Impact of the Impairment in 360-Degree Videos on Users VR Involvement and Machine Learning-Based QoE Predictions

MUHAMMAD SHAHID ANWAR[1], **JING WANG**[1], **(Member, IEEE),**
SADIQUE AHMAD[2], **WAHAB KHAN**[1,3], **ASAD ULLAH**[4],
MUDASSIR SHAH[5], **AND ZESONG FEI**[1], **(Senior Member, IEEE)**
[1]School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China
[2]Department of Computer Science, Bahria University Karachi Campus, Karachi 75260, Pakistan
[3]Department of Electrical Engineering, University of Science and Technology Bannu, Khyber Pakhtunkhwa 28100, Pakistan
[4]Department of CS/IT, Sarhad University of Science and Information Technology, Peshawar 25000, Pakistan
[5]College of Electronics Science and Technology, Xiamen University, Xiamen 361005, China

Corresponding author: Jing Wang (wangjing@bit.edu.cn)

ABSTRACT Current extended virtual reality (VR) applications use 360-degree video to boost viewers' sense of presence and immersion. The quality of experience (QoE) effectiveness of 360-degree video in VR has often been related to many aspects. The four significant aspects to take into account when evaluating QoE in the VR are a sense of presence and immersion, acceptability, reality judgment, and attention captivated. In this manuscript, we subjectively investigate the impact of 360-degree videos QoE-affecting factors, including quantization parameters (QP), resolutions, initial delay, and different interruptions (single interruption and two interruptions) on these QoE-aspects. We then design a Decision Tree-based (DT) prediction models that predict users' VR immersion, acceptability, reality judgment, and attention captivated based on subjective data. The accuracy performance of the DT-based model is then analyzed with respect to mean absolute error (MAE), precision, accuracy rate, recall, and f1-score. The DT-based prediction model performs well with a 91% to 93% prediction accuracy, which is in close agreement with the subjective experiment. Finally, we compare the performance accuracy of the proposed model against existing Machine learning methods. Our DT-based prediction model outperforms state-of-the-art QoE prediction methods.

INDEX TERMS Quality of Experience, 360-degree videos, virtual reality, decision tree, machine learning.

I. INTRODUCTION

Virtual Reality (VR) has received significant attention due to the advancement of multimedia and computing technology. Filmmakers and industries have also started to work on VR technologies and applications. The education, immersive telepresence, health industry, sports, and telehealth have quickly commercialized to meet consumer satisfaction and demand. 360-degree video is one of the critical VR application to offer an interactive experience to users. The quality of experience (QoE) evaluation and modeling of 360-degree videos in VR is a fledging yet hot topic. The 360-degree video should have a high resolution to meet end-users satisfaction. Besides, to offer excellent end users experience, the 360-degree video delivery should be smooth, and there

The associate editor coordinating the review of this manuscript and approving it for publication was Tai-hoon Kim.

should not be any delay and interruption during playback. This makes the QoE evaluation of high-quality 360-degree video in VR more challenging. Therefore, it is inevitable to get deep understanding and knowledge of factors that affect the QoE in terms of various aspects for 360-degree VR videos.

QoE is the degree of delight or annoyance of the user of an application or service. The research on factors that affect VR QoE and different QoE aspects is foremost for the investigation of QoE of 360-degree videos in VR. Four significant VR QoE aspects are immersion, acceptability, reality judgment, and attention captivated. Users' immersion in VR is the condition in which the virtual environment replaces users' real-world surrounding, and the viewers completely lose awareness of the fact that he/she are really in the virtual environment. Witmer and Singer [1] characterized the immersion as a subjective measure of being in one

environment or place, even when the viewer is physically presented in another. Acceptability is the fact that, to what extent the virtual environment is acceptable to the user [2]. Focusing on the QoE aspect, the immersion and acceptability could be limited. Therefore, reality judgment is another aspect that can contribute to the QoE evaluation in VR, i.e., to what extent does the user feel that they are in virtual world [3], and to what extent is this experiment real. Attention captivated is another significant QoE aspect that shows how much the users' attention is captivated by the virtual environment while watching 360-degree videos [3]. The users QoE in terms of these aspects can be affected by many factors during watching 360-degree videos in VR. Therefore, we investigate the impact of various encoding parameters, initial delay, and different interruptions on these aspects. Existing studies on QoE aspects and factors evaluation are still limited. Most of the existing literature focuses on perceptual quality [4]–[8], cybersickness aspects [9], [10], and presence aspect [5], [6], [11]. These mentioned existing researches will be elaborated in Section II in detail. To the knowledge of authors, no existing study has investigated the impact of these factors. In our work, we investigate the impact of encoding parameters, initial delay, and interruptions on different QoE aspects. These QoE aspects are immersion, acceptability, reality judgment, and attention captivation in 360-degree VR videos. How much the influence of these factors affects the end-users QoE is still unclear, but the expectations are much higher.

Subjective assessment is a famous and well known QoE evaluation method. In subjective QoE evaluation, subjects directly record their score during the test for the viewed videos. The recorded scores are arranged in a dataset for training and evaluating purposes. In recent years, different machine-learning (ML) algorithms have been utilized for QoE prediction of multimedia. The intent of using ML is to model the unknown target variables from observations. Different ML techniques have been proposed for the QoE prediction of 360-degree video in VR [5], [10]. Still, there is a lot of room to build and propose a better model that can predict the QoE of 360-degree video in VR. In this manuscript, we aim to evaluate the impact of different resolution, quantization parameters (QP), initial delay, and different interruptions on four significant VR QoE aspects (i.e., immersion, acceptability, reality judgment, and attention captivated). Besides, we propose a Decision Tree-based (DT) QoE prediction model, four different datasets obtained from subjective experiments are applied to the proposed model to predict the QoE for 360-degree VR video in four different aspects. To this aim, our study focuses on five key QoE-affecting factors and four significant QoE aspects.

The contribution of our work is threefold.

- First, we choose three videos; the first video is encoded in three different resolution (fHD, 2.5K, and 4K), another video is encoded in four different quantization parameters (22, 28, 34, and 40). We simulate the initial delay of 5-second, single interruption of 5-second, and two interruptions of 5-second each within the third video to cover the various affecting factors that influence the end-users QoE in VR.

- Second, we conducted a subjective test on 34 subjects and evaluated the influence of different resolution, QP, initial delay, and different interruptions on four QoE aspects, i.e., immersion, acceptability, reality judgment, and attention captivated in a virtual environment.

- Third, we propose a DT-based model for QoE prediction in terms of these four QoE aspects. The proposed model is trained on four different datasets of QoE aspects obtained from the subjective experiment, and the QoE is predicted. The prediction accuracy of the proposed DT-based model is then compared against the existing methods.

The remainder of this manuscript is arranged as follows: Section II gives an overview of the related work significant to the subject of this manuscript. The subjective experiment methodology and analysis are included in Section III. The proposed QoE prediction model is explained in detail in Section IV. The accuracy performance comparison of the proposed model is shown in Section V. Section VI includes the conclusion of this manuscript.

II. RELATED WORK

This section includes an overview of the existing work related to the scope of our study. Several QoE-affecting factors, QoE aspects, and different machine learning (ML) based QoE prediction models will be discussed in detail in this section.

Several research work have been published recently in the field of VR [12]–[17]. 360-degree video is one of the essential application of VR that facilitates the user with an interactive virtual environment. Many subjective assessment methods have been suggested for QoE evaluation by the International Telecommunication Union (ITU). Different subjective assessment methods are applied for 360-degree videos in [18]–[20]. The authors in [2], [5]–[7] investigated the impact of resolution, bitrate, QP, and content characteristics on QoE in terms of perceptual quality in 360-degree videos. The significant influence of stalling on perceptual quality [2], [8], presence [6], and cybersickness [10] evaluated in detail. The authors in [8] elaborated that multiple stalling in a video badly affect the perceptual quality of 360-degree VR video than single stalling. Regarding perceptual quality, the significant effect of encoding parameter, rendering device, gender, users familiarity with VR, and users interest have a significant impact on 360-degree VR QoE [5]. The subjective study in [7] evaluates the effect of rendering device, content type, and encoding parameter on the users' profile. Their study claims that end-users are less sensitive about resolution and QP when watching a 360-degree video of their interest. Regarding cybersickness and presence aspect, the impact of content type, camera motion, and the number of moving targets in a video under various stalling was investigated by [10], they conclude that fast video, video recorded with
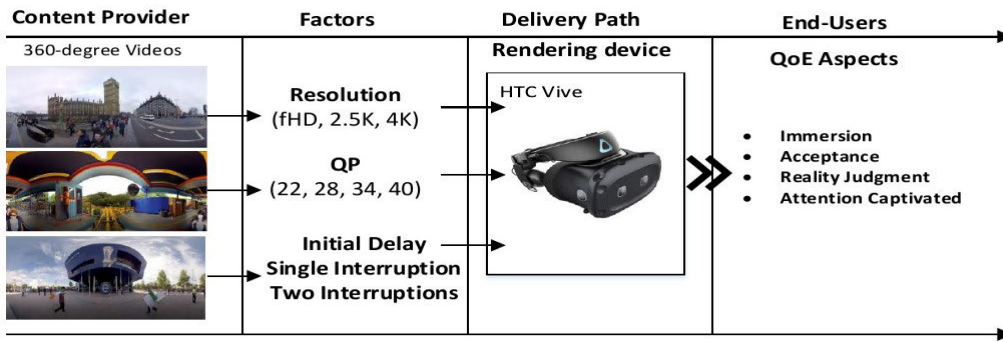
M. S. Anwar *et al.*: Impact of the Impairment in 360-Degree Videos on Users VR Involvement and ML-Based QoE Predictions

IEEE*Access*

FIGURE 1. Framework of our subjective experiment.

vertical motion, and video having multiple moving targets induce more sickness than other factors. Besides, viewers feel more cybersickness than other factors addressed in their work. Besides, viewers feel more presence while watching a medium video than slow and fast motion video. Several existing literature evaluate the impact of bitrate [8], [20], frame rate [21], QP [7], [22], and resolution [4], [9], [21], [23] on QoE have shown significant impact. These studies mentioned above focus on the impact of factors on QoE in terms of perceptual quality, cybersickness, and presence. In our work, we addressed the significant impact of various factors on QoE in terms of immersion, acceptability, reality judgment, and attention captivated.

Several QoE prediction models have been proposed in the literature for the improvement of video QoE. ML techniques have been widely used for the QoE prediction of traditional video [24], [25] and 360-degree videos [5], [10], [26]–[28]. The authors in [8] subjectively investigate the impact of the different stalling event under varying bitrates, and the interaction between bitrate and stalling were also addressed. Their study proposed Bayesian Inference Method (BIM) to predict the QoE of 360-degree videos in terms of perceptual quality. Another work carried out in [10] approached a neural network-based QoE prediction model of 360-degree video in terms of cybersickness. Their proposed ANN-based model achieved 90% prediction accuracy in terms of cybersickness prediction. A comprehensive study in [5] explores the supervised machine learning algorithm; their work proposed Logistic Regression (LR) based QoE model in terms of perceptual quality for 360-degree videos. Furthermore, they compare the prediction accuracy of the proposed model against k-nearest Neighbour (KNN), DT, and Support Vector Machine (SVM). Their proposed model performs well with 86% perceptual quality prediction accuracy. The research work in [35] proposed a VR model that captures the quality-of-service (QoS) of VR users in small cell networks (SCNs). The proposed multi-attribute utility theory model jointly addresses the essential VR metrics, including processing delay, transmission delay, and tracking delay. Another work in [29] utilized a deep reinforcement learning technique and suggested a model for 360-degree videos called DRL360 that can adapt to changing features

and optimize different QoE objectives dynamically. Their proposed model estimates future viewport and bandwidth. The proposed method improves the prediction accuracy by 20% to 30% over existing methods.

These studies above mainly focus on perceptual quality and cybersickness prediction. To the best of our knowledge, no previous study proposed a model that can predict the QoE in terms of immersion, acceptability, reality judgment, and attention captivated using ML. This manuscript collected and arranged four different datasets from the subjective experiment; the established dataset fits into the DT-based QoE prediction model. The DT model is trained on four datasets separately, and then based on the training dataset, the QoE is predicted in terms of immersion, acceptability, reality judgment, and attention captivated. DT is a simple but powerful learning algorithm and has been practically applied for many classification tasks. DT has produced an understandable classification model with excellent accuracy in various application domains. The purpose of using the DT model is that this algorithm requires less effort for data preparation during pre-processing. Besides, it does not need normalization and scaling of data. Furthermore, another advantage of using DT is that the missing values in the data do not impact the decision tree creation process to any considerable extent.

III. SUBJECTIVE EXPERIMENT METHODOLOGY AND ANALYSIS

This section includes the QoE-affecting factors and aspects and the methodology used in our subjective experiment. The complete framework our study is shown in Figure 1.

A. AFFECTING FACTORS AND QoE-ASPECTS

In this subsection, we take into account the influence of encoding parameters (i.e., QP and resolution), interruptions (i.e., single and two interruptions), and initial delay on four QoE-aspects, namely immersion and presence, acceptability, reality judgment, and attention captivated.

Both resolution and QP plays a significant role, and any change in these two factors can influence the end user's QoE. We investigate the impact of four QP, i.e., 22, 28, 34, and 40 and three different resolution, i.e., fHD, 2.5K, and 4K on users' QoE. The initial delay is another factor that may

IEEE *Access*

M. S. Anwar *et al.*: Impact of the Impairment in 360-Degree Videos on Users VR Involvement and ML-Based QoE Predictions

TABLE 1. Detailed specifications of source video.

Factors	Vidoes	SI/TI	Frame Rate	Video ID (YouTube)
QP	London Tour	54.9/61.7	30	SNX8B_OE8IY
Resolution	Roller coaster	48.9/16.7	30	8IsB−P8nGSM
Initial Delay Single Interruption Two Interruptions	PAKvsIND Cricket	44.3/52.4	30	GvfJmFX5csI

affect the end-users' experience when it occurs. We evaluate the impact of 5-second initial delay on four different QoE aspects. Similarly, any interruption during 360-degree videos playback can badly affect the viewer's experience. Interruptions may occur single or multiple times in a video, and the number of interruptions and duration of interruption while watching a 360-degree video in VR have different affect on end-user. Therefore, we aim to evaluate the impact of single interruption of 5-second and two interruptions of 5-second each on four QoE aspects to investigate the QoE of 360-degree videos in VR in terms of immersion and presence, acceptability, reality judgment, and attention captivated.

B. SUBJECTIVE EXPERIMENT

We downloaded three videos from YouTube, having a wide range of Spatial (SI) and Temporal (TI) indexes and frame rate 30fps shown in Table 1. HTC Vive [1] is used as an HMD device having 2160 × 1200 resolution and 110 degrees field of view (FoV). Virtual desktop software is used as a 360-degree video player. A total of 34 subjects participated in the subjective test, including 17 males and 17 females subjects. Each source video is cut into a 1-minute duration ad audio tracks discarded to bypass acoustic information. Out of three source videos, one video is encoded in four different QP i.e., 22, 28, 34, and 40, while other videos is encoded in three resolution, i.e, fHD (1920 × 1080), 2.5K (2560 × 1400), and 4K (3480× 1920) by using FFMPEG [2] software tool. We simulate 5-second initial delay, 5-second single interruption, and two interruptions of 5-second each in a video using AviSynth[3] software tool. The interruptions are not fixed and can occur at a different location in a video so that the viewers can experience a real-world scenario. The viewers are unaware of the interruptions locations. Therefore, in total, we obtained 10 test videos, including four QP, three resolutions, one initial delay, one single interruption, and one multiple interruption video.

Before the actual test, the subjects were exposed to a training session. We instruct the subjects about the test procedure and devices to help them to adjust the HMD according to their head size. During the subjective experiment, each subject watched ten videos and allowed to rest for two minutes

---

[1]www.htcvive.com

[2]www.ffmpeg.org

[3]www.avisynth.nl: It is a script-based tool used for editing and processing videos. The scripting language is simple yet powerful, and complex filters can be generated from basic operations to build a sophisticated palette of useful and unique effects
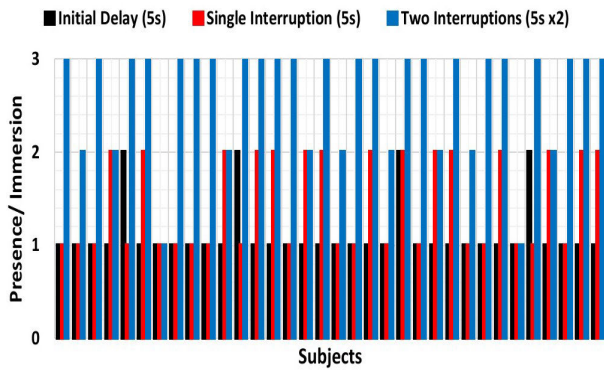
after watching five test videos. The total test duration was almost eight hours. After watching each video the subjects were asked to record their score to evaluate the immersion (ITQ) [1], acceptability [2], reality judgment [3], and attention captivation [3]. The questionnaires we used in the subjective experiment are given below.

- Q1: Immersion/ Presence: To what extent did you immerse in the virtual world? (1-Fully immersed, 2- Occasionally immersed, 3- Not immersed).
- Q2: Acceptability: Is this viewing acceptable to you? (1- Acceptable, 2- Slightly acceptable, 3- Not acceptable).
- Q3: Attention captivated: To what extent your attention captivated by the virtual environment? (1- Completely, 2- Occasionally 3- Not at all).
- Q4: Reality judgment: To what extent did the VR experience seem real to you? (1- Very real, 2- Occasionally real, 3- Not real).

C. SUBJECTIVE RESULTS AND EVALUATION

In this subsection, we explain the subjective result obtained from the experiment. The significant impact of QP, resolution, initial delay, and interruptions on immersion, acceptability, reality judgment, and attention captivated will be discussed in detail in this subsection.

1) IMPACT ON VR IMMERSION

The effect of initial delay, single interruption, and two interruptions on user's immersion is depicted in Figure 2, while Figure 3 presents the effect of three different resolution on users' immersion. It can be seen that users feel more immersion in VR when there is an initial delay of 5-seconds in a video as compared to interruptions while users' immersion is badly affected when there is a single interruption of 5-second occurs while watching a 360-degree video in VR. Users immersion level is lower when two interruptions occur in a single video clip compared to single interruption. This reveals that users are less tolerant of interruption and less sensitive about the initial delay. Thus multiple interruptions should be avoided to improve the QoE of end-users in terms of immersion. From Figure 3 it can be noted that the viewers are comfortable to watch 2.5K and 4K video while slightly uncomfortable and feel less immersion when watching a 360-degree video with fHD resolution. The impact of four different QP values on users' immersion is presented in Figure 4. Almost all users are satisfied with QP22 and feel higher immersion than other QP value.

M. S. Anwar *et al.*: Impact of the Impairment in 360-Degree Videos on Users VR Involvement and ML-Based QoE Predictions

IEEE *Access*

FIGURE 2. Impact of Initial delay and interruptions on users' immersion (1-fully immersed, 2- occasionally immersed, 3- not immersed).

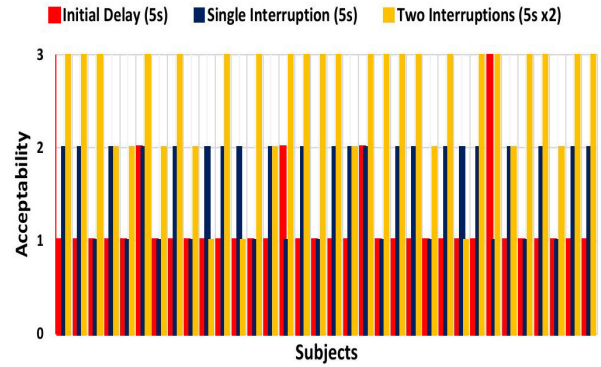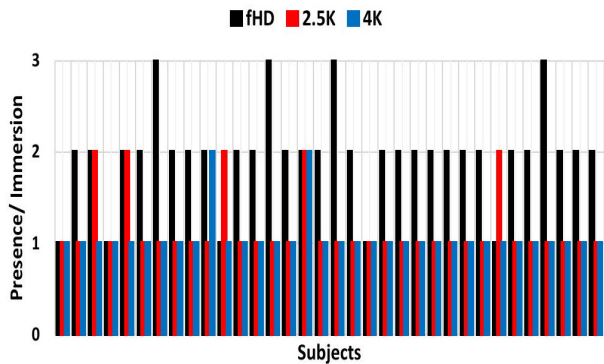**FIGURE 3.** Impact of resolution on users' immersion (1-fully immersed, 2- occasionally immersed, 3- not immersed).

FIGURE 4. Impact of QP on users' immersion (1-Fully immersed, 2- occasionally immersed, 3- not immersed).

FIGURE 5. Impact of initial delay and interruptions on users' acceptability (1- acceptable, 2- slightly acceptable, 3- not acceptable).
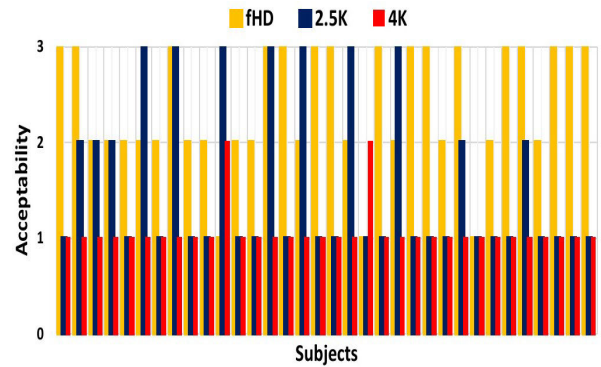

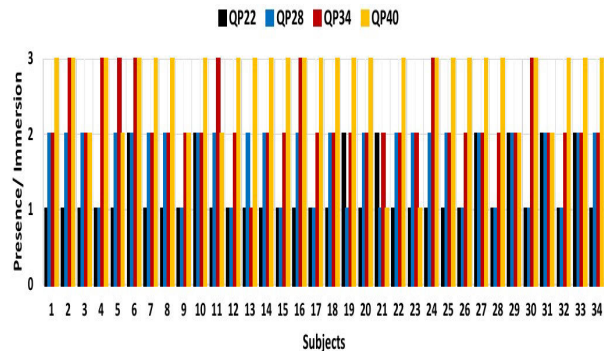**FIGURE 6.** Impact of resolution on users' acceptability (1- acceptable, 2- slightly acceptable, 3- not acceptable).
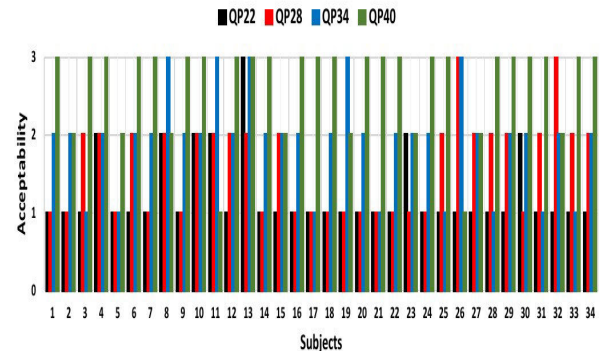

FIGURE 7. Impact of QP on Users' acceptability (1- acceptable, 2- slightly acceptable, 3- not acceptable).

Maximum users fees disturbance and recorded less immersion in VR when watching a video with QP40. Therefore, the higher the QP value, the less will be the users' immersion in virtual reality.

2) IMPACT ON VR ACCEPTABILITY

Figure 5 shows the impact of initial delay and interruptions on users' acceptability while Figure 6 present the impact of different resolution on acceptability. Most users agree to watch and accept the 360-degree video with an initial delay of 5-second and single interruption of 5-second compared to

two interruptions occurs in a single video. This suggests that viewers are willing to watch the video until the end of the session when there is an initial delay or single interruption. At the same time, most of the viewers feel annoying and want to quit the session when multiple interruptions occur in a single video. On the other hand users' acceptability rate is higher when watching a 360-degree video with 4K resolutions shown in Figure 6. As usual, the lower resolution badly affects the viewers' acceptability in VR. Mostly viewers show less acceptability when watching fHD video. The impact of four different QP values on users' VR acceptability is presented in Figure 7, where almost all users agree and
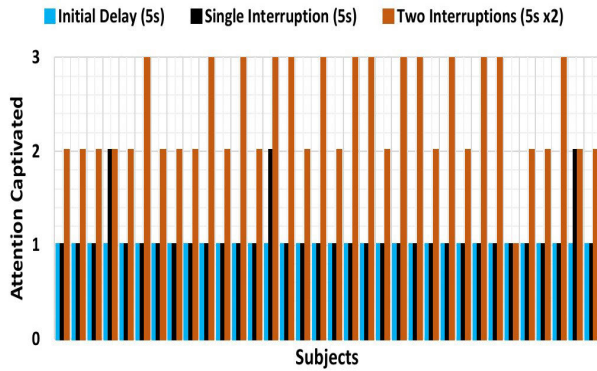
IEEE *Access*

M. S. Anwar *et al.*: Impact of the Impairment in 360-Degree Videos on Users VR Involvement and ML-Based QoE Predictions

FIGURE 8. Impact of initial delay and interruptions on attention captivation (1- completely, 2- occasionally 3- not at all).
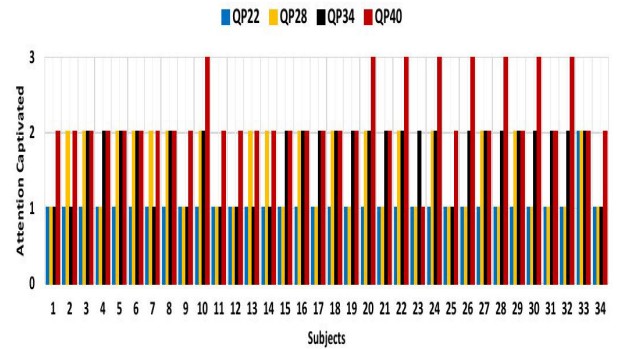

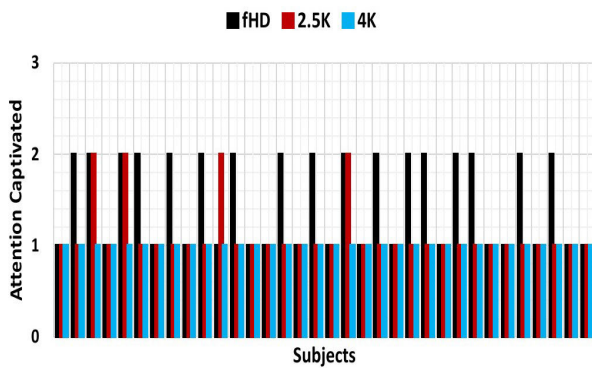**FIGURE 10.** Impact of QP on users' attention captivation (1- completely, 2- occasionally 3- not at all).

FIGURE 9. Impact of resolution on attention captivation (1- completely, 2- occasionally 3- not at all).
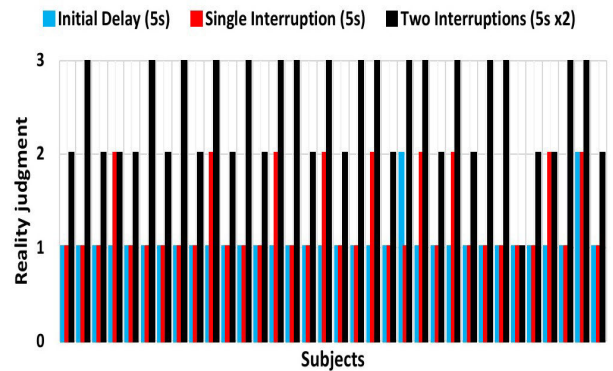

FIGURE 11. Impact of initial delay and interruptions on reality judgment (1- very real, 2- occasionally real, 3- not real).

comfortable to watch the video until the end of the session when watching a video with QP22. Most of the users rejected the videos with QP40 and wanted to quit the session before it ends. Therefore, it suggests that 360-degree video should not be encoded with QP40 when rendering in VR HMD to offer satisfactory QoE in terms of acceptability.

3) IMPACT ON ATTENTION CAPTIVATION
Figure 8 shows the impact of initial delay and different interruptions on users' QoE in terms of attention captivation. It can be seen that when two interruptions each with 5-second duration occurs in 360-degree video divert the viewers' attention in VR and badly affect the QoE. Viewers are less sensitive about the initial delay of 5-second and single interruption of 5-second but less tolerant when multiple interruptions occur in a single video. Figure 9 show the impact of different resolution on users attention captivation, where all users are totally captivated by the virtual environment. We have observed some exciting outcomes about fHD video, users are comfortable and feel captivated while watching fHD video, which has never seen in previous QoE aspect cases. This suggests that lower resolution up to fHD do not affect the users' attention and viewers feel completely captivated while watching in VR. The effect of different QP values on users attention is presented in Figure 10. Again it is interesting to observe that very few users' attention disturbed by higher

QP value compared to other QoE aspects. In the case of QP 22, almost all users show higher attention captivation in VR. In contrast, the attention captivation score of all users recorded lower than 3 in case of QP28 and QP34. Thus, we can conclude that very less number of viewers are not happy with QP40 when it comes to attention captivation in the virtual environment while almost all users agree with QP22, QP28, and QP34. These outcomes are totally different from other QoE aspects. Therefore, higher QP value up to 34 is acceptable in case of attention captivation in VR.

4) IMPACT ON REALITY JUDGMENT
The impact of initial delay and interruptions is shown in Figure 11, while the impact of different resolutions on end-users QoE in terms of reality judgment is presented in Figure 12. Same like other aspects, the impact of two interruptions in a single video on the reality judgment aspect is profound. Viewers' virtual experience is significantly affected by multiple interruptions compared to initial delay and single interruption. The viewers feel virtually in the real world in case of both initial delay and single interruption. From Figure 12, it is observed that users feel more reality in VR in case of 2.5K and 4K resolution. Most of the users occasionally feel real in VR when they watch fHD video, while out of 34 users, only four users' reality judgment is affected by fHD video and scored ''3'' (not real). The MOS score
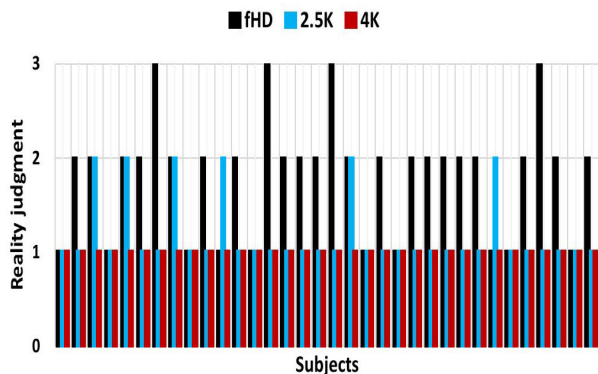
M. S. Anwar *et al.*: Impact of the Impairment in 360-Degree Videos on Users VR Involvement and ML-Based QoE Predictions

IEEEAccess·

FIGURE 12. Impact of resolution on reality judgment (1- very real, 2- occasionally real, 3- not real).
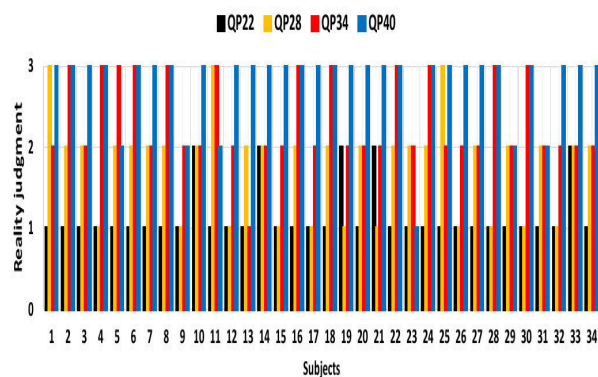

FIGURE 13. Impact of QP on reality judgment (1- very real, 2- occasionally real, 3- not real).

of all subjects is recorded ''1'' (very real) for 4K resolution. While most of the subjects recorded their score ''1'' and few rated ''2'' (occasionally real) for 2.5K resolution. Thus, for better reality judgment, it is suggested that 360-degree videos should be encoded in 2.5K and higher resolution. From these observations, it is suggested that 2.5K and higher resolution provide a fully immersive experience and complete virtual world that feels real to the viewers.

On the other hand, the impact of different QP values on reality judgment is shown in Figure 13. Most of the subjects record their MOS score ''1'' (very real) when they are shown a video with QP22. At the same time, most of the users scored ''3'' (not real) when they watch the video with QP40. Furthermore, the subjects reality judgment score in VR decreases with the increase in QP. It shows that viewers feel more real in VR when they are shown 360-degree VR video encoded with QP22 compared to other QP values. This reveals that the end-users reality judgment about 360-degree video encoded with QP22 is higher in a virtual environment. Besides, the higher the QP value, the lower the users' reality judgment recorded in VR. It is observed that 360-degree video encoded with QP28, QP34, and QP40 affects the users' virtual experiment. In the reality judgment aspect, users only accept the video encoded with QP22 while feeling annoyed with higher QP values, which result in poor QoE.

From the above results and observations, it is concluded that viewers prefer initial delay compared to interruption when occurring during the playback. Besides, for a satisfactory level of QoE of 360-degree video in VR, the videos 360-degree videos should be offered in lower QP to meet the end-users satisfaction. These findings in our study can be helpful to improve the QoE of 360-degree video service over the Internet and VR applications.

To analyze the effect of different factors on these QoE aspects, we carried out a Kruskal-Wallis test on results obtained from the subjective experiment. Note that the Kruskal-Wallis test's purpose is to determine whether there are any statistical impacts of different factors on QoE aspects. It is a rank-based nonparametric test used to determine any statistically significant differences between these independent variables on a dependent variable. Therefore, this test is carried out to calculate if there are any statistical impacts of different factors (used as independent variables) on QoE aspects (dependent variables). The critical value $\alpha = 16.9190$ with 9 degrees of freedom; if x^2 is greater than the critical value, we reject the null hypothesis. The H value recorded 8.346, 11.783, 7.938, and 6.784 for immersion, acceptability, reality judgment, and attention captivated, respectively. Therefore we do not reject the null hypothesis, and there are no significant differences among the factors.

After datasets arrangement, we evaluate all four datasets shown in Figure 14. The total number of samples counts shown against distribution in three categories representing the subjects satisfaction score.

IV. DECISION TREE-BASED QoE PREDICTION

The decision tree is a supervised machine learning algorithm used to train the model. The task of DT is to predict the target variables from labeled data. The proposed method builds a binary tree based on the feature and threshold that produce the maximum information gain at each node. After the dataset is split on features, the information gain is based on the decrease in entropy. DT creates a set of partitions from original data so that the best class can be achieved by making if-then-else decision rules inferred from the data features. For a training vector $x_i \in R^n$, i.e, $X = [x_1, x_2, \ldots, x_n]^t$ and a label vector $y \in R^l$, i.e, $Y = [y_1, y_2, \ldots, y_l]^t$, where $1 \leq i \leq n$ and $1 \leq j \leq l$. Based on the training vectors and corresponding label vectors, the DT periodically partitions the space to grouped the same labeled samples together. Let Q represent the data at node m. For each candidate split $\theta = (j, t_m)$ comprising of a feature j and threshold t_m that partition the data (Q) into two subset $Q_{left}(\theta)$ and $Q_{right}(\theta)$, can be calculated as

$$Q_{\text{left}}(\theta) = (x, y) \mid x_j \leq t_m \tag{1}$$

$$Q_{\text{right}}(\theta) = Q \backslash Q_{\text{left}}(\theta) \tag{2}$$

where the left division is performed by a division operator '\'. The impurity function $H(\cdot)$ is used to calculate the impurity at
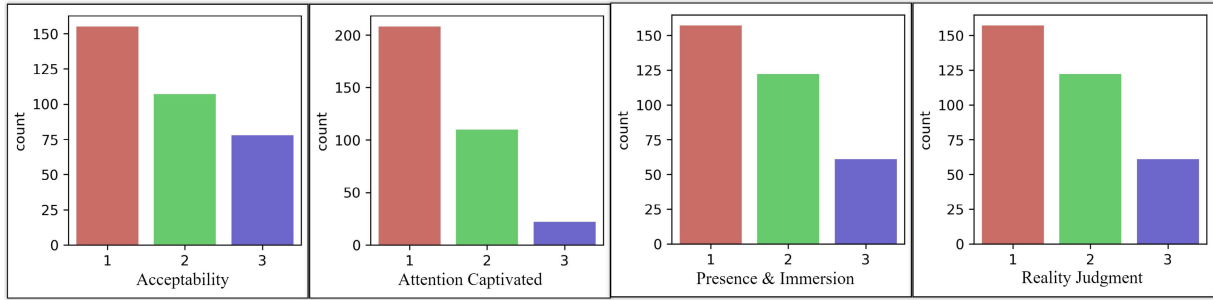
IEEE *Access*

M. S. Anwar *et al.*: Impact of the Impairment in 360-Degree Videos on Users VR Involvement and ML-Based QoE Predictions

FIGURE 14. Dataset evaluations (total number of samples vs distributions in three categories).

Algorithm 1 Learning and Classification of Decision Tree-Based Model

Input:
- Dataset D, a set of training data and associated class labels.
- *attribute_list*, set the candidate attributes.
- *Attribute_selection_method*, select the attribute that best classify the data tuples into individual classes

Output: Classification with a Decision Tree

Method:

1: create a node N;
2: **if** tuples in D are all of the same class C **then**
3: return N as leaf node labeled with class C;
4: **if** *attribute_list* is empty **then**
5: return N as leaf node with labeled with majority class in D;‖ majority voting
6: apply **attribute_selection_method** (*D, attribute_list*)
7: label node N with *splitting_criterion*;
8: select the "entropy" as an attribute selection measure
9: perform the *pruning* process by controlling the maximum depth of the tree with *max_depth =5*
10: **for each** outcome j of *splitting criterion*
11: split the tuples and expand subtrees for each split
12: let *Dj* be the set of data tuples in *D* satisfying outcome j; // a partition
13: **if** *Dj* is empty **then**
14: connect a leaf labeled with the class having majority in D to node N;
15: **else** connect the node returned by **Generate decision tree** (*Dj, attribute list*) to node N;
   end for
16: return N;

---

node m, the choice of which depend on the task being solved (classification)

$$G(Q, \theta) = \frac{n_{\text{left}}}{N_m} H\left(Q_{\text{left}}(\theta)\right) + \frac{n_{\text{right}}}{N_m} H\left(Q_{\text{right}}(\theta)\right) \quad (3)$$

where $Q_{leff}(\theta)$ and $Q_{\text{right}}(\theta)$, are the partition of DT that partition the data (Q) into two subsets.

To minimize the impurity, the parameters are selected,

$$\theta^* = \text{argmin}_\theta \, G(Q, \theta) \quad (4)$$

The two subset $Q_{leff}(\theta^*)$ and $Q_{\text{right}}(\theta^*)$ are then repeated until the maximum allowed depth (i.e, $N_m < \text{min}_{\text{samples}}$, which is 5 in our case) is reached. Maximum allowable depth is a pruning technique to improve the performance by reducing the tree branches with lower importance. This process reduces the complexity performance by reducing over-fitting.

The training observations proportion in mth region from kth class in node m is calculated as

$$p_{mk} = 1/N_m \sum_{x_i \in R_m} I(y_i = k) \quad (5)$$

where m denotes a region R_m with N_m observations.

The Gini impurity function across k class is

$$H(X_m) = \sum_{k=1}^{K} p_{mk}(1 - p_{mk}) \quad (6)$$

where a less value of $H(X_m)$ indicates that node m holds predominantly observations from a single class.

The entropy is calculated as

$$H(X_m) = -\sum_{k=1}^{K} p_{mk} \log(p_{mk}) \quad (7)$$

and the misclassification in node m is given by

$$E(X_m) = 1 - \max_m \{p_{mk}\} \quad (8)$$

where X_m represents the node m of training data.

A. EXPERIMENTS AND EVALUATION

The proposed DT-based QoE prediction model is implemented in python. The ten variables ($x_1, x_2, \ldots, x_{10}$) of four different factors are given as an input variables (independent variables) to a DT model. After the subjective experiment, we have obtained four different datasets from four significant QoE aspects, i.e., acceptability, presence and immersion, reality judgment, and attention captivated. The proposed model is applied separately to four different datasets achieved from the subjective experiment. The DT model studies the training data and then predicts the QoE on testing data. 80% of the dataset is used as training and 20% as testing. The final QoE is then predicted in three different classes,
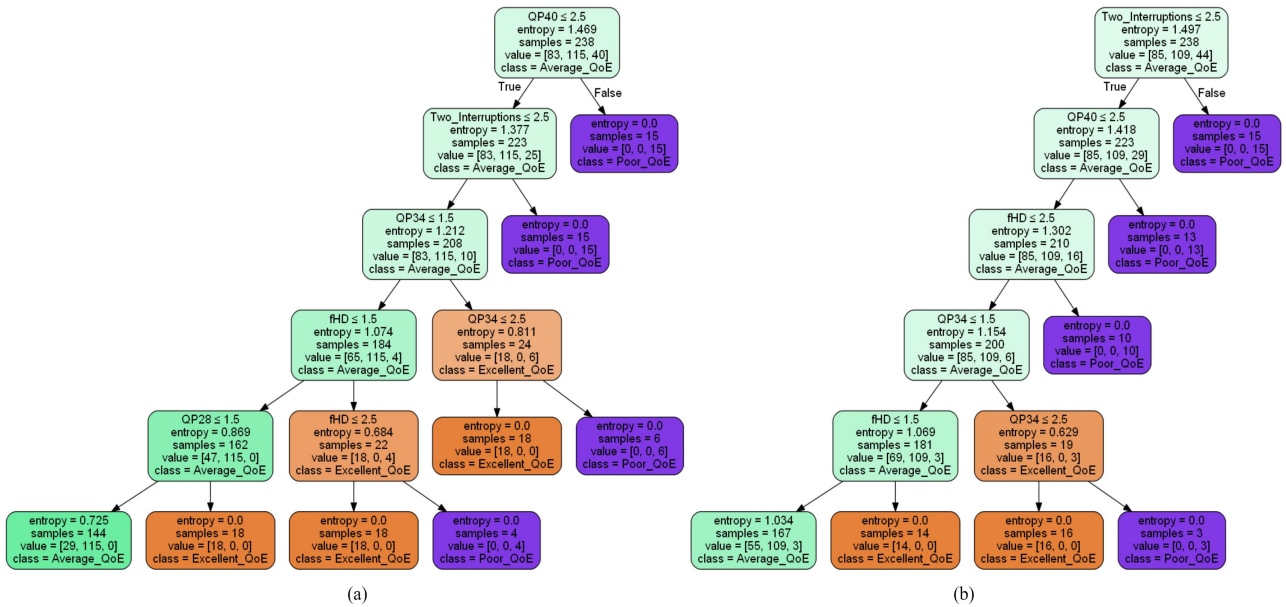
M. S. Anwar *et al.*: Impact of the Impairment in 360-Degree Videos on Users VR Involvement and ML-Based QoE Predictions

IEEE *Access*

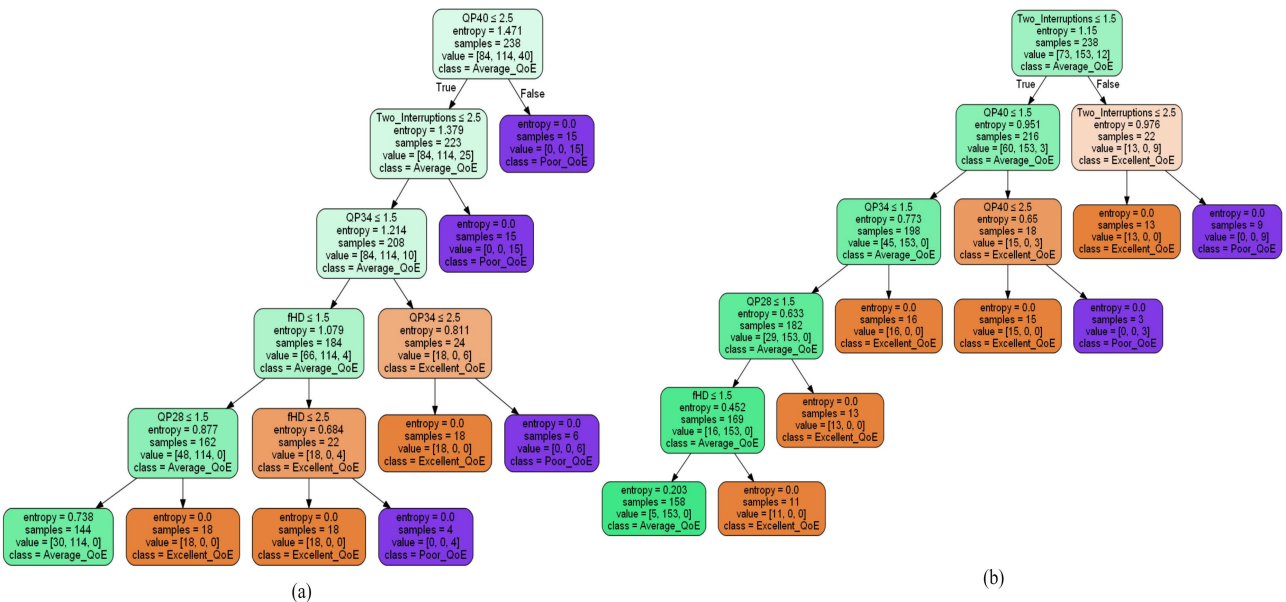**FIGURE 15.** Decision tree graph (a) Immersion and presence (b) Acceptability.

FIGURE 16. Decision tree graph (a) Reality judgment (b) Attention captivated.

i.e., 1= excellent QoE, 2= average QoE, and 3= poor QoE in terms of immersion, acceptability, reality judgment, and attention captivated shown in DT graphs in Figure 15 and Figure 16. The learning and classification process of our proposed DT-based QoE prediction model is shown in Algorithm 1.

We used fivefold cross-validation to overcome the train/test procedure limitations. The fivefold cross-validation is applied to all observations of data for testing and training. Figure 17 shows the accuracy of the proposed model, which is slightly improved train/test split with fivefold cross-validation for all four QoE aspects.

V. ACCURACY AND PERFORMANCE COMPARISON

Table 2 presents the parameters selected for DT-based prediction. The strategy applied to select the best split at each node, i.e., "best." for pruning purposes, the depth of the tree was selected 5, which means that the node will expand until all leaves contain less than the selected minimum number of sample. Entropy is the criteria for calculating information gain and is the measure of node's impurity. The accuracy of the DT-based QoE prediction model for immersion and presence, acceptability, reality judgment, and attention captivated is evaluated in terms of the accuracy rate, precision, recall, f1-score, and mean absolute error (MAE) shown in Table 3.
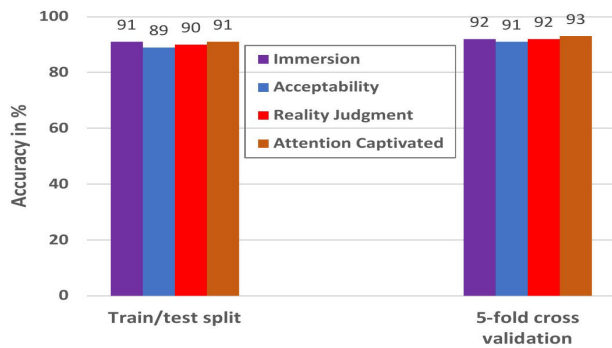
IEEE *Access*

M. S. Anwar *et al.*: Impact of the Impairment in 360-Degree Videos on Users VR Involvement and ML-Based QoE Predictions

FIGURE 17. Fivefold cross validation performance improvement.

- **Precision** indicates the number of correctly predicted positive results against the overall predicted positive case.
- **Recall** represents the number of correctly predicted positive case in a dataset.
- **f1-score** is the weighted harmonic mean and accuracy measure of the test.
- **MAE** compute the mean difference between actual and predicted values. The error difference is proportional to the absolute difference between computed and actual value.

TABLE 2. DT Parameter selection.

Parameters	Selection
criterion	entropy
max_depth	5
splitter	best
random_state	1

TABLE 3. Performance comparison of the proposed QoE prediction Model.

Method	Precision	Recall	f1-Score	MAE	Accuracy
Immersion and Presence	0.94	0.92	0.91	0.08	92%
Acceptability	0.95	0.91	0.89	0.09	91%
Reality Judgment	0.94	0.90	0.93	0.08	92%
Attention Captivated	0.96	0.89	0.94	0.07	93%

For further validation of the proposed model, we compare the prediction accuracy of the proposed DT-based QoE prediction model against state-of-the-art methods shown in Table 3. The proposed model performs well in terms of all four QoE aspects, the accuracy percentage ranging from 91% to 93%. The existing machine learning model [32] based on neural network, naive Bayes, and DT achieved 84% to 88% classification rate. Another model based on semi-supervised learning method [33] with 0.84 f-score, and [31] with 0.75 f1-score and classification rate up to 74%. Recently proposed model published in [34] for video-on-demand quality prediction. They tested SVM, random forest, and Back Propagation Neural Network (BPNN) and achieved f1-score ranging from 0.75 to 0.89. In [5], four supervised ML algorithms (LR, SVM, KNN, and DT) were used for the QoE prediction of 360-degree in VR, while in [30] proposed random forest-based QoE model. These all existing methods achieved lower classification rate than our model.

TABLE 4. Performance comparison of the proposed QoE prediction model against state of the art supervised learning models.

Model Reference	f1-Score	Classification Accuracy (%)
[5]	0.78 − 0.86	79 − 86
[30]	0.73	74
[31]	0.33 − 0.69	44 − 74
[32]	−	84 − 88
[33]	0.84	−
[10]	0.90	90
[34]	0.75 − 0.89	−
(Proposed Model)	0.89 − 0.94	91 − 93

VI. CONCLUSION

This manuscript has investigated the critical QoE-affecting factors that affect the end-users' QoE of 360-degree video in VR. The influence of encoding parameter, initial delay, and interruptions on four significant QoE-aspects, namely immersion, acceptability, reality judgment, and attention captivated was evaluated. The experimental results show that lower resolution and higher QP badly affect the users QoE in terms of these four aspects. Viewers prefer initial delay over interruptions and feel annoying when the interruption occurs in a video, while the users' frustration increases when two interruptions occur in a single video clip. Furthermore, we proposed a DT-based QoE prediction model for 360-degree videos in VR. The prediction model performs well, and the achieved accuracy ranging from 91% to 93%, which is in close agreement to the subjective experiment. The prediction performance was evaluated in terms of precision, recall, f1-score, and MAE. Finally, the accuracy performance of the proposed model is compared against the state-of-the-art methods. Our proposed model performs well against the existing methods.

There are few limitations to our work; we fixed the duration of initial delay and interruptions to 5-seconds so that all users can experience the same disturbance. In reality, the disturbance duration can be shorter or broader in case of interruptions during the playback. Also, there can be more than two interruptions in reality. However, our findings show that users prefer initial delay over interruptions. Therefore, the more interruptions in a video, the more end-users QoE will be degraded. In our future work, we aim to evaluate the effect of different 360-degree projection schemes such as equirectangular, cubic map on end-user QoE. Besides, we will consider different objective metrics in our next study and will build a QoE model for 360-degree VR video. The findings in this manuscript and our future work expected to be helpful to improve the QoE of 360-degree VR videos and other VR applications.

REFERENCES

[1] B. G. Witmer and M. J. Singer, "Measuring presence in virtual environments: A presence questionnaire," *Presence, Teleoperators Virtual Environ.*, vol. 7, no. 3, pp. 225–240, Jun. 1998.

[2] H. T. T. Tran, N. P. Ngoc, C. T. Pham, Y. J. Jung, and T. C. Thang, "A subjective study on QoE of 360 video for VR communication," in *Proc. IEEE 19th Int. Workshop Multimedia Signal Process. (MMSP)*, Oct. 2017, pp. 1–6.

M. S. Anwar *et al.*: Impact of the Impairment in 360-Degree Videos on Users VR Involvement and ML-Based QoE Predictions

IEEE *Access*

[3] R. M. Baños, C. Botella, A. Garcia-Palacios, H. Villa, C. Perpiña, and M. Alcañiz, "Presence and reality judgment in virtual environments: A unitary construct?" *CyberPsychol. Behav.*, vol. 3, no. 3, pp. 327–335, Jun. 2000.

[4] 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. 10th ACM Multimedia Syst. Conf.*, 2019, pp. 232–242.

[5] M. Shahid 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.

[6] M. S. Anwar, J. Wang, A. Ullah, W. Khan, S. Ahmad, and Z. Li, "Impact of stalling on QoE for 360-degree virtual reality videos," in *Proc. IEEE Int. Conf. Signal, Inf. Data Process. (ICSIDP)*, Dec. 2019, pp. 1–6.

[7] M. S. Anwar, J. Wang, A. Ullah, W. Khan, Z. Li, and S. Ahmad, "User profile analysis for enhancing QoE of 360 panoramic video in virtual reality environment," in *Proc. Int. Conf. Virtual Reality Visualizat. (ICVRV)*, Oct. 2018, pp. 106–111.

[8] M. S. Anwar, J. Wang, A. Ullah, W. Khan, S. Ahmad, and Z. Fei, "Measuring quality of experience for 360-degree videos in virtual reality," *Sci. China Inf. Sci.*, vol. 63, no. 10, pp. 1–15, Oct. 2020.

[9] 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. Qual. Multimedia Exp. (QoMEX)*, May 2017, pp. 1–6.

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

[11] R. Schatz, A. Sackl, C. Timmerer, and B. Gardlo, "Towards subjective quality of experience assessment for omnidirectional video streaming," in *Proc. 9th Int. Conf. Qual. Multimedia Exp. (QoMEX)*, May 2017, pp. 1–6.

[12] Z. Yan and Z. Lv, "The influence of immersive virtual reality systems on online social application," *Appl. Sci.*, vol. 10, no. 15, p. 5058, Jul. 2020.

[13] Z. Lv, D. Chen, R. Lou, and H. Song, "Industrial security solution for virtual reality," *IEEE Internet Things J.*, early access, Jul. 1, 2020, doi: 10.1109/JIOT.2020.3004469.

[14] Z. Lv, "Virtual reality in the context of Internet of Things," *Neural Comput. Appl.*, vol. 32, no. 13, pp. 9593–9602, Jul. 2020.

[15] Z. Li, J. Wang, Z. Yan, X. Wang, and M. S. Anwar, "An interactive virtual training system for assembly and disassembly based on precedence constraints," in *Proc. Comput. Graph. Int. Conf.*, Cham, Switzerland: Springer, Jun. 2019, pp. 81–93, doi: 10.1007/978-3-030-22514-8_7.

[16] Z. Li, J. Wang, M. S. Anwar, and Z. Zheng, "An efficient method for generating assembly precedence constraints on 3D models based on a block sequence structure," *Comput.-Aided Des.*, vol. 118, Jan. 2020, Art. no. 102773.

[17] Z. Li, S. Zhang, M. S. Anwar, and J. Wang, "Applicability analysis on three interaction paradigms in immersive VR environment," in *Proc. Int. Conf. Virtual Reality Vis. (ICVRV)*, Oct. 2018, pp. 82–85.

[18] H. Ahmadi, O. Eltobgy, and M. Hefeeda, "Adaptive multicast streaming of virtual reality content to mobile users," in *Proc. Thematic Workshops ACM Multimedia-Thematic Workshops*, 2017, pp. 170–178.

[19] A.-F. Perrin, C. Bist, R. Cozot, and T. Ebrahimi, "Measuring quality of omnidirectional high dynamic range content," *Proc. SPIE*, vol. 10396, Sep. 2017, Art. no. 1039613.

[20] B. Zhang, J. Zhao, S. Yang, Y. Zhang, J. Wang, and Z. Fei, "Subjective and objective quality assessment of panoramic videos in virtual reality environments," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jul. 2017, pp. 163–168.

[21] H. Duan, G. Zhai, X. Yang, D. Li, and W. Zhu, "IVQAD 2017: An immersive video quality assessment database," in *Proc. Int. Conf. Syst., Signals Image Process. (IWSSIP)*, May 2017, pp. 1–5.

[22] H. T. T. Tran, N. P. Ngoc, C. M. Bui, M. H. Pham, and T. C. Thang, "An evaluation of quality metrics for 360 videos," in *Proc. 9th Int. Conf. Ubiquitous Future Netw. (ICUFN)*, Jul. 2017, pp. 7–11.

[23] A. Singla, S. Fremerey, W. Robitza, P. Lebreton, and A. Raake, "Comparison of subjective quality evaluation for HEVC encoded omnidirectional videos at different bit-rates for UHD and FHD resolution," in *Proc. Thematic Workshops ACM Multimedia-Thematic Workshops*, 2017, pp. 511–519.

[24] L. Yu, T. Tillo, and J. Xiao, "QoE-driven dynamic adaptive video streaming strategy with future information," *IEEE Trans. Broadcast.*, vol. 63, no. 3, pp. 523–534, Sep. 2017.

[25] F. Chiariotti, S. D'Aronco, L. Toni, and P. Frossard, "Online learning adaptation strategy for DASH clients," in *Proc. 7th Int. Conf. Multimedia Syst. (MMSys)*, 2016, pp. 1–12.

[26] J. Kim, W. Kim, H. Oh, S. Lee, and S. Lee, "A deep cybersickness predictor based on brain signal analysis for virtual reality contents," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 10580–10589.

[27] C. Li, M. Xu, X. Du, and Z. Wang, "Bridge the gap between VQA and human behavior on omnidirectional video: A large-scale dataset and a deep learning model," in *Proc. ACM Multimedia Conf. Multimedia Conf. (MM)*, 2018, pp. 932–940.

[28] R. I. T. D. C. Filho, M. C. Luizelli, S. Petrangeli, M. T. Vega, J. V. D. Hooft, T. Wauters, F. D. Turck, and L. P. Gaspary, "Dissecting the performance of VR video streaming through the VR-EXP experimentation platform," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 15, no. 4, pp. 1–23, Jan. 2020.

[29] Y. Zhang, P. Zhao, K. Bian, Y. Liu, L. Song, and X. Li, "DRL360: 360-degree video streaming with deep reinforcement learning," in *Proc. IEEE INFOCOM-IEEE Conf. Comput. Commun.*, Apr. 2019, pp. 1252–1260.

[30] M. J. Khokhar, T. Ehlinger, and C. Barakat, "From network traffic measurements to QoE for Internet video," in *Proc. IFIP Netw. Conf. (IFIP Netw.)*, May 2019, pp. 1–9.

[31] B. Peng, J. Lei, H. Fu, L. Shao, and Q. Huang, "A recursive constrained framework for unsupervised video action clustering," *IEEE Trans. Ind. Informat.*, vol. 16, no. 1, pp. 555–565, Jan. 2020.

[32] S. Mustafa and A. Hameed, "Perceptual quality assessment of video using machine learning algorithm," *Signal, Image Video Process.*, vol. 13, no. 8, pp. 1495–1502, Nov. 2019.

[33] S. Yan, Y. Guo, Y. Chen, F. Xie, C. Yu, and Y. Liu, "Enabling QoE learning and prediction of WebRTC video communication in WiFi networks," 2015.

[34] M. Bhat, J.-M. Thiesse, and P. Le Callet, "A case study of machine learning classifiers for real-time adaptive resolution prediction in video coding," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2020, pp. 1–6.

[35] M. Chen, W. Saad, and C. Yin, "Virtual reality over wireless networks: Quality-of-service model and learning-based resource management," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5621–5635, Nov. 2018.

MUHAMMAD SHAHID ANWAR received the M.Sc. degree (Hons.) in telecommunications technology from Aston University, Birmingham, U.K., in 2012, with the research dissertation end-to-end multimedia transmission in a heterogeneous network. He is currently pursuing the Ph.D. degree in information and communication engineering with the School of Information and Electronics, Beijing Institute of Technology, China. His research interests include the area of 360-degree videos, virtual reality (VR), and quality of experience (QoE) evaluations. He currently focus on machine learning-based QoE prediction of 360-degree videos. He has published many research papers in peer-reviewed journals and conferences in the area of 360VR videos and QoE.

JING WANG (Member, IEEE) received the B.S. degree in communication and information system and the Ph.D. degree in electronic engineering from the Beijing Institute of Technology (BIT) in 2002 and 2007, respectively. She is currently an Associate Professor with BIT, where she is also with the Research Institute of Communication Technology. She has ever been a Visiting Scholar with The Chinese University of Hong Kong (CUHK). Her research interests include speech and audio signal processing, multimedia quality assessment, and mobile communication. She is a member of Audio Engineering Society (AES) and of China Computer Federation (CCF); a Senior Member of China Institute of Communications (CIC) and of Chinese Institute of Electronics (CIE), and an Expert Member of Digital Audio and Video Coding Standard Workgroup (AVS), China.

SADIQUE AHMAD received the Ph.D. degree in computer science and technology from the Beijing Institute of Technology, China. He is currently working as a Senior Assistant Professor with the Department of Computer Science, Bahria University, Karachi, Pakistan. He has published 24 research papers in peer-reviewed journals and conferences. His research interests include deep learning and image processing. He has worked on developing new measurement techniques for the prediction of students' Cognitive Skills during cognitive tasks (i.e., measurement of student's performance during the interview, any written examination, and class activities) (transfer learning). The main focus of his current work is to recognize emotions (e.g., frustration, stress, and anxiety) using videos of student's specific activities, such as interviews, written examinations, and final year project presentation.

WAHAB KHAN received the M.Sc. degree in electrical engineering from the COMSATS Institute of Information Technology, Pakistan. He is currently pursuing the Ph.D. degree in information and communication engineering with the Beijing Institute of Technology China. Since 2013, he has been a Lecturer with the Department of Electrical Engineering, University of Science and Technology Bannu, Pakistan. His research interests include wireless channel measurement and modeling, satellite communications, and wireless sensor networks.

ASAD ULLAH received the Ph.D. degree in information and communication engineering from the Beijing Institute of Technology, Beijing, China. He is currently working as an Assistant Professor with the Department of CS/IT, Sarhad University of Science and Information Technology, Peshawar, Pakistan. His research interests include digital image processing, computer vision, and pattern recognition, and have several publications in these areas.

MUDASSIR SHAH received the master's degree in electronic science and technology from the University of Electronic Science and Technology of China in 2020. He is currently pursuing the Ph.D. degree in electronic science and technology with the College of Electronic Science and Technology, Xiamen University, China. His research interest includes medical imaging, and currently focusing on the applications of supervised and unsupervised machine learning methods for mass spectrometry imaging data analysis.
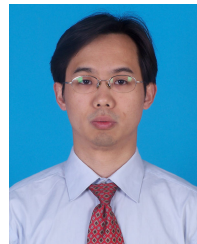
ZESONG FEI (Senior Member, IEEE) received the B.Eng. and Ph.D. degrees in electronic engineering from the Beijing Institute of Technology (BIT), China, in 1999 and 2004, respectively. He is currently a Professor with the Research Institute of Communication Technology (RICT), BIT. From 2012 to 2013, he was a Visiting Scholar with The University of Hong Kong, Hong Kong. He has published over 30 manuscripts in the IEEE journals. His current research interests include error control coding, physical-layer security, cooperative communications, and quality of experience. He is a Senior Member of the Chinese Institute of Electronics and the China Institute of Communications. He is currently serving as a Lead Guest Editor for *Wireless Communications and Mobile Computing* and *China Communications* Special Issue on Error Control Coding.

• • •