Quality of Multimedia Experience Prediction using Peripheral Physiological Signals

Sowmya Vijayakumar[1], Ronan Flynn[1], Peter Corcoran[2], and Niall Murray[1]

[1]*Department of Computer and Software Engineering, Technological University of the Shannon, Athlone, Ireland*
[2] *Electrical and Electronic Engineering, University of Galway, Galway, Ireland*

Abstract

This paper proposes the utilization of physiological signals for quality of experience (QoE) assessment by employing machine learning and deep learning classifiers. Accurately predicting user QoE by analysing physiological signals holds significant potential in diverse fields, including human-computer interaction, healthcare, and education. To predict various QoE factors from physiological signals, the experiments were conducted on two datasets: SoPMD Dataset 1 and SoPMD Dataset 2. The bidirectional long-short-term memory (BLSTM), support vector machine, k-nearest neighbour and random forest algorithms were evaluated using fused electrocardiogram and respiration signals to predict subjective QoE scores, including perceived quality levels, user preference, and the sense of presence. The results demonstrate the effectiveness of the models, with BLSTM emerging as the top-performing algorithm across most experiments, achieving high classification F1-scores. These findings suggest that the physiological signals can be effectively used in the classification of subjective QoE scores.

Keywords: Deep Learning, Machine Learning, Physiological Signals, Quality of Experience.

1 Introduction

In today's digital era, the quality of user experience (QoE) has become a crucial aspect of the success of applications and services. QoE, a multidimensional concept, is defined as "*the degree of delight or annoyance of the user of an application or service, resulting from the fulfilment of his or her expectations with respect to the utility and/or enjoyment of the application or service in the light of the users' personality and current state*" [Möller et al., 2014]. Traditional QoE assessment approaches that rely on subjective ratings and self-reports may be biased and time-consuming. On the other hand, objective approaches utilize various metrics to predict perceived quality, but they may not accurately reflect the perceived QoE. Therefore, developing reliable and accurate QoE models is essential to effectively capture users' perception of quality.

Recent studies have explored the use of physiological signals for QoE assessment, presenting a promising approach to understanding users' experiences. Wearables enable continuous monitoring of physiological signals in facilitating convenient and real-time QoE assessment. Analysing these signals provides insights into users' cognitive and affective states, enhancing our understanding of their experiences. While previous QoE assessment studies [Engelke et al., 2017] relied on statistical correlations, recent research has explored the potential of machine learning (ML) and deep learning (DL) techniques for predicting QoE using physiological signals [Vijayakumar et al., 2022]; [Perrin et al., 2015]. Despite the widespread use of artificial intelligence (AI) techniques in various domains, their application in QoE assessment using physiological signals is still limited. This study aims to investigate the effectiveness of physiological signals in predicting QoE, employing ML and DL models to enhance the accuracy of physiology-based QoE predictions.

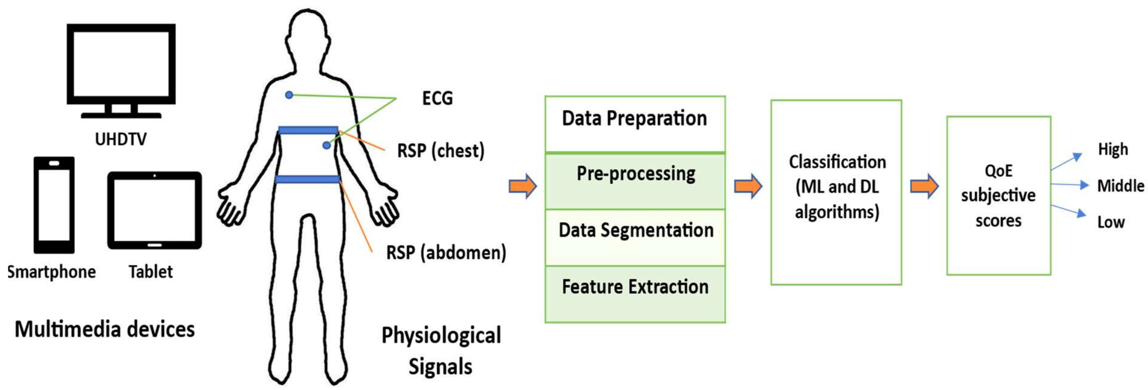

Figure 1: The proposed architecture of the QoE prediction from physiological signals for typical multimedia consumption scenarios using AI techniques.

2 Methods

Fig. 1 outlines the methodology of physiology-based QoE prediction. The process involves steps including data pre-processing, segmentation, feature extraction, and classification.

The methodology is applied to two datasets, namely SoPMD Dataset 1 [Perrin et al., 2015] and SoPMD Dataset 2 [Perrin et al., 2016]. These datasets consist of physiological signals and various QoE factors, including perceived quality levels, user preference, and the sense of presence (SoP). The stimuli employed in these datasets were carefully designed to induce different levels of immersion by varying factors such as audio sound systems, compression levels, resolution, and device type. In the SoPMD Dataset 1, a QoE database was developed for the analysis of perceived Sense of Presence (SoP). The audio-visual stimuli were configured to induce low, middle and high levels of immersion based on the video quality (level of compression), resolution (UHD, HD and SD), and sound reproduction (mono, stereo and 5.1). On the other hand, the SoPMD Dataset 2 is a multimodal database that investigates QoE across three commonly used devices: iPad, iPhone and UHD TV. Each dataset includes a total of 540 data instances collected from 20 participants, including a set of 27 videos or trials. During each trial, three physiological signals, electroencephalogram (EEG), electrocardiogram (ECG), and respiration (RSP) were recorded at a sampling rate of 250 Hz. At the end of each trial, the participants provided QoE ratings using a 9-point scale, ranging from 1 to 9. In our study, ECG and RSP signals are used to predict QoE because, in contrast to EEG signals, they can be measured non-invasively using wearable devices, allowing for long-term ecologically valid monitoring. For classification, rating values from 1 to 3 were categorized as low class, 4 to 6 as middle class, and 7 to 9 as high class.

This study aimed to predict QoE subjective ratings as a low, middle, and high class using ECG and RSP signals. The signal from the 60-second stimulus period was processed for both ECG and RSP signals. The processed signal was then used for segmentation and feature extraction. For segmentation, a sliding window of 3 seconds with 50% overlap was applied to divide the 60-second signals into 40 frames. From each of these frames, a total of 19 features were extracted from various domains such as time, frequency, and non-linear. These features included mean, median, minimum, maximum, standard deviation, variance, first-degree difference, second-degree difference, normalized mean, normalized minimum, normalized maximum, normalized first-degree difference, normalized second-degree difference, skewness, kurtosis, power spectrum, an average of the gradients, sample entropy, and Hurst component. The models evaluated were bidirectional long-short-term memory (BLSTM) [Graves et al., 2005], support vector machine (SVM), k-nearest neighbour (KNN) and random forest (RF). These models were chosen based on our preliminary analysis [Vijayakumar et al., 2022], which indicated that they performed well for the QoE classification task. Since the classes in both datasets were imbalanced, the Synthetic Minority Oversampling TEchnique, SMOTE [Chawla et al., 2002] was applied to the training set to balance the class distribution, excluding

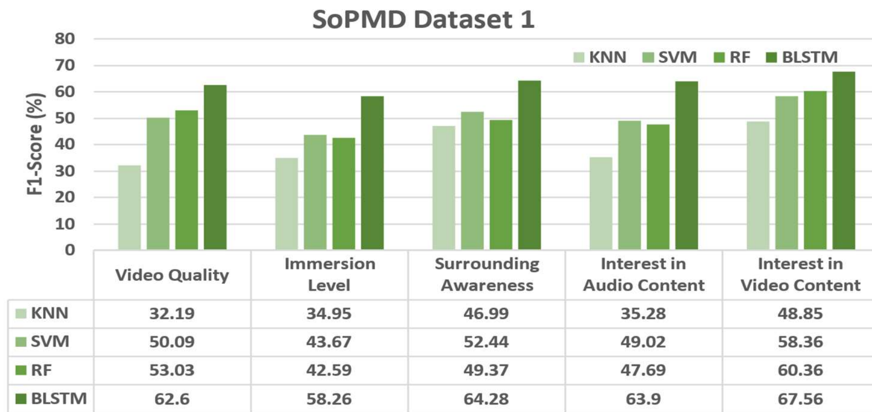

Figure 2: Comparison of F1-Scores of shallow ML models and BLSTM model for 3-class classification of QoE subjective scores from the fusion of ECG and RSP signals of the SoPMD Dataset 1.

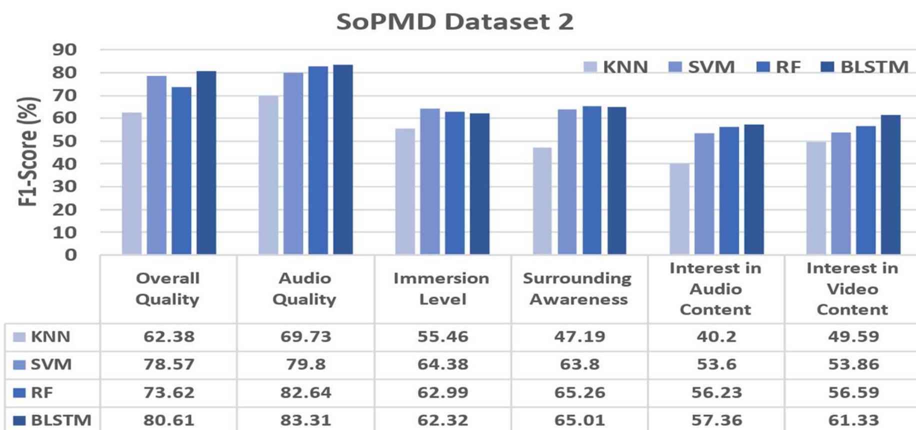

Figure 3: Comparison of F1-Scores of shallow ML models and BLSTM model for 3-class classification of QoE subjective scores from the fusion of ECG and RSP signals of the SoPMD Dataset 2.

the test set. The training and testing sets were standardized using the mean and standard deviation of the training samples. To assess the performance of the models, stratified 10-fold cross-validation was employed. Each fold consisted of a 90% training set and a 10% testing set. The final performance was calculated by averaging the results from the ten folds. The performance metrics evaluated were accuracy, F1-score, precision, and recall.

3 Results

The study implemented ML and DL algorithms on SoPMD Dataset 1 to classify subjective ratings for five QoE factors, namely, level of immersion, perceived video quality, interest in audio content, interest in video content and surrounding awareness. The 3-class classification results of SoPMD Dataset 1 are shown in Figure 2. The BLSTM model outperformed the ML models, achieving classification accuracies and F1-scores ranging between 58% and 67% for classifying different QoE factors. The results were encouraging, indicating the effectiveness of the BLSTM framework for QoE assessment in this dataset.

Similarly, in SoPMD Dataset 2, the ML and DL algorithms were used to predict QoE ratings for six factors, namely, level of immersion, perceived overall quality, perceived audio quality, interest in audio content, interest in video content and surrounding awareness. The 3-class classification results of SoPMD Dataset 2, as depicted in

Figure 3, demonstrate the effectiveness of the BLSTM and RF classifiers in classifying QoE ratings based on physiological signals, with F1-scores ranging from 57% to 83%. Notably, the performance of perceived quality factors, such as overall and audio quality, was higher compared to SoPMD Dataset 1, achieving F1-scores of 80.61% and 83.31%, respectively. Similarly, the assessment of surrounding awareness in SoPMD Dataset 2 yielded comparable results to SoPMD Dataset 1. However, SOPMD Dataset 1 exhibited slightly better performance in predicting content preference ratings when compared to SoPMD Dataset 2.

Overall, the results demonstrate the effectiveness of the BLSTM algorithm in both datasets in classifying QoE ratings based on physiological signals, while also highlighting the differences in performance between the two datasets in terms of perceived quality factors and content preference rating prediction. The proposed framework in the study effectively reduces the number of parameters and improves classification performance through the use of a sliding window approach and feature extraction techniques. However, further investigation and critical analysis are necessary to fully evaluate the effectiveness, generalizability, and robustness of the proposed framework.

4 Conclusion

This study demonstrates the potential of using physiological signals and AI techniques to accurately predict user QoE. The results obtained from experiments conducted on two datasets, SoPMD Dataset 1 and SoPMD Dataset 2, highlight the effectiveness of incorporating physiological signals in predicting subjective QoE scores. These findings emphasize the importance of integrating physiological signals and ML and DL techniques to advance QoE prediction and enhance user experiences in multimedia applications.

Acknowledgements

This research was funded by the Irish Research Council under grant number GOIPG/2021/357.

References

[Chawla et al., 2002] Chawla, N. V., K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. 2002. "SMOTE: Synthetic Minority Over-Sampling Technique." *Journal of Artificial Intelligence Research* 16: 321–57.

[Engelke et al., 2017] Engelke, U. et al. 2017. "Psychophysiology-Based QoE Assessment: A Survey." *IEEE Journal of Selected Topics in Signal Processing* 11(1): 6–21.

[Graves et al., 2005] Graves, A., and Jürgen, S. 2005. "Framewise Phoneme Classification with Bidirectional LSTM and Other Neural Network Architectures." *Neural Networks* 18(5): 602–10.

[Möller et al., 2014] Möller, S., and Raake, A. eds. 2014. *Quality of Experience: Advanced Concepts, Applications and Methods*. Cham: Springer International Publishing. http://link.springer.com/10.1007/978-3-319-02681-7.

[Perrin et al., 2015] Perrin, A. M. et al. 2015. "Multimodal Dataset for Assessment of Quality of Experience in Immersive Multimedia." In *Proceedings of the 23rd ACM International Conference on Multimedia - MM '15*, Brisbane, Australia: ACM Press, 1007–10. http://dl.acm.org/citation.cfm?doid=2733373.2806387.

[Perrin et al., 2016] Perrin, A. M. et al. 2016. "Towards Prediction of Sense of Presence in Immersive Audiovisual Communications." *Electronic Imaging* 2016(16): 1–8.

[Vijayakumar et al., 2022] Vijayakumar, S. et al. 2022. "BiLSTM-based Quality of Experience Prediction using Physiological Signals." *QoMEX 2022,* Germany: IEEE, 1-4.