluxData*. Accessed: Dec. 2022. [Online]. Available: https://www.influxdata.com/time-series-platform/telegraf/

[43] *Grafana*. Accessed: Dec. 2022. [Online]. Available: https://grafana.com/

[44] G. Kougioumtzidis, V. Poulkov, Z. Zaharis, and P. Lazaridis, "Machine learning for QoE management in future wireless networks," in *Proc. 34th Gen. Assem. Scientific Symp. Int. Union Radio Sci. (URSI GASS)*, Rome, Italy, Aug. 2021, pp. 1–4.

[45] *Whitepaper on the VR-Oriented Bearer Network Requirement*, Huawai, Shenzhen, China, Sep. 2016.

[46] R. D. Levie, "Curve fitting with least squares," *Crit. Rev. Anal. Chem.*, vol. 30, no. 1, pp. 59–74, Jun. 2010.

[47] *Python*. Accessed: Dec. 2022. [Online]. Available: https://www.python.org/
[48] B. V. Vishwas and A. Patel, *Hands-on Time Series Analysis With Python*. New York, NY, USA: Apress, 2020.
[49] *Tensorflow*. Accessed: Dec. 2022. [Online]. Available: https://www.tensorflow.org/learn
[50] *Keras*. Accessed: Dec. 2022. [Online]. Available: https://keras.io/

GEORGIOS KOUGIOUMTZIDIS received the B.Sc. degree in electronics engineering from the Alexander Technological Educational Institute of Thessaloniki, Thessaloniki, Greece, in 2007, the M.Sc. degree in wireless communication systems from the Open University of Cyprus, Nicosia, Cyprus, in 2017, and the M.A. degree in acoustic design and multimedia from Hellenic Open University, Patras, Greece, in 2018. He is currently pursuing the Ph.D. degree with the Faculty of Telecommunications, Technical University of Sofia, Sofia, Bulgaria. From 2013 to 2020, he was with the Hellenic Telecommunications Organization S.A (OTE Group). He also holds an Early Stage Researcher (ESR) position in European Union's Horizon 2020 MOTOR5G project. His research interests include the quality of experience (QoE) enhancement in mobile and wireless networks, open radio access networks, machine learning and artificial neural networks, and extended reality and holographic telepresence communications.

ATANAS VLAHOV received the B.Sc. and M.Sc. degrees in telecommunications from the Technical University of Sofia (TUS), Bulgaria, in 2019 and 2021, respectively, where he is currently pursuing the Ph.D. degree with the Faculty of Telecommunication. His current research interests include the field of telecommunications are related to open radio access networks and software defined networks.

VLADIMIR K. POULKOV (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees from the Technical University of Sofia (TUS), Sofia, Bulgaria. He has more than 30 years of teaching, research, and industrial experience, managing numerous educational, research and development and engineering projects in the field of telecommunications. He has been the Dean of the Faculty of Telecommunications, TUS, and the Vice Chairman of the General Assembly of the European Telecommunications Standards Institute (ETSI). Currently, he is the Head of the "Teleinfrastructure" Research and Development Laboratory, TUS; the Head of the Intelligent Communication Infrastructures Laboratory, Research and Development and Innovation Consortium, Sofia Tech Park; the Chairman of the "Cluster for Digital Transformation and Innovation," Bulgaria. He is the author of many scientific publications, tutoring the B.Sc., M.Sc., and Ph.D. courses in the field of information transmission theory and wireless access networks.

PAVLOS I. LAZARIDIS (Senior Member, IEEE) received the M.Eng. degree in electrical engineering from the Aristotle University of Thessaloniki, Greece, in 1990, the M.Sc. degree in electronics from Université Pierre and Marie Curie (Paris 6), Paris, France, in 1992, and the Ph.D. degree from École Nationale Supérieure des Télécommunications (ENST) Paris and Université Paris 6, in 1996. From 1991 to 1996, he was involved in research with France Télécom, and teaching with ENST Paris. In 1997, he became the Head of the Antennas and Propagation Laboratory, Télédiffusion de France/the France Télécom Research Center (TDF—C2R Metz). From 1998 to 2002, he was a Senior Examiner with the European Patent Office (EPO), The Hague, The Netherlands. From 2002 to 2014, he was involved in teaching and research with the Alexander Technological Educational Institute of Thessaloniki, Greece, and Brunel University, London, U.K. He is currently a Professor of electronics and telecommunications with the University of Huddersfield, U.K. He has been involved in several international research projects, such as EU Horizon 2020 MOTOR5G and RECOMBINE, NATO-SfP ORCA, and he has published over 200 research articles and several national and European patents. He is a member of the IET (MIET), a Senior Member of URSI, and a fellow of the Higher Education Academy (FHEA). He is serving as an Associate Editor for IEEE Access.

ZAHARIAS D. ZAHARIS (Senior Member, IEEE) received the B.Sc. degree in physics, the M.Sc. degree in electronics, the Ph.D. degree in antennas and propagation modeling for mobile communications, and the Diploma degree in electrical and computer engineering from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 1987, 1994, 2000, and 2011, respectively. From 2002 to 2013, he was with the Administration of the Telecommunications Network, Aristotle University of Thessaloniki. Since 2013, he has been with the School of Electrical and Computer Engineering, Aristotle University of Thessaloniki. He has been involved in several European research projects, such as Horizon 2020 MOTOR5G, Horizon 2020 RECOMBINE, and Horizon Europe 6G-ICARUS. He is the author of 82 scientific journal articles, 109 international conference papers, one national patent, five book chapters, and one textbook. His current research interests include the design and optimization of antennas and microwave circuits, signal processing on smart antennas, the development of evolutionary optimization algorithms, and neural networks. He is a member of the Technical Chamber of Greece. Recently, he was elected the Chair of the Electron Devices/Microwave Theory and Techniques/Antennas and Propagation Joint Chapter of the IEEE Greece Section. He is currently serving as an Associate Editor for IEEE Access.