IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 11, 2022, accepted January 30, 2022, date of publication February 7, 2022, date of current version February 24, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3149592

A Survey on Multimedia Services QoE Assessment and Machine Learning-Based Prediction

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

[1]Faculty of Telecommunications, Technical University of Sofia, 1000 Sofia, Bulgaria
[2]School of Electrical and Computer Engineering, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece
[3]School of Computing and Engineering, University of Huddersfield, Huddersfield HD1 3DH, U.K.

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

This research was supported by the European Union, partially 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 partially through the Horizon 2020 Marie Skłodowska-Curie Research and Innovation Staff Exchange Programme "Research Collaboration and Mobility for Beyond 5G Future Wireless Networks (RECOMBINE)" under Grant Agreement no. 872857.

ABSTRACT The groundbreaking evolution in mobile and wireless communication networks design in recent years, in combination with the advancement of mobile terminal equipment capabilities, has led in an exponential growth of mobile internet technologies, and arose an ever-growing demand for innovative multimedia services. The highly demanding in terms of network resources over-the-top media services, as well as the emergence of new and complex mobile multimedia services such as video gaming, ultra-high-definition video, and extended reality, requires the enhancement of end-users' perceived quality of experience (QoE). QoE has garnered much research interest in recent years, and has emerged as a key component in the evaluation of network services and operations. As a result, a QoE-aware network planning approach is getting increasingly favored, and novel design challenges, such as how to quantify and measure QoE, have arisen. In this regard, a paradigm shift in network implementations is being envisioned, in which the focus will be on machine learning (ML) methodologies for developing QoE prediction models, directly related to end-user's personalized experience. In this survey, an analysis on application-oriented, ML-based QoE prediction models for the goal of QoE management for multimedia services is presented. In addition, an examination of the state-of-the-art ML-based QoE predictive models and some of the innovative techniques and challenges related to multimedia services quality assessment with focus on extended reality and video gaming applications are outlined.

INDEX TERMS Extended reality, machine learning, quality of experience (QoE), QoE prediction models, multimedia services, video streaming, video gaming, virtual reality.

I. INTRODUCTION

Mobile and wireless communication networks have grown to be among the most noteworthy modern achievements, revolutionizing the way people communicate and share information, and facilitating improvements in every aspect of everyday life, including education, media and entertainment, entrepreneurship, transportation, healthcare, security and emergency services, while also positively contributing to the economic and social growth both for the developed and developing countries. The number of mobile terminal devices has seen exponential increase in recent years, considerably escalating mobile data traffic. In 2015 there were about 7.5 billion mobile subscriptions, overcoming for the first time the then world's population of 7.3 billion people [1], and this number is expected to rise to 8.8 billion subscriptions by the end of 2026 [2]. Mobile video transmission undisputedly is one of the critical services of the fifth generation (5G) of mobile networks, and will be as well for the forthcoming beyond 5G (B5G) era, accounting for the majority of mobile data traffic. There will be a considerable rise in mobile internet traffic, with video traffic accounting for 82 percent

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

of net consumer traffic by 2022, based on the Cisco Visual Networking Index projection [3]. As a result, mobile video transmission constitutes a substantial research field in the design of wireless communication systems.

The observed growth of multimedia services, owing mainly to the rising acceptance and practice of video streaming services like Netflix, live TV streaming, or Youtube, has sparked new revenue opportunities for communication service providers (CSPs), mobile network operators (MNOs), and over-the-top (OTT) providers. It has also highlighted new challenges in terms of operational efficacy, as the provision of high-quality video to end-users is critical to the long-term viability of these services [4]. Therefore, the notion of quality of experience (QoE) has gained prominence, and enormous exertions from study groups in industry as well in academia have been invested on offering reliable services, with enhanced personalized user experience [5]. Assessing and predicting an end-user's QoE of a multimedia stream, is the first step towards optimizing mobile streaming service delivery and the implementation of efficient QoE management. This also allows gaining a deeper understanding of how network's technical characteristics that make up quality of service (QoS), affect service quality as perceived by the end-users [6]. Nevertheless, ensuring high levels of QoE is a difficult task due to a variety of factors, such as different types of terminal devices, varied service demand motifs, altering media contents, fluctuating broadcast and network states, and considerable spatial and temporal variance in the efficiency of the content distribution networks (CDNs) [7]. Extensive research has been conducted in order to improve the provision of multimedia services and heighten end-user QoE, with the largely utilized strategies relying on either quality-based network resource allocation and quality-based routing, which can be regarded as network optimization procedures, or client-based adaptive video streaming [8].

The primary target of QoE management is linked with the optimization of end-user's QoE, while utilizing network resources efficiently, upholding at the same time a satisfied customer base, and averting customer churn. The proper QoE management for a particular application, requires to understand and identify a set of influencing factors, both subjective and objective, under the standpoint of numerous components in the service supply sequence. The derived QoE models define the features to be examined and assessed, with the primary aim to be the development of efficient QoE optimization techniques that can effectively tackle the challenges of QoE management [9]. Aside from the challenges posed on QoE by the mobility and the necessity of attaining seamless session continuity and seamless horizontal and vertical handover [10], the unremitting emergence of new and complex mobile multimedia services, including 3D video streaming [11], [12], video gaming, ultra-high definition (UHD) video, augmented reality (AR), virtual reality (VR) and mixed reality (MR), introduces additional complexity to the QoE provisioning procedure [13]. Limitations deriving from both terminal equipment capabilities and transmission channel characteristics, have as well a clear influence on the QoE perception of the end-user within the context of wireless communication systems [14]. Therefore, there is a need for developing reliable and accurate QoE models in order to appropriately estimate end-user's QoE, and implement QoE-aware network control and management. Models with that competence typically factor into the equation numerous network and application level QoS parameters, with the target of predicting end-user's QoE by associating these parameters with QoE influencing factors [15].

The procedure of QoE management can be divided within three main stages as follows: understanding and modeling QoE, monitoring and estimating QoE, and adapting and controlling QoE [16]. Typically, QoE management relies on subjective evaluation and deterministic adaptation. In subjective methods, a preset group of end-users score the incoming multimedia streaming using the mean opinion score (MOS) scale [17]. The MNO would monitor user feedback and progressively adjust the service to the needs of the users. The drawback of subjective methods is that they are costly, time consuming and laborious, and can only be conducted offline, because of the time necessary for the subjective evaluations and network adjustments. Objective methods on the other hand, utilize QoE models with a large number of characteristics to predict users' QoE, with the majority of extant assessments relying on three major classification methods: the psychophysical approach, the reference-based algorithm, and the input data-based algorithm [18], [19]. Nevertheless, as the volume and diversity of multimedia streaming services grows exponentially, terminal equipment and network states necessitate real-time, precise, and adaptive QoE management. As a result, the typical techniques of QoE management become unattainable. To address this, significant research activity in recent years have adopted the methods of artificial intelligence (AI) and machine learning (ML) in QoE management [20]. ML improves the accuracy of QoE models, aids in QoE monitoring, and provides a fast optimization feedback loop for adaptive streaming applications [21]. Moreover, ML offers a theoretical and methodological framework for quantifying the correlation between QoE and QoS [22]. Yet, selecting the optimal ML model for a particular type of application is an open research issue in and of itself.

This survey presents an analysis of the most prominent current and evolving approaches regarding multimedia services QoE assessment, and a comprehensive examination of the state-of-the-art ML-based QoE prediction models. The structure of the survey is as follows: for the shake of completeness, the QoE definition within the context of multimedia services and a spherical examination of the QoE influencing factors are given in *Section II*. Subsequently, a complete QoE assessment methodology is presented and includes the following: 1) gathering, classification and analysis of all the significant quality metrics with regard to subjective and objective QoE assessment; 2) examination of the methodologies for evaluating the QoE metrics' operation; and 3) analysis of the mathematical models for QoE/QoS correlation.

In Section III and IV, we define the specific QoE assessment aspects and offer a thorough analysis of the QoE influencing factors in extended reality and video gaming applications respectively, underlining the discrimination among such evolving technologies and conventional video streaming applications. Moreover, in Section V, we stress the significance of ML in implementing effective QoE prediction models, describe the ML methodologies and assay its main algorithms. Furthermore, in Section VI, we review state-of-the-art ML-based QoE predictive models and provide a comparative analysis centering on video streaming, extended reality and video gaming applications. Finally, Section VII includes final remarks and conclusions. The main contributions of this survey can be summarized in the following: 1) according to the best of the authors' knowledge, this is the first endeavor to present a complete hands-on guide on multimedia services QoE assessment that unlike existing surveys, includes besides conventional video streaming, extended reality and video gaming applications; and 2) up to this date, this is the first survey to provide a comparative examination of ML-based QoE prediction models that focus in particular on extended reality and video gaming applications.

II. MULTIMEDIA SERVICES QoE ASSESSMENT METHODOLOGY

In recent years, QoE has drawn a lot of attention, and has been recognized as an essential element in evaluating network operational efficiency. A substantial amount of research effort has been gone into understanding, measuring, and modeling QoE for a range of multimedia services. The QoE approach aims to maximize the perceived user experience while reducing the impact on network resources, as well as to improve the level of quality in multimedia services, whilst maintaining efficient and cost-effective network operations [23].

A. QoE DEFINITION

Quality evaluation has gotten increasingly complicated as the operational sophistication of services and systems has increased, owing to the exponentially growth number of factors involved. Formerly, QoS-centric metrics that consist the network's key performance indicators (KPIs), and include parameters such as throughput, packet loss, latency and jitter, were widely employed to quantify the degree of satisfaction from a communication service. Since QoS metrics are not directly and explicitly related to an end-user's perceived gratification and overall experience with a service, user-centric metrics called key quality indicators (KQIs) have been deployed for the quality evaluation. Thus, QoE is a subjective indicator that incorporates human parameters, as it connects customer perception, expectations, and experience, with application and network efficiency, allowing for a more holistic understanding of quality as experienced by the end-users [24]. According to ITU-T [25] and the Qualinet white paper [17], QoE considers the user's subjective perception and expectations toward a given service and may be defined as "the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and/or enjoyment of the application or service in the light of the users' personality and current state" [17].

B. QoE INFLUENCING FACTORS

QoE influencing factors (IFs) have been described as the factual condition or adjustment of every feature of a user, system, service, application, or context, that can affect the user's experienced quality [17]. The IFs include, among other things, the application or service's type and characteristics, the usage context, the accomplishment of user's expectations for an application or service, the cultural background of the user, the socioeconomic aspects and psychological portrait of the user, and finally, the emotional condition of the user [25]. These IFs can be organized in three broad classes, as human-related, system-related and context-related [17]. In addition, a content-related IF class was added for video applications [26] as depicted in Fig. 1.

Human-related IFs refer to any variant attribute of a human user like motivation, attention level and emotional state, or any invariant trait such as age, gender, and visual and hearing sharpness. The demographic and socioeconomic context, the physical and mental constitution, or the emotional state of the user, may be also described by the human-related IFs [5].

System-related IFs refer to the impact of the parameters that operate at the technical level. They are linked to properties such as delay, transmission, packet loss, coding, storage, video buffering strategies, system hardware, rendering, and reproduction and display of media, which are linked to the transmission network, the end-devices, and the application layer of a communication link [27].

Context-related IFs consider the environmental factors associated with the user, such as the user's location, transient information like mobility, social factors like the presence or involvement of other individuals, and the purpose of using the service, such as for entertaining or educational reasons [28].

Content-related IFs consider the video streaming distinguishing features, such as the encoding rate, format, resolution, playback length, video quality, and video age, type and popularity [26].

C. SUBJECTIVE QoE ASSESSMENT

QoE assessment may be performed using two methods: the subjective and objective evaluation. Subjective assessment methodologies rely on receiving information from human assessors, who are subjected to a variety of tests or stimuli. Objective assessment models on the contrary, can be seen as the mean for evaluating QoE based exclusively on objective quality metrics.

In subjective QoE assessment a group of assessors is subjected to varying degrees of quality, which result in a form of explicit or implicit reaction from their side. Usually, quantitative approaches originating from related disciplines, like psychophysics and psychometrics are employed

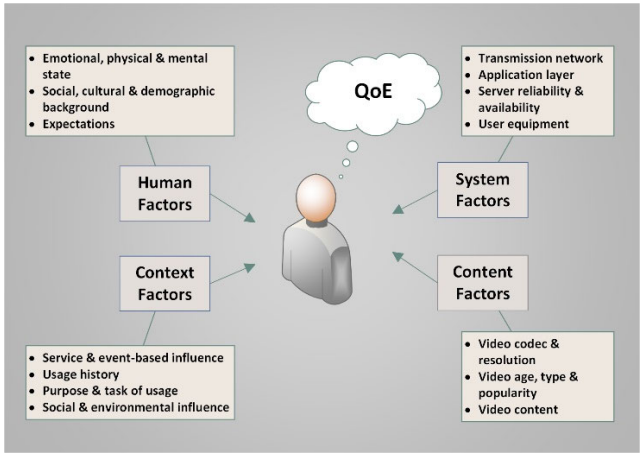

FIGURE 1. QoE influencing factors.

to acquire information on evaluators' ratings, utilizing scores that characterize their awareness on the degree of quality they experienced. Furthermore, qualitative approaches like focus groups, interviews, or profile assessments are employed as well, particularly to determine which IFs contribute to QoE, and in what capacity. The subjective evaluations are usually performed in a controlled laboratory environment, and need meticulous planning on which variables and IFs should be included in the procedure of assessing, monitoring and controlling quality [29]. Typically, the assessors rate a number of perceived quality aspects on a MOS scale, using a numeric value ranging from 1 to 5 (i.e., bad to excellent) [30] as depicted in Table 1, and report their ability to run a service and their level of satisfaction through survey methods like interviews, focus groups and questionnaires [31]. The MOS scale is calculated by averaging the perceptual video quality ratings acquired from the assessors. In the event of double stimulus tests, the differential mean opinion score (DMOS) is utilized, which is determined as the arithmetic difference between the ratings assigned to the processed video, and the ratings assigned to the source video. Video services constitute one of the most challenging QoE assessment, and therefore several methods were developed to perform subjective evaluation of video quality [32]–[35] as they are displayed in Table 2.

TABLE 1. Mean opinion score scale.

MOS	Quality	Impairment
5	Excellent	Imperceptible
4	Good	Perceptible but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

1) SINGLE SEQUENCE ASSESSMENT

Single stimulus (SS), known also as *absolute category rating (ACR)*, is a type of class ruling in which test sequences are displayed one at a time and scored independently on a five-level class scale. According to this approach, following each display, the assessors are requested to rate the quality of the delivered content. The required number of replications is specified by recreating the identical test settings at different point of times in the test.

Absolute category rating-hidden reference (ACR-HR), is a method of class ruling in which test sequences are shown one at a time and assessed independently on a class scale. A reference version of every test, which is called hidden reference condition, need to be included and displayed in the current test process, along with the rest test stimulus. A DMOS scale will be calculated throughout the data analysis on each test sequence and its corresponding reference, which is referred to as hidden reference. The ACR-HR approach should be used only with reference video that an expert deems to be of ''good'' or ''excellent'' quality on the five-level category scale, and it may not be appropriate for analyzing uncommon impairments that occur in the first and final second of the video sequence.

TABLE 2. Subjective assessment metrics.

Assessment method	Sequence type	Reference sequence	Rating
SS/ACR	Single	Test sequence	Five level category scale
ACR-HR	Single	Test & hidden reference sequence	Five level category scale
DCR/DSIS	Double	Test & reference sequence	Five level category scale
DSCQS	Double	Test & reference sequence	Five level category scale
PC	Double	Test & reference sequence	Five level category scale
SAMVIQ	Multiple	Test sequence	Continuous scale
SSCQE	Single, long sequence	Test sequence	Continuous scale
SDSCE	Double, long sequence	Test & reference sequence	Continuous scale

2) DOUBLE SEQUENCE ASSESSMENT

Degradation category rating (DCR), known as well as double stimulus impairment scale (DSIS), specifies that the test sequences are delivered in pairs, with the initial stimulus always be the source reference, whereas the second one to be the same source channeled through the system under evaluation. After viewing these two sequences on each session, the assessors provide a subjective judgment of the impairment sequence on the five-level category scale. The DCR method may be utilized to assess the accuracy of broadcasting systems as well as the fidelity of high-quality systems.

Double stimulus continuous quality scale (DSCQS), is regarded to be very effective in cases where it is not feasible to obtain test stimulus settings that reflect the complete extent of quality impairments. This approach is cyclic, in regards

that the assessors are required to evaluate a pair of sequences from an identical source, one straight from the source, the other through the system under evaluation and then they are tasked with evaluating the quality of both. The assessors are provided with a set of sequence pairs in internally random order, as well as random impairments encompassing all needed combinations, in sessions lasting up to half an hour. The mean scores for every test condition and test sequence are evaluated at the end of the sessions.

Pair comparison method (PC) approach indicates that the test sequences are provided in pairs, with the identical sequence provided initially via the first system under evaluation and then via the next one. The systems under evaluation (A, B, C, etc.) are commonly brought together in all possible n(n – 1) arrangements AB, BA, CA, and so on. As a result, all pairs of sequences should be shown in both potential orders (e.g., AB, BA). Following every pair, a decision is taken as to which element in the pair is preferable in the context of the testing procedure. The number of replications does not typically need to be addressed for the PC technique, as the approach itself requires repeated presentation of the same conditions applied in different sequence pairs.

3) MULTIPLE SEQUENCE ASSESSMENT

Subjective assessment methodology for video quality (SAMVIQ), is a subjective, non-interactive approach for assessing the video quality of multimedia applications. This approach may be used for a variety of applications, including among others algorithm selection, audiovisual system performance ranking, and video quality level evaluation during an audiovisual connection. A continuous quality scale is used in this approach, in which each assessor adjusts a slider on a continuous scale ranging from 0 to 100, which is divided in five quality levels that are defined in a linear fashion (excellent, good, fair, poor, bad).

4) LONG SEQUENCE ASSESSMENT

Single stimulus continuous quality evaluation (SSCQE), was initially intended to undertake time-efficient subjective quality assessments of digital services, as it eliminates the majority of the challenges experienced when utilizing traditional double stimulus approaches to evaluate the visual quality of digital systems. In this approach, the actual quality of a lengthier sequence is rated continually over time, using a slider on a scale from 0 to 100. Samples are collected at regular time intervals, yielding a quality curve over time rather than individual quality grading.

Simultaneous double stimulus for continuous evaluation (SDSCE) was designed on the basis of the SSCQE, by applying minor changes to the way the sequences are presented to the assessors and the rating scale. According to this method, a panel of assessors examines a pair of sequences simultaneously, one as the reference and the other as the test condition. Assessors are prompted to score the accuracy of the video information and the changes between the two sequences, through moving a slider on a handset-voting device. When the fidelity is flawless, the slider should be at the top of the scale ranging from 0 to 100, and when the fidelity is minimal, the slider should be at the lowest end of the scale.

The majority of subjective assessments are carried out in a laboratory setting. *Crowdsourcing* environments on the other hand, are gaining traction among researchers, since QoE assessment of multimedia applications may be relocated from traditional laboratory conditions to the internet, providing researchers with a valuable technique for accessing a worldwide pool of participants [36]. Consequently, a diverse and heterogenous set of users, terminal equipment and software configurations may be taken into consideration, while assessment may be performed in the assessors' real-life surroundings. Shorter turnaround times and decreased remuneration costs for test volunteers due to the large volume of participants, are also enticing to researchers [37]. There are already several commercial crowdsourcing systems accessible to perform online user surveys, including Crowdflower, Crowdsource, Microtask, and Amazon Mechanical Turk [38]. Furthermore, internet-based crowdsourcing platforms like Quadrant of Euphoria [39], crowdMOS [40], and QualityCrowd [41] demonstrate a methodological technique to deploying subjective assessments that may be carried out via a web browser [37]. These systems enable popular crowdsourcing platforms and assessment approaches like ACR and DCR [38].

The subjective assessment approach yields the most reliable findings due to direct data collection from end-users. The major disadvantages of the subjective assessment methods on the other hand, emanate on the fact that they are costly, time consuming, unable to be utilized in real time, and not repeatable. Because of these limitations, a strong motivation for the deployment of objective methods that predict the subjective perceived quality based only on physical attributes, was emerged [24].

D. OBJECTIVE QoE ASSESSMENT

The objective models are described as a method for assessing subjective quality purely on the basis of objective quality measurements or indices [42]. Namely, these models are anticipated to produce an estimation that is close to the rating acquired by subjective assessment methods. The advantages of the objective approach are its ease of implementation and modification, as researchers need only to be concerned with the measurable QoS factors and related mathematical models. The disadvantage of the objective assessment lies on its inaccuracy, as the obtained QoE is merely an approximation, rather than a precise value of the perceived quality of the end-user [43]. Over the years, researchers have explored methodologies and approaches for estimating image, video, and audio quality as perceived by end-users, and have devoted significant effort to the creation of metrics and models that can objectively predict the quality of a multimedia service. These metrics make use of audio, image, and video features to estimate the quality, and according to the quantity of source

information available, they are classified as full reference, reduced reference, and no-reference [44].

Full reference (FR) metrics have both the reference and the outcome sequences accessible, and in consequence, comprehensive subjective and objective associations of the videos are possible. Such metrics are appropriate for conventional broadcasting and television systems [26]. In terms of human perception accuracy, FR metrics that perform a frame-by-frame examination between the source and the affected sequence produce the better outcome. The structural similarities (SSIM) [45], video quality model (VQM) [46], and peak signal to noise ratio (PSNR) [47], are examples of such metrics. However, these metrics need access to the source data and are computationally demanding. As a result, they are unsuitable for real-time assessment, but preferable for benchmarking.

Reduced reference (RR) metrics utilize the same group of features to calculate the reference and outcome sequences. To get the quality evaluation, only a subset of partial parameters from the prototype input and output sequences is required [48]. These features may be at the application layer, such as bit-rate and frame-rate, as well as at the network layer, such as packet loss. RR methods are appropriate for real-time transport networks with limited computational and transmission bandwidth. Furthermore, they are well-matched to conditions in which the prototype input sequence is intricate to transport and store, or when computational power is constrained [38].

No reference (NR) metrics have only the outcome sequence supplied and therefore, the quality must be assessed without reference. The computational requirements of NR methods are the lightest in comparison with the other methods accompanied with efficient time response, but they are unable to deliver an accurate evaluation along a wide variety of video conditions [20]. These metrics are more appropriate to online services where just the outcome sequence is provided to the end-users. In mobile video streaming services for instance, it is difficult to discern if the discordance in quality is attributable to the quality of the reference or the in-between parts of the communication network [26].

1) APPLICATIONS OF OBJECTIVE QUALITY MODELS

Objective quality evaluation models can be used for a number of applications, including planning, lab-testing, and monitoring [49].

Planning consists of evaluating the perceived quality of services provided by networks and systems prior to implementation. Because it is not employed in a real-time setting, real-time inputs to the objective model are not necessary.

Lab-testing consists of evaluating the perceived quality of services of networks and systems in the laboratory, whilst equipment is being deployed.

Monitoring is the process of evaluating the perceived quality of services provided by operational networks and systems. The required information is gathered from the network, and analyzed to indicate the impairment of user's experience quality.

2) CLASSIFICATION OF OBJECTIVE QUALITY MODELS

Based on the application, objective quality assessment methodologies as depicted in Table 3 can be divided into five categories: media-layer models, packet-layer models, bitstream models, hybrid models, and planning models [49]–[51]:

TABLE 3. Objective assessment methods.

Method category	Input information	Primary application
Media-layer Model	Media signal	Quality benchmarking
Packet-layer model	Packet header information	In-service non-intrusive monitoring
Bitstream model	Packet header and payload information	In-service non-intrusive monitoring
Hybrid model	Combination of any	In-service non-intrusive monitoring
Planning model	Quality design parameters	Network planning, terminal/ application designing

Media-layer models accept as input actual media audio-visual signals and consider the codec compression and channel parameters. They estimate QoE using advanced perceptually-based psychophysical models that compare the output impaired signal to the input source signal (FR/RR models), or merely analyze the output impaired signal (NR model). The main applications of FR models include QoE evaluation in laboratory settings, like codec comparison and optimization, because such techniques estimate QoE using both the impaired received signal and the original source signal. RR/NR models on the other hand, can be used to monitor QoE either at the mid-point or end-points of an internet protocol television (IPTV) framework.

Packet-layer models predict QoE using merely packet header information. Since they do not parse the packet payload information, it is difficult to add parameters of QoE linked to media content into such models, even though they require a relatively modest computational efficiency overhead. Packet-layer models are mostly utilized as network probes at network mid-points or end-points.

Bitstream models accept both encoded bitstream information and packet header information as input. As a result, these models may be thought of as a hybrid of packet and media layer models. Because the bitstream-layer model uses solely the received packet information of the impaired signal, it may be utilized to monitor QoE at the mid-point or end-points of an IPTV framework.

Hybrid models, as the name indicates, are a mixture of the aforesaid stated models that employ as much information and data as possible to estimate QoE.

Planning models comprise the quality planning characteristics of networks or terminals to determine their input. They typically necessitate prior understanding of the system under test. These models can be used for network planning, as well as terminal and application design.

3) OBJECTIVE QUALITY METRICS

There are several objective quality evaluation methodologies dedicated in audio, image and video applications that differ in terms of intricacy, operation, and association with subjective quality evaluation [19], [29]. The following as shown in Table 4 are a few of the more representative objective metrics for multimedia services [38].

TABLE 4. Objective quality assessment metrics.

Metrics	Model Basis	Primary application
PSNR	Differentiation of original and distorted signal	Images and video
SSIM	Luminance, contrast, and structure comparison	Images
MS-SSIM	Weighted comparison of image characteristics and video luminance	Images and video
MPQM	Spatio-temporal HVS, contrast sensitivity and masking effect	Video
VQM	Structural and temporal parameters, perception-based characteristics	Video
VIF	Ratio of distorted image information to reference image information	Images
VSNR	Visual masking and summation approaches	Images and video
MOVIE	Space-time domain evaluation of spatial distortions and temporal impairments	Video
VMAF	Video quality degradation caused by compression and rescaling	Video
STRRED	Computation of the distortion between an impaired and a reference video sequence	Video
STRREDopt	Computational efficient variant of STRRED	Video
SpEED-QA	Mean-subtracted pixel values of frames and frame difference	Video
NR-P	Use of the decoded representation of the video to determine the quality	Video
NR-B	Features read from the encoded bitstream	Video
BRISQUE	Scene statistics of regionally normalized luminance coefficients	Images
NIQE	Quality-aware set of statistical characteristics	Images
PIQE	Psychovisually-based fidelity criteria	Images

Peak signal to noise ratio (PSNR) is a simple and widespread objective image and video quality evaluation method. PSNR assesses the variance among the prototype and impaired signals, by computing the mean squared error among the pair of signals, and the ratio among the greatest potential power of a signal and the power of degrading noise [52]. Though PSNR shows a poor connection with subjective evaluations, and is unsuitable for usage in real-time [53], it is nonetheless widely employed in video quality analysis, since it is simple to be calculated and provides an initial approximation of quality.

Structural similarity index (SSIM), which is commonly utilized for visual quality assessment (VQA), can be described as a still image sequence quality evaluation methodology that considers the visual masking phenomena. SSIM, which is based on the human visual system (HVS) concept, resolves some of the shortcomings of PSNR, like the responsiveness to alterations in brightness and contrast [54]. It relates the impaired video sequence with the source sequence in three ways: luminance, contrast and structure. These three parameters combine to provide the SSIM output. SSIM operation is more related with subjective QoE assessment, since the luminance and contrast evaluations are congruent with their masking effects [55].

Multiscale-SSIM (MS-SSIM) is a broadening of single-scale SSIM, which was first introduced for still images evaluation and then expanded to video. It considers the image signal's sampling density, the proximity among the viewer and the image, and the perceptual capabilities of the viewer's HVS [56]. MS-SSIM quantifies the impact of every scale with variant weights, in order to assess their comparative significance. It may as well be used in video applications, by taking into account the frame-by-frame luminance component of the video, and computing the average of the frame level quality ratings [57].

Moving picture quality measure (MPQM) is considered the main utilized metric for assessing moving picture quality, through modeling the spatio-temporal HVS framework and using a filter bank technique. To define visual detection, it takes into account two aspects of human perception, the masking effect and the contrast sensitivity. Unlike SSIM, MPQM assesses the quality of video sequences rather than single frame pictures, since it incorporates the effects of network transmission-related parameters on video quality [58].

Video quality metric (VQM) offers standardized and non-standardized techniques to evaluating perceived video quality, by considering temporal as well as structural parameters. The implementation of VQM entails collecting characteristics centered on perception, computing video quality features, and integrating these parameters for creating the model [46]. Because of the VQM high degree of correlation with subjective assessments from viewers, the American national standards institute (ANSI) has adopted it as a national standard [59].

Visual information fidelity (VIF), was introduced in the assessment of still image quality by contrasting two types of information, including the reference image that goes straight via the HVS, and the impaired image that initially goes via the distortion channel. The VIF metric consists of the ratio of the impaired image to reference image information [60].

Visual signal-to-noise ratio (VSNR) is a quality evaluation metric that has been suggested for still images, and has shown potential effectiveness for evaluating video quality. It identifies near-threshold and suprathreshold aberrations of human vision, in order to reduce the suprathreshold issue in HVS. To validate detectable aberrations, VSNR employs the visual masking and summation approaches. In the context of VQA, VSNR is applied frame-by-frame to the luminance component of the video and calculated as the average of the frame level ratings [61].

Motion-based video integrity evaluation (MOVIE) identifies video distortion in the space-time domain, instead of identifying them separately in the space and time domains, by calculating motion tracks of the video objects and conducting spatio-temporal evaluations of distortion. It is made up of two parts: the spatial MOVIE index, that evaluates spatial distortions, and the temporal MOVIE index, that evaluates temporal impairments. The aggregate of these two indices yields the concluding MOVIE score for a video sequence [62]. MOVIE correlates highly with subjective quality assessments, but its high computing cost prevents it from being employed in real-time applications [63].

Video multimethod assessment fusion (VMAF) has been created by the video streaming provider Netflix [64]. The metric is designed to be as robust as feasible in terms of association with subjective evaluations across the many types of material available on Netflix. Its primary focus is on video quality degradation caused by compression and rescaling. VMAF generates a quality score by first calculating values from four different NR and FR measures, which are then fused into a single quality score using a support vector regression (SVR) method. The following are the four metrics contained in the SVR: i) anti-noise signal-to-noise ratio (ANSNR); ii) detail loss measure (DLM); iii) VIF; iv) motion information [64].

Spatio-temporal reduced-reference entropic differencing (STRRED) computes the distortion between an impaired and a reference video sequence by constructing a Gaussian scale mixture (GSM), using the wavelet coefficients of the frames and frame differences [65]. These GSMs provide an indicator of each stream's spatial and temporal information, which may be compared in terms of entropy to evaluate the quality deterioration of the distorted stream.

STRREDopt is a computationally efficient variant of STRRED, in which only the best performing sub band of STRRED is calculated and utilized for quality assessment [66]. As a result, it is not necessary to calculate the whole steerable filter bank.

Spatial efficient entropic differencing for quality assessment (SpEED-QA) shares a similar methodology with the STRRED metric, with the exception that the GSMs are based on mean-subtracted pixel values of frames and frame differences instead of wavelet coefficients [66].

Pixel-based methods (NR-P) use the decoded representation of the video to determine the quality of a received stream. To estimate video quality, one or more visual and temporal artifacts are examined [67]. The following are the most commonly used features: blurriness, noise, blockiness, motion intensity (MI), jerkiness, spatial information (SI), and temporal Information (TI).

Bitstream methods (NR-B) seek to evaluate the received stream's quality based on features read from the encoded bitstream. This typically is accomplished using standard network and encoding techniques. The following are the most important features: packet loss ratio (PLR), bitrate, framerate (FR), quantization parameter (QP), scene complexity (SC), level of motion (LoM) and resolution (Res) [68].

Blind/referenceless image spatial quality evaluator (BRISQUE), does not calculate distortion-specific characteristics like ringing, blur, or blocking, but rather employs scene statistics of regionally normalized luminance coefficients to assess probable losses of naturalness in the image owing to the presence of distortions, resulting in a holistic measure of quality [69]. The inherent characteristics are derived from an empirical distribution of locally normalized luminance and outcomes of locally normalized luminance using a spatial natural scene statistic method.

Natural image quality evaluator (NIQE) relies on the development of a quality-aware set of statistical characteristics that stem from a space domain natural scene statistic (NSS) framework [70]. These characteristics are generated from a corpus of undistorted natural images. NIQE uses only quantifiable deviations from statistical regularities seen in natural images with no training on human-rated distorted images and no exposure to distorted images.

Psychovisually-based image quality evaluator (PIQE) assesses picture quality based on two psychovisually-based fidelity criteria, blockiness and similarity [71]. The blockiness index quantifies the patterned square artifact produced as a byproduct of JPEG and MPEG lossy discreet cosine transform (DCT)-based compression method. The similarity metric evaluates the amount of perceptible detail that remains after compression. The blockiness and similarity are merged into a single PIQE index which is used to evaluate quality.

4) OBJECTIVE QUALITY METRICS EVALUATION

The efficacy of objective quality metrics is typically evaluated with use of the following metrics [50]:

Pearson correlation coefficient (PCC) is defined as the linear relationship among the projected objective quality and the subjective MOS scores. It assesses a metric's prediction accuracy, namely its ability to predict subjective quality evaluations with a limited margin of error. The PCC for N data pairs (x_i, y_i), with $\bar{x}$ and $\bar{y}$ being the means of the respective data sets, is provided by:

$$PCC = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2}\sqrt{\sum ((y_i - \bar{y})^2}}, \tag{1}$$

The value of PCC fluctuates between $[-1, 1]$ as follows: 1 signifies full positive linear correlation, 0 signifies no connection and -1 signifies whole negative linear correlation.

In way to correlate the objective metric quality evaluations to the subjective quality ratings, the PCC is often computed after applying a nonlinear regression, using a logistic function.

Spearman rank order correlation coefficient (SROCC) is the coefficient of correlation among the projected objective quality and subjective MOS ratings. It assesses a metric's projection monotonicity, that is the level to which its projections coincide with the relative magnitudes of subjective quality scores. The SROCC is defined as follows:

$$SROCC = \frac{\sum (X_i - X')((Y_i - Y')}{\sqrt{(X_i - X')^2}\sqrt{(Y_i - Y')^2}}, \tag{2}$$

where X_i and Y_i are the xi and yi ranks, respectively. The midranks are represented by X' and Y'. SROCC has a value between $[-1, 1]$, in which 1 indicates that X is a monotonically growing function of Y and -1 indicates that X is a monotonically declining function of Y.

The outlier ratio (OR) is defined as the percentage of projections that fall outside of a span of ± 2 times the subjective outcomes' standard deviations. It assesses prediction consistency, viz how well the metric pertains projection accuracy. If N is the whole number of data points and N' is the number of determined outliers, the OR is described as follows:

$$OR = \frac{N'}{N}. \tag{3}$$

Root mean square error (RMSE) for N data points x_i, $i = 1, \ldots, N$, with $\bar{x}$ being the mean of the data set, is defined as:

$$RMSE = \sqrt{\frac{1}{N}\sum (x_i - \bar{x})^2}. \tag{4}$$

RMSE reflects the degree of data dispersion, and therefore, the smaller the RSME value, the better the prediction accuracy.

Mean absolute error (MAE) calculates the average degree of inaccuracies in a series of estimations devoid of taking into account their direction. It is the average of the absolute variances among estimation and actual examination of the test sample, where any distinct variances are given equal weight [72]. MAE is defined as follows:

$$MAE = \frac{1}{n}\sum_{j=1}^{n} \left|y_j - \hat{y}_j\right|. \tag{5}$$

Mean squared error (MSE) is a model assessment metric that is frequently used with regression models. A model's mean squared error with regard to a test set is the average of the squared prediction errors over all instances in the test set. The prediction error for an instance is the difference between the true and predicted values [73]. MSE is defined as follows:

$$MSE = \frac{\sum_{i=1}^{n} (y_i - \lambda(x_i))^2}{n}. \tag{6}$$

Median absolute error (MedAE) is a metric resistant to outliers. The loss is computed by averaging the absolute variances between the target and the prediction. If $\hat{y}$ is the projected value of the i_{th} sample and y_i is the actual value, then the median absolute error calculated across n samples is defined as follows [74]:

$$MedAE(y, \hat{y}) = median\left(\left|y_1 - \hat{y}_1\right|, \ldots, \left|y_n - \hat{y}_n\right|\right). \tag{7}$$

Kendall rank correlation coefficient (KRCC) is a non-parametric test that determines the level of dependency between a pair of variables. If we examine two samples, a and b, each with a sample size of n, the total number of pairings with a, b is $n(n-1)/2$. The KRCC value is computed exploiting the following equation [75]:

$$\tau = \frac{C - D}{\frac{1}{2}n(n-1)}, \tag{8}$$

where C is the number of concordant pairs, D is the number of discordant pairs, and $|\tau| \leq 1$.

R^2 is a statistical metric in a regression model that determines how much of the variation in the dependent variable can be explicated by the independent variable. R^2 indicates how well the data fit the regression model. R^2 is defined as follows [76]:

$$R^2 = 1 - \frac{SSR}{SST}, \tag{9}$$

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

E. QoE/QoS CORRELATION MODELS

Among the numerous studies on the IFs, considerable research efforts have concentrated on discovering the correlation between QoS parameters and QoE, with many examples arguing that a user's perceived quality is mostly determined by QoS [44]. Evaluating the association among the parameters of QoS and their individual and mutual effect on QoE, requires establishing a correlation model between QoE and QoS. Several correlation models have been suggested in the literature, with the two more characteristic include: a) the general models, where QoS is just one component in a series of QoE determining variables; and b) the particular models, which examine the impact of QoS parameters on QoE [77]. The concept underlying QoS/QoE mapping is to determine QoE values from a collection of measurable input parameters. The objective parameters of QoS refer to the degree of the service adequacy, and include the network KPIs. QoE can be derived from these metrics through a QoS/QoE mapping course that employs suitable mathematical models. However, subjective user-centric factors that cannot be assessed directly from the network but might impact the end-user's overall experience, are also included in the QoE influencing factors [78]. Comprehending the correlation among QoS parameters that rely on network and QoE as perceived by end-users is critical in QoE management procedure, particularly for CSPs having control over network resource scheduling and supplying processes [9].

Considering that subjective and objective quality indices often have distinct ranges, an appropriate mapping function is necessary to convert the objective video quality (VQ) into

the predicted subjective score (MOSp). Mapping functions are classified as linear or non-linear. When both objective and subjective assessments are similarly scaled, an identical numerical difference correlates to an equal perceived difference in quality across the entire span, so the linear mapping function may be utilized [79].

$$MOSp = a + b \cdot VQ, \tag{10}$$

where a1 and a2 parameters may be defined by employing a linear fit between the VQ values and the associated MOS scores. Afterwards, in order to assess the objective metric, MOSp values and predicted scores must be correlated mathematically to the actual results. Nonetheless, objective quality scales are seldom consistent, hence the linear mapping function may give a downwards evaluation of overall performance. Nonlinear mapping functions solve this problem, which is why they are commonly utilized in most applications. Nonlinear mapping functions typically provide substantially stronger correlations than their linear equivalents [24]. Logistic (6), Cubic (7), Exponential (8), Logarithmic (9) and Power (10) functions are the most often used mapping functions in the literature. All these different mapping function types correlate to different QoS and QoE parameter measurements [79]:

$$MOSp = \frac{a}{1 + \exp[-b \cdot (VQ - c)]}, \tag{11}$$

$$MOSp = a + b \cdot VQ + c \cdot VQ^2 + d \cdot VQ^3, \tag{12}$$

$$MOSp = a \cdot \exp(b \cdot VQ) + c \cdot \exp(d \cdot VQ), \tag{13}$$

$$MOSp = a - b\,|\log(VQ)|\,, \tag{14}$$

$$MOSp = a \cdot VQ^b + c. \tag{15}$$

A number of generic modeling methodologies for the development of QoE/QoS correlation models for multimedia services may be found through a literature review, including Choquet integral [80], as well as the widely used exponential and logarithmic approaches [29].

IQX hypothesis is an exponential approach that defines QoE as a properly parameterized negative exponential function of an individual QoS degradation parameter. In principle, QoE is a function of n impact factors I_j, $1 \leq j \leq n$:

$$QoE = \Phi(I_1, I_2, \ldots I_n). \tag{16}$$

IQX hypothesis centers on an individual influence component, $I = QoS$, in effort to obtain the primary correlation $QoE = f(QoS)$ [81]. At large, the subjective sensibility of the QoE becomes more acute as the experienced quality increases. If the QoE is very high, even a minor disruption will significantly reduce the QoE, but if the QoE is already low, a new disruption is not felt as strongly. In light of this, it is inferred that the change in QoE is dependent on the existing level of QoE, given the same amount of change in QoS value, but with the opposite sign [44]. Assuming a linear relationship on the level of QoE, we end up to the differential equation:

$$\frac{\partial QoE}{\partial QoS} = \sim -(QoE - \gamma). \tag{17}$$

The solution to this equation is determined as an exponential function that represents the IQ hypothesis underlying relationship:

$$QoE = \alpha \cdot e^{-\beta \cdot QoS} + \gamma. \tag{18}$$

Weber-Fechner Law (WFL) is a logarithmic approach that marks the beginning of psychophysics as a scientific field. In principle, the WFL relates the human sensory system's perceptual capacities, with the awareness of barely noticeable variations among two degrees of an evident stimulus. Regarding human senses, such a barely noticeable variation can be demonstrated as a constant proportion of the initial stimulus magnitude. Experiments on the sense of touch for example, have revealed that humans can perceive a rise in the heaviness of an item in their hands in case it is raised by roughly 3%, regardless of its absolute value [82]. The differential equation that expresses this is as follows:

$$\frac{\partial Perception}{\partial Stimulus} \sim -\frac{1}{Stimulus}. \tag{19}$$

Consequently, the resultant mathematical relationship has logarithmic form, and may be utilized to characterize the interdependence between stimulus and perception. In the domain of QoE, common stimuli have been demonstrated to be most commonly QoS features in the application level, that are immediately experienced by end-users [83].

III. SPECIFIC QoE ASSESSMENT ASPECTS FOR EXTENDED REALITY

Extended reality (XR) encompasses all real-and-virtual mixed environments, as well as accompanying human-machine interactions, enabled by computer technology and wearable devices. It involves a series of characteristic forms, including virtual reality, augmented reality, and mixed reality, as well as the regions interpolated between them [84]. Virtuality levels may vary from partially sensory inputs, to completely immersive virtual reality. The expansion of human sensations, particularly those related to the senses of existence, as represented by virtual reality, and the development of cognition, as expressed by augmented reality, is a major element of XR applications.

A rendered representation of a given visual and audio scenario is referred to as *virtual reality (VR)*. An observer or user moving within the application's limitations, receives the rendered imitating visual and aural sensory stimuli of the actual world as naturally as possible. VR applications typically necessitate the use of a head mounted display (HMD), to entirely substitute the user's field of view with a simulated visual component, as well as the use of headphones, to supply the user with the associated audio. Moreover, a degree of head and motion tracking of the user is generally required, to allow the simulated visual and audio features to be updated,

ensuring that from the user's perspective, objects and sound sources stay consistent with the user's motions.

When a user is supplied with additional information, artificially created objects, or content overlaid on their current environment, this is referred to as *augmented reality (AR)*. This type of additional information or content typically consists of visual and/or audible features, either by direct observation of their current environment, without intermediate sensing, processing, or rendering, or by indirect observation, by relaying the perception of their potentially enhanced or processed environment through sensors.

Mixed reality (MR) refers to a more sophisticated version of AR, in which certain virtual components are integrated into the physical environment to create the illusion that they are part of the real scene.

A. IMMERSION AND PRESENCE

Significant concepts used in the context of XR include *immersion*, which refers to the sensation of being surrounded by the virtual environment, and *presence*, which refers to the sensation of being physically and spatially placed in the virtual environment [84].

The sense of presence is crucial not only in VR experiences, but also in immersive AR applications. To establish presence in AR, the virtual content and actual environment must be seamlessly integrated. Moreover, the virtual content, such as in VR, must correspond to the user's expectations. It is envisaged that users would be unable to distinguish virtual items from actual objects in truly immersive AR, and in particular in MR [84]. The awareness for the user is essential for VR and especially for AR, but also essential is the awareness for the environment. This involves parameters such as the safe zone discovery, dynamic obstacle warning, geometric and semantic environment parsing, environmental lighting, and world mapping. In the case of AR, the user may wear a see-through HMD to see 3D computer-generated items superimposed on its real-world perspective. The see-through feature may be achieved with either an optical, or a video see-through HMD [85].

1) COGNITIVE AND PERCEPTIVE PRESENCE

There are two kinds of presence, cognitive and perceptive presence:

Cognitive presence refers to the degree to which the virtual environment prevails over the actual environment as the ground for cognition. This only covers the abstract concept of the virtual environment, not the virtual reality system or the technology utilized to display it [86].

Perceptive presence refers to the presence of a user's senses and it can be achieved with deception of the senses, including sight, hearing, touch, and smell. In order to attain perceptual presence, the XR devices are employing positional tracking based on movement to accomplish deception of the user's senses, in particular those related to the audio-visual system. The system's objective is to keep the user's sense of presence intact [84].

The illusion of being in a stable spatial place, of self-embodiment, of physical interaction and of social communication are all key components of perceptive presence [87]. From a technical standpoint, the most essential component is the illusion of being in a stable spatial place, which may be divided into three main categories: visual presence, auditory presence, and sensory or haptic presence.

2) VISUAL PRESENCE

The formulated technical requirements for *visual presence* include parameters like tracking, latency, persistence, resolution and optics [88]:

Tracking requires: 1) 6 degrees of freedom, which enables tracking of the user's head in rotational and translational movements; 2) 360 degrees tracking, which track the user's head regardless of the direction the user is facing; 3) sub-centimeter tracking accuracy; 4) quarter-degree-accurate rotation tracking; 5) no jitter or shaking, as image on the screen must remain completely steady; 6) suitable tracking volume for room-scale games and experiences, with about 2m cubes of area to roam around while still being monitored, and lower tracking volume for seated games and experiences; and 7) frequent update rates in order to be able to operate with the latest XR Viewer Pose.

Latency requires: 1) less than 20ms motion-to-photon latency, and less than 20ms of overall latency, which is determined as the time between the head movement and the change in the display; 2) minimization in the time of pose-to-render-to-photon, as render to photon should be less than 50ms in order to avoid incorrectly rendered content; 3) fusion of optical tracking and inertial measurement unit (IMU) data; 4) minimization of the loop including tracker, CPU, GPU, display, and photons; and 5) minimization of interaction delays and age of content depending on the application.

Persistence requires: 1) low persistence, in which pixels turn on and off every 2-3ms to avoid smearing and motion blur, where pixel persistence is defined as the amount of time per frame that the display is actually illuminated rather than black; and 2) 90Hz and beyond display refresh rate to minimize visible flicker.

Resolution requires: 1) spatial resolution, with no apparent pixel structure, in which the pixels are not visible, as low resolution and low pixels per inch (PPI) can create pixelation and the users to feel as though they are gazing through a screen door; and 2) temporal resolution, because regardless the use of asynchronous time warping, a continuous frame rate of 90Hz or above is needed in order to provide comfortable, engaging VR that genuinely generates presence.

Optics require: 1) wide field of view (FoV), as the extension of the observable world at any given moment, with typically needing 100-110 degrees of FoV; 2) comfortable eyebox, which is defined as the minimum and maximum eye-lens distances at which an image may be comfortably seen via the lenses; and 3) high quality calibration and correction, with rectification for distortion and chromatic aberration being an exact match for the lens characteristics.

TABLE 5. Virtual reality QoE influencing factors.

Human factors	System factors				Context factors
	Content-related	Media/content-related	Network/transmission-related	Hardware-related	
Vision and hearing	Spatial audio	Compression	Delay	Head-mounted display	Physical context
Simulator sickness	Spatial depth (3D)	Video	Bandwidth	Headphones	Temporal context
Immersion	Spatiotemporal complexity	Audio	Loss	Decoder performance	Social context
Expectations and expertise		Storage and transport		Head-tracking	Task context
		Bitrate		Field of view	
		Resolution		Display resolution	
		Frame rate		Refresh rate	
		Audio sample rate			
		Coding delay			

B. USER INTERACTION DELAY AND AGE OF CONTENT

Aside from the sense of presence and immersion, the age of the content and user interaction delay are critical for both immersive and non-immersive interactive experiences, namely the set of experiences in which user interaction with the scene affects its content, as follows [84]:

User interaction delay refers to the time elapsed between the instant a user action is started and the instant in which the content creation engine takes this action under consideration. This is the time period between the instant the user interacts with the game, and the instant in which the game engine analyses such a player reaction within the context of gaming.

Age of content refers to the amount of time that elapses among the time content is generated, and the time it is shown to the user. Within the context of gaming, this is the time period between the generation of a video frame from the game engine, and the moment that this frame is eventually shown to the player.

Moreover, *roundtrip interaction delay* is described as the sum of the user interaction delay and age of content, which is significant in the case of raster-based split rendering in cloud gaming applications, where part of the rendering is completed in an XR server and the service generates a frame buffer as the rendering outcome of the state of the content. In cloud gaming applications, the processes of user interaction delay and age of content listed below, contribute to the roundtrip interaction delay as follows [84]:

User interaction delay contributes with: 1) capture of user interaction in game client; 2) delivery of user interaction to the game engine (network delay); and 3) processing of user interaction by the game engine/server.

Age of content contributes with: 1) creation of one or several video buffers by the game engine/server; 2) encoding of the video buffers into a video stream frame; 3) delivery of the video frame to the game client (network delay); 4) decoding of the video frame by the game client; and 5) presentation of the video frame to the user (framerate delay).

C. VIRTUAL REALITY QoE INFLUENCING FACTORS

VR QoE IFs as shown in Table 5, are categorized as human influencing factors, system influencing factors (content-related, media/codec-related, network/transmission-related, hardware-related), and context influencing factors, as follows [89]:

1) VIRTUAL REALITY QoE HUMAN INFLUENCING FACTORS

VR QoE human IFs include vision and hearing, simulator sickness, immersion, and expectations and expertise [89]:

Vision and hearing: in the human eye, visual anomalies can arise and may have a detrimental impact on the user experience. Hearing impairments can cause attenuation of hearing over the whole audible frequency range or at specific frequencies and effect QoE as well [89].

Simulator sickness: simulator sickness, also known as cybersickness, virtual reality sickness or visually induced motion sickness, is caused by visual stimuli and can cause symptoms such as fatigue, perspiration, vertigo, or nausea [90]. Aside from technical factors, individual factors, contextual factors and covariate constructs can all have an effect on the intensity of simulator sickness [91]. The simulator sickness questionnaire (SSQ), VR sickness predictor (VRSP), VR sickness assessment (VRSA) and visual comfort assessment (VCA) are the most widely used metrics for measuring simulator sickness.

Immersion: individuals differ in their proclivity to experience immersion and their level of competence in utilizing VR equipment [89].

Expectations and expertise: the degree of experience in utilizing VR systems may influence how capable users are in using the systems to achieve a certain objective [89].

2) VIRTUAL REALITY QoE SYSTEM INFLUENCING FACTORS

VR QoE system IFs are further categorized in four classes, namely content-related, media/codec-related, network/transmission-related and hardware-related, as follows [89]:

a: CONTENT-RELATED

Spatial audio: spatial audio entails the use of loudspeaker or headphone-based spatial audio reproduction techniques for generating the illusion of immersion in VR applications [89].

Spatial depth (3D): stereoscopic video content is based on the depth-dependent disparity resulting from the two slightly different perspectives shown to the two eyes [92].

Spatiotemporal complexity: the complexity of a video image is indicated by spatial perceptual information. The amount of change in the video image is indicated by temporal perceptual information. High spatiotemporal complexity may cause a large level of simulator sickness [93].

b: MEDIA/CODEC-RELATED

Compression: video and audio codecs are employed to compress raw scene data so that it may be saved offline or streamed over a network, therefore preserving bandwidth and resources [89].

Video: conventional video codecs may be incompatible for some spatial representations in VR content. New video coding methods, such as versatile video coding (VVC) are under development and will greatly enhance the transport quality [89].

Audio: additional data for user head rotation should be added where a large number of static points is authored in order to display a spatial auditory scene compatible with listener motions. Direct sound, early reflections, and late reverberation should also all be considered [89].

Storage and transport: a method for encoding 360° movies while preserving over 80% of the bitrate uses a pyramid geometry, in which when the viewing direction is changed, the network condition and the user's orientation are used to determine which stream should be fetched [94].

Bitrate: the amount of audio or video bits transferred or processed per unit of time is referred to as the bitrate. Under the same encoding conditions, better resolution, higher frame rates and lower compression typically result in increased bitrate [89].

Resolution: the quantity of discrete pixels included in video content that may be depicted in every dimension, is represented by the video resolution. Because pixels are distributed in a 360° viewing radius around the viewer, increased resolution is necessary for VR compared to 2D video [89].

Frame rate: the frame rate is the frequency at which successive images, known as frames, are depicted. In a VR content, the frame rate should be exactly the same as the refresh rate of the HMD's display to improve QoE, otherwise it can cause artefact including frame fluctuation, frame dropouts and frame manipulation, which produce jerkiness and result in decreased QoE [95]. In the case of 360° videos, adding motion interpolation to content with a lower frame rate than the HMD's display refresh rate, is an effective way to improve QoE [96].

Audio sample rate: the number of audio samples conveyed each second, is known as the sample rate, given in hertz (Hz). This element is the same in VR services as it is with regular streaming services [89].

Coding delay: VR-related applications need very extremely low coding delays. Perceptual thresholds are available for television broadcasting [97], but VR introduces additional challenges owing to the immersive experience and sensorimotor coupling in six degrees of freedom [89].

c: NETWORK/TRANSMISSION-RELATED

Delay: stringent latency constraints are critical in VR applications for offering a good immersive VR experience. Examples of delay are queuing delay, over-the-air delay, and buffering delay. Delay is generally the major cause of excessive motion-to-photon delay, which causes simulator sickness. It is also the source of poor presentation quality, such as extended initial loading delay and stalling [98].

Bandwidth: immersive experiences with VR streaming environments necessitate a large amount of data. If the needed bandwidth is not provided, long delays and packet loss can be created [89].

Loss: the effect of packet loss on the VR experience is determined by the scheme of transmission. Packet loss in reliable transmission methods results in packet retransmissions, which adds to the total delay. In unreliable transmission, packet loss may result in the loss of portions of frames or complete frames, degrading audiovisual quality, which may be manifested as phenomena such as video freezing and tiling artefacts [89].

d: HARDWARE-RELATED

Head-mounted display: HMD wearing comfort may have a significant influence on ultimate VR QoE. To enhance this, it is critical to take into consideration the device's weight, size, heat dissipation, resolution, refresh rate, and so on [89]. Because of the importance of HMD in overall QoE, a complete evaluation framework of these devices must be taken into account [99].

Headphones: the frequency response of headphones is an important component in QoE. Neutral headphones or headphones with adjusted frequency response may be able to effectively express the listener's spatial audio experience.

Decoder performance: the capabilities of the decoder determine the ultimate resolution of the video to be transmitted and decoded in the display device [89].

Head-tracking: It is critical to acquire user locations and motion information in order to enable interaction between users and the environment. This is often accomplished, with the inertial measurement unit integrated in the HMD [89].

Field of view: FoV is the size of the viewable environment at any one time. With a broader FoV it is likely to experience immersion. FoV is the solid angle that a human can see through the HMD lenses. While a broad FoV might improve immersion, it can also induce simulator sickness [100].

Display resolution: display resolution is a fundamental feature of a screen that represents the number of pixels per inch that a panel can handle [89].

Refresh rate: the refresh rate is determined as the number of times per second that a display receives a new image from the GPU. A decreased refresh rate might contribute to greater processing delay and VR sickness, which is characterized by screen glitches [89].

3) VIRTUAL REALITY QoE CONTEXT INFLUENCING FACTORS

VR QoE context IFs include physical context, temporal context, social context, and task context [89]:

Physical context: physical context factors are associated with the setting in which a user is interacting with VR services. The user's experience may be impacted by background sounds, by whether the HMD device is wireless or linked to a fixed processing unit, by the ambient temperature of the room, or by the quantity of sunlight that enters the environment.

Temporal context: the frequency and duration of usage are characteristic temporal context factors. A VR device may not be able to withstand prolonged use, as simulator sickness symptoms such as dizziness, loss of spatial awareness, nausea, and eye discomfort generally worsen with increased use duration.

Social context: Considerations such as the popularity of VR content and how VR services are accessed are examples of social context variables. Interaction with a group of other individuals for instance, may have an impact on the user.

Task context: The VR experience is determined by the user's intentions for using the VR service. These are known as task context factors. For instance, the QoE for streaming type VR, such as 360-degree VR, would be significantly different for gaming or for social VR.

IV. SPECIFIC QoE ASSESSMENT ASPECTS FOR VIDEO GAMING

Apart from audio, video, and web browsing, online video games that operate over IP-based networks, are gaining increasing attention and popularity. Assessing the QoE of online gaming applications is a necessary condition for the management of gaming services [101]. Game providers strive to enhance their users' experiences by guaranteeing increased levels of platform and transmission operation, introducing novel methods of interaction and also more intriguing interfaces, or developing innovating game concepts. Video games can be described as a rule-centered structure with a changeable and computable after-effect, in which the player puts effort to affect the outcome and feels emotionally tied to it, and where the consequences of the action are negotiated [102].

Video games may be executed on personal computers (PC games), consoles (console games), mobile devices like smartphones and tablets (mobile games), and can operate independently on a device or on a server connected to the internet (online games). Regarding online games, they can be divided in instances in which the interface software, interface device, and backend platform constitute a physical system, while the game operates in a remote location (multiplayer games), or instances in which only the device and interface software constitute the physical system, and the control execution, game logic, and rendering occur at a remotely in the cloud (cloud games) [103]. One explanation for the difficulty in understanding the gaming's QoE, is attributed in that video gaming can be seen as an interaction among people and machines, rather than a mere media provision, therefore traditional methodologies for measuring transmission effect on media provision do not apply. Moreover, besides the content of game, the backend platform on which the game is built, user interface concerning both hardware and software, transmission channels involved, and also the user's attributes, may all have a major influence on user-perceived QoE.

A. TAXONOMY OF VIDEO GAMING QoE ASPECTS

The taxonomy of QoE aspects include the following [104]:

Aesthetics and appeal: aesthetics relate to the sensory experience elicited by a system, as well as the degree to which this experience is consistent with a user's aims and attitude. The term ''system personality'' states end-users' impressions of system attributes generated from technological and gameplay aspects. The product's appearance, physical qualities, and degree to which it inherits distinctive, unique and unexpected traits, all add to its attraction.

Interaction quality: the degree to which all functional and structural components of the game create a positive player experience is referred to as interaction quality. This definition takes playability into account as a requirement for positive player experience, or as a technological and structural foundation for it, but not as directly as the experience of the player.

Playing quality: playability may be thought of as a subset of playing quality. This is described as a player's capacity to learn, comprehend, and control a game instinctively. Usability may not deal with problems such as entertainment, engagement, or plot, all of which are inextricably linked to creative as well as technological issues.

Engagement: involvement, immersion, presence, flow, and absorption, are concepts that outline engaging experiences when playing video games, as follows: i) involvement is a mental condition that occurs as a result of the user's psychophysical state and attention being directed toward a coherent collection of stimuli, substantively connected actions, or occurrences; ii) immersion is a psychological condition in which the users perceive themselves to be surrounded by, considered part, and interconnecting with an environment that continuously delivers incentives. Immersion is a term used in video games to characterize a player's degree of participation, and it is split into three stages: engagement, engrossment and total immersion [105]; iii) presence is the mental sensation of ''being there,'' conveyed by surroundings that excite consciousness, attracts interest, and promotes direct engagement; iv) flow is the pleasant feeling that emerges as a consequence of an adequate equilibrium of obstacles and abilities in a target-oriented setting, as well as the satisfaction of the need for competence. It is a unique feeling that results from the accomplishment of a certain objective [106]; and v) cognitive

TABLE 6. Video gaming QoE influencing factors.

Human factors	System factors				Context factors
	Game-related	Playing device-related	Network/transmission-related	Compression-related	
Experience	Game genre	Device portability	Delay	Frame rate	Physical environment factors
Intrinsic and extrinsic motivation	Game mechanics and rules	Handheld device size	Jitter	Resolution	Social context
Static and dynamic human factors	Temporal and spatial accuracy	Input modalities	Bandwidth	Rate controller modes	Service factors
Human vision	Temporal and spatial video complexity	Output modalities	Packet loss	Group of pictures	Novelty
	Pace	Display		Motion range search	
	Visual perspective of the player			Audio compression	
	Aesthetics and design characteristics				
	Learning difficulty				

absorption is a broad term that refers to intense engagement with a game. It is founded on the following interconnected ideas: the absorption as a personality characteristic, flow condition and cognitive engagement as a concept [107].

Positive and negative effect: positive effects can take various forms, and they are typically the objective of any gaming activity. This definition of fun associates positive emotions such as joy, involvement, satisfaction, enthusiasm, amusement, fulfillment, euphoria, enthusiasm, and material expertise. Frustration and boredom on the other hand might be considered negative effects.

Player experience: player experience describes the degree of joy or exacerbation felt by the player after the gaming session. It includes intensity, immersion, favorable and unfavorable consequences, difficulty, ability, and flow.

Acceptability: acceptability, defines how readily the system is used by a user. A purely economic metric that compares the number of prospective users with the size of the target group may indicate acceptability. Acceptability is impacted by prices, accessibility, player experience, and service conditions.

B. VIDEO GAMING QoE INFLUENCING FACTORS

Video gaming QoE IFs as shown in Table 6, are categorized as human influencing factors, system influencing factors (game-related, playing device-related, network/transmission-related, compression-related) and context influencing factors, as follows [108]:

1) VIDEO GAMING QoE HUMAN INFLUENCING FACTORS

Video gaming human IFs include experience, intrinsic and extrinsic motivation, static and dynamic human factors, and human vision [108]:

Experience: experience with gaming in general, is used to differentiate user groups on the basis of an average time spent playing or encountering a particular game or game genre. These attributes are linked and vary dynamically with the competency of the user [108].

Intrinsic and extrinsic motivation: since the wide variety of available gaming types provide various kinds of enjoyment as well as incentives in engaging with them, intrinsic and extrinsic motivation can have a noteworthy effect on QoE [109].

Static and dynamic human factors: static human factors refer to a player's static attributes like age, gender and mother tongue, whereas dynamic human factors are emotional, like tedium, diversion, interest, and so forth [108].

Human vision: the properties of the visual stimuli influence the visual perception. The susceptibility of a user to video/network abnormalities varies depending on the person. Sensitivity to frame rate as an encoding parameter, for instance, is determined by the user's critical flicker fusion threshold [110].

2) VIDEO GAMING QoE SYSTEM INFLUENCING FACTORS

Video gaming QoE system IFs are further categorized in four classes, namely game-related, playing device-related, network/transmission-related, and compression-related, as follows [108]:

a: GAME-RELATED

Game genre: in the experimental design, genre classification may be employed as a basic criteria of content selection. Although various game interactions can be components of a specific game genre, the game itself is not adequate to define the game's susceptibility to technological factors [108].

Game mechanics and rules: game mechanics and rules have a major bearing and decide on the results of the game, being unique for any game [108].

Temporal and spatial accuracy: the time necessary to perform an action is characterized as temporal accuracy, whereas the level of accuracy taken to accomplish the interaction properly, is spatial accuracy [111].

Temporal and spatial video complexity: video complexity is essential for streaming services, particularly when taking into consideration encoding parameters like bitrate. Video content with high-complexity is more susceptible in the influence of network factors such as bandwidth, packet loss and encoding artifacts [108].

Pace: pace relates to the speed of gameplay and must be viewed as a speed in one game type or one game genre, which implies that two games with the same temporal complexity may not have similar paces. Pace should be regarded as an influencing parameter, particularly when studying temporal factors like delay and frame rate [108].

Visual perspective of the player: games are categorized into three types based on the camera's perspective: first-person linear perspective, third-person linear perspective and third-person isometric perspective. In cloud gaming, game perspectives are highly essential and an interplay with video coding can be expected [112].

Aesthetics and design characteristics: the game design that the player experiences is typically defined by design connoisseurs. There is no established categorization for game designs, nevertheless they certainly have a major effect on player experience [113].

Learning difficulty: when aiming for a quick interactive assessment, the time necessary to acquire knowledge on how to play a game is a crucial requirement [108].

b: PLAYING DEVICE-RELATED

Portability: the ongoing popularity of portable gaming devices shows that for a set of users, the value of mobility exceeds the constraints of a portable device [108].

Size: the dimensions of hand-held equipment have been found to impact the evaluations of playing test participants. Unless it is the subject of a research, it should thus stay consistent [114].

Input modalities: modalities used for gaming input vary greatly regarding feedback, speed, and precision. Nonetheless, various controllers can be used interchangeably, which will likely impact the *game's* experience [108].

Output modalities: the obtainability of output modalities, as well as their technological characteristics, limit the perceivable experience. Individuals using a VR headset report better degrees of immersion than users of the identical game simulation utilizing a traditional 2D screen [115].

Display: the viewing distance, display size, brightness, contrast, sharpness, screen resolution, refresh rate, and color, all have a major impact on perceived video quality. If the frame rate is high, the display size as well as the refresh rate of the display might result in greater quality [116].

c: NETWORK/TRANSMISSION-RELATED

Delay: the delay experienced by an end-user relates to the time elapsed between the execution of user commands and the appearance of a visible game event. The impact of delay on QoE is heavily influenced by game parameters [108].

Jitter: jitter has a discernible impact on the online and cloud gaming experiences [117]. Jitter could also cause a less smooth visual appearance of the game depending on the client implementation, since frames are shown at fluctuating time intervals [118].

Bandwidth: the effect of bandwidth constraints on QoE has been shown to be significant in the context of cloud gaming. Depending on the technique used to overcome restricted bandwidth, this may result in buffering delay, packet loss, and video artifacts caused by video compression [119].

Packet loss: packet loss has a substantial effect on QoE in gaming applications, with values as low as 1% resulting in a considerable deterioration in end-user's perceived experience. Substantial packet loss severely degrades the graphics quality, resulting in a reduced frame rate and an unsatisfactory gaming experience [120].

d: COMPRESSION-RELATED

Frame rate: the frame rate has a major effect on a gamer's effectiveness, and as a result on QoE. Experiencing the distinction among significantly high frame rates, is heavily dependent on a gamer's eye skills, gaming setup, game features and most notably game pace. When examining the influence of frame rate on QoE, it is necessary to consider display characteristics like refresh rate, and display size [108].

Resolution: the encoding resolution, whilst having moderate effect on a gamer's performance, is a critical parameter in impacting the quality and performance of video in every streaming application. Greater resolution is needed for consumers with a broad bandwidth, rather than enhancing other encoding factors like QP [121].

Rate controller modes: video streaming rates are controlled using a number of techniques to achieve a specific quality level with limited bandwidth available. Three types of rate controllers can be identified: the constant quantization parameter (CQP); constant rate factor (CRF); and constant bitrate (CBR) [108].

Group of pictures (GoP): in a video sequence, the GoP structure defines the alignments of the inter and intra frames (I, B, and P). The distance between two anchor frames determines the GoP value. The interval among two I-frames determines the GoP length, which is designed to minimize propagation error while maintaining video compression [108].

Motion range search: a motion estimation process effects coding efficiency and, by extension, the general performance of a gaming video service. The motion span analysis must be defined in addition to the motion assessment technique, depending on the video material [108].

Audio compression: an alike impact with the video compression feature is expected for audio as well. Nevertheless, unlike video compression, audio compression has not yet been exposed to equal level of research examination in the context of cloud gaming [108].

3) VIDEO GAMING QoE CONTEXT INFLUENCING FACTORS

Video gaming context IFs include physical environment factors, social context, service factors, and novelty [108]:

Physical environment factors: physical environment factors include room features such as size, acoustics, and lighting, as well as usage scenario such as in-house, or moving [108].

Social context: social context refers to the player's interactions with other players, possible concurrent operations and concerns about privacy and security that may be especially significant in multi-player games [108].

Service factors: customer satisfaction with online game services is influenced by service factors such as ease of access, availability, and cost, which is especially likely for cloud gaming applications [122].

Novelty: novelty indicates that improving user experience in the introduction of new technology does has an influence on quality scores, not due to genuine enhancement in learning or accomplishment, but because of increased interest in new technologies and services [123].

V. MACHINE LEARNING METHODOLOGIES

ML refers to the domain of computer theory, which allows algorithms to extract models directly from data without having to explicitly construct them, drawing inferences and estimations from input samples [124]. ML is a branch of artificial intelligence (AI) that in recent years has seen an unparalleled rise in its utilization in applications that solve complex problems and allow automation, across a wide range of domains, including telecommunication networks. This is mostly attributed to the massive volumes of available data, major advancements in ML methods, and latest progress in the capacity of computational resources. Without a question, ML algorithms are being constantly implemented to provide solutions to a wide range of complex issues in mobile and wireless communication networks control and management [125].

ML is utilized in a wide range of computational applications, when the creation and implementation of efficient explicit methods is infeasible. ML applications may include solutions to problems in which traditional approaches require extensive fine-tuning, or lengthy lists of rules, for which a single ML algorithm may provide code simplicity and outperform the conventional processing techniques. Furthermore, due to the capacity of ML algorithms to adapt to new data, ML methods can address a series of highly complicated issues that lie in fluctuating computational environments, for which traditional approaches offer no viable solutions, and also to gain insight about vast volumes of data.

Since there are many different types of ML algorithms, it is important to categorize them into broad classes, based on: 1) whether they are trained under human supervision and a learning signal or feedback is available to the learning system (supervised, unsupervised, and reinforcement learning); 2) whether they can learn incrementally on the fly (batch and online learning); and 3) whether they function by comparing new data points to known data points, or by identifying patterns in the training data and constructing a prediction model (instance-based and model-based learning) [126].

A. SUPERVISED LEARNING

In supervised learning (SL) the algorithm supplies examples of the inputs and their intended outcomes. The intended answers termed as labels, are part of the algorithm-fueled training data set. Supervised models train algorithms with use of a predetermined quantity of labeled data [127]. When there is input data and an intended label, the algorithm uses them to compute a label-data pair. The objective is for a function to be derived that incorporates mappings between input data and the output labels, using example data-label pairs as training dataset. In a specific scenario, where the algorithm only knows a portion of the sample data-label pairs, and parts of the intended output labels of incoming data are absent, the related learning model is referred to as semi-supervised learning [128].

SL models are categorized into two types based on whether the data they are called upon to evaluate are discrete or continuous: 1) *classification*, in which inputs with discrete values are split into two or more classes, and during training it is created a model that allocates the unobserved inputs either to one class in single-label classification mode, or more classes in multi-label classification mode; and 2) *regression*, which aids assessing the correlation between variables with continuous values, and allows the prediction of an output variable based on the value of a single or multiple predictor variables. The main SL algorithms are depicted in Table 7.

TABLE 7. Main supervised learning algorithms.

Method	Type	Description
Linear regression (LR)	Regression	Linear model of the connection between a scalar output and independent variables
Support vector machine (SVM)	Classification & Regression	A hyperplane or set of hyperplanes constructed in a high-dimensional space
K-nearest neighbors (KNN)	Classification & Regression	Non-parametric instance-based learning approach
Decision trees (DT)	Classification & Regression	Mapping observations to the target value
Random forest (RF)	Classification & Regression	Ensemble of decision trees to improve performance

B. UNSUPERVISED LEARNING

In unsupervised learning (UL) the algorithm does not include any labels, letting the pattern embodied in the input data to be discovered on its own, as the system attempts to learn without the assistance of an instructor. Unsupervised models are employed when it is required from the ML algorithms to infer a function from unlabeled data, in order to identify their hidden structure [124]. The accuracy of the structure cannot

be assessed, since the training samples provided through the learning process are unlabeled, and therefore, they are not benchmarked. The problem of density estimation in statistics is a basic instance of UL, although UL algorithms cover many additional issues that require the synopsis and elucidation of significant aspects of the unlabeled data. Because UL is not hinged on labeled data for training, it is appropriate for applications where the objective is unknown, or scalability is critical. The main UL algorithms can be seen in Table 8.

TABLE 8. Main unsupervised learning algorithms.

Method	Type	Description
K-means clustering	Clustering	Distance-based clustering approach
Expectation-maximization	Latent variable learning	Statistical model of a maximum likelihood iterative approach
Principal component analysis	Latent variable learning	Orthogonal transformation of possibly correlated training samples into uncorrelated variables
Independent component analysis	Latent variable learning	Decomposition of multivariate variables into a series of additive, statistically distinct & non-Gaussian subcomponents

C. REINFORCEMENT LEARNING

In reinforcement learning (RL) only feedback on the performance of the algorithm in a dynamic setting is provided, in terms of rewards and penalties. The learning system, termed as an agent, must then learn for itself the optimal method, known as a policy, to maximize reward over time. The RL methodology is based on the area of behaviorist psychology [124]. In RL models, the agent uses a trial-and-error method because it lacks a comprehensive model of the surrounding environment, and so does not know the consequences of an action. The environment alters its condition and produces reward and punishment information as the agent performs different actions. The agent subsequently modifies its actions dynamically in response to the acquired state and reward, seeking to maximize the benefit of its efforts. The main RL algorithms are included in Table 9.

D. ARTIFICIAL NEURAL NETWORKS

Artificial Neural Networks (ANNs) refer to computational modeling techniques that have gained widespread acceptance for modeling difficult real-world issues across many fields of applications. ANNs can be specified as computational structures made up of densely interlinked fundamental processing components, known as artificial neurons or nodes, that can perform huge numbers of parallel computations for implementing data processing and knowledge representation [129]. ANNs resemble the interaction amongst neurons in the human brain in two ways: 1) the network obtains its knowledge from the environment via a learning process; and 2) the interneuron connection strengths, namely synaptic weights,

TABLE 9. Main reinforcement learning algorithms.

Method	Type	Description
K-armed bandit	Model-based	Mimics a decision-making state where rewards are based on a stationary probability distribution linked with the actions
Markov decision process	Model-based	Decision-making framework in a discrete-time stochastic environment of Markov state transitions
Temporal-difference learning	Model-free	Mix of Monte Carlo techniques and dynamic programming that gathers knowledge from raw experience
State-action-reward-state-action	Model-free	Updates the Q-function based on interactions with the environment
Q-learning	Model-free	Modifies the Q-function based on the maximum reward

are utilized to preserve the acquired knowledge [130]. The input of each artificial neuron in a typical ANN model is a real valued signal, and its output is subjected to various nonlinear processes, such as activation functions [131]. In order to regulate the pace of the learning activity, artificial neurons and their connections generally utilize a weighting factor. Furthermore, artificial neurons are arranged in layers, where different layers transform their inputs in different ways, and input signals go from the initial to the final layer through a number of hidden layers.

Deep neural networks (DNNs) are ANNs with multiple hidden layers between the input and the output layers, opposed to ANNs with only one hidden layer, that are referred to as shallow ANNs [132]. The main objective of DNNs is to approximate complicated functions by combining basic and specified actions of units or neurons. An objective function of this kind may be of practically any sort, including classification, regression, or control. Depending on the model's structure, the functions are generally specified by a weighted blend of a certain collection of hidden units that have a non-linear activation function. These procedures, adjunct with the units in the output, are referred to as layers. A DNN learns multiple levels of representation and abstraction, by modeling high-level data abstractions via numerous nonlinear transformations. Recent advances in computational capability, the broad disposal of data for the training of a DNN, and the advent of efficient DNN training methods, are the major incentives that have facilitated the shift from traditional, shallow ANNs to DNN [131]. The main ANNs algorithms are depicted in Table 10.

VI. MACHINE LEARNING QoE PREDICTION MODELS

In this section of the survey, we analyze and classify ML-based QoE prediction approaches. The process of predicting end-users' QoE consists the first stage in the optimization of the multimedia streaming service provision. What is more, QoE prediction offers a deeper insight in the way that the technical parameters of the communication network impact the quality of service as it is perceived by

TABLE 10. Main artificial neural networks algorithms.

Method	Learning type	Description
Multilayer perceptron	Supervised, unsupervised, reinforcement	Data modeling using simple correlations
Deep neural networks (DNNs)	Supervised, unsupervised, reinforcement	Modeling of complicated operations using a weighted combination of a set of hidden levels with a non-linear activation function
Recurrent neural networks (RNN)	Supervised, unsupervised, reinforcement	Modeling of sequential data & dynamic temporal characteristics
Random neural networks	Supervised, unsupervised, reinforcement	Modeling of neurons interaction by exchanging excitatory and inhibitory spiking signals
Convolutional neural networks (CNN)	Supervised, unsupervised, reinforcement	Modeling spatial data & mapping multi-dimensional features
Generative adversarial networks (GANs)	Unsupervised	Generation of data and creation of realistic artifacts from a target distribution
Restricted Boltzmann machines (RBMs)	Unsupervised	Probabilistic generative models that extract features from their input data
Deep reinforcement learning	Reinforcement	Modeling & controlling high-dimensionality scenarios, under complex, changeable, & heterogeneous environments

the end-users. The classification of the state-of-the-art predictive models is application-oriented, as it includes solutions concerning video streaming, virtual reality and video gaming applications. We provide a thorough comparative analysis for each application genre, aiming to outline the distinction between conventional video streaming services and the emerging virtual reality and video gaming applications, with regard to the differentiation in the factors that have a significant influence on QoE, as well as in the metrics used to evaluate QoE.

A. VIDEO STREAMING

The comparative analysis of QoE prediction models for video streaming services as it is depicted in Table 11, includes approaches concerning video streaming applications of dynamic HTTP (DASH) video, HTTP adaptive streaming (HAS) video, HTTP video, H.264/AVC video, mobile video, YouTube video, 4K ultra-high definition (UHD) video and 5G video.

The predictive models for DASH video streaming [133]-[135] take into consideration the technical characteristics and network's conditions that impacting video quality and consequently QoE. The assessment of QoE relies on subjective metrics such as MOS and ACR [134], [135], as well as objective metrics including the FR MS-SSIM, the RR STRRED and the NR NIQE metric [133]. The highest prediction accuracy is achieved through the utilization of a model based on a combination of RNN and long short-term memory (LSTM) algorithms [133], which succeeds to reflect the nonlinearity and complicated temporal dependence owing to adaptive streaming speed adjustments of QoE. In [134], a QoE video DASH metric approach is presented that relies on three-dimensional convolutional neural networks (3D CNN) and LSTM, and utilizes the ridge regression technique to provide a QoE metric, which dynamically describes the correlation among the input characteristics vector and the MOS value. In [135], adaptive bitrate streaming (ABS) algorithms are analyzed, and an ML model based on decision tree regression (DTR), multi-linear regression (MLR) and random forest regression (RFR) is provided to evaluate QoE in DASH video streaming with respect to network metrics.

In the case of HAS video, the prediction models examine IFs that derive from end-user's traffic pattern characteristics [136] and incorporate both forward and backward dependence of the continuous QoE prediction [137]. The most accurate model however [138], is an end-to-end and unified predictive approach based on deep learning (DL) as a mix of CNN and LSTM that uses the MOS metric to assess QoE. In [136], the predictive model utilizes three distinct component selection methods and six different classifiers, by employing SL techniques, whereas, in [137], the inputs from perceptual visual quality metrics, rebuffering, and temporal memory-related data are analyzed, with use of bidirectional LSTM (BLSTM). As we can observe in the Table 11, both the models that employ ANNs methods attain higher prediction accuracy than the model based on SL algorithms.

For the QoE prediction in HTTP video streaming, objectivity-aware and psychology-aware impacting parameters are considered [139], and the influence of buffering and initial delay is examined [140]. Moreover, in [139], the characteristics of video content, encoding settings, network transmission metrics, and playout buffer parameters are taken into account, while in [140], the proposed model demonstrates that buffering pattern descriptors, particularly those associated with the occurrence of the last stalling event, have a clear effect on QoE. Both the approaches use subjective metrics to assess QoE and are based on SL algorithms. The model that employs the SVM algorithm achieves high QoE prediction accuracy, whereas the model based on the M5P tree model manage to substantially reduce the prediction errors.

In the QoE prediction for the H.264/AVC video, the models take under consideration cross-layer, application layer, video content and terminal equipment features [141], as well as lossy compression distortion and network transmission distortion [142]. As we can see in Table 11, these methods offer similar prediction accuracy, although they rely on different approaches. Both models use subjective as well as objective metrics to assess QoE, but the model in [141] is based on the employment of a feed-forward ANN with strong approximation capabilities, i.e., the radial function network (RBFN), whereas, in [142], an SL model based on DT algorithm is employed, which utilizes a set of basic

TABLE 11. Comparative table of QoE prediction models for video streaming services.

Application	ML technique	Influencing factors	Assessment metrics	Prediction accuracy	Reference
DASH video streaming	LSTM	Short time subjective quality (STSQ), playback indicator (PI), time elapsed since last rebuffering	STRRED, MS-SSIM, NIQE	0.907 to 0.985 PCC, 0.875 to 0.971 SROCC	[133]
DASH video streaming	C3D, LSTM	Avg. bitrate, proportion of bitrate, frame rate (FPS), avg. playback interruption length, avg. rebuffering, initial buffering, avg. bitrate switching count, variance in proportion of bitrate	MOS	0.9124 PCC, 0.9170 to 0.9465 SROCC	[134]
DASH video streaming	DTR, multi-linear regression, RFR	Round trip time (RTT), throughput, number of packets per video segment, rate-based, buffer-based, & hybrid ABS algorithms, number of stalls	MOS, ACR	72.37 to 87.63%	[135]
HAS video streaming	LR, linear discriminant analysis, KNN, DT, Gaussian naive Bayes, SVM	Bitrate, FPS, resolution (Res), device, application, SI, TI, QP, user profile, gender, duration	ACR	73.5 to 86%	[136]
HAS video streaming	BLSTM	STSQ, PI, time elapsed since last video impairment	STRRED	0.894 PCC, 0.830 SROCC	[137]
HAS video streaming	DL as combination of CNN & LSTM	Text, video, categorial information, continuous information, sequence data	MOS	88.74%	[138]
HTTP video streaming	SVM	SI, TI, brightness (Br), color information (CI), encoding bitrate (EBR), FPS, Res, PLR, initial buffering delay, rebuffering time ratio (RTR)	MOS	91.3%	[139]
HTTP video streaming	M5P	Stalling patterns, initial playback delay, video duration, frames-per-second, content class	MOS, DMOS, ACR	25 to 50% prediction errors reduction	[140]
H.264/AVC video streaming	RBFN	Bitrate, FPS, Res, PLR, screen size, SI, TI, Br, CI	MOS, PSNR	0.89 PCC, 0.28 RMSE	[141]
H.264/AVC video streaming	DT	Avg. quantization parameter, avg. bits/pixel in intra frames, avg. bits/pixel in inter frames, avg. ratio of bits/inter frame to bits/intra frame in the same pictures set, % of successfully received slices, % of correctly decoded frames, avg. burst length	MOS, SSIM, VQM	88.9 to 90.5%	[142]
Mobile video streaming	Random neural network	Encoder quantization parameter, PLR, Mean burst length (MBL), content class	MOS, VQM	0.39 RMSE, 0.90 R^2	[143]
Mobile video streaming	Multiclass incremental SVM	Delay, packet loss, rate, video type, movement, Res, video size, mean bitrate, FPS, frame lost, audio rate, audio lost, buffer time, vlc-catching, starting video time, lag between image and audio, image quality, audio quality	MOS	89%	[144]
Mobile video streaming	RF	RSSI, RSRP, RSRQ, SSSP, total reference signal power, CQI, MCS index, CINR, frame delay, frame skips, blurriness	MOS, PSNR	75 to 85%	[145]
Mobile video streaming	DNN	Visual quality, loading, stalling, overall quality, & 89 network parameters of mobile video transmission	MOS	0.8686 RMSE, 0.7609 MAE	[146]
Mobile video streaming	TCN	Short time subjective quality, PI, number of rebuffering events, time elapsed since the last video impairment	STRRED	0.820 to 0.892 PCC, 0.733 to 0.885 SROCC, 4.81 to 6.97 RMSE	[147]
Mobile video streaming	DL as combination of word embedding, C3D & representation learning	Video, text, categorical information, continuous values	MOS, ACR	90.94%	[148]
Mobile video streaming	Hierarchical & K-means clustering	Encoding bitrate, PLR, FPS, content classification	MOS, PSNR, single stimulus impairment scale (SSIS)	0.215 to 0.251 RMSE	[149]
YouTube video streaming	KNN, DT, RF	Video identifier, video lifetime, upload time, time between video uploading &	MOS	0.408 to 0.712 RMSE	[150]

TABLE 11. *(Continued.)* **Comparative table of QoE prediction models for video streaming services.**

		data collection, category description, view number, favorites count, likes number, dislikes number, shares number, discussion density, video total duration			
YouTube video streaming	BSVR	Delay, packet loss, rate, FPS, audio rate, video size, mean bitrate, Res, audio lost, frame lost, buffering, vlc-catching, video type, movement	MOS	0.47 RMSE	[151]
4K UHD video streaming	DL based on CNN	Brightness, colorfulness, RMS contrast, sharpness, image bitrate, resolution, JPEG compression quality, noise, JPEG artifacts, aliasing, lens and motion blur, over-sharpening, wrong exposure, color fringing, over-saturation	MOS	78%	[152]
5G video streaming	LR, SVR	Access node downlink (DL) throughput, access node uplink (UL) throughput, user equipment (UE) DL throughput, UE UL throughput, DL CQI, UL CQI measured at UE, measured RTT at the UE, Smoothed RTT using a moving average, statistic metrics	MOS	0.1 to 0.15 MSE	[153]

characteristics extracted from the compressed bitstream and network to predict QoE.

The prediction models for mobile video streaming [143]–[149] evaluate the impact of cross-layer IFs, including QoS components from the application layer as well as the physical layer. In [143], a no reference cross-layer end-to-end estimation model for mobile video perceptual quality is presented, based on random neural networks. In [144] an online QoE prediction model is proposed, capable of classifying user perception of video streaming services, based on incremental multiclass SVM (multiclass-iSVM) algorithm, which examines the efficacy of incremental learning in handling large scale dynamic data and improving QoE prediction accuracy. In [145], radio measurements of the wireless communication channel are considered, including the received signal strength indicator (RSSI), reference signal received power (RSRP), reference signal received quality (RSRQ), secondary synchronization signal power (SSSP), total reference signal power, channel quality indicator (CQI), modulation coding scheme (MCS) index, and carrier to interference plus noise ratio (CINR). In [146] a QoE estimation model with large-scale QoE dataset for mobile video streaming based on DNN is proposed, and is designed to learn the correlations among network characteristics and the subjective QoE scores. In [147], CNN-QoE (a model for continuously prediction of QoE) is proposed, and is based on temporal convolutional network (TCN). The CNN-QoE utilizes the benefits of TCN to overcome the computational complexity limitations of LSTM-based QoE models, whilst also providing architectural enhancements to increase QoE prediction accuracy. In [148], DeepQoE (an end-to-end framework for QoE estimation) is presented. DeepQoE is based on a combination of DL techniques, including word embedding, 3D CNN and representation learning. In [149], a QoE prediction model based on hierarchical and K-means clustering algorithms is proposed, which besides QoS parameters takes into account the video content classification in estimating QoE in the case of multipath video streaming over heterogeneous networks. For the QoE assessment of the aforementioned models both subjective and objective metrics are utilized as shown in Table 11. The highest prediction accuracy among the ANNs and ML implementations is achieved with the approach in [148], where word embedding and 3D CNN are utilized to capture generalized characteristics. These characteristics are then aggregated and incorporated into a neural network for representation learning, and subsequently the learned representation is used as input for classification or regression operations.

The prediction of YouTube video streaming QoE relies on quantifying the relationship between social context factors, user engagement characteristics and QoE [150], together with evaluating the impact of QoS and quality of application (QoA) factors [151]. In [150], the proposed model relies on boosting support vector regression (BSVR) with the goal to examine the efficacy of integrating many learners rather than the traditional individual learner for enhancing QoE prediction performance. In [151], the prediction model is based on SL algorithms, aiming to analyze the relationship between social context factors, user engagement characteristics and QoE, and calculate the end-to-end QoE for a specific element of user. Both these approaches employ the MOS metric for the QoE assessment and utilize SL algorithms. The highest prediction accuracy is achieved through the combination of KNN, DT and RF algorithms of the model in [151].

The QoE prediction model [152] for UHD video streaming takes raw RGB pixel images as input and operates in the spatial domain. The no-reference image quality assessment (NR-IQA) model is based on CNNs architecture to classify the images within the MOS classes.

TABLE 12. Comparative table of QoE prediction models for extended reality services.

Application	ML technique	Influencing factors	Assessment metrics	Prediction accuracy	Reference
3D video streaming	ANN with gradient decent	QP, content type, PLR, MBL	MOS, VQM	0.008 MSE, 0.92 R^2	[154]
AR still images overlay	Linear regression, bound linear regression, LR	EEG	MOS, DMOS, ACR-HR, BRISQUE	0.21 to 0.83 PCC, 0.33 to 1.36 MSE, 0.30 to 0.80 MAE, 0.26 to 0.71 MedAE	[155]
Stereoscopic videos	K-means clustering	Depth video histogram, SI & TI for the luminance component, depth pixels average time presence	PSNR, SAMVIQ, VQM, SSIM	95.4%	[156]
Stereoscopic videos	C3D, SVR	Automatically captured local spatiotemporal features	MOS, PSNR, SSIM	0.9478 to 0.9503 PLC, 0.9231 to 0.9426 SROCC, 0.7883 to 0.8038 KRCC, 0.3333 to 0.3514 RMSE	[157]
Stereoscopic videos	CNN	Spatiotemporal feature pooling strategy	MOS, PSNR, SSIM	0.9301 PCC, 0.9334 SROCC	[158]
Tele-immersive applications	FFNN	FPS, perceptual evaluation of speech quality, synchronization, interactivity	CMOS	Not specified	[159]
VR video streaming	DTR	Delay, packet loss, TCP throughput, tiling scheme, startup delay, quality level (bitrate), quality switches, stall time	K-fold cross-validation	Residual error ≤ 0.03922 for over 90% of the cases	[160]
VR 360-degree video	INN	Bandwidth, packet loss, latency, quality score, immersion score, non-spinning sensation score, global score	MOS	0.922 PCC	[161]
VR 360-degree video	C3D, LSTM	User perceived video quality, quality variation within viewport, quality variation across segments, miss ratio, rebuffering	BBA, BOLA, viewport only, viewport plus	~90%	[162]
VR 360-degree video	LR, ANN based on SGD	QP, Res, rendering device, gender, user's interest, user's familiarity with VR, perceptual quality, cybersickness	MOS, VRSP, VRSA, VCA	86%	[163]
VR 360-degree video	ANN based on SGD	Fast, medium & slow video, fixed, horizontal, & vertical camera motion, none, single, & multiple number of moving targets, cybersickness, perceptual quality, presence, stalling events	MOS, ACR, IPQ, SSQ, VRSP, VRSA, VCA	90%	[164]
VR 360-degree video	DT	Immersion & presence, acceptability, reality judgment, attention captivated	MOS	91 to 93%	[165]

In the model for 5G video streaming QoE prediction [153], the network data analytics function links network statistics with application measurements and the resulting QoE. The model is based on SL algorithms that utilize the statistical analysis of network-level characteristics to predict QoE in terms of MOS scale.

As we can observe in Table 11, the QoE models' implementations are almost equally divided between utilizing ANNs and ML algorithms. The prediction accuracy however is improved with the use of ANNs and specifically DNNs. Moreover, the majority of the models utilizes subjective metrics for the QoE evaluation and maps QoS characteristics in QOE values.

B. EXTENDED REALITY

In Table 12 it is depicted the comparative analysis of QoE prediction models for XR applications. The analysis includes applications focused on 3D video streaming, AR images,

stereoscopic videos, tele-immersiveness, VR video streaming and VR 360-degree video.

The model for 3D video streaming QoE prediction [154] uses the relations between QoS and QoE so as to map the parameters that lie in the network statistics and coding in MOS values. The predictive model evaluates QoE in a mobile 3D video streaming scenario relied on the development of an ANN with gradient descent optimization algorithm.

In the QoE prediction model for AR still images overlay [155], a comparison among user ratings, NR objective picture quality measurements and the human subject dry electrode electroencephalography (EEG) signals is introduced, in order to discover significant connections between QoS inputs and aggregated user ratings as MOS values with regard to spherical images. The predictive model is based on SL techniques.

As for the models for QoE prediction for stereoscopic videos [156]–[158], the stereoscopic video quality assessment (SVQA) evaluates the impact of spatiotemporal parameters using both subjective and objective metrics. In [156], the proposed model is built upon the K-means clustering algorithm, which employs customized content clustering via spatiotemporal activity within depth layers, based both on FR and NR metrics. In [157] a stereoscopic video quality assessment (SVQA) model is proposed, formed on 3D CNN and SVR. The model is designed to collect local spatiotemporal information in an automatic manner and take into account global temporal clues. In [158], a no reference SVQA technique is developed, relied on an end-to-end dual stream DNN (EDN). Since the stereoscopic videos contain left and right views, the EDN consists of two sub-networks comprising of two CNNs with the same set up and parameters shared between them. The EDN analyzes the perceptual quality for every image patch pair in the left and right pivotal frames of the stereoscopic video. Distortion-related and data-driven features are learnt end-to-end, by integrating multiple convolutions, max-pooling, and fully-connected layers with regression in the model's architecture. Next, a spatiotemporal pooling method is used on these image patch pairings to assess the overall stereoscopic video quality. As we can see in Table 12, although all the approaches achieve prediction accuracy $>90\%$, the highest value is put through the implementation of an UL solution [156], which subsequently uses discriminant analysis (DA) to predict opinion ratings for each cluster, utilizing video quality metrics such as PSNR.

In the case of QoE prediction for tele-immersive conference applications, the predictive model [159] embodies the link among 4-dimensional objective quality metrics and tele-immersive application QoE, stated with regard to CMOS values. The suggested model is based on a feed forward neural network (FFNN).

For the VR video streaming application, PERCEIVE [160] (a two-stage predictive approach) estimates the perceived quality of adaptive VR videos as they are streamed over mobile networks. The predictive model is based on SL DTR algorithm and provides an estimation of video playout performance by utilizing network QoS metrics as predictors. In a subsequent stage, it models and estimates the end-user perceived quality using the expected VR video playout performance metrics.

The QoE predictive models for VR 360-degree video applications [161]–[165], take into consideration the effect of various parameters on QoE, such as the network transmission characteristics, the physiological psychology and cognitive neurology features, the cybersickness, the degree of familiarity with VR and the level of interest in 360 degree video. In [161], a VR QoE prediction approach incorporating online, offline and mixed scenarios is proposed. Formed on network transmission characteristics, this framework creates a subjective assessment technique and an objective QoE evaluation model. It uses four dimensions for subjective assessment. In the objective assessment section, an improved two-step neural network (INN) algorithm is utilized by combining physiological psychology and cognitive neurology features. In VR transmission, this model reflects the inherent connection among the original input network characteristics and the resulting perception. In [162], the authors propose Mosaic, which is a new approach that mixes a neural network-based viewport estimation utilizing 3D CNN and LSTM, with a rate control system that dispenses rates to distinct tiles in the 360-degree frame. Mosaic models optimization as a multi-choice knapsack issue and solves it utilizing a greedy approach. Moreover, it creates an end-to-end testbed with standards compliant constituents and takes into account two cutting-edge algorithms for ordinary nontiled adaptive video streaming, the buffer-based algorithm (BBA) and buffer occupancy-based Lyapunov algorithm (BOLA), and two variants of tiled adaptive video streaming algorithms, i.e., the viewport only and viewport plus. In [163], a QoE prediction model for VR 360-degree videos is suggested, which utilizes the LR algorithm and an ANN based on the stochastic gradient descent (SGD) optimization algorithm. The suggested model takes into account two key aspects of QoE, the perceptual quality and cybersickness. Furthermore, it offers two additional QoE-influencing parameters for the QoE evaluation, the degree of familiarity with VR and the level of interest in 360-degree video. In terms of cybersickness, the QoE prediction model uses performance comparing methods including VRSP, VRSA and VCA. In [164], a QoE prediction method based on ANN optimized with SGD is developed, which may estimate the level of cybersickness impacted by 360-degree videos under the influence of several stalling events in VR applications. The cybersickness level is evaluating and predicted with the use of metrics such as SSQ, VRSP, VRSA, and VCA, and the user's sense of presence is evaluated with the igroup presence questionnaire (IPQ). In [165], a QoE estimation method relied on DT algorithm is presented, which subjectively explore the influence of QoE-affecting factors in VR 360-degree videos, such as quantization parameters (QP), resolutions, initial delay, and different interruptions. The model predicts the four most significant VR QoE factors which include immersion, acceptability,

reality judgment, and attention captivated, based on subjective data. As we can see in Table 12, the evaluation methods rely both on subjective and objective metrics, and the implementations are based on ANNs as well as SL algorithms. The highest accuracy is achieved with use of the DT algorithm and the evaluation of four significant VR QoE factors that include immersion, acceptability, reality judgment and attention captivated.

As we can see in Table 12, the majority of the QoE prediction models for extended reality applications utilize ANNs and in particular DNNs solutions. These implementations achieve better prediction accuracy values compared with ML algorithms, with the exception of the utilization of an UL solution for stereoscopic video applications. Moreover, the QoE assessment is based on both subjective and objective metrics, as well as on metrics for the evaluation of simulator sickness that impacts VR applications.

C. VIDEO GAMING

The Table 13 contains the comparative analysis of QoE predictive models for video gaming applications. The analysis is centered on 3D media platform, computer-generated imagery, gaming video streaming and massively multiplayer online role-playing games (MMORPGs) applications.

The QoE prediction model [166] for the 3D media platform for interactive multiplayer video games applications simulates a tele-immersive interactive multiplayer video game. The monitoring parameters include Prometheus [177] derived metrics, application-level metrics and MOS values calculated based on the frame rate and PSNR. The model implements a cognitive network optimizer (CNO) formulated as an RL agent, based on a set of actual monitoring factors such as infrastructure, application-level, and QoE metrics.

For the computer-generated imagery for gaming video streaming services applications, the QoE prediction model [167] takes into account the effect of the unique features of gaming video content when compared to conventional video services, including ultra-high motion, specific motion patterns, synthetic and repetitive content. The QoE evaluation is based on the FR VMAF metric as ground truth and the model's implementation relies on the utilization of a CNN.

In the case of gaming video streaming applications, the QoE predictive models [168]–[175] investigate the impact of IFs that lie on frame-level such as blur, naturalness, blockiness and complexity, as well as the effect of spatiotemporal features and psychometric parameters. In [168], NR-GVQM, a no reference gaming video quality measurement based on the SVR algorithm is developed. SVR's training exploits nine frame-level indexes as input features and VMAF scores as the ground truth. NR- GVQM offers low complexity as it utilizes characteristics that can be obtained exclusively in real-time. In [169], two no reference ML-based lightweight QoE prediction methods for gaming video streaming are proposed, i.e., the NR-GVSQI quality index and the NR-GVSQE quality estimator. The models' design is formed on SVR, Gaussian process regression (GPR), ANN, and RF. Because of their low intricacy, both models may be utilized as the first stage of a real-time optimized online gaming QoE management framework, even on thin clients. In [170], an approach to increase the video quality on compressed gaming content is proposed, based on super-resolution generative adversarial networks (SRGAN), which employs a DNN in conjunction with an adversarial network to generate better resolution images. The suggested approach includes a modified loss function together with changes in the generator network, like layer levels and skip connections, to enhance the flow of information in the network, which was proven to considerably increase perceived quality. In [171], ''nofu'', a no reference lightweight video quality module for gaming content is developed, based on the RFR algorithm. The suggested method predicts video quality scores using only the recorded video, and focuses on features that are easy to calculate. Moreover, it employs as few features as feasible in order to create a model capable of making real-time QoE predictions. In [172], a method to create a CNN-based quality metric to evaluate the quality of gaming video is proposed. The CNN is trained using the objective quality model VMAF as ground truth, and fine-tuned using subjective picture quality evaluations. What is more, a new temporal pooling approach based on frame-level predictions is presented to predict gaming video quality. In [173], a real-time reduced reference gaming video quality evaluation methodology is proposed. The methodology is formed on low-complexity psychometric curve-fitting approach. The ML techniques that the model utilizes include DTR and ANNs. The suggested solution chooses the most relevant objective features with the least amount of complexity. Following that, the link between these features and the ground-truth quality is modeled using HVS psychometric perception. In [174], DEMI is presented. This is a QoE estimation model based on CNN and RF that considers both gaming and non-gaming videos. In this model, the CNN in the is trained using an objective metric, allowing the CNN to learn video artifacts like blurriness and blockiness. Following that, the model is fine-tuned using blockiness and blurriness scores from a small image quality dataset. Finally, to estimate video quality, an RF is utilized for pool frame-level estimations and temporal information of videos. The model's low complexity makes it suitable for real-time applications. In [175], ERAQUE is developed, an efficient hard-rank quality estimator for gaming video streaming based on CNN. The estimation model includes a hard pairwise ranking loss, which allows the model to focus more on distinguishing alike pairs, as well as an effective adapted model distillation, which incurs insubstantial performance loss. For the QoE assessment of the aforementioned models, a combination of subjective and objective metrics is utilized. As we can observe in the Table 13, the majority of the approaches use ANNs implementations and achieve high levels of prediction accuracy, but the highest accuracy value was put over through the use of VMAF scores as ground truth and the implementation

TABLE 13. Comparative table of QoE prediction models for video gaming services.

Application	ML technique	Influencing factors	Assessment metrics	Prediction accuracy	Reference
3D media platform for interactive multiplayer video games	RL	Transmitted network packet loss, received network packet loss, bitrate, bitrate (aggregated), FPS, FPS (aggregated), consumed profile, number of produced profiles, output data bytes, working frames per second, theoretic load percentage	MOS, PSNR	Not specified	[166]
Computer-generated imagery for gaming video streaming services	CNN	DLM, mean co-located pixel difference, ANSNR	VMAF, VIF	3.11 to 7.50 RMSE, 0.937 to 0.987 SROCC	[167]
Gaming video streaming	SVR	SI, TI, noise, bluriness, Blockiness, contrast	MOS, VMAF, BRISQUE, NIQE, PIQE, PSNR, SSIM, STRREDOpt, SpEED-QA	0.89 to 0.98 PCC	[168]
Gaming video streaming	SVR, gaussian process regression, ANN, RF	SI, TI, Res, bitrate, blockiness, blockloss, blur, contrast, exposure, flickering, interlacing, noise, slicing, spatial activity, temporal activity	VMAF, MOS, BRISQUE, BIQI, NIQE, SpEED-QA, STRRED	0.905 PCC, 0.913 SROCC	[169]
Gaming video streaming	SRGAN	Blurriness, blockiness	MOS, ACR, VMAF, PIQE, NIQE	0.64 to 0.77 PCC	[170]
Gaming video streaming	RFR	SI, TI, staticness, blockiness, blockmotion, blurriness, type of motion	VMAF, BRISQUE, NIQE, SSIM. PSNR, SpEED-QA, STRRED	0.91 to 0.96 PCC, 0.75 to 0.82 KRCC, 0.91 to 0.95 SROCC, 0.22 to 0.42 RMSE	[171]
Gaming video streaming	CNN	Video fragmentation, video unclearness, temporal complexity, TI,	MOS, DMOS, ACR, PSNR, SSIM, VMAF, SpEED-QA, STRREDOpt, PIQE, BRISQUE, NIQE	0.968 SROCC, 0.30 RMSE	[172]
Gaming video streaming	DTR, ANN	SI, TI, SC, level of motion, blurriness, noise blockiness, jerkiness, motion,	MOS, ACR, PSNR, SSIM, VQM, VMAF, SpEED-QA	0.953 PCC, 0.004 MSE	[173]
Gaming video streaming	CNN, RF	Fragmentation (Blockiness), Unclearness (Blurriness)	VMAF, PSNR, SSIM, BRISQUE, NIQE, PIQE, ACR, MOS	0.93 PCC, 0.92 SROCC	[174]
Gaming video streaming	CNN	Bitrate, Res, FPS, duration	ACR, MOS, VMAF, PSNR, SSIM, MS-SSIM	0.964 PCC, 0.964 SROCC, 0.843 KRCC, 2.638 RMSE	[175]
MMORPGs	Linear regression, partial least squares, ridge regression, SVR with linear kernel & radial basis function kernel, RF, gradient boosting machine	Delay, packet loss, jerkiness, FPS, gender, age, experience, social context, action categories	MOS, ACR	0.62 to 0.86 RMSE	[176]

of the SVR algorithm [168], the training of which relies on the Gaussian kernel.

The model for QoE prediction centering on MMORPGs [176], examines system, user, and context elements and assesses their influence on QoE. It also addresses certain methodological issues linked to assessing gaming QoE and delves deeper into a collection of quality metrics beyond MOS, like the percentages of users rating the gameplay scenario, and acceptance measures. The model is developed with use of a series of SL techniques.

As we can observe in Table 13, the implementations for the QoE prediction in the case of video gaming applications rely both on ANNs and ML methods, with the models of each category achieving equally high levels of prediction accuracy.

The QoE assessment however, is based on utilization of FR, RR and NR objective metrics rather than mappings of QoS parameters in QoE values, unlike the practice that is the norm in conventional video content.

VII. CONCLUSION

QoE has received a lot of research interest in the last years, and has been acknowledged as an important factor in determining network operating efficiency. Understanding measuring, and modeling QoE for a variety of multimedia services has gained significance, and CSPs have made considerable efforts in providing dependable services with better personalized end-user experience. In this regard, the first stage in optimizing a mobile multimedia streaming service delivery is evaluating and predicting the end-user's QoE, which helps in acquiring a better understanding of how the technical aspects of a network impact multimedia service quality as experienced by end-users. Nevertheless, QoS metrics are not immediately and clearly connected to an end-user's gratification and perceived experience, thus user-centric KQIs metrics have been developed to assess quality. Understanding and identifying a range of subjective and objective influencing factors for KQIs, which may be categorized as human-related, system-related, context-related, and content-related, is fundamental for appropriate QoE management. The parameters that are necessary to be monitored and assessed are defined in the QoE models, with the objective of implementing efficient QoE optimization approaches, capable of efficiently addressing QoE management issues. The quality assessment includes two sorts of approaches, the subjective and objective assessment. Subjective assessment techniques rely on human assessors, whereas objective techniques are regarded as a way for measuring subjective quality based solely on objective quality metrics. Due to the exceptionally growing number of factors involved, QoE assessment has become an increasingly complicated issue, hence a variety of ML solutions have been proposed in the last years in order to tackle this problem. Since ML increases the accuracy of QoE models, assists in QoE monitoring, and provides the methodological basis for measuring the relationship between QoS and QoE, the research community has adopted ML-based approaches to achieve real-time, precise, and adaptable QoE management frameworks.

In this survey, we define QoE within the context of multimedia services, and provide a spherical analysis of the QoE IFs. Moreover, we gather and analyze the more significant quality metrics, both subjective and objective, as well as the methods for evaluating their performance, and the mathematical models for correlating QoS parameters with QoE. In addition, we look at the specific QoE features in extended reality and video gaming applications, and highlight the distinction between these emerging technologies and the conventional video streaming services. We also provide a comprehensive analysis of the quality influencing factors for both extended reality and video gaming applications. Furthermore, we underline the importance of ML in the development of efficient QoE predictive models, describe its techniques, and analyzed the more significant algorithms concerning SL, UL, RL, and ANNs. Finally, we examine state-of-the-art ML-based QoE prediction models for video streaming, extended reality, and video gaming applications. Since QoE is a concept that has been introduced in recent years in mobile and wireless networks design, especially in the case of the emerging applications of extended reality and video gaming, this survey focuses on the latest research outcomes, and brings together the main publications of recent years, with the majority of which concentrating in the last three years. Although choosing the best-suited ML model for a certain sort of application is still an open research question, the trend of recent years, enhanced by the vast amount of available data and high computational capabilities, tends towards solutions implemented by using ANNs, and more specifically DNNs. The major contributions of this survey can be found in the following two points: 1) to the best of the authors' knowledge, this is the first endeavor to present a complete hands-on guide on multimedia services QoE assessment, which, contrary to existing surveys, includes extended reality and video gaming applications in addition to conventional video streaming; and 2) up to this date, this is the first survey to provide a comparative study of ML-based QoE prediction models that particularly focus on extended reality and video gaming applications.

APPENDIX
LIST OF THE MAIN ACRONYMS AND ABBREVIATIONS

3D	Three dimensions
5G	Fifth generation
ACR	Absolute category rating
ACR-HR	Absolute category rating-hidden reference
AI	Artificial intelligence
ANN	Artificial Neural Network
ANSNR	Anti-noise signal-to-noise ratio
AR	Augmented reality
B5G	Beyond 5G
BLSTM	Bidirectional long short-term memory
BRISQUE	Blind/referenceless image spatial quality evaluator
BSVR	Boosting support vector regression
C3D	3D convolutional neural network
CDN	Content distribution network
CI	Color information
CINR	Carrier to interference plus noise ratio
CNN	Convolutional neural network
CQI	Channel quality indicator
CSP	Communication service provider
DASH	Dynamic adaptive streaming over HTTP
DCR	Degradation category rating
DL	Deep learning
DLM	Detail loss measure
DMOS	Differential mean opinion score
DNN	Deep neural network

DSCQS	Double stimulus continuous quality scale
DT	Decision tree
DTR	Decision trees regression
EEG	Electroencephalography
FFNN	Feed-forward neural network
FoV	Field of view
FPS	Frame rate
FR	Full reference
GAN	Generative adversarial network
GoP	Group of pictures
GSM	Gaussian scale mixture
HAS	HTTP adaptive streaming
HMD	Head mounted display
HVS	Human visual system
IF	Influencing factor
IQA	Image quality assessment
KNN	K-nearest neighbor
KPI	Key performance indicator
KQI	Key quality indicator
KRCC	Kendall rank correlation coefficient
LR	Linear regression
LSTM	Long short-term memory
MAE	Mean absolute error
MBL	Mean burst length
MCS	Modulation coding scheme
ML	Machine learning
MMORPGs	Massively multiplayer online role-playing games
MNO	Mobile network operator
MOS	Mean opinion score
MOSp	Predicted mean opinion score
MOVIE	Motion-based video integrity evaluation
MPQM	Moving picture quality measure
MR	Mixed reality
MSE	Mean squared error
MS-SSIM	Multiscale-SSIM
NIQE	Natural image quality evaluator
NR	No reference
NR-B	No reference bitstream
NR-P	Pixel-based no reference
OR	Outlier ratio
OTT	Over-the-top
PCC	Pearson correlation coefficient
PI	Playback indicator
PIQE	Psychovisually-based image quality evaluator
PLR	Packet loss ratio
PSNR	Peak signal to noise ratio
QoE	Quality of experience
QoS	Quality of service
QP	Quantization parameter
RBFN	Radial function network
RBM	Restricted Boltzmann machine
Res	Resolution
RF	Random forest
RFR	Random forest regression
RL	Reinforcement learning
RMSE	Root mean square error
RNN	Recurrent neural network
RR	Reduced reference
RSRP	Reference signal received power
RSRQ	Reference signal received quality
RSSI	Received signal strength indicator
RTR	Rebuffering time ratio
RTT	Round-trip time
SAMVIQ	Subjective assessment methodology for video quality
SC	Scene complexity
SDSCE	Simultaneous double stimulus for continuous evaluation
SGD	Stochastic gradient descent
SI	Spatial information
SL	Supervised learning
SpEED-QA	Spatial efficient entropic differencing for quality assessment
SRGAN	Super-resolution generative adversarial network
SROCC	Spearman rank order correlation coefficient
SSIM	Structural similarity index
SSSP	Secondary synchronization signal power
SSQ	Simulator sickness questionnaire
STRRED	Spatio-temporal reduced-reference entropic differencing
STSQ	Short time subjective quality
SVM	Support vector machine
SVQA	Stereoscopic video quality assessment
SVR	Support vector regression
TCN	Temporal convolutional network
TI	Temporal information
UHD	Ultra-high definition
UL	Unsupervised learning
VCA	Visual comfort assessment
VIF	Visual information fidelity
VMAF	Video multimethod assessment fusion
VQA	Video quality assessment
VQM	Video quality metric
VR	Virtual reality
VRSA	Virtual reality sickness assessment
VRSP	Virtual reality sickness predictor
XR	Extended reality

REFERENCES

[1] *IMT Traffic Estimates for the Years 2020 to 2030*, document Rep. ITU-R M.2370-0, ITU-R, Jul. 2015.

[2] *Ericsson Mobility Report*, Ericsson, Stockholm, Sweden, Nov. 2020.

[3] *Cisco Visual Networking Index: Forecast and Trends, 2017–2022*, Cisco, San Jose, CA, USA, 2018.

[4] A. Ahmad, A. Floris, and L. Atzori, "QoE-aware service delivery: A joint-venture approach for content and network providers," in *Proc. 8th Int. Conf. Quality Multimedia Exper. (QoMEX)*, Jun. 2016, pp. 1–6.

[5] K. Bouraqia, E. Sabir, M. Sadik, and L. Ladid, "Quality of experience for streaming services: Measurements, challenges and insights," *IEEE Access*, vol. 8, pp. 13341–13361, 2020.

[6] V. Menkovski, A. Oredope, A. Liotta, and A. C. Sánchez, "Predicting quality of experience in multimedia streaming," in *Proc. 7th Int. Conf. Adv. Mobile Comput. Multimedia (MoMM)*, 2009, pp. 52–59.
[7] A. A. Barakabitze, N. Barman, A. Ahmad, S. Zadtootaghaj, L. Sun, M. G. Martini, and L. Atzori, "QoE management of multimedia streaming services in future networks: A tutorial and survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 1, pp. 526–565, 1st Quart., 2020.
[8] M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hoßfeld, and P. Tran-Gia, "A survey on quality of experience of HTTP adaptive streaming," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 469–492, 1st Quart., 2014.
[9] S. Baraković and L. Skorin-Kapov, "Survey and challenges of QoE management issues in wireless networks," *J. Comput. Netw. Commun.*, vol. 2013, pp. 1–28, Mar. 2013.
[10] S. Baraković, J. Baraković, and H. Bajrić, "QoE dimensions and QoE measurement of NGN services," in *Proc. 18th TELFOR*, Belgrade, Serbia, Nov. 2010, pp. 1–4.
[11] Y. Liu, J. Liu, A. Argyriou, and S. Ci, "3DQoE-oriented and energy-efficient 2D plus depth based 3D video streaming over centrally controlled networks," *IEEE Trans. Multimedia*, vol. 20, no. 9, pp. 2439–2453, Sep. 2018.
[12] Y. Liu, S. Ci, H. Tang, Y. Ye, and J. Liu, "QoE-oriented 3D video transcoding for mobile streaming," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 8, no. 3s, pp. 1–20, Sep. 2012.
[13] R. Stankiewicz and A. Jajszczyk, "A survey of QoE assurance in converged networks," *Comput. Netw.*, vol. 55, pp. 1459–1473, May 2011.
[14] M. G. Martini, C. W. Chen, Z. Chen, T. Dagiuklas, L. Sun, and X. Zhu, "Guest editorial QoE-aware wireless multimedia systems," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 7, pp. 1153–1156, Aug. 2012.
[15] N. Barman and M. G. Martini, "QoE modeling for HTTP adaptive video streaming—A survey and open challenges," *IEEE Access*, vol. 7, pp. 30831–30859, 2019.
[16] T. Hobfeld, R. Schatz, M. Varela, and C. Timmerer, "Challenges of QoE management for cloud applications," *IEEE Commun. Mag.*, vol. 50, no. 4, pp. 28–36, Apr. 2012.
[17] K. Brunnström, S. A. Beker, K. D. Moor, A. Dooms, S. Egger, M. N. Garcia, T. Hossfeld, S. Jumisko-Pyykkö, C. Keimel, M. C. Larabi, and B. Lawlor, "Qualinet white paper on definitions of quality of experience," in *Proc. Novi Sad, Serbia, 5th Qualinet Meeting*, Mar. 2013, pp. 1–24.
[18] R. Huang, X. Wei, L. Zhou, C. Lv, H. Meng, and J. Jin, "A survey of data-driven approach on multimedia QoE evaluation," *Frontiers Comput. Sci.*, vol. 12, no. 6, pp. 1060–1075, Aug. 2018.
[19] Y. Chen, K. Wu, and Q. Zhang, "From QoS to QoE: A tutorial on video quality assessment," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 1126–1165, 2nd Quart., 2015.
[20] M. T. Vega, C. Perra, F. De Turck, and A. Liotta, "A review of predictive quality of experience management in video streaming services," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 432–445, Jun. 2018.
[21] M. Claeys, S. Latré, J. Famaey, T. Wu, W. Van Leekwijck, and F. D. Turck, "Design and optimisation of a (FA)Q-learning-based HTTP adaptive streaming client," *Connection Sci.*, vol. 26, no. 1, pp. 25–43, Jan. 2014.
[22] S. Aroussi and A. Mellouk, "Survey on machine learning-based QoE-QoS correlation models," in *Proc. Int. Conf. ComManTel*, Da Nang, Vietnam, Apr. 2014, pp. 200–204.
[23] L. Skorin-Kapov, M. Varela, T. Hoßfeld, and K.-T. Chen, "A survey of emerging concepts and challenges for QoE management of multimedia services," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 14, no. 2s, pp. 1–29, Apr. 2018.
[24] M. Alreshoodi and J. Woods, "Survey on QoE-QoS correlation models formultimedia services," *Int. J. Distrib. Parallel Syst.*, vol. 4, no. 3, pp. 53–72, May 2013.
[25] *Vocabulary for Performance, Quality of Service and Quality of Experience*, document Rec. ITU-T P.10/G.100, ITU-T, Nov. 2017.
[26] P. Juluri, V. Tamarapalli, and D. Medhi, "Measurement of quality of experience of video-on-demand services: A survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 401–418, 1st Quart., 2016.
[27] I. Sousa, M. P. Queluz, and A. Rodrigues, "A survey on QoE-oriented wireless resources scheduling," *J. Netw. Comput. Appl.*, vol. 158, May 2020, Art. no. 102594.
[28] Y. Wang, P. Li, L. Jiao, Z. Su, N. Cheng, X. S. Shen, and P. Zhang, "A data-driven architecture for personalized QoE management in 5G wireless networks," *IEEE Wireless Commun.*, vol. 24, no. 1, pp. 102–110, Feb. 2017.
[29] 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.
[30] *Mean Opinion Score (MOS) Terminology*, document Rec. ITU-T P.800.1, ITU-T, Jul. 2006.
[31] *Human Factors (HF); Quality of Experience (QoE) Requirements for Real-Time Communication Services*, Standard ETSI TR 102 643 (V1.0.1), ETSI, Tech. Rep., Dec. 2009.
[32] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, document Rec. ITU-R BT.500-13, ITU-R, Jan. 2012.
[33] *Subjective Video Quality Assessment Methods for Multimedia Applications*, document Rec. ITU-T P.910, ITU-R, Apr. 2008.
[34] *Methods for Subjective Determination of Transmission Quality*, document Rec. ITU-T P.800, ITU-T, Aug. 1996.
[35] *Methods for the Subjective Assessment of Video Quality, Audio Quality and Audiovisual Quality of Internet Video and Distribution Quality Television in Any Environment*, document Rec. ITU-T P.913, ITU-T, Mar. 2016.
[36] K.-T. Chen, C.-C. Wu, Y.-C. Chang, and C.-L. Lei, "A crowdsourceable QoE evaluation framework for multimedia content," in *Proc. 17th ACM Int. Conf. Multimedia (MM)*, 2009, pp. 491–500.
[37] T. Hossfeld, C. Keimel, M. Hirth, B. Gardlo, J. Habigt, K. Diepold, and P. Tran-Gia, "Best practices for QoE crowdtesting: QoE assessment with crowdsourcing," *IEEE Trans. Multimedia*, vol. 16, no. 2, pp. 541–558, Feb. 2014.
[38] M. Yang, S. Wang, R. N. Calheiros, and F. Yang, "Survey on QoE assessment approach for network service," *IEEE Access*, vol. 6, pp. 48374–48390, 2018.
[39] K.-T. Chen, C.-J. Chang, C.-C. Wu, Y.-C. Chang, and C.-L. Lei, "Quadrant of euphoria: A crowdsourcing platform for QoE assessment," *IEEE Netw.*, vol. 24, no. 2, pp. 28–35, Mar. 2010.
[40] F. Ribeiro, D. Florencio, C. Zhang, and M. Seltzer, "CROWDMOS: An approach for crowdsourcing mean opinion score studies," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2011, pp. 2416–2419.
[41] C. Keimel, J. Habigt, C. Horch, and K. Diepold, "QualityCrowd—A framework for crowd-based quality evaluation," in *Proc. Picture Coding Symp.*, Krakow, Poland, May 2012, pp. 245–248.
[42] A. Takahash, "Framework and standardization of quality of experience (QoE) design and management for audiovisual communication services," NTT, Tokyo, Japan, Tech. Rep., 2009, pp. 1–5, vol. 7, no. 4.
[43] Y. Wang and P. Zhang, *QoE Management in Wireless Networks*. New York, NY, USA: Springer, 2016.
[44] M. Fiedler, T. Hossfeld, and P. Tran-Gia, "A generic quantitative relationship between quality of experience and quality of service," *IEEE Netw.*, vol. 24, no. 2, pp. 36–41, Mar. 2010.
[45] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Process., Image Commun.*, vol. 19, no. 2, pp. 121–132, 2004.
[46] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Trans. Broadcast.*, vol. 50, no. 3, pp. 312–322, Sep. 2004.
[47] R. Serral-Graciá, E. Cerqueira, M. Curado, M. Yannuzzi, E. Monteiro, and X. Masip-Bruin, "An overview of quality of experience measurement challenges for video applications in IP networks," in *Proc. 8th Int. Conf. WWIC*, Jun. 2010, pp. 252–263.
[48] W. Song, D. W. Tjondronegoro, and M. J. Docherty, "Understanding user experience of mobile video: Framework, measurement, and optimization," in *Mobile Multimedia–User and Technology Perspectives*. Rijeka, Croatia: InTech, 2012, pp. 3–30.
[49] *Reference Guide to Quality of Experience Assessment Methodologies*, document Rec. ITU-T G.1011, ITU-T, Jul. 2016.
[50] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, "Objective video quality assessment methods: A classification, review, and performance comparison," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 165–182, Jun. 2011.

[51] A. Takahashi, D. Hands, and V. Barriac, "Standardization activities in the ITU for a QoE assessment of IPTV," *IEEE Commun. Mag.*, vol. 46, no. 2, pp. 78–84, Feb. 2008.

[52] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video," *Electron. Lett.*, vol. 44, no. 13, pp. 800–801, Jun. 2008.

[53] K. Piamrat, C. Viho, J.-M. Bonnin, and A. Ksentini, "Quality of experience measurements for video streaming over wireless networks," in *Proc. 6th Int. Conf. Inf. Technol., New Generat.*, Apr. 2009, pp. 1184–1189.

[54] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 2366–2369.

[55] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[56] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2003, pp. 1398–1402.

[57] C. G. Bampis and A. C. Bovik, "Learning to predict streaming video QoE: Distortions, rebuffering and Memory," Mar. 2017, *arXiv:1703.00633*.

[58] C. J. Van den Branden Lambrecht and O. Verscheure, "Perceptual quality measure using a spatiotemporal model of the human visual system," in *Proc. Electron. Imag., Sci. Technol.*, San Jose, CA, USA, Mar. 1996, pp. 450–461.

[59] M. Wichtlhuber, G. Wicklein, S. Wilk, W. Effelsberg, and D. Hausheer, "RT-VQM: Real-time video quality assessment for adaptive video streaming using GPUs," in *Proc. 7th Int. Conf. Multimedia Syst.*, May 2016, pp. 1–11.

[60] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.

[61] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2284–2298, Sep. 2007.

[62] K. Seshadrinathan and A. C. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 335–350, Feb. 2009.

[63] A. Rehman, K. Zeng, and Z. Wang, "Display device-adapted video quality-of-experience assessment," *Proc. SPIE*, vol. 9394, Mar. 2015, Art. no. 939406.

[64] A. Aaron, Z. Li, M. Manohara, J. Y. Lin, E. C.-H. Wu, and C.-C.-J. Kuo, "Challenges in cloud based ingest and encoding for high quality streaming media," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 1732–1736.

[65] R. Soundararajan and A. C. Bovik, "Video quality assessment by reduced reference spatio-temporal entropic differencing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 4, pp. 684–694, Apr. 2012.

[66] C. G. Bampis, P. Gupta, R. Soundararajan, and A. C. Bovik, "SpEED-QA: Spatial efficient entropic differencing for image and video quality," *IEEE Signal Process. Lett.*, vol. 24, no. 9, pp. 1333–1337, Sep. 2017.

[67] A. R. Reibman, S. Sen, and J. Van der Merwe, "Analysing the spatial quality of internet streaming," in *Proc. VPQM*, Scottsdale, AZ, USA, Jan. 2005.

[68] T. Yamada, Y. Miyamoto, and M. Serizawa, "No-reference video quality estimation based on error-concealment effectiveness," in *Proc. Packet Video*, Nov. 2007, pp. 288–293.

[69] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.

[70] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a completely Blind' image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, Nov. 2012.

[71] R. W. Chan and P. B. Goldsmith, "A psychovisually-based image quality evaluator for JPEG images," in *Proc. SMC Conf. IEEE Int. Conf. Syst., Man Cybern. Cybern. Evolving Syst., Hum., Organizations, Their Complex Interact.*, Oct. 2000, pp. 1541–1546.

[72] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Climate Res.*, vol. 30, no. 1, pp. 79–82, Dec. 2005.

[73] C. Sammut and G. I. Webb, *Encyclopedia of Machine Learning*. Boston, MA, USA: Springer, 2011.

[74] R. Bonnin, *Machine Learning for Developers*, Birmingham, U.K.: Packt Publishing, Oct. 2017.

[75] E. Szmidt and J. Kacprzyk, "The Spearman and Kendall rank correlation coefficients between intuitionistic fuzzy sets," in *Proc. EUSFLAT*, Aix-Les-Bains, France, Jul. 2011, pp. 521–528.

[76] J. M. Wooldridge, "A note on computing r-squared and adjusted r-squared for trending and seasonal data," *Econ. Lett.*, vol. 36, no. 1, pp. 49–54, May 1991.

[77] S. Aroussi, T. Bouabana-Tebibel, and A. Mellouk, "Empirical QoE/QoS correlation model based on multiple parameters for VoD flows," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2012, pp. 1963–1968.

[78] *Definition of Connectivity and QoE/QoS Management Mechanisms—Intermediate Report*, document G NORMA, Deliverable D5.1, Nov. 2016.

[79] 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.*, Jul. 2012, pp. 57–62.

[80] Y. Liu, J. Liu, Z. Xu, and S. Ci, "Choquet integral based QoS-to-QoE mapping for mobile VoD applications," in *Proc. IEEE/ACM 24th Int. Symp. Quality Service (IWQoS)*, Jun. 2016, pp. 1–6.

[81] T. Hoßfeld, P. Tran-Gia, and M. Fiedler, "Quantification of quality of experience for edge-based applications," in *Proc. ITC*, Jan. 2007, pp. 361–373.

[82] 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.*, May 2010, pp. 1–5.

[83] P. Reichl, B. Tuffin, and R. Schatz, "Logarithmic laws in service quality perception: Where microeconomics meets psychophysics and quality of experience," *Telecommun. Syst.*, vol. 52, no. 2, pp. 587–600, 2013.

[84] *5G; Extended Reality (XR) in 5G*, Standard ETSI TR 126 928 (V16.0.0), ETSI, Tech. Rep., Nov. 2020.

[85] J. P. Rolland, R. L. Holloway, and H. Fuchs, "Comparison of optical and video see-through, head-mounted displays," *Telemanipulator Telepresence Technol.*, vol. 2351, pp. 293–307, Dec. 1995.

[86] D. Nunez and E. Blake, "Cognitive presence as a unified concept of virtual reality effectiveness," in *Proc. 1st Int. Conf. Comput. Graph., Virtual Reality Visualisation (AFRIGRAPH)*, 2001, pp. 115–118.

[87] T. C. Ching. (Aug. 2016). *The Concept of Presence in Virtual Reality*. [Online]. Available: https://choongchingteo.medium.com/the-concept-of-presence-in-virtual-reality-6d4332dc1a9c

[88] Oculus. (Sep. 2014). *Oculus Shares 5 Key Ingredients for Presence in Virtual Reality*. [Online]. Available: https://www.roadtovr.com/oculus-shares-5-key-ingredients-for-presence-in-virtual-reality/

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

[90] 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.

[91] A. Kopyt and J. Narkiewicz, "Technical factors influencing simulator sickness," *Sci. Lett. Rzeszow Univ. Technol. Mech.*, vol. 30, no. 85, pp. 455–467, 2013.

[92] A. Woods, "Understanding crosstalk in stereoscopic displays," in *Proc. 3DSA*, Tokyo, Japan, May 2010, pp. 19–21.

[93] A. Singla, S. Göring, A. Raake, B. Meixner, R. Koenen, and T. Buchholz, "Subjective quality evaluation of tile-based streaming for omnidirectional videos," in *Proc. MMSys*, New York, NY, USA, Jun. 2019, pp. 232–242.

[94] D. P. E. Kuzyakov. *Next-Generation Video Encoding Techniques for 360 Video and VR*. Accessed: Sep. 17, 2021. [Online]. Available: https://code.facebook.com/posts/1126354007399553/

[95] F. Hofmeyer and S. Fremerey, "Impacts of internal HMD playback processing on subjective quality perception," in *Proc. IS&T*, Burlingame, CA, USA, Jan. 2019, pp. 1–6.

[96] S. Fremerey, F. Hofmeyer, S. Göring, and A. Raake, "Impact of various motion interpolation algorithms on 360° video QoE," in *Proc. QoMEX*, Berlin, Germany, Jul. 2019, pp. 1–3.

[97] *Requirements for Operational Monitoring of Video-to-Audio Delay in the Distribution of Television Programs*, document Rec. ITU-T J.248, ITU-T, Jun. 2008.

[98] 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)*, May 2017, pp. 1–6.

[99] L. Zhang, H. Dong, and A. E. Saddik, "Towards a QoE model to evaluate holographic augmented reality devices," *IEEE Multimedia*, vol. 26, no. 2, pp. 21–32, Apr. 2019.

[100] VR Lens Lab. (2016). *Field of View for Virtual Reality Headsets Explained*. [Online]. Available: https://vr-lens-lab.com/field-of-view-for-virtual-reality-headsets/

[101] S. Moller, S. Schmidt, and S. Zadtootaghaj, "New ITU-T standards for gaming QoE evaluation and management," in *Proc. 10th Int. Conf. Quality Multimedia Exper. (QoMEX)*, May 2018, pp. 1–6.

[102] J. Juul, *Half-Real: Video Games Between Real Rules and*. Cambridge, MA, USA: The MIT Press, 2005.

[103] S. Müller, S. Schmidt, and J. Beyer, "Gaming taxonomy: An overview of concepts and evaluation methods for computer gaming QoE," in *Proc. QoMEX*, Klagenfurt am Wörthersee, Austria, Jul. 2013, pp. 236–241.

[104] *Subjective Evaluation Methods for Gaming Quality*, document Rec. ITU-T P.809, ITU-T, Jun. 2018.

[105] E. Brown and P. Cairns, "A grounded investigation of game immersion," in *Proc. CHI*, Vienna, Austria, Apr. 2004, pp. 1297–1300.

[106] M. Hassenzahl, "User experience (UX): Towards an experiential perspective on product quality," in *Proc. IHM*, Metz, France, Sep. 2008, pp. 11–15.

[107] R. Agarwal and E. Karahanna, "Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage," *MIS Quart.*, vol. 24, no. 4, pp. 665–694, Dec. 2000.

[108] *Influence Factors on Gaming Quality of Experience*, document Rec. ITU-T G.1032, ITU-T, Oct. 2017.

[109] C. M. Frederick and R. M. Ryan, "Differences in motivation for sport and exercise and their relations with participation and mental health," *J. Sport Behav.*, vol. 16, no. 3, Sep. 1993.

[110] J. Davis, Y.-H. Hsieh, and H.-C. Lee, "Humans perceive flicker artifacts at 500 Hz," *Sci. Rep.*, vol. 5, no. 1, pp. 1–4, Feb. 2015.

[111] M. Claypool and K. Claypool, "Latency and player actions in online games," *Commun. ACM*, vol. 49, no. 11, pp. 40–45, 2006.

[112] M. Claypool and K. Claypool, "Perspectives, frame rates and resolutions: It's all in the game," in *Proc. 4th Int. Conf. Found. Digit. Games (FDG)*, 2009, pp. 42–49.

[113] R. Hunicke, M. LeBlanc, and R. Zubek, "MDA: A formal approach to game design and game research," in *Proc. AAAI*, San Hose, CA, USA, Jul. 2004, p. 1722.

[114] J. Beyer, V. Miruchna, and S. Moller, "Assessing the impact of display size, game type, and usage context on mobile gaming QOE," in *Proc. 6th Int. Workshop Quality Multimedia Exper. (QoMEX)*, Sep. 2014, pp. 69–70.

[115] I. Hupont, J. Gracia, L. Sanagustín, and M. A. Gracia, "How do new visual immersive systems influence gaming QoE? A use case of serious gaming with Oculus Rift," in *Proc. QoMEX*, Pilos, Greece, May 2015, pp. 1–6.

[116] S. Winkler, "Issues in vision modeling for perceptual video quality assessment," *Signal Process.*, vol. 78, no. 2, pp. 231–252, Oct. 1999.

[117] P. Quax, A. Beznosyk, W. Vanmontfort, R. Marx, and W. Lamotte, "An evaluation of the impact of game genre on user experience in cloud gaming," in *Proc. IEEE Int. Games Innov. Conf. (IGIC)*, Sep. 2013, pp. 216–221.

[118] M. Ries, P. Svoboda, and M. Rupp, "Empirical study of subjective quality for massive multiplayer games," in *Proc. 15th Int. Conf. Syst., Signals Image Process.*, Jun. 2008, pp. 181–184.

[119] Z.-Y. Wen and H.-F. Hsiao, "QoE-driven performance analysis of cloud gaming services," in *Proc. IEEE 16th Int. Workshop Multimedia Signal Process. (MMSP)*, Sep. 2014, pp. 1–6.

[120] K.-T. Chen, Y.-C. Chang, H.-J. Hsu, D.-Y. Chen, C.-Y. Huang, and C.-H. Hsu, "On the quality of service of cloud gaming systems," *IEEE Trans. Multimedia*, vol. 16, no. 2, pp. 480–495, Feb. 2014.

[121] M. Claypool, K. Claypool, and F. Damaa, "The effects of frame rate and resolution on users playing first person shooter games," in *Proc. Electron. Imag.*, San Jose, CA, USA, Jan. 2006, Art. no. 607101.

[122] H.-E. Yang, C.-C. Wu, and K.-C. Wang, "An empirical analysis of online game service satisfaction and loyalty," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 1816–1825, Mar. 2009.

[123] T. Meline, *A Research Primer for Communication Sciences and Disorders*. Boston, MA, USA: Pearson, Sep. 2009.

[124] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. New York, NY, USA: Springer, Aug. 2006.

[125] R. Boutaba, M. A. Salahuddin, N. Limam, S. Ayoubi, N. Shahriar, F. Estrada-Solano, and O. M. Caicedo, "A comprehensive survey on machine learning for networking: Evolution, applications and research opportunities," *J. Internet Services Appl.*, vol. 9, no. 1, pp. 1–99, Jun. 2018.

[126] A. Géron, *Hands-on Machine Learning With Scikit-Learn, Keras and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 2nd ed. Sebastopol, CA, USA: O'Reilly Media, Sep. 2019.

[127] S. Suthaharan, "Supervised learning algorithms," in *Machine Learning Models and Algorithms for Big Data Classification*. New York, NY, USA: Springer, 2016, pp. 183–206.

[128] J. Wang, C. Jiang, H. Zhang, Y. Ren, K.-C. Chen, and L. Hanzo, "Thirty years of machine learning: The road to Pareto-optimal wireless networks," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 1472–1514, 3rd Quart., 2020.

[129] I. A. Basheer and M. Hajmeer, "Artificial neural networks: Fundamentals, computing, design, and application," *J. Microbiolog. Methods*, vol. 43, no. 1, pp. 3–31, Dec. 2000.

[130] S. Haykin, *Neural Networks and Learning Machines*, 3rd ed. Upper Saddle River, NJ, USA: Pearson, 2009.

[131] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier networks," in *Proc. AISTATS*, Fort Lauderdale, FL, USA, Apr. 2011, pp. 315–323.

[132] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[133] N. Eswara, S. Ashique, A. Panchbhai, S. Chakraborty, H. P. Sethuram, K. Kuchi, A. Kumar, and S. S. Channappayya, "Streaming video QoE modeling and prediction: A long short-term memory approach," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 3, pp. 661–673, Mar. 2020.

[134] L. Du, L. Zhuo, J. Li, J. Zhang, X. Li, and H. Zhang, "Video quality of experience metric for dynamic adaptive streaming services using DASH standard and deep spatial-temporal representation of video," *Appl. Sci.*, vol. 10, no. 5, p. 1793, Mar. 2020.

[135] R. U. Mustafa, S. Ferlin, C. E. Rothenberg, D. Raca, and J. J. Quinlan, "A supervised machine learning approach for DASH video QoE prediction in 5G networks," in *Proc. Q2SWinet*, Alicante, Spain, Nov. 2020, pp. 57–64.

[136] R. Shalala, R. Dubin, O. Hadar, and A. Dvir, "Video QoE prediction based on user profile," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, Mar. 2018, pp. 588–592.

[137] T. N. Duc, C. M. Tran, P. X. Tan, and E. Kamioka, "Bidirectional LSTM for continuously predicting QoE in HTTP adaptive streaming," in *Proc. 2nd Int. Conf. Inf. Sci. Syst.*, Mar. 2019, pp. 156–160.

[138] L. Liu, H. Hu, Y. Luo, and Y. Wen, "When wireless video streaming meets AI: A deep learning approach," *IEEE Wireless Commun.*, vol. 27, no. 2, pp. 127–133, Apr. 2020.

[139] L. Qian, H. Chen, and L. Xie, "SVM-based QoE estimation model for video streaming service over wireless networks," in *Proc. Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Oct. 2015, pp. 740–755.

[140] P. Casas and S. Wassermann, "Improving QoE prediction in mobile video through machine learning," in *Proc. 8th Int. Conf. Netw. Future (NOF)*, Nov. 2017, pp. 1–7.

[141] Y. Kang, H. Chen, and L. Xie, "An artificial-neural-network-based QoE estimation model for video streaming over wireless networks," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Aug. 2013, pp. 264–269.

[142] A. Hameed, R. Dai, and B. Balas, "A decision-tree-based perceptual video quality prediction model and its application in FEC for wireless multimedia communications," *IEEE Trans. Multimedia*, vol. 18, no. 4, pp. 764–774, Apr. 2016.

[143] E. Danish, M. Alreshoodi, A. Fernando, B. Alzahrani, and S. Alharthi, "Cross-layer QoE prediction for mobile video based on random neural networks," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Jan. 2016, pp. 227–228.

[144] Y. B. Youssef, M. Afif, R. Ksantini, and S. Tabbane, "A novel online QoE prediction model based on multiclass incremental support vector machine," in *Proc. IEEE 32nd Int. Conf. Adv. Inf. Netw. Appl. (AINA)*, May 2018, pp. 334–341.

[145] D. Minovski, C. Ahlund, K. Mitra, and P. Johansson, "Analysis and estimation of video QoE in wireless cellular networks using machine learning," in *Proc. 11th Int. Conf. Quality Multimedia Exper. (QoMEX)*, Jun. 2019, pp. 1–6.

[146] X. Tao, Y. Duan, M. Xu, Z. Meng, and J. Lu, "Learning QoE of mobile video transmission with deep neural network: A data-driven approach," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 6, pp. 1337–1348, Jun. 2019.

[147] T. N. Duc, C. T. Minh, T. P. Xuan, and E. Kamioka, "Convolutional neural networks for continuous QoE prediction in video streaming services," *IEEE Access*, vol. 8, pp. 116268–116278, 2020.

[148] H. Zhang, L. Dong, G. Gao, H. Hu, Y. Wen, and K. Guan, "DeepQoE: A multimodal learning framework for video quality of experience (QoE) prediction," *IEEE Trans. Multimedia*, vol. 22, no. 12, pp. 3210–3223, Dec. 2020.

[149] Z. Deng, Y. Liu, J. Liu, X. Zhou, and S. Ci, "QoE-oriented rate allocation for multipath high-definition video streaming over heterogeneous wireless access networks," *IEEE Syst. J.*, vol. 11, no. 4, pp. 2524–2535, Dec. 2017.

[150] F. Laiche, A. B. Letaifa, I. Elloumi, and T. Aguili, "When machine learning algorithms meet user engagement parameters to predict video QoE," *Wireless Pers. Commun.*, vol. 116, no. 3, pp. 2723–2741, Sep. 2020.

[151] Y. B. Youssef, M. Afif, R. Ksantini, and S. Tabbane, "A novel QoE model based on boosting support vector regression," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2018, pp. 1–6.

[152] A. Ashiquzzaman, S. M. Oh, D. Lee, H. Jung, T.-W. Um, and J. Kim, "Deeplearning convolutional neural network based QoE assessment module for 4K UHD video streaming," in *Proc. SIMULTECH*, Prague, Czech Republic, Jul. 2019, pp. 392–397.

[153] S. Schwarzmann, C. Cassales Marquezan, M. Bosk, H. Liu, R. Trivisonno, and T. Zinner, "Estimating video streaming QoE in the 5G architecture using machine learning," in *Proc. 4th Internet-QoE Workshop QoE-Based Anal. Manage. Data Commun. Netw. Internet-(QoE)*, 2019, pp. 7–12.

[154] K. Almohammadi, "Quality prediction model based on artificial neural networks for mobile 3D video streaming," *Int. J. Comput. Sci. Inf. Secur.*, vol. 17, no. 11, pp. 107–115, Nov. 2019.

[155] B. Bauman and P. Seeling, "Spherical image QoE approximations for vision augmentation scenarios," *Multimedia Tools Appl.*, vol. 78, no. 13, pp. 18113–18135, Jan. 2019.

[156] H. Malekmohamadi, W. A. C. Fernando, and A. M. Kondoz, "Automatic QOE prediction in stereoscopic videos," in *Proc. IEEE Int. Conf. Multimedia Expo. Workshops*, Jul. 2012, pp. 581–586.

[157] J. Yang, Y. Zhu, C. Ma, W. Lu, and Q. Meng, "Stereoscopic video quality assessment based on 3D convolutional neural networks," *Neurocomputing*, vol. 309, pp. 83–93, Oct. 2018.

[158] W. Zhou, Z. Chen, and W. Li, "Stereoscopic video quality prediction based on end-to-end dual stream deep neural networks," in *Proc. PCM*, Hefei, China, Sep. 2018, pp. 482–492.

[159] N. R. Veeraragavan, H. Meling, and R. Vitenberg, "QoE estimation models for tele-immersive applications," in *Proc. Eurocon*, Jul. 2013, pp. 154–161.

[160] 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.*, Jun. 2018, pp. 154–161.

[161] Z. Fei, F. Wang, J. Wang, and X. Xie, "QoE evaluation methods for 360-degree VR video transmission," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 1, pp. 78–88, Jan. 2020.

[162] 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.

[163] 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.

[164] 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.

[165] 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.

[166] P. Athanasoulis, E. Christakis, K. Konstantoudakis, P. Drakoulis, S. Rizou, A. Weit, A. Doumanoglou, N. Zioulis, and D. Zarpalas, "Optimizing QoE and cost in a 3D," in *Proc. MMEDIA*, Lisbon, Portugal, Feb. 2020, pp. 1–6.

[167] M. Utke, S. Zadtootaghaj, S. Schmidt, and S. Müller, "Towards deep learning methods for quality assessment of computer-generated imagery," May 2020, *arXiv:2005.00836*.

[168] S. Zadtootaghaj, N. Barman, S. Schmidt, M. G. Martini, and S. Moller, "NR-GVQM: A no reference gaming video quality metric," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Dec. 2018, pp. 131–134.

[169] N. Barman, E. Jammeh, S. A. Ghorashi, and M. G. Martini, "No-reference video quality estimation based on machine learning for passive gaming video streaming applications," *IEEE Access*, vol. 7, pp. 74511–74527, 2019.

[170] N. J. Avanaki, S. Zadtootaghaj, N. Barman, S. Schmidt, M. G. Martini, and S. Moller, "Quality enhancement of gaming content using generative adversarial networks," in *Proc. 12th Int. Conf. Quality Multimedia Exper. (QoMEX)*, May 2020, pp. 1–6.

[171] S. Goring, R. R. R. Rao, and A. Raake, "Nofu—A lightweight no-reference pixel based video quality model for gaming content," in *Proc. 11th Int. Conf. Quality Multimedia Exper. (QoMEX)*, Jun. 2019, pp. 1–6.

[172] M. Utke, S. Zadtootaghaj, S. Schmidt, S. Bosse, and S. Möller, "NDNetGaming–development of a no-reference deep CNN for gaming video quality prediction," *Multimedia Tools Appl.*, pp. 1–23, Jul. 2020.

[173] S. Van Damme, M. T. Vega, J. Heyse, F. D. Backere, and F. D. Turck, "A low-complexity psychometric curve-fitting approach for the objective quality assessment of streamed game videos," *Signal Process., Image Commun.*, vol. 88, Oct. 2020, Art. no. 115954.

[174] S. Zadtootaghaj, N. Barman, R. R. R. Rao, S. Goring, M. G. Martini, A. Raake, and S. Moller, "DEMI: Deep video quality estimation model using perceptual video quality dimensions," in *Proc. IEEE 22nd Int. Workshop Multimedia Signal Process. (MMSP)*, Sep. 2020, pp. 1–6.

[175] S. Wen, S. Ling, J. Wang, X. Chen, L. Fang, Y. Jing, and P. L. Callet, "Subjective and objective quality assessment of mobile gaming video," 2021, *arXiv:2103.05099*.

[176] M. Suznjevic, L. Skorin-Kapov, A. Cerekovic, and M. Matijasevic, "How to measure and model QoE for networked games?" *Multimedia Syst.*, vol. 25, no. 4, pp. 395–420, May 2019.

[177] Prometheus. *Prometheus Monitoring System & Time Series Database*. Accessed: Jul. 20, 2021. [Online]. Available: https://prometheus.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 the 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, Thessaloniki. He also holds an Early Stage Researcher (ESR) position in the European Union's Horizon 2020 MOTOR5G Project. His research interests include QoE enhancement in future wireless networks, open radio access networks, machine learning, and extended reality and holographic telepresence communications.

VLADIMIR 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 in the field of telecommunications. He has been the Dean of the Faculty of the Telecommunications, TUS; and the Vice Chairman of the General Assembly of the European Telecommunications Standards Institute (ETSI). He is currently a Professor. He is also the Head of the ''Teleinfrastructure'' Research and Development Laboratory, TUS; and the Chairman of the Cluster for Digital Transformation and Innovation, Bulgaria. He has successfully managed numerous industrial, engineering, research and development, and educational projects. He has authored many scientific publications and is tutoring the B.Sc., M.Sc., and Ph.D. courses in the field of information transmission theory and wireless access networks. He is a fellow of the European Alliance for Innovation.

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, where he has been with the Department of Electrical and Computer Engineering, since 2013. His current research interests include design and optimization of antennas and microwave circuits, signal processing on smart antennas, development of evolutionary optimization algorithms, and neural networks. He is a member of the Technical Chamber of Greece. He is also serving as an Associate Editor for IEEE Access.

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 the École Nationale Supérieure des Télécommunications (ENST) Paris and Université Paris 6, in 1996. From 1991 to 1996, he was involved in research at France Télécom and teaching at 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 at the ATEI 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, and NATO-SfP ORCA. He has published over 150 research articles, and several national and European patents. He is a member of IET (MIET), a Senior Member of URSI, and a fellow of the Higher Education Academy (FHEA). He is also serving as an Associate Editor for IEEE Access.