Subjective QoE of 360-Degree Virtual Reality Videos and Machine Learning Predictions

MUHAMMAD SHAHID ANWAR[1], **JING WANG**[1], **(Member, IEEE), WAHAB KHAN**[1,2],
ASAD ULLAH[3], **SADIQUE AHMAD**[4], **AND ZESONG FEI**[1], **(Senior Member, IEEE)**
[1]School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China
[2]Department of Electrical Engineering, University of Science and Technology Bannu, Bannu 28100, Pakistan
[3]Faculty of Computing, Riphah International University, Faisalabad 38000, Pakistan
[4]Faculty of Engineering Sciences and Technology, Iqra University, Karachi 75500, Pakistan

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

ABSTRACT 360-degree video provides an immersive experience to end-users through Virtual Reality (VR) Head-Mounted-Displays (HMDs). However, it is not trivial to understand the Quality of Experience (QoE) of 360-degree video since user experience is influenced by various factors that affect QoE when watching a 360-degree video in VR. This manuscript presents a machine learning-based QoE prediction of 360-degree video in VR, considering the two key QoE aspects: perceptual quality and cybersickness. In addition, we proposed two new QoE-affecting factors: user's familiarity with VR and user's interest in 360-degree video for the QoE evaluation. To aim this, we first conduct a subjective experiment on 96 video samples and collect datasets from 29 users for perceptual quality and cybersickness. We design a new Logistic Regression (LR) based model for QoE prediction in terms of perceptual quality. The prediction accuracy of the proposed model is compared against well-known supervised machine-learning algorithms such as k-Nearest Neighbors (kNN), Support Vector Machine (SVM), and Decision Tree (DT) with respect to accuracy rate, recall, f1-score, precision, and mean absolute error (MAE). LR performs well with 86% accuracy, which is in close agreement with subjective opinion. The prediction accuracy of the proposed model is then compared with existing QoE models in terms of perceptual quality. Finally, we build a Neural Network-based model for the QoE prediction in terms of cybersickness. The proposed model performs well against the state of the art QoE prediction methods in terms of cybersickness.

INDEX TERMS Quality of Experience, 360-degree video, virtual reality, perceptual quality, machine learning.

I. INTRODUCTION

360-degree videos (360°), also known as panoramic, omni-directional, or spherical videos, are an emerging multimedia technology that offers immersive viewing experience to end-users. Unlike 2D traditional videos, 360° videos require additional bandwidth for a satisfactory viewing experience that covers a wide area. Different manufactured Head-Mounted-Displays (HMDs), also known as VR headset, can be used to enjoy the extra immersive experience that 360° videos offer. These videos require much higher bandwidth [1], [2] and may result in poor user's QoE while transmitting over a bandwidth-constrained network. To offer excellent end-user QoE, these videos should have high resolution and quality

that result in bulk size. Therefore, for efficient streaming and satisfactory QoE of 360° video, it is significant to get an in-depth understanding of QoE-affecting factors.

QoE counts on how users observe the overall value of service, and therefore it depends on subjective measures. The research on QoE-affecting factors is of great importance for QoE evaluation. Two significant QoE aspects are perceptual quality and cybersickness. Perceptual quality indicates the satisfaction degree of a user for a video quality watched with HMD. Cybersickness is the feeling of nausea and dizziness when watching 360° video in a virtual environment. Since cybersickness is an effect that does not happen while watching traditional videos on a regular screen (as compared to VR glasses or goggles), this is one of the critical differences of viewer's experience between watching traditional videos and 360° video. Previous studies show that the traditional

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

video affected by compression and rendering device due to blurring artifact and blocking may result in QoE degradation [3]–[7]. The existing study on QoE-affecting factors and QoE aspects of are still limited [8]–[10]. Most existing researches focus on 1) the perceptual quality by considering the affecting factors such as gender, usage history, and rendering device, and 2) the cybersickness with the effect of content motion [9]–[12]. Besides these factors, the researcher should consider the user's interest in the 360° video he/she is watching. Another affecting factor is the users' familiarity with VR that can impact the users' judgement about the quality of the video. Users having regular VR experience may have a different opinion about the video quality compared to first-time and rarely viewers. To the best of the authors' knowledge, no previous research studies have taken into account the effect of user's interest in a video and user's familiarity with VR on cybersickness and perceptual quality. Besides, the impact of six (QP, resolution, rendering device, gender, user's interest, and user's familiarity with VR) factors considered in this manuscript has not been investigated collectively on perceptual quality.

There are many studies and various approaches in the literature for the QoE prediction of videos [13]–[16], mainly including three categories: subjective QoE assessment, objective assessment, and data-driven QoE method. Subjective QoE assessment is the most promising method that directly reflects the viewers' judgment. Objective assessment method uses a mathematical tool and mostly based on reference-based classification methods or human visual system. The data-driven method realizes QoE modelling and prediction by analyzing the collected dataset. The subjective QoE method indicates to directly ask the participants to judge the QoE while watching videos. The subjects need to give his/her score after watching a video in a subjective experiment. Based on the subjective score and collected dataset, the researchers can use the recorded data as a training set for the prediction of QoE through various Machine Learning (ML) algorithms. In recent years, ML-based QoE prediction for multimedia streaming has emerged. The intents of applying ML is to model an unknown target concept from observations. ML algorithms have been used to predict the QoE for traditional videos [13], IPTV services [14], online video service provisioning [15], mobile video transmission [16], and 3D-immersive media streaming [17], however, little is known in the area of 360° video. This manuscript intent to investigate the impact of six key QoE-affecting factors on users' perceptual quality and the effect of three (gender, users interest, and users familiarity with VR videos) QoE-affecting factors on users' cybersickness level. Besides, four supervised machine-learning algorithms such as LR, kNN, DT, and SVM are applied to the subjective dataset to predict the QoE in terms of perceptual quality which has not been used collectively for the QoE prediction of 360° video in VR. We proposed the LR-based model for the prediction of QoE in terms of perceptual quality. Moreover, for the prediction of QoE in terms of cybersickness, we propose an ANN-based

model. To aim this, we focus on 360° videos QoE evaluation in terms of two aspects and six QoE-affecting factors. Our main contributions are fourfold.

- First, we subjectively investigate the impact of six QoE-affecting factors such as quantization parameter (QP), resolution, rendering device, gender, user's interest, and user's familiarity with VR on perceptual quality. Besides, we subjectively evaluate the impact of three QoE-affecting factors (gender, user's interest, and user's familiarity with VR) on cybersickness. To aim this, we conducted a subjective experiment on 96 video samples and obtained a dataset from 29 subjects to investigate the QoE. The statistical result from the subjective experiment provides various findings on QoE impact factors in terms of perceptual quality and cybersickness.

- Second, we present two new QoE-affecting factors such as user's familiarity with VR and user's interest in 360° video. The impact of these two influence factors is evaluated on users' QoE in a subjective experiment.

- Third, we proposed the LR-based model for QoE prediction in terms of perceptual quality. The proposed model is trained on perceptual quality (MOS) dataset obtained from subjective experiment, and then QoE is predicted in terms of perceptual quality based on training data. The performance accuracy of the LR model is then compared against well-known supervised ML algorithms such as kNN, SVM, DT, and existing QoE prediction models. The prediction accuracy of the proposed model is higher compared to kNN, SVM, DT, and existing QoE models. To the best of the authors' knowledge, this is the first study that collectively applies the four supervised ML methods (LR, kNN, DT, and SVM) for the QoE prediction of 360° video in VR.

- Fourth, we use an ANN-based model for QoE prediction in terms of cybersickness. To predict the QoE, The ANN model is trained and tested on cybersickness data obtained from the subjective test. The prediction accuracy of the proposed model is then compared against existing cybersickness prediction models. The proposed ANN-based model achieves higher QoE prediction accuracy in terms of cybersickness.

The remainder of this manuscript is structured as follows: The related work significant to the subject of this manuscript is explained in detail in Section II. In Section III-A, we describe the QoE influence factors of 360° video in detail while, in Section III-B, the detailed description of the subjective experiment and test video generation is presented. The subjective results and analysis are illustrated in Section IV, and the ML-based QoE prediction is presented in detail in Section V. Section VI includes the experiments evaluation and performance comparison of ML algorithms. Section VII presents the QoE prediction in terms of cybersickness based on ANN. Finally, Section VIII concludes the manuscript.

II. RELATED WORK

In this section, we enumerate the related work relevant to the scope of our study. First, we present the related literature in detail about affecting factors that affect the QoE in various aspects. Then, we discuss the machine learning-based QoE-prediction methods founded in the literature of traditional videos and 360° videos.

A. 360-DEGREE VIDEOS QoE-AFFECTING FACTORS AND SUBJECTIVE EVALUATIONS

360° videos contain considerable information as compared to traditional ones, and thus these videos have higher resolutions and are encoded at higher bitrates. Besides, to fully immerse in the virtual environment and cover the wide-area, an HMD device is used to watch such kind of videos, which creates many challenges for the researchers to analyze the QoE and consider different factors that may affect the QoE. A substantial number of approaches have been applied to evaluate the QoE of traditional 2D videos, and many QoE-affecting factors and aspects are considered. Although a considerable number of studies have been published for the QoE of traditional video [18]–[23], only a few studies on 360° videos [24]–[30] have been recently published. However, much more attention is needed to investigate QoE-affecting factors and aspects. In [12], authors evaluate the QoE in terms of simulator sickness for omnidirectional videos; their study emphasizes the impacts produced by different HMD device. The authors in [31] investigate the direction and velocity of head motion and those of object motion in VR. Besides, their study analyzes the content feature such as object motion and background complexity to know the feeling of objects under diverse conditions by collecting sickness level ratings from 80 subjects. Regarding the perceptual quality of the 360-degree video, the authors in [32] subjectively evaluate the effect of encoding parameters, content type, and rendering device on QoE by taking into account the user's profile. Their study concludes that viewers tolerate the encoding parameters when they watch an interesting 360-degree video in VR compared to a non-interesting video. Besides, they reveal that viewers are concerned about device type and recorded higher MOS scores while watching in HTC Vive compared to Google cardboard.

Authors in [33] collected subjective score by considering three influence factors, including projection scheme, QP, and video characteristics. The study builds a QoE model using linear regression based on their subjective scores. To minimize the cybersickness level in 360° video, the authors proposed a method by applying image processing and a wearable 360 camera [34]. Their study reveals that a sudden speed change in the translational and vibrational camera can result in severe cybersickness. Most of the existing studies either investigated the QoE in terms of perceptual quality by considering the head motion, object motion in the video, background complexity, and projection scheme. Most importantly, from the previous literature, the cybersickness aspect is evaluated by taking into account the content motion such as low, medium, and fast videos.

The most comprehensive study on QoE-affecting factors and QoE aspects for 360° videos in VR, and closest to our work is presented by Huyen T.T *et al.* in [35]. Their study investigated the subjective QoE in terms of presence, acceptability, perceptual quality, and cybersickness. The authors considered the effect of four factors on QoE such as encoding parameters, content characteristics, rendering device and mode. Their results show a lower presence score strongly affected by content characteristics compared to perceptual quality. Furthermore, the study analyzed the impact of low, medium, and fast motion content on the user's cybersickness level. The authors concluded that cybersickness is a significant problem while watching fast motion videos in VR. The authors extend their work in a recent subjective study carried out in [11], which presents the content motion effect in more detail. In [11], the content motion type such as an object, camera, and background motion is analyzed. The study concluded that the impact of content motion on cybersickness is serious while the impact of rendering device is insignificant. In terms of cybersickness aspect, the above two studies focused on the content motion and device type. As compared to these studies above, we investigate the impact of all six QoE-affecting factors on perceptual quality. Among them, two new QoE factors (user's interest and user's familiarity with VR) were also analyzed. Besides, three influence factors, gender, user's interest, and user's familiarity with VR, which induce cybersickness in a virtual environment, are also investigated during the subjective experiment. These three factors have shown to have a substantial influence on the user's cybersickness level in a virtual environment.

B. ML-BASED QoE PREDICTION METHODS

The study of the literature shows that ML techniques have extensively been used for the QoE prediction of VR videos [36]–[38]. The previous researches in [39]–[42] presents the capability of various ML algorithms to classify the values for QoE-affecting factors, such as video quality and stalling. A Deep Neural Network (DNN) model is proposed to interpret the relation between subjective QoE score and network parameters [16]. The viewers were shown pre-installed traditional videos from a dataset of 80K samples with 89 features and the subjects were asked to give their opinion about video quality, stalling, loading, and overall quality of the video. For a YouTube QoE in cellular networks, a dataset of a distributed network and QoE measurement in a real user's smartphone is collected in [43]. The authors proposed ML models to predict the four QoE-affecting factors such as initial delays, video quality switches, number of stalling, and stalling frequency. The study also approaches multiple ML techniques for the prediction of MOS score and user engagement. These techniques include random forest, DT, SVM, kNN, and Naive Bayes (NB), the performance of random forest classifier is higher than other ML algorithms.
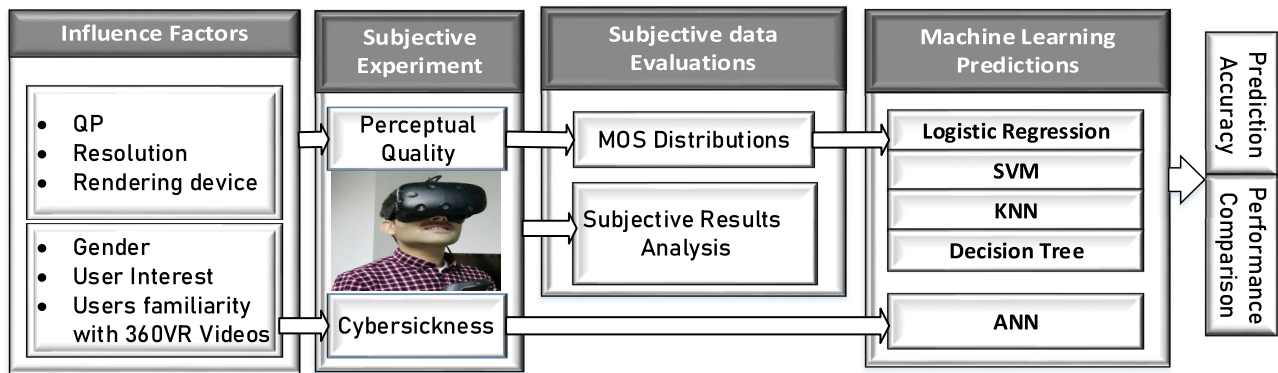
FIGURE 1. Subjective QoE Evaluation and Machine Learning Prediction Framework.

TABLE 1. Influence Factors and QoE Aspects.

Factors	Features	QoE Aspects
QPs	22, 28, 34, 40	Perceptual Quality
Resolutions	HD, fHD, 2.5K,4K	Perceptual Quality
Rendering Device	HTC Vive, Google Cardboard	Perceptual Quality
Gender	Male, Female	Perceptual Quality, Cybersickness
user's interest	Interested , Not Interested	Perceptual Quality, Cybersickness
User's Familiarity with VR	First time, Rarely, weekly	Perceptual Quality, Cybersickness

In a recent study conducted in [44], a data analysis model is proposed to analyze the video based on MOS collected from various datasets. The dataset includes the MOS value of three QoE factors such as initial buffering, stalling time, and stalling ratio. The QoE is predicted by the combination of two ML methods, LR and K-means clustering. The study concludes that their proposed technique achieved 97% performance improvement as compared to the existing prediction methods. The above-mentioned machine learning techniques for the QoE estimation and prediction are applied to only traditional videos, but not for 360° videos. Unfortunately, there is still limited work on ML-based QoE prediction of 360° video in VR.

Regarding 360° videos, The authors in a recent study [45] subjectively evaluate the effect of various stalling events on end-users QoE under three different bitrate levels (1Mbps, 5Mbps, and 15Mbps). Besides, the interaction between stalling event and bitrate level was investigated. They proposed a Bayesian Inference Method (BIM) to predict the end-user QoE of 360-degree video in VR. Their study concluded that the adverse effect of multiple stalling in a single video sequence is more profound when the presentation quality level approaches to the high and low end. Another study in [46] proposed Movement (HM) saliency prediction method by using a deep reinforcement learning approach. The study also proposes offline-DHP to predict HM map of 360° videos and online-DHP approach to predict HM position of one subject at the next frame, based on the 360° video and HM scanpath until the current frame. Authors claim that their proposed method is significant in both offline and online prediction of head movement.

Recent work carried out in [47] approaches a deep reinforcement learning method and proposed a DRL360 model for 360° videos that can adjust to dynamically varying features in environments and optimize different QoE objectives. The proposed technique utilizes DNN method to predict the future viewport and bandwidth, and assign the rate for tile by ACTOR-CRITIC algorithm. The model outperforms the existing methods by 20%-30%. These aforementioned machine-learning methods mostly focused on head movement and future viewport prediction. In this manuscript, we have collected a dataset from a subjective test, and the established dataset fits into various ML algorithms. Four supervised ML techniques LR, kNN, SVM, and DT, are trained with a perceptual dataset while ANN is trained with cybersickness data obtained from the subjective experiment and QoE are predicted. The Subjective QoE evaluation and machine learning-based prediction framework used in this manuscript are shown in Figure 1.

III. DESCRIPTION AND EXPERIMENTAL SETUP
A. 360-DEGREE VIDEO QoE INFLUENCE FACTORS
To investigate the influence of various factors on two QoE aspects such as user's perceptual quality and cybersickness, we choose six factors that affect the end-user QoE. These factors are QP, resolution, rendering device, gender, user's interest, and user's familiarity with VR. The impact of these factors on the user's perception and cybersickness is recoded during a subjective experiment. QoE-affecting factors considered in our study are shown in Table 1, with features description and QoE aspects.
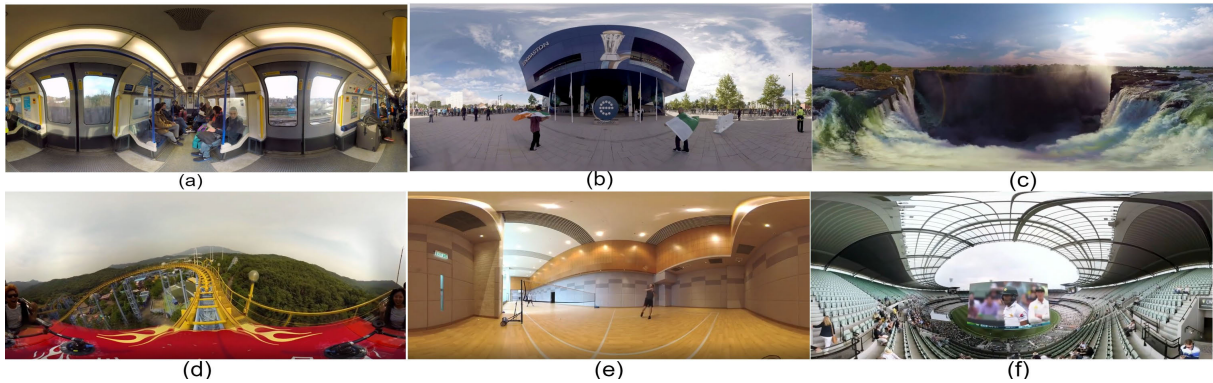
FIGURE 2. Source Videos Example Frames.

1) QP AND RESOLUTIONS

Both encoding parameters (QP and resolution) plays a vital role in video quality. Changing QP and resolutions of video affect the end-users' QoE. In this study, we evaluate the impact of various QPs and resolutions on users' QoE in terms of perceptual quality during a subjective test. To investigate the influence of these factors, we choose four QP (22, 28, 34, and 40) and four resolution (HD, fHD, 2.5K, and 4K).

2) RENDERING DEVICE

Different HMD devices have a distinct effect on viewers while watching 360° video in VR. Different manufactured HMD devices are easily accessible to the users, which creates many challenges for researchers to evaluate the QoE because investigating the user's opinion in one device may be different in another type of HMD. Therefore, in this work, we used two different types of rendering devices (HTC Vive and Google Cardboard) during the subjective experiment. Both devices render 360° video differently; HTC Vive is directly connected with desktop PC while Google Cardboard with a smartphone and the impact of both HMDs on the user's perceptual quality recorded.

3) GENDER

In our subjective study, we include both male and female users. The effect of cybersickness (dizziness, nausea eye strain, headache, disorientation, and vomiting) on different genders is recorded on a five-point scale ranging from "1 to 5", where "5" means very dizzy and "1" not dizzy.

4) USER's INTEREST

The user's interest in a video plays a vital role while investigating perceptual QoE in virtual environment. User personal interest in a video significantly affects the user's QoE. Since viewers immerse themselves in a virtual environment, if the user is not interested in a video he or she is watching, their opinion will be different in both cases. User's tolerance about perceptual video quality varies according to their interest. A user may be less sensitive and more tolerant when watching an interesting video. Even a lower quality of the video may less influence the viewer's opinion as compared to a

TABLE 2. 360° Source Videos Detail.

Index	Description	YouTubeID	Frame Rate
(a)	London Tour	SNx8B_oE8IY	30
(b)	PAKvsIND Cricket	GvfJmFX5csI	30
(c)	Victoria Fall	WsMjBMxpUTc	24
(d)	Roller Coaster	8lsB-P8nGSM	30
(e)	Boring Badminton	Ngj6C_RMK1g	30
(f)	Cricket ENGvsPAK	qngOiJQaRak	25

non-interesting video. Similarly, in the case of cybersickness, the more user not interested in the content, more will he or she feel boredom, which also results in high dizziness or nausea. Therefore, it is essential to investigate the user interest in each 360° video during a subjective experiment to evaluate the QoE.

5) USER's FAMILIARITY WITH VR

The viewer's prior experience about VR affects the end-users opinion. The judgment of those viewers having the first experience of watching 360° video in the virtual environment is different about video content from those of regular or rarely users, which affect the viewer's opinion. In this study, we have users with different prior VR experience and familiarity (users watching the first time, rarely, and weekly).

B. SUBJECTIVE USER STUDY

Total twenty-nine subjects participated in the subjective experiment, including nineteen males and ten females subjects. To ensure the QoE of 360° videos in different HMDs, we select two devices, HTC Vive and Google Cardboard for the subjective test in a virtual environment. The resolution of HTC Vive is 2160 × 1200 with a field of view (FoV) of 110 degrees and is directly connected with Desktop PC. Virtual Desktop software is installed in a PC used as a 360° video player. Google Cardboard with 90-degree FoV covers 75% of human FoV used with Samsung Galaxy A7 mobile phone with 5.7 inches display and 1920 × 1080 resolutions. Six different types of 360° videos are downloaded from YouTube. Figure 2 shows their example frames and the details of source videos are given in Table 2. The source videos are

diverse contents that cover a wide range of SI (spatial) and TI (temporal) indexes. The SI and TI of source videos are shown in Figure 3. Each source video reduced into a short clip of 1-minute duration. Each source video is encoded in four different QP (22, 28, 34 and 40) with four resolutions 3480×1920 (4K), 2560×1400 (2.5K), 1920×1080 (fHD), and 1280×720 (HD). We use FFMPEG[1] software tool with H.264 (libx264), to create the 16 different videos stream with four QP and four resolution. Moreover, the audio track of the videos is discarded during the encoding process to bypass the impacts of acoustic information. Total, there are 96 videos used in our experiment.
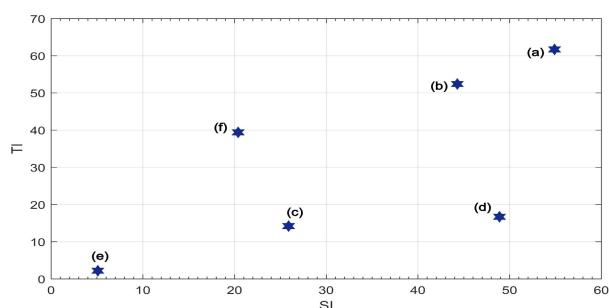
FIGURE 3. Spatial and Temporal Indexes of Source Videos.

Before the actual test, Snellen (20/20) and Ishiara charts were used to check the visual acuity and color vision of all subjects. All the subjects were reported normal. Participants were exposed to a training session before the test to enlighten the subjects with the actual test procedure and to assist them to adjust the HMD device according to their head size [12]. The subjects allowed to sit on a swivel chair to move their heads freely and to cover the wide region of 360° video. A five-point ACR scale were used according to ITU [48] corresponding to the experience of "excellent 5", "good 4", "fair 3", "poor 2", and "bad 1".

During the test, subjects were randomly divided into two groups. Each user of a group watches 48 samples of videos. The purpose of dividing subjects into two groups is to avoid the negative impact of the long experiment. Each subject in a group takes 1.5 hours, including 15 minutes break after every 16 samples (48 samples in each group). Therefore total test time was almost 40 h. After watching each sample, the subjects were asked four questions. The questions included in our subjective test are listed below.

- Q1: How do you perceive the quality of video on a 5-point scale? (1=bad, 5= excellent).
- Q2: How do you rate the level of cybersickness (dizziness, nausea eye strain, headache, disorientation, and vomiting) during watching in VR? (5= very dizzy, 1= not dizzy).
- Q3: Was the video interesting or not-interesting to you? (Yes or no).

- Q4: What is your prior experience of watching 360° videos in VR? (Watching first time, rarely, weekly).

We have recorded five subjects with first time 360° VR videos experience, eighteen users with rarely and six subjects with weekly experience during the subjective experiment. Each subject had to speak their score loudly so that the test organizer could note down on a paper to save time. The average rating of each user for each question is used as a MOS of that question. After conducting the subjective experiment, we applied the outlier removal scheme [49], and no subject was reported.

IV. SUBJECTIVE RESULTS AND ANALYSIS

The impact of six factors (QP, resolution, rendering device, gender, user's interest and user's familiarity with VR) on perceptual quality were investigated while the influence of three factors (gender, user's interest, and user's familiarity with VR videos) on cybersickness in VR was also evaluated. The results of a subjective experiment and their description are presented below.

A. IMPACT ON PERCEPTUAL QUALITY

1) USER's INTEREST

The subjective results show exciting outcomes of how the user's interest influence the perceptual quality and cybersickness while watching 360° video in VR. In the case of a user interested in a video, Figure 4(a), 5(a), and 6(a) show the subject perceptual quality score against QP and resolution. Figure 4(b), 5(b), and 6(b) depict the perceptual quality score of the users watching a non-interesting video. In all three Figures (Figure 4, 5, and 6), it can be seen that users are less sensitive about the quality of the content for resolution and QP when watching interesting contents and at the same time more conscious about the quality of content for resolution and QP in non-interesting case. Therefore, it is suggested that the user's interest has an essential role in evaluating the QoE of 360° videos.

2) USER's FAMILIARITY WITH VR

The subject's prior experience of watching 360° video in a virtual environment has a significant effect on perceptual quality. Figure 4 depicts the perceptual quality score of the subjects watching 360° video for the first time in VR. The figure shows first-time viewers' MOS scores against encoding parameters. Figure 5 interprets the impact of a user watching rarely with HMD on perceptual quality against encoding parameters. The influence of user watching weekly 360° video in VR on perceptual quality is shown in Figure 6. Among all these three figures, the result of Figure 4 is exciting in the case of viewers having the first experience of a virtual environment. In Figure 4 (first time users), the MOS is higher in both interesting and non-interesting scenarios than rarely and weekly users, shown in Figures 5 and 6, respectively. One possible reason may be that users watching the first time in VR are more passionate as compared to rarely and weekly
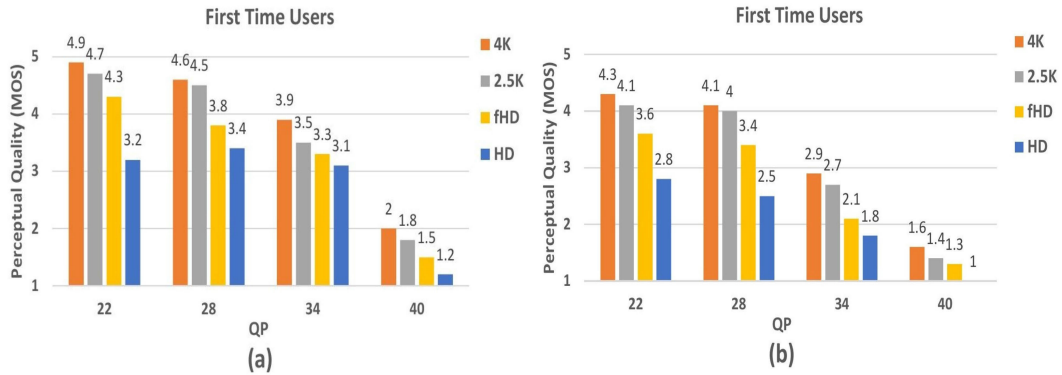
FIGURE 4. Subject's perceptual quality against QP and resolution having first time experience of watching 360° video with HMD (a) Interesting 360° videos (b) Non-Interesting 360° videos.
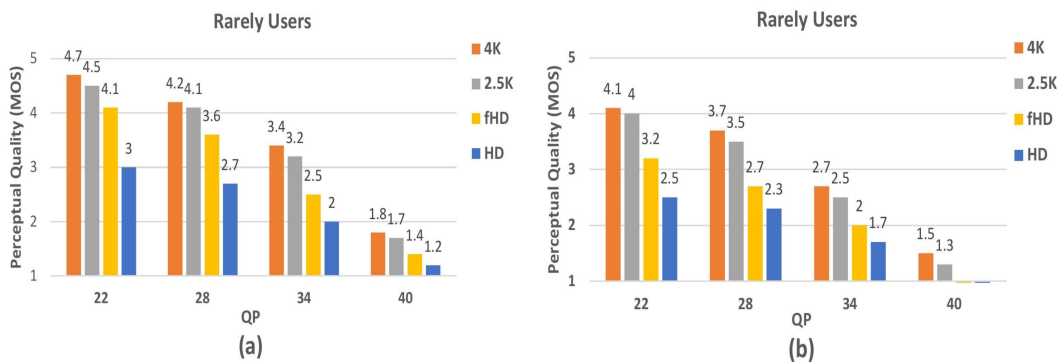

FIGURE 5. Subject's perceptual quality against QP and resolution watching rarely 360° video with HMD (a) Interesting 360° videos (b) Non-Interesting 360° videos.
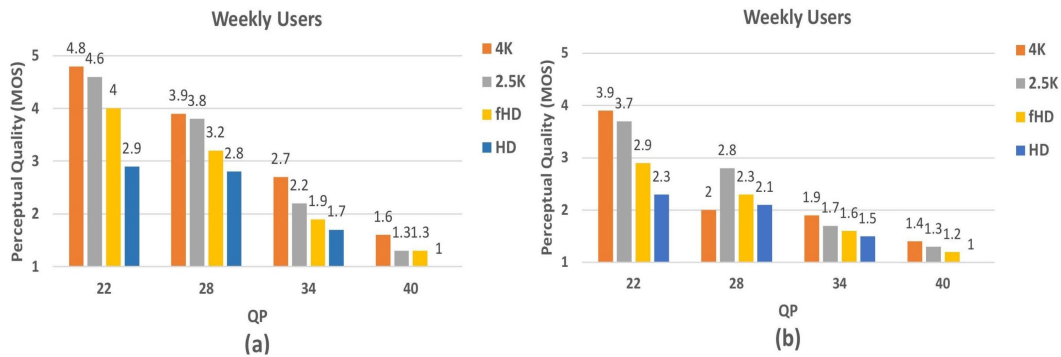

FIGURE 6. Subject's perceptual quality against QP and resolution watching weekly 360° video with HMD (a) Interesting 360° videos (b) Non-Interesting 360° videos.

users because there is a clear difference between users' MOS score. Therefore, it recommends that the user's prior experience of watching 360° video with VR has a notable impact on perceptual quality.

3) QP AND RESOLUTIONS

The impact of QP and resolution on perceptual quality is evident in Figures 4, 5, and 6. In all three figures, the MOS values for 4K is higher than other resolutions, even in interesting and

non-interesting cases. Similarly, the MOS score for QP22 is higher than all other QPs in all three figures. In most cases, the MOS score for HD resolutions is lower than 3, which suggests that the content provider should avoid HD resolution when encoding 360° video. Besides, the MOS score for QP22 and QP28 in all cases is higher than 3MOS. Therefore, it is clear that viewers feel comfortable at QP22 and QP28, even watching both interesting and non-interesting videos.

Another notable fact from all three Figures (Figure 4, 5, and 6) is that users having weekly experience are more tolerant and less sensitive with given QP and resolution when watching interesting 360° video. At the same time, these weekly users are very conscious about resolution and QP. Therefore, we can say that weekly users are regular viewers of VR videos and their MOS results are significant. Interestingly these viewers are also susceptible to video resolution in both interesting and non-interesting cases when they watch fHD and HD virtual reality videos. Therefore, based on these observations, content providers should take into account the regular VR viewers' requirements and interests, which meets the end user's satisfaction and expectations.

4) RENDERING DEVICE
Figure 7 shows the impact of two rendering devices on perceptual quality against four resolutions. From the subjective test, the MOS value for HTC Vive is recorded higher than Google Cardboard. For 4K resolution, the HTC Vive MOS value is 4.8 while the MOS is recorded 4.3 when subjects watch videos with 4K resolution in Google Cardboard. It can be seen that the average MOS difference between these rendering devices at 4K resolution videos is 0.5. It is also observed that the average MOS difference between these both devices is 0.1 when subjects watch 360° video with HD resolution, which is negligible. It means that viewers QoE is almost the same for both devices in case of HD resolutions. From Figure 7, we can conclude that the MOS score of subjects is higher in all four-resolution cases when they watch through HTC Vive. Therefore, these results suggest that the rendering device also has an impact on viewers QoE. Hence, the HMD device type should be taken into account while investigating the QoE of 360° videos.
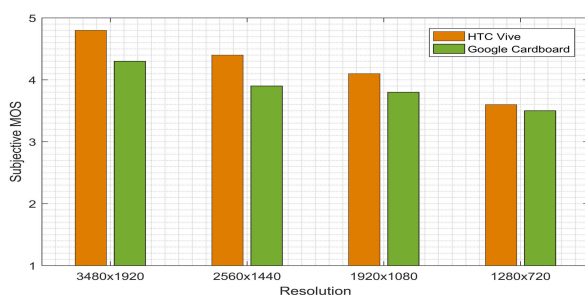

FIGURE 7. Impact of rendering device on Perceptual Quality.

5) GENDER
Figure 8 depicts the average impact score of males and females based on the user's familiarity with VR on perceptual quality. One can see that the perceptual quality of males is higher for the first time and rarely users compared to female viewers. Interestingly, in the case of weekly users, the MOS value for the perceptual quality of females is higher. It suggests that the perceptual quality judgment of female
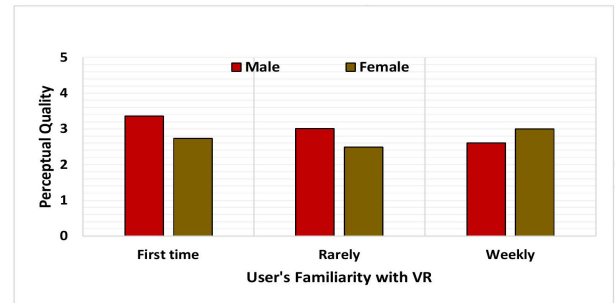

FIGURE 8. Impact of Gender on Perceptual Quality based on their familiarity with VR.

users watching 360° video in VR regularly (weekly) is higher than males.

B. IMPACT ON CYBERSICKNESS
1) USER's FAMILIARITY WITH VR
Figure 9 depicts the impact of the user's familiarity with VR videos on cybersickness. Each bar represents the cybersickness level (SSQ score) of each user for each sample video based on their familiarity with VR (a) first time users (b) weekly users, and (c) rarely users. Figure 9 (a) shows that users watching the first time in VR feels more cybersickness as compared to rarely users and weekly users shown in Figure 9 (b) and Figure 9 (c). Weekly users are less affected by cybersickness as compared to rarely users and first-time users. Interestingly the perceptual quality score of first-time users is higher as we see earlier in Figure 4, this might be because subjects having the first experience of VR are more excited than other users.

2) USER's INTEREST
Similarly, the impact of the user's interest on cybersickness is shown in Figure 10. Each bar represents the cybersickness level of each user for each sample video based on their interest (a) interested and (b) not interested. It can be seen that the subject's level of cybersickness is higher while watching non-interesting 360° video (the content they are not interested in watching). Hence, it is crucial to consider the user's interest while evaluating the level of cybersickness in a virtual environment.

3) GENDER
The subjective results reveal that different genders have a distinct level of cybersickness while watching videos in a virtual environment. Females report a higher level of cybersickness and feel very dizzy than males. During the subjective test, out of 10 female feels very dizzy and recorded higher cybersickness shown in Figure 11. Similarly, each bar represents the cybersickness level of each user for each sample video based on their gender (a) male and (b) female. It recommends that females are more prone to cybersickness than males, which affect QoE. Therefore keeping in mind the impact of gender on cybersickness, the researcher and content provider
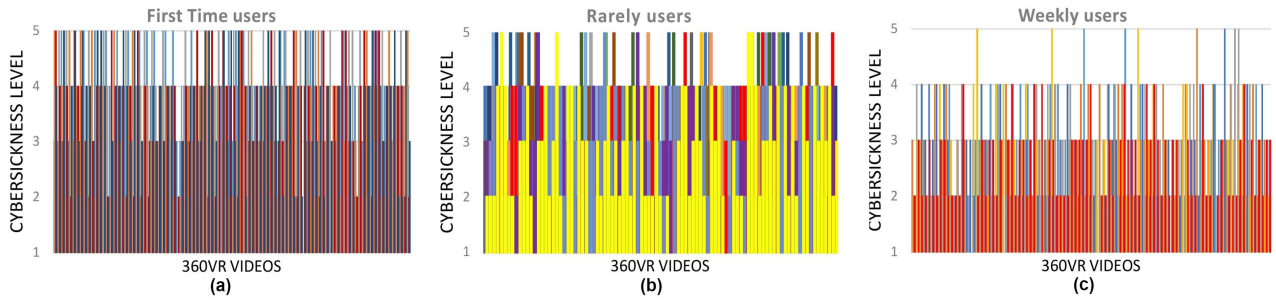
FIGURE 9. Impact of user's familiarity on Cybersickness (SSQ score from 1 to 5 (5 = very dizzy, 1 = not dizzy)).
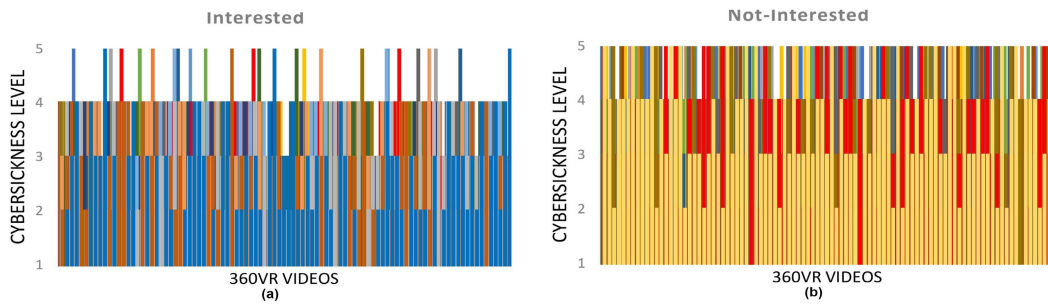

FIGURE 10. Impact of user's interest on Cybersickness (SSQ score from 1 to 5 (5 = very dizzy, 1 = not dizzy)).
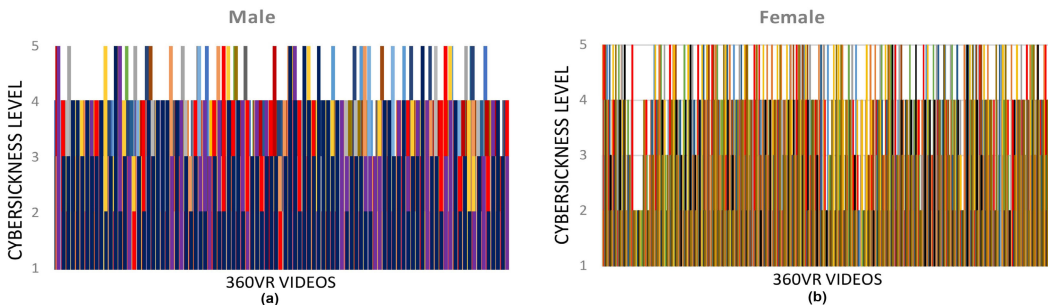

FIGURE 11. Impact of Gender on Cybersickness (SSQ score from 1 to 5 (5 = very dizzy, 1 = not dizzy)).

should take into account the gender before evaluating QoE for 360-degree videos. To easily understand the effect of all three factors on cybersickness, Figure 12 depicts the average impact of the user's familiarity, user's interest, and gender on cybersickness. Each bar represents the average SSQ score of users based on their familiarity, interest, and gender.

V. MACHINE LEARNING QoE PREDICTION

To predict the QoE of 360° video in VR, we choose supervised machine learning techniques. The machine learning technique is an algorithm that can predict output by giving input independent variables. It takes a known dataset as an input and its known response (output) to learn the classification model. The learning algorithm then trains the model to predict the response to a new dataset. We collect the dataset from a subjective experiment for training the model. To assess the machine learning to predict the QoE, we modelled and

designed four classification methods in python, including LR, kNN, SVM, and DT.

A. LOGISTIC REGRESSION (LR)

LR is a classification algorithm used when the target variable is dichotomous (1 or 0). We use binary LR for the prediction of QoE. Binary LR is a classification method that predicts whether the end-users' QoE is "satisfactory" or "unsatisfactory." In binary classification, the output dependent variables are dichotomous, having only two possible values (binary). Therefore, we distribute the sample dataset into binary values "0" and "1" to label the output response. For classification purposes, we categorize the six factors shown in Table 1 and distribute the MOS values (independent variables) into "0" as unsatisfactory QoE and "1" as satisfactory QoE (output dependent variables). The MOS >= 3 is "1," and MOS < 3 is
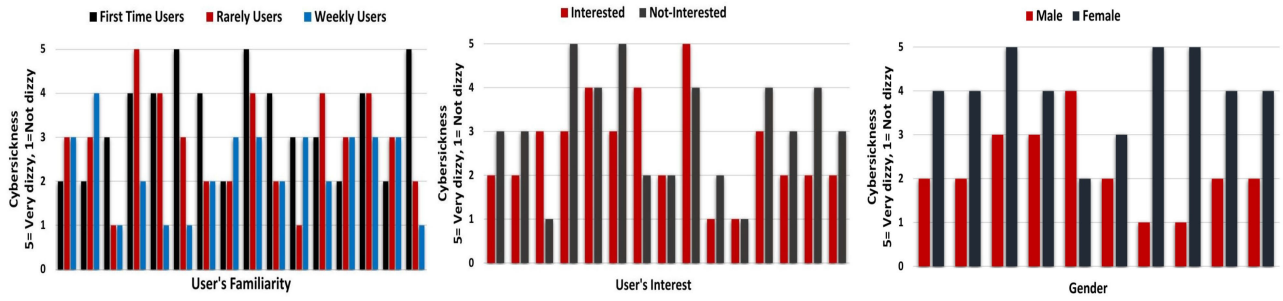
FIGURE 12. Average Impact of User's familiarity, User's interest, and Gender on Cybersickness.
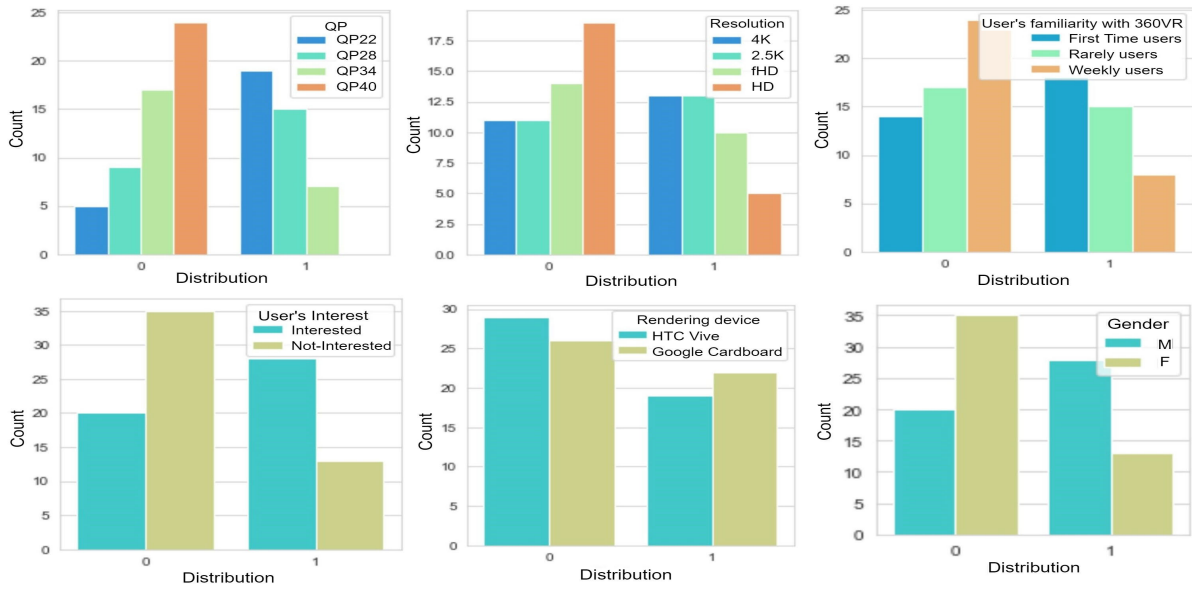

FIGURE 13. Factors Evaluation and MOS Distributions (0 means MOS<3 and 1 means MOS >=3).

considered "0". Figure 13 shows the MOS distribution of all factors. Note that, only the MOS values of perceptual quality is considered in this section. We do not include the subjective results of cybersickness in this section. The cybersickness prediction will be explained in detail in Section VIII.

LR model establishes the relationship between output dependent binary values and a group of input variables to model the probability of logit transform with independent input variables. Considering that Y is the output binary (dependent) variable indicates unsatisfactory/satisfactory QoE with [0, 1], and p is the probability of Y to be 1 then, $p = P(Y = 1)$ and considering that $\{x_1, x_2, \ldots x_k\}$ is a group or set of input variables (independent) shown in Table 3. Then the LR of Y on $\{x_1, x_2, \ldots x_k\}$ estimates the parameter value for regression coefficient $\{\beta_0, \beta_1, \ldots, \beta_k\}$ via maximum likelihood estimation written as,

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k \quad (1)$$

where $\{\beta_1, \ldots, \beta_k\}$ is the corresponding regression coefficient for each factor, and β_0, is the intercept parameter. P is

the probability of QoE being "satisfactory," while $1 - P$ is the probability of QoE being "unsatisfactory."

$$\frac{1-p}{p} = \frac{1}{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)} \quad (2)$$

$$\frac{1}{p} = \frac{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k) + 1}{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)} \quad (3)$$

The probability that $P(Y = 1)$ can be calculated by the equation written as,

$$p = \frac{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)} \quad (4)$$

Applying the seventeen input independent variables shown in Table 3. Thus rewriting equation (1) as,

$$\begin{aligned}\text{logit}(p) \\ = \log\left(\frac{p}{1-p}\right) &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 \\ &+ \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9 + \beta_{10} x_{10} + \beta_{11} x_{11} \\ &+ \beta_{12} x_{12} + \beta_{13} x_{13} + \beta_{14} x_{14} + \beta_{15} x_{15} + \beta_{16} x_{16} + \beta_{17} x_{17}\end{aligned}$$

$$(5)$$

TABLE 3. Categorization of Features.

Factors	Features	Variables
QPs	QP22	x_1
	QP28	x_2
	QP34	x_3
	QP40	x_4
Resolution	4K	x_5
	2.5K	x_6
	fhd	x_7
	HD	x_8
Rendering Device	HTC Vive	x_9
	Google Cardboard	x_{10}
Gender	Male	x_{11}
	Female	x_{12}
User's Interest	Interested	x_{13}
	Not-Interested	x_{14}
User's familiarity with VR	First time	x_{15}
	Rarely	x_{16}
	Weekly	x_{17}

FIGURE 14. Receiver Operating Characteristic Curve (ROC).

Therefore, considering all factors equation (4) can be written as,

$$p = \frac{\exp\left(\beta_0 + \beta_1 x_1 + \cdots + \beta_{17} x_{17}\right)}{1 + \exp\left(\beta_0 + \beta_1 x_1 + \cdots + \beta_{17} x_{17}\right)} \quad (6)$$

Based on equation (6), the probability that QoE is "satisfactory" or "unsatisfactory" can be predicted, i.e., when MOS >= 3, the equation predicts 1 (satisfactory) and when MOS < 3, it predicts 0 (unsatisfactory).

VI. EXPERIMENTS EVALUATIONS AND PERFORMANCE COMPARISON

We used python programming language for machine learning algorithms. Three steps are taken to accomplish the classification and prediction task: Dataset collection, training and testing of machine learning model, and evaluation of the prediction. The dataset obtained from a subjective experiment and the binary LR model trained on labelled data by using the scikit machine-learning library. The algorithm examines the training data and predicts the QoE on testing data. We used 70% of the subjective dataset for training and 30% for testing. The estimated parameter values for the regression coefficients and intercept are shown in Table 4.

TABLE 4. Estimated regression coefficient and constant values.

Coefficient	Value	Coefficient	Value
β_0	−0.86964138	β_9	0.53123255
β_1	1.11907364	β_{10}	−1.40087393
β_2	1.18276753	β_{11}	0.4663474
β_3	0.439487	β_{12}	0.05049784
β_4	−0.87041256	β_{13}	−1.38648662
β_5	0.38731463	β_{14}	−0.43298509
β_6	0.31586766	β_{15}	−0.43665629
β_7	−0.41540454	β_{16}	−0.30642439
β_8	−1.15741913	β_{17}	−0.563217

In order to evaluate the ability of binary LR model, two essential evaluation curves, Receiver Operating Characteristic Curve (ROC) and Area Under Curve (AUC) are applied,
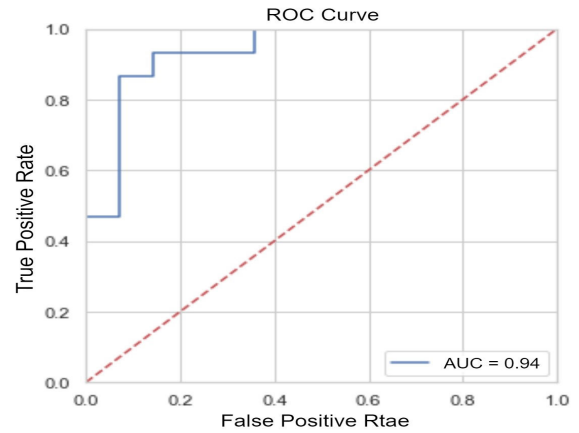
as shown in Figure 14. The solid line interprets the ROC curve of a binary logistic model, while the area under curve (AUC) is 0.94, which is > 0.5. It indicates that the model performs well as the more area under the curve, the more the model performs well.

A. K-NEAREST NEIGHBORS (kNN)

The k-nearest-neighbors algorithm is a supervised machine learning classification technique that uses a group of labelled data. Based on these labelled data, the algorithm learns how to label the other points. To label a new data point, the kNN algorithm search at the nearest label point that is closest (neighbor) to that new data point having those neighbours vote. Therefore, whatever label the most of the nearest neighbour have, will be the label for that new data point. *K* is the number of neighbours the algorithm checks.

We used the scikit-learn library in python to implement the kNN algorithm. To find the best value of *K*, we execute a loop for *k*, ranging from 1 to 40, to allow the algorithm to check up to 40 neighbours. To do so, we first calculate the mean error rate for all the predicted outcomes, where the range of *K* is kept from 1 to 40. In each iteration the mean error for the predicted outcome of the test set is calculated and then added to the list of errors. The mean error value against *K* ranging from 1 to 40 is shown in Figure 15. It is observed that the most optimal value of *K* is 19, 29, 30, 31, 35, and 37 for the given dataset, where the mean error is recorded zero.

B. SUPPORT VECTOR MACHINE (SVM)

Support Vector Machine (SVM) classifies data points by separating them with a linear decision boundary called a hyperplane. SVM split all data points of one class from another class and identify best separating hyperplane that maximizes the boundary between classes of training data points. SVM algorithm takes a supervised training (labelled) data and determines an optimal hyperplane to identify the most significant possible distance to minimize the upper bound. In SVM, support vectors are the coordinates of data that are closest to
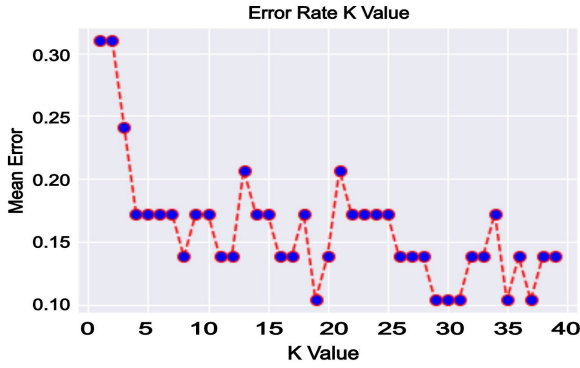
FIGURE 15. Error rate for K values.

the best separating boundary (hyperplane), which gives the complete suitable information for SVM classification. The prediction performance of the SVM algorithm also depends on the kernel parameters. For optimization, we tune the hyper-parameters using grid search. Two functions of kernel parameters, linear and radial basis function (rbf) are applied. In addition, the different parameter of c is tested. Finally, we achieved the highest accuracy at $c=1000$ and kernel= rbf. The algorithm performs well in rbf case and achieved 79% accuracy while 72% recorded with linear.

C. DECISION TREE (DT)

Decision tree (DT) is a machine-learning technique used for classification problems. DT split the dataset into a smaller subset and incrementally construct a binary tree using the features and thresholds that returns the maximum information gain at each node. After splitting the dataset, the information gain depends on the decrease in entropy because the more data is impure, the more entropy will be. Gini index is the most prominent to measure the information gain. Therefore, the probability of getting the desired QoE outcome, the algorithm weight the entropy calculated for that desired QoE value. The outcome is a tree with decision nodes and leaf nodes. Figure 16 shows the DT graph generated on the subjective experiment dataset. The final decision is the prediction of the algorithm on the bases of training data that classify the users' QoE in class 1 (satisfactory) and class 0 (unsatisfactory).

We used 70% of subjective data as a training set and 30% as a test set to check the accuracy and performance of machine-learning QoE prediction algorithms. We compared the four ML classification algorithms' performance with respect to confusion matrix, accuracy rate, precision, recall, f1-score, and MAE.

1) CONFUSION MATRIX
The confusion matrix is a performance measurement for ML classification that provides the classification of the correct match rates for predicted against actual class. It gives the four different combinations of actual and predicted values.
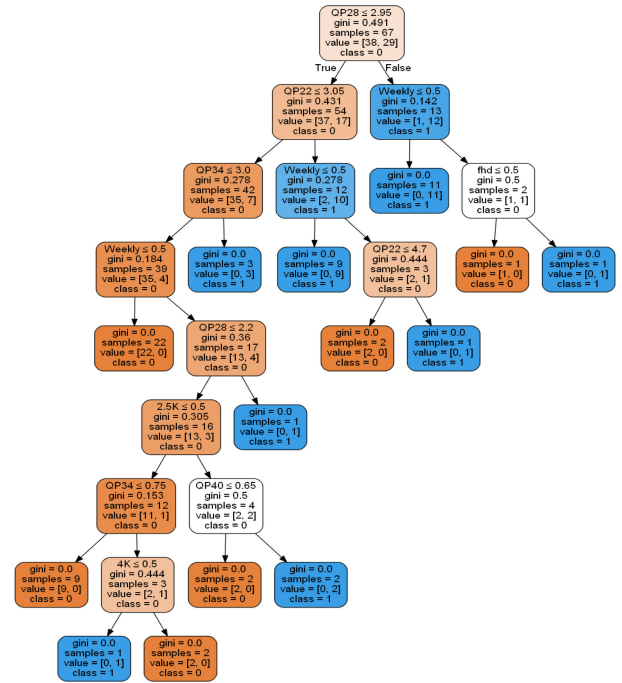

FIGURE 16. Decision Tree Graph.

True positive of confusion matrix interpret correct outcome while false negative reflects the miss percentage. LR recorded 86% accuracy while kNN and DT 82% and 79%, respectively. SVM accuracy with rbf kernel is 79%, and with the linear kernel is 72%. A confusion matrix is prominent to measure precision, recall, and F1-score.

2) PRECISION
Precision is measured

$$\text{Precision} = \frac{TP}{TP + FP} \tag{7}$$

3) RECALL
It is the measure of how many of the actual positive a model predicted, through labeling it true positive. That is,

$$\text{Recall} = \frac{TP}{TP + FN} \tag{8}$$

4) F1-SCORE
F1 is the accuracy measure of test and weighted harmonic mean of the precision and recall. The final value of the F1 measure is between 1 and 0. The closer the final F1-score to 1, the better the accuracy of the test. The equation to calculate the F1-measure is.

$$\text{F1} = 2 \times \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{9}$$

5) OVERALL ACCURACY
The percentage of the correct prediction is the overall accuracy that is,

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

where, TP is true positive, TN is true negative, FP is false positive and FN is false negative.

6) MEAN ABSOLUTE ERROR (MAE)

Mean absolute error calculates the average difference between actual values and predicted values. The difference error is proportional to the absolute difference between calculated and actual values. The following equation can measure MAE.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} \left| Y_{actual_i} - Y_{\text{calculated } i} \right| \qquad (11)$$

The calculated value of precision, recall, f1-score, MAE, and overall accuracy is shown in detail in Table 5.

TABLE 5. Performance comparison of ML algorithms for QoE prediction in terms of perceptual quality.

Method	Precision	Recall	f1-Score	MAE	Accuracy
kNN	0.85	0.83	0.83	0.17	82%
SVM	0.80	0.79	0.79	0.21	79%
LR	0.87	0.86	0.86	0.14	**86%**
DT	0.85	0.79	0.78	0.21	79%

For validation purpose, we choose the linear regression-based model [33], Back Propagation Neural Network (BPNN) [50], and Bayesian Inference Method (BIM) [45] as the prediction performance method. We compared these methods from the perspective of Pearson Linear Correlation Coefficient (PLCC) and Spearman's Rank-order Correlation Coefficient (SRCC) on the test set. Furthermore, the proposed QoE method is compared against the objective quality metrics designed for 360° videos such as S-PSNR, WS-PSNR, and CPP-PSNR [51]. The PLCC and SRCC are calculated to evaluate the performance comparison of the proposed models. Figure 17 shows the PLCC and SRCC to compare the prediction performance of the proposed model. It can be seen that our proposed LR-based model delivers the best performance in predicting the subject's QoE.
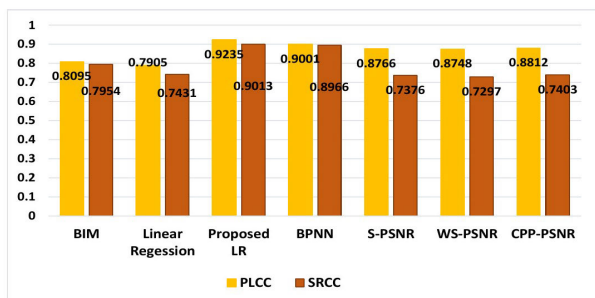

FIGURE 17. PLCC and SRCC Performance Comparison of the Proposed Model against existing QoE methods.

VII. CYBERSICKNESS PREDICTION BASED ON ANN

After subjectively analyzing the impact of user's interest, user's familiarity with VR, and gender on QoE in terms of cybersickness, we build a model based on ANN that can assist content providers to keep in mind the viewer's preferences and interests. It can also help researchers to take into account the mentioned QoE-affecting factors while evaluating and predicting the cybersickness level in virtual reality. In our ANN model, we used four layers based on the Stochastic Gradient Descent (SGD) shown in Figure 18. Our model comprises one input layer, two hidden layers, and one output layer. $X_1, \ldots, X_7$ are input neurons, which is seven in our model. We use seven features of three QoE-affecting factors (user's interest, user's familiarity with VR and gender) as input nodes and five output nodes, $Y_1, \ldots, Y_5$ represent QoE in terms of cybersickness which is {1, 2, 3, 4, 5}. We used two hidden layers, $h_1^1, h_2^1, \ldots, h_n^1$ represents neurons in first hidden layer while $h_1^2, h_2^2, \ldots, h_n^2$ indicates the neurons in second hidden layer.
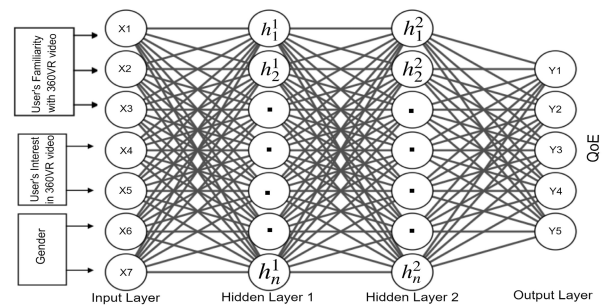

FIGURE 18. The structure of Artificial Neural Network (ANN).

The aim of using ANN for cybersickness prediction is to map these influence factors into a range from 1 to 5. We use a high-level neural network Keras library that runs on the top of TensorFlow. Keras offers the SGD optimizer with a learning rate. For optimization purpose, we firstly fix the SGD learning rate to 0.2, the number of neurons is then adapted in the hidden layer. Finally, we get the 64 neurons in the first hidden layer and 32 neurons in a second hidden layer that performs well with a prediction accuracy of 85%. We used 1000 epochs (iteration) during training the model. We also noticed that the final accuracy of the network varies with the tuning of learning rate. Therefore, we test the prediction accuracy of the proposed model against different learning rates. Figure 19 depicts the change in prediction accuracy with different learning rates.

To check the validity of the proposed ANN-based QoE prediction model, we compare the performance accuracy of our model against well known QoE prediction methods in terms of cybersickness. The PLCC and SRCC are computed to compare the prediction result of the proposed method and exiting QoE models. We choose the existing QoE prediction model in terms of cybersickness such as VR sickness predictor (VRSP) [31], VR sickness assessment (VRSA) network [52], and deep learning visual comfort assessment (VCA) method [53] as a performance comparing methods. Figure 20 shows the computed PLCC and SRCC of the proposed and existing QoE models in terms of cybersickness. It can be seen that our
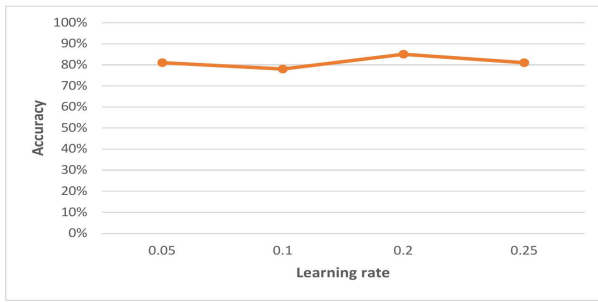
FIGURE 19. Cybersickness prediction with different learning rates in ANN.
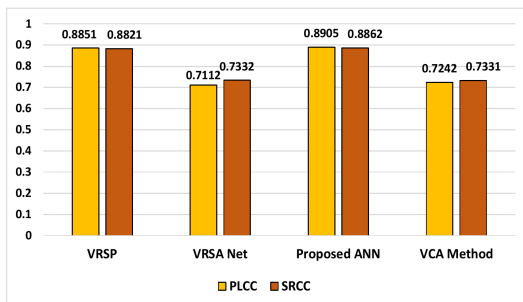

FIGURE 20. PLCC and SRCC Performance Comparison of the Proposed Model against existing QoE methods in terms of Cybersickness.

proposed ANN-based model delivers the best performance in predicting the subject's QoE in terms of cybersickness.

VIII. CONCLUSION

This manuscript has utilized the machine learning algorithms for the QoE prediction of 360° videos in virtual reality. Six QoE influence factors such as QPs, resolution, rendering device, gender, user's familiarity with VR videos, and user's interest in 360° video are considered. The QoE is investigated subjectively on 96 video samples in two important QoE aspects, perceptual quality, and cybersickness. It is suggested that the user's interest has a vital role in evaluating the QoE. Similarly, the subject's cybersickness level is higher while watching a non-interesting 360° video. Besides, the user's prior experience of watching 360° video in VR has a notable impact on perceptual quality. Regarding the effect of gender on cybersickness, female users reported more severe cybersickness than males. The results from the subjective dataset used as training data for the prediction of QoE. Four supervised machine-learning algorithms are trained on subjective data, including LR, kNN, DT, and SVM. We proposed a binary LR technique for the QoE prediction of 360° video. We compared the prediction results of the proposed method against kNN, SVM, and DT with respect to accuracy rate, recall, f1-score, precision, and mean absolute error (MAE). The proposed method performed well and recorded an 86% accuracy rate against other supervised machine learning techniques, which is in close agreement with subjective opinion. We also build a neural network model for QoE prediction in terms of cybersickness. We observe that ANN performs

well at a 0.2 learning rate. Finally, the prediction accuracy of the proposed LR-based model is compared against the state of the art QoE methods in terms of perceptual quality, and the accuracy performance of the ANN model is compared against well known QoE models in terms of cybersickness. Both proposed models perform well against existing QoE methods.

In our future work, we will take into account deep learning methods to predict QoE for more influencing factors such as content type (fast, slow, medium), number of moving targets in the content, camera motion, and exposure time. Besides, few other QoE aspects will be considered, such as presence, acceptability, usability, and immersion.

REFERENCES

[1] D. V. Nguyen, H. T. T. Tran, A. T. Pham, and T. C. Thang, "An optimal tile-based approach for viewport-adaptive 360-degree video streaming," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 9, no. 1, pp. 29–42, Mar. 2019.

[2] M. Hosseini and V. Swaminathan, "Adaptive 360 VR video streaming: Divide and conquer," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Dec. 2016, pp. 107–110.

[3] Y. Liu, S. Dey, F. Ulupinar, M. Luby, and Y. Mao, "Deriving and validating user experience model for DASH video streaming," *IEEE Trans. Broadcast.*, vol. 61, no. 4, pp. 651–665, Dec. 2015.

[4] M. Seufert, F. Wamser, P. Casas, R. Irmer, P. Tran-Gia, and R. Schatz, "YouTube QoE on mobile devices: Subjective analysis of classical vs. Adaptive video streaming," in *Proc. Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Aug. 2015, pp. 43–48.

[5] P. Spachos, W. Li, M. Chignell, A. Leon-Garcia, L. Zucherman, and J. Jiang, "Acceptability and quality of experience in over the top video," in *Proc. IEEE Int. Conf. Commun. Workshop (ICCW)*, Jun. 2015, pp. 1693–1698.

[6] H. T. Tran, N. P. Ngoc, T. Hoßfeld, and T. C. Thang, "A cumulative quality model for HTTP adaptive streaming," in *Proc. 10th Int. Conf. Qual. Multimedia Exper. (QoMEX)*, May/Jun. 2018, pp. 1–6.

[7] W. Song and D. W. Tjondronegoro, "Acceptability-based QoE models for mobile video," *IEEE Trans. Multimedia*, vol. 16, no. 3, pp. 738–750, Apr. 2014.

[8] A. Steed, S. Frlston, M. M. Lopez, J. Drummond, Y. Pan, and D. Swapp, "An 'In the Wild' experiment on presence and embodiment using consumer virtual reality equipment," *IEEE Trans. Vis. Comput. Graphics*, vol. 22, no. 4, pp. 1406–1414, Apr. 2016.

[9] H. T. T. Tran, T. H. Hoang, P. N. Minh, N. P. Ngoc, and T. C. Thang, "A perception-based quality metric for omnidirectional images," in *Proc. IEEE Int. Conf. Consum. Electron.-Asia (ICCE-Asia)*, Jun. 2019, pp. 12–14.

[10] 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.*, Jun. 2019, pp. 232–242.

[11] H. T. T. Tran, N. P. Ngoc, C. T. Pham, Y. J. Jung, and T. C. Thang, "A subjective study on user perception aspects in virtual reality," *Appl. Sci.*, vol. 9, no. 16, p. 3384, Aug. 2019.

[12] 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 Exper. (QoMEX)*, May 2017, pp. 1–6.

[13] V. Vasilev, J. Leguay, S. Paris, L. Maggi, and M. Debbah, "Predicting QoE factors with machine learning," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2018, pp. 1–6.

[14] R. Huang, X. Wei, Y. Gao, C. Lv, J. Mao, and Q. Bao, "Data-driven QoE prediction for IPTV service," *Comput. Commun.*, vol. 118, pp. 195–204, Mar. 2018.

[15] U. Bulkan and T. Dagiuklas, "Predicting quality of experience for online video service provisioning," *Multimedia Tools Appl.*, vol. 78, pp. 18787–18811, Feb. 2019.

[16] 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.

[17] A. Doumanoglou, D. Griffin, J. Serrano, N. Zioulis, T. K. Phan, D. Jimenez, D. Zarpalas, F. Alvarez, M. Rio, and P. Daras, "Quality of experience for 3-D immersive media streaming," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 379–391, Jun. 2018.

[18] T. Zhao, Q. Liu, and C. W. Chen, "QoE in video transmission: A user experience-driven strategy," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 1, pp. 285–302, 1st Quart., 2017.

[19] 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.

[20] H. Yuan, X. Hu, J. Hou, X. Wei, and S. Kwong, "An ensemble rate adaptation framework for dynamic adaptive streaming over HTTP," *IEEE Trans. Broadcast.*, vol. 66, no. 2, pp. 251–263, Jun. 2020.

[21] 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.

[22] 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.

[23] J. B. Husić, S. Baraković, E. Cero, N. Slamnik, M. Oćuz, A. Dedović, and O. Zupčić, "Quality of experience for unified communications: A survey," *Int. J. Netw. Manage.*, vol. 30, no. 3, May 2020.

[24] C. Li, M. Xu, S. Zhang, and P. Le Callet, "State-of-the-art in 360° video/image processing: Perception, assessment and compression," 2019, *arXiv:1905.00161*. [Online]. Available: http://arxiv.org/abs/1905.00161

[25] C.-L. Fan, W.-C. Lo, Y.-T. Pai, and C.-H. Hsu, "A survey on 360° video streaming: Acquisition, transmission, and display," *ACM Comput. Surv.*, vol. 52, no. 4, p. 71, 2019.

[26] 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.

[27] X. Liu, Q. Xiao, V. Gopalakrishnan, B. Han, F. Qian, and M. Varvello, "360° innovations for panoramic video streaming," in *Proc. 16th ACM Workshop Hot Topics Netw.*, 2017, pp. 50–56.

[28] H. Yuan, S. Zhao, J. Hou, X. Wei, and S. Kwong, "Spatial and temporal consistency-aware dynamic adaptive streaming for 360-degree videos," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 1, pp. 177–193, Jan. 2020.

[29] I.-S. Comsa, R. Trestian, and G. Ghinea, "360° mulsemedia experience over next generation wireless networks-a reinforcement learning approach," in *Proc. 10th Int. Conf. Qual. Multimedia Exper. (QoMEX)*, May/Jun. 2018, pp. 1–6.

[30] C. Perfecto, M. S. Elbamby, J. Del Ser, and M. Bennis, "Taming the latency in multi-user VR 360°: A QoE-aware deep learning-aided multicast framework," *IEEE Trans. Commun.*, vol. 68, no. 4, pp. 2491–2508, Apr. 2020.

[31] J. Kim, W. Kim, S. Ahn, J. Kim, and S. Lee, "Virtual reality sickness predictor: Analysis of visual-vestibular conflict and VR contents," in *Proc. 10th Int. Conf. Qual. Multimedia Exper. (QoMEX)*, May 2018, pp. 1–6.

[32] 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.

[33] S.-H. Yao, C.-L. Fan, and C.-H. Hsu, "Towards quality-of-experience models for watching 360° videos in head-mounted virtual reality," in *Proc. 11th Int. Conf. Qual. Multimedia Exper. (QoMEX)*, Jun. 2019, pp. 1–3.

[34] S. Kasahara, S. Nagai, and J. Rekimoto, "First person omnidirectional video: System design and implications for immersive experience," in *Proc. ACM Int. Conf. Interact. Exper. TV Online Video*, 2015, pp. 33–42.

[35] 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.

[36] 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.

[37] N. Padmanaban, T. Ruban, V. Sitzmann, A. M. Norcia, and G. Wetzstein, "Towards a machine-learning approach for sickness prediction in 360 stereoscopic videos," *IEEE Trans. Vis. Comput. Graphics*, vol. 24, no. 4, pp. 1594–1603, Apr. 2018.

[38] H. G. Kim, W. J. Baddar, H.-T. Lim, H. Jeong, and Y. M. Ro, "Measurement of exceptional motion in VR video contents for VR sickness assessment using deep convolutional autoencoder," in *Proc. 23rd ACM Symp. Virtual Reality Softw. Technol.*, Nov. 2017, pp. 1–7.

[39] G. Dimopoulos, I. Leontiadis, P. Barlet-Ros, and K. Papagiannaki, "Measuring video QoE from encrypted traffic," in *Proc. ACM Internet Meas. Conf. (IMC)*, 2016, pp. 513–526.

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

[41] T. Mangla, E. Halepovic, M. Ammar, and E. Zegura, "emimic: Estimating HTTP-based video QoE metrics from encrypted network traffic," in *Proc. Netw. Traffic Meas. Anal. Conf. (TMA)*, Jun. 2018, pp. 1–8.

[42] M. Seufert, P. Casas, N. Wehner, L. Gang, and K. Li, "Stream-based machine learning for real-time QoE analysis of encrypted video streaming traffic," in *Proc. 22nd Conf. Innov. Clouds, Internet Netw. Workshops (ICIN)*, Feb. 2019, pp. 76–81.

[43] S. Wassermann, P. Casas, M. Seufert, and F. Wamser, "On the analysis of youtube QoE in cellular networks through in-smartphone measurements," 2019.

[44] Q. Wang, H.-N. Dai, D. Wu, and H. Xiao, "Data analysis on video streaming QoE over mobile networks," *EURASIP J. Wireless Commun. Netw.*, vol. 2018, no. 1, p. 173, Dec. 2018.

[45] A. Muhammad, W. Jing, A. Ullah, K. Wahab, A. Sadique, and F. Zesong, "Measuring quality of experience for 360° videos in virtual reality," *Sci. China Inf. Sci.*, to be published. [Online]. Available: http://engine.scichina.com/publisher/scp/journal/SCIS/doi/10.1007/s11432-019-2734-y?slug=fulltext

[46] M. Xu, Y. Song, J. Wang, M. Qiao, L. Huo, and Z. Wang, "Predicting head movement in panoramic video: A deep reinforcement learning approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 11, pp. 2693–2708, Nov. 2019.

[47] 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 Conf. Comput. Commun. (INFOCOM)*, Apr. 2019, pp. 1252–1260.

[48] *Methods for the Subjective Assessment of Video Quality, Audio Quality and Audiovisual Quality of Internet Video and Distribution Quality Television in Any Environment*, Standard ITU-T Rec. P.913, International Telecommunication Union, 2014.

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

[50] L. Wenjing, Y. Peng, W. Ruiyi, F. Lei, D. Ouzhou, and Q. Xuesong, "Quality of experience evaluation of HTTP video streaming based on user interactive behaviors," *J. China Univ. Posts Telecommun.*, vol. 24, no. 3, pp. 24–32, 2017.

[51] E. Upenik, M. Rerabek, and T. Ebrahimi, "On the performance of objective metrics for omnidirectional visual content," in *Proc. 9th Int. Conf. Qual. Multimedia Exper. (QoMEX)*, May 2017, pp. 1–6.

[52] H. G. Kim, H.-T. Lim, S. Lee, and Y. M. Ro, "VRSA net: VR sickness assessment considering exceptional motion for 360° VR video," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1646–1660, Apr. 2018.

[53] H. Jeong, H. G. Kim, and Y. M. Ro, "Visual comfort assessment of stereoscopic images using deep visual and disparity features based on human attention," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 715–719.

MUHAMMAD SHAHID ANWAR received the M.Sc. degree 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. He has published many research articles in peer-reviewed journals and conferences in the area of 360VR videos and QoE. His research interests include the area of 360-degree videos, virtual reality (VR), and Quality-of-Experience (QoE) evaluations. His current research interest includes machine learning-based QoE prediction of 360-degree videos.

JING WANG (Member, IEEE) received the B.S. degrees in communication and information system in 2002 and 2007, respectively, and the Ph.D. degree in electronic engineering from the Beijing Institute of Technology (BIT). She is currently an Associate Professor with BIT, where she is also with the Research Institute of Communication Technology. She has 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 the Audio Engineering Society (AES) and the China Computer Federation (CCF), a Senior Member of the China Institute of Communications (CIC) and the Chinese Institute of Electronics (CIE), and an Expert Member of the Digital Audio and Video Coding Standard Workgroup (AVS) in China.

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, 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 School of Computing, Riphah International University, Faisalabad, Pakistan. His research interests include digital image processing, computer vision, and pattern recognition; and have several publications in these areas.

SADIQUE AHMAD received the Ph.D. degree in computer science and technology from the Beijing Institute of Technology, China. He is currently working as an Assistant Professor with the School of Computer Sciences and Technology, Iqra University, Karachi. He has published 22 research articles 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). His main focus of his 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.

ZESONG FEI (Senior Member, IEEE) received the B.Eng. degree in electronic engineering and the Ph.D. degree in electronic engineering from the Beijing Institute of Technology (BIT), China, in 1999 and 2004, respectively. From 2012 to 2013, he was a Visiting Scholar with The University of Hong Kong, Hong Kong. He is currently a Professor with the Research Institute of Communication Technology (RICT), BIT. His current research interests include error control coding, physical-layer security, cooperative communications, and Quality of Experience. He has published over 30 manuscripts in the IEEE journals. 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 Special Issue on Error Control Coding of *Wireless Communications and Mobile Computing* and *China Communications*.

. . .