

OBJECTIVE QUALITY METRICS ASSESSMENT FOR CLOUD GAMING

Jasmina Baraković Husić^{1,3}, Sara Kozić¹, Sabina Baraković²

Abstract: This paper aims to provide objective quality metrics assessment for cloud gaming using machine learning algorithms. Three classification algorithms (i.e., Random Forest, Random Three and J-48) have been used for the development of models for objective quality assessment of two metrics: blurriness and blockiness. The results indicate that Random Forest has the best performance in this experimental case of objective quality metrics assessment for cloud gaming. Future research activities will cover comparison of a broad range of objective quality metrics and machine learning algorithms while using larger dataset to enhance the results significance.

Keywords: cloud gaming, objective quality assessment, blockiness, blurriness, machine learning

INTRODUCTION

Cloud gaming is a new approach of games consuming, which will experience growth in the future [1]. Due to the increased interest, several cloud gaming platforms have been introduced so far, such as Google Stadia, Microsoft xCloud, Amazon Luna, etc. Games are hosted on the cloud platform instead of gaming devices or consoles. Gamers send commands from any connected device to a virtual server hosted on a cloud platform, which processes them, and subsequently delivers the results to the gamer. The final goal is to allow users to enjoy games seamlessly across all connected devices. Cloud gaming is therefore a more cost-effective and easier approach to game consumption. Although the benefits are nearly countless, cloud gaming requires consistent and flawless connectivity [2].

5G is expected to be a key enabler for the success of cloud gaming, which is the first use case to be investigated at a scale. 5G will provide lower latency, higher speeds, and greater capacity to improve the gaming experience and its quality [3]. However, the current market is still dominated by downloading games to a gaming console or smartphone. Furthermore, the market is divided into intense market serving regular and hardcore gamers and intense market serving casual and occasional gamers [1]. Although the future of cloud gaming seems to be optimistic, there are still numerous technical challenges that need to be solved for gamers to be provided with a satisfactory gaming experience and its quality.

When it comes to assessing cloud gaming quality, there are both subjective and objective approach [4]. Due to its resource-intensive and time-consuming demands, subjective quality assessment may not be feasible in many applications. To address the drawbacks of subjective testing, there has been growing interest in using objective quality assessment methods that rely on application performance metrics [5]. Because cloud gaming is relatively recent application, several performance metrics are still in their infancy. Despite that, objective performance metrics are preferred due to their practicality and speed. Moreover, the amount of application-related data is constantly increasing while creating the need to automate their analysis by using machine learning algorithms [6], [7]. Therefore, this paper seeks to objectively evaluate the quality of two metrics (i.e., blockiness and blurriness) by using machine learning algorithms.

Blockiness and blurriness are important metrics to consider when assessing the quality of cloud gaming because they can significantly impact the user experience. Blockiness and blurriness occur when compression techniques are utilized to reduce the size of video content transmitted over network [8]. Blockiness occurs after employment of compression techniques and block-based processing manifesting itself as a discontinuity between adjacent block in video frames. On the other side, blurriness occurs when spatial details in the video are lost due to compression resulting in a reduction of sharpness [9], [10]. These issues can significantly impact the user experience [11], making it difficult to notice details in the

¹ Faculty of Electrical Engineering, University of Sarajevo, Bosnia and Herzegovina

² Faculty of Traffic and Communication, University of Sarajevo, Bosnia and Herzegovina

³ Correspondence email: jbarakovic@etf.unsa.ba

© 2023 Author(s). This is an open access article licensed under the Creative Commons Attribution License 4.0. (<http://creativecommons.org/licenses/by/4.0/>).

game environment. This can lead to annoyance, lack of satisfaction and, users abandoning the service. Therefore, when assessing the quality of cloud gaming, it is important to evaluate the level of blockiness and blurriness of the video stream. This paper contributes to addressing these issues and thereby can help cloud gaming providers to improve the quality of their service, leading to increased user satisfaction and retention.

Following this introduction, the paper is structured in the following manner. Section 1 offers an overview of the background and related works. Section 2 describes the research methodology used to setup and conduct experiments. Section 3 provides results and their discussion in form of models for objective quality assessment of blockiness and blurriness. Section 5 provides conclusions and guidelines for future research.

1. BACKGROUND

1.1. Gaming video quality assessment

Video games allude to interactive games running on multiple platforms. They can be classified into different genres, such as strategy, fighting, racing, action, board, adventure, and more. The primary area of focus for research on gaming video quality has been streamed gaming (e.g., online gaming, interactive cloud gaming) and passive gaming (e.g., recorded gameplay, passive live broadcast) [12].

There are typically two main approaches for quality evaluation: subjective assessments and the creation of objective metrics and models. Subjective evaluation of gaming video quality is necessary, but expensive task, which is not suitable for evaluating quality in real-time. This results in a need for objective quality assessment metrics. Three categories exist for classifying objective Image Quality Assessment (IQA) and Video Quality Assessment (VQA) metrics, i.e., Full-Reference (FR), Reduced-Reference (RR), and No-Reference (NR). FR metrics requires complete reference information (e.g., Structural Similarity (SSIM), Peak Signal to Noise Ratio (PSNR), Video Multimethod Assessment Fusion (VMAF)). RR metrics are calculated based on piece of reference information (e.g., Spatio Temporal-Reduced Reference Entropic Differences (ST-RRED)). NR metrics predict quality without any reference information (e.g., Blind Image Quality Index (BIQI), Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE), Natural Image Quality Evaluator (NIQE)) [13].

Gaming videos are different forms of content, given that they are created by computers and feature artificial content that is perceived in a unique way. Previous studies have highlighted distinctions between video content that includes gaming versus those that do not [14], [15]. For example, in contrast to natural videos, the top-

rated quality evaluation metrics demonstrated a weaker correlation between predicted and actual gaming quality [14]. Furthermore, various metrics (e.g., PSNR, SSIM, BIQI, NIQE, BRISQUE) for evaluating the quality of gaming videos have been assessed in [13] using GamingVideoSET dataset [16]. Additionally, blockiness and blurriness have been considered in [17] as NR metrics representing video artefacts most encountered in the spatial domain when using lossy encoding.

While FR and RR metrics may demonstrate superior performance compared to NR metrics, it is not possible to utilize them for assessing the quality of gaming videos as there is no reference information available. Additionally, NR metrics designed for real-life video are not an appropriate choice for gaming content. Machine learning has led to the creation of multiple NR metrics for evaluating the quality of gaming videos, such as No-Reference Gaming Video Streaming Quality Estimator (NR-GVSQE), No-reference Gaming Video Streaming Quality Index (NR-GVSQI) [17] or No Reference Game Video Quality Metric (NR-GVQM) [18].

1.2. Machine learning in gaming video quality assessment

Advances in machine learning algorithms have contributed to the construction of a broad spectrum of metrics and models to objectively evaluate quality of gaming video. These metrics are based on machine learning algorithms, which use various influence factors, including compression artefacts, jitter, or packet loss. Since this paper aims to perform objective quality assessment, we present a concise survey of relevant literature that utilized machine learning techniques for this purpose. Table I gives a non-exhaustive review of related works that have been carefully chosen based on their topicality.

Two machine learning models, NR-GVSQI and NR-GVSQE, have been presented in [17]. These models are based on Support Vector Regression (SVR), Neural Networks (NN), Random Forest (RF), and Gaussian Process (GP), that are typical machine learning algorithms utilized for video quality modelling and prediction. Another machine learning based model, NR-GVQM, has been proposed in [18]. NR-GVQM is NR gaming video quality metric developed by training SVR that has the highest performance in quality assessment for different applications. Furthermore, an additional quality assessment model, DEMI, has been presented in [19]. DEMI is deep learning model built in three sequential steps: (i) the training of Convolutional Neural Network (CNN) relies on objective metrics for learning video artefacts, (ii) fine-tuning is performed on the model using image quality dataset, with a specific focus on addressing blockiness and blurriness, (iii) the features are fused by RF to facilitate the prediction of video quality. To boost the progress of gaming video VQA models development, several VQA models (BRISQUE, Two Level Video Quality Model (TLVQM), VIDEVAL, RAPIQUE) have been evaluated in [12] including new one called

GAME-VQP. This NR model has been proposed based on both CNN features and natural video statistics and has significantly outperformed the others. CNN has been also used in the development process of a novel NR video performance indicator, so-called NDNNetGaming, which can evaluate the visual fidelity of gaming videos affected by compression artefacts [20]. Similarly, comparison between two machine learning techniques, i.e., Decision Regression Tree (DRT) and Artificial Neural Network (ANN), and psychometric curve fitting approach has been conducted in [21]. Although the curve fitting approach is a hopeful strategy for reaching the desired computational complexity and scalability, it shows lower accuracy than ANN and DRT. Beside these interesting insights, a new FR performance metric is proposed, i.e., Game Video Streaming Quality Metric (GVSQM). Addressing the prediction challenges in the absence of a reference video, an additional video quality model, called nofu, is developed by [22]. This NR model is based on RF algorithm and outperforms VMAF in prediction accuracy.

This brief review in Table I served as the basis to define the research methodology to be used in this paper. Blockiness and blurriness have been considered as compression artefacts measures [17]. Therein lies the motivation for more detailed research of these quality metrics using several machine learning algorithms.

Table I: Literature review on gaming quality metrics assessment by machine learning algorithms

REF.	YEAR	TYPE	METRIC/ MODEL	ML
[12]	2022	Passive gaming	GAME-VQP	SVR
[17]	2019	Passive gaming	NR-GVSQI, NR-GVSQE	SVR, NN, GP, RF
[18]	2018	Interactive gaming	NR-GVQM	SVR
[19]	2020	Interactive gaming	DEMI	CNN, RF
[20]	2020	Passive gaming	NDNetGaming	CNN
[21]	2020	Passive gaming	GVSQM	DRTs, ANNs
[22]	2020	Passive gaming	Nofu	RF

Legend: ML (Machine Learning), NR-GVQM (No Reference Gaming Video Quality Metric), SVR (Support Vector Regression), RF (Random Forest), DRTs (Decision Regression Trees), CNN (Convolutional Neural Network), VQP (Video Quality Prediction), NN (Neural Network), ANN (Artificial Neural Networks), GP (Gaussian Process), NDNNetGaming (No-reference Deep CNN for Gaming), NR-GVSQE (No-Reference Gaming Video Streaming Quality Estimator), NR-GVSQI (No-reference Gaming Video Streaming Quality Index), GVSQM (Game Video Streaming Quality Metric).

2. RESEARCH METHODOLOGY

2.1. Experiment environment

Experiment environment included the following components: (i) Steam gaming platform serving as a dedicated server [23], (ii) Steam Link application allowing users to stream Steam games from dedicated server [24], (iii) Wireless N300 Easy Setup router providing Wireless Fidelity (Wi-Fi) connection, (iv) Wireshark network protocol analyser [25], and (v) Aiseesoft Screen Recorder serving as a screen recording software [26].

Steam gaming platform (as server side) was installed on HP EliteBook 855 G8 Notebook PC with the following specification: processor AMD Ryzen 7 PRO 5850U with Radeon Graphic 1.90 GHz, installed RAM 32.0GB, Windows 10 Business. Steam Link application (as client side) was installed on a desktop computer with the Intel(R) Core (TM) i3-6006U processor with a base clock speed of 2.00GHz, the 4.00GB of installed RAM, and Windows 10 Pro operating system.

This experiment environment was used to create sixteen (16) different scenarios for recording network traffic at the client side. Four different parameters of Stream Link application, i.e., games genre, video options, bandwidth, and framerate, were combined for this purpose. Games genre was manipulated at four levels (i.e., racing game, strategy game, puzzle game, action game), and video option at three levels (i.e., Balanced, Beautiful, Fast). Both bandwidth and framerate were manipulated at two levels (i.e., bandwidth (Automatic, 3Mbps), framerate (Automatic, 60fps)).

2.2. Eperiment procedure

Experiment procedure included seven steps: (i) run Steam Link application and connect to Steam dedicated server, (ii) configure performance parameters and select a game, (iii) start recording network data using Wireshark, (iv) start screen recording using Aiseesoft Screen Recorder, (v) play racing games, strategy games, or puzzle games for 30 seconds and action games for 60 seconds, (vi) stop running recordings and save the videos, and (vii) exit Steam Link application.

The recorded network data were processed by Video Quality Measurement Tool (VQMT) [27] and MATLAB [28] to be prepared for analysis by Waikato Environment for Knowledge Analysis (WEKA) tool [29]. Several features were extracted from the recorded network data, i.e., frame duration, source address, destination address, and frame size, which was further used for objective quality metrics assessment.

In case of blockiness metric, the resulting .csv file consists of 300 samples and four prior mentioned attributes which

were used to classify samples into three classes, class 1 (209 samples), class 2 (61 samples), and class 3 (30 samples). The probabilities of these classes were as follows: $p_{class1}=0.697$, $p_{class2}=0.203$, $p_{class3}=0.100$. Similarly, in case of blurriness metric, the resulting .csv file contains again 300 samples and four previously mentioned attributes used for classification of samples into three classes, i.e., class 1 (135 samples), class 2 (120 samples), and class 3 (45 samples). The probabilities of these classes were as follows $p_{class1}=0.45$, $p_{class2}=0.4$, $p_{class3}=0.15$. These classes for both blockiness and blurriness were used for attribute selection and decision making by different machine learning techniques.

3. RESULTS AND ANALYSIS

3.1. Results of quality assessment for blockiness

A comparison of three classification algorithms (i.e., Random Forest, Random Tree, and J-48) in terms of the number of correctly and incorrectly classified instances is presented in Table II. The Random Forest algorithm correctly classified 78.33% of the samples. The J-48 algorithm achieved accuracy of 78%, while the Random Tree algorithm achieved accuracy of 73% in classification of samples. This indicates that the Random Forest and J-48 algorithms provide similar classification accuracy, with Random Forest having a slight advantage.

The following quality metrics are used to compare various classification algorithms in Table III: (i) True Positive Rate (TPR), (ii) False Positive Rate (FPR), (iii) precision, (iv) recall, (v) F-measure, (vi) Receiver Operating Characteristic (ROC) area. TPR or recall is described as true positives divided by cumulative number of actual positives. FPR is determined as accuracy on the actual negative instances. Precision is the ratio of true positives to the sum of true positives and false positives. F-measure is a performance metric that considers both precision and recall and reaches its maximum value when precision and recall are equal. ROC area plots TPR against FPR at different classification thresholds [30].

The Random Forest algorithm performs best in terms of detecting TPR, whereas J-48 is better in FPR. Moreover, the Random Forest algorithm exhibits the lowest percentage of FPR detection, while the Random Tree algorithm shows the lowest percentage of TPR detection. In terms of precision, the Random Forest algorithm achieves the highest performance, whereas the Random Tree algorithm has the lowest precision percentage. As the recall is presented as a TPR, the results for these two metrics are the same. The F-measure has the highest percentage for the Random Forest algorithm and the lowest for the Random Tree algorithm. Regarding the ROC area, the behaviour is consistent with the previous finding, with the Random Forest algorithm producing the

best result and the Random Tree algorithm producing the worst result.

Based on the comparison of all metrics for the considered classification algorithms shown on Figure 1, it can be concluded that the Random Forest and J-48 algorithms performed the best. Specifically, the Random Forest algorithm had a slightly better performance.

Table II: Accuracy of classification algorithms for blockiness

	RANDOM FOREST	RANDOM TREE	J-48
Number of Correctly Classified Instances	235	219	234
Number of Incorrectly Classified Instances	65	81	66
Total Number of Instances	300	300	300

Table III: Comparison of classification algorithms for blockiness

	RANDOM FOREST	RANDOM TREE	J-48
TPR	0.783	0.730	0.780
FPR	0.319	0.293	0.261
Precision	0.771	0.732	0.768
Recall	0.783	0.730	0.780
F-measure	0.773	0.731	0.772
ROC area	0.820	0.711	0.783

Legend: TPR (True Positive Rate), FPR (False Positive Rate), ROC (Receiver Operating Characteristic).

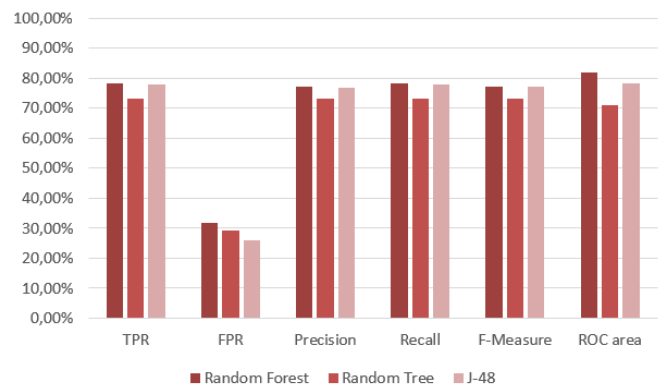


Figure 1: Comparison of classification algorithms for blockiness

3.2. Results of quality assessment for blurriness

The classification accuracy of considered machine learning algorithms (i.e., Random Forest, Random Tree, and J-48) is presented in Table IV indicating that

Random Forest algorithm gives the best results, correctly classifying 67.33% of samples. On the other hand, J-48 and Random Tree algorithms correctly classify 64.33% and 63% of samples, respectively.

Table V presents a comparison of classification algorithms using the same six quality metrics: TPR, FPR, precision, recall, F-measure, and ROC area. In terms of TPR, the Random Forest algorithm provides the highest accuracy rate of 67.3%, while the Random Tree algorithm falls slightly behind with the accuracy rate of 63%. The same applies to FPR.

Table IV: Accuracy of classification algorithms for blurriness

	RANDOM FOREST	RANDOM TREE	J-48
Number of Correctly Classified Instances	202	189	193
Number of Incorrectly Classified Instances	98	111	107
Total Number of Instances	300	300	300

Table V: Comparison of classification algorithms for blurriness

	RANDOM FOREST	RANDOM TREE	J-48
TPR	0.673	0.630	0.643
FPR	0.216	0.254	0.226
Precision	0.673	0.633	0.663
Recall	0.673	0.630	0.643
F-measure	0.672	0.631	0.643
ROC area	0.825	0.695	0.743

Legend: TPR (True Positive Rate), FPR (False Positive Rate), ROC (Receiver Operating Characteristic).

Therefore, it can be observed that the Random Forest algorithm outperforms other algorithms in terms of correctly classifying negative samples, with the lowest number of misclassifications observed relative to the total number of negative samples, while the Random Tree algorithm has a moderately higher value. The Random Forest algorithm also demonstrates the highest precision, while the Random Tree algorithm has a comparatively lower percentage of precision. As for blockiness, the results for TPR and blockiness are the same. The Random Forest algorithm has the highest percentage of F1-measure and ROC area, while the Random Tree algorithm has the lowest percentage for both metrics. Figure 2 illustrates how different classification algorithms compare in terms of considered quality metrics.

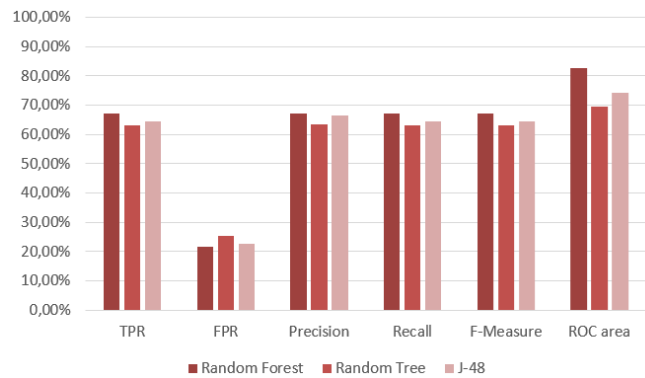


Figure 2: Comparison of classification algorithms for blurriness

3.3. Analysis of results for blockiness and blurriness

A comparative analysis of quality assessment models based on machine learning for gaming video is presented in [31]. It is important to notice that blockiness and blurriness have been mainly considered as influence factors rather than objective quality metrics [17], [18], [19], [21], [22]. Although not related to cloud gaming, an exception is a study that considers the relationship between blockiness and blurriness, and subjective quality ratings for mobile multiparty audio-visual telemeetings [32]. Blockiness and blurriness here are analysed as objective quality metrics for cloud gaming being assessed by several machine learning algorithms (i.e., Random Forest, Random Three, and J-48).

Among others, the Random Forest algorithm was found to provide the best classification results for both blockiness (accuracy: 78.33%) and blurriness (accuracy: 67.33%). This is in line with the findings of [17], [19], [22], where the Random Forest algorithm showed to be a successful solution in most of existing models, even though they did not use the same features or metrics for quality assessment. Moreover, recent research study reported that Random Forest is positioned as the most promising algorithm for evaluation of cloud gaming key quality indicators [33]. Therefore, it can be concluded that the Random Forest algorithm is a versatile and effective machine learning algorithm that can be applied to various attribute sets and metrics.

Machine learning algorithms bring several advantages when evaluating objective quality metrics, including scalability, adaptability, automation, etc. [34]. However, it is important to mention that machine learning algorithms are only as good as the data used to train the algorithms, representing various network conditions (e.g., packet loss, jitter, latency), games-specific factors (e.g., multiplayer support, game genre, pace), playing device (e.g., display, device size, device portability), compression-specific factors (e.g., frame rate, resolution, audio compression), context factors (e.g., novelty, physical environment), and human factors (e.g., human vision, experience) [31].

Moreover, it is important to analyse different machine learning algorithms (e.g., Support Vector Machine [35]) and their potential limitations, such as overfitting or bias. To gain more comprehensive understanding of quality assessment for cloud gaming, they should be used in conjunction with other quality metrics (e.g., WPSNR (Weighted Peak Signal-to-Noise Ratio), IFWPSNR (Importance Factor Weighted Peak Signal-to-Noise Ratio)) [36]. All these limitations represent a solid basis for future work that should be directed toward an interdisciplinary research effort to develop a reliable method for assessing multimodal quality, experience, or perception of cloud gaming [11].

Objective quality metrics assessment of cloud gaming has several implications related to user experience, performance of cloud gaming platforms, user satisfaction, competitive advantage, and future development. Firstly, it helps ensure that the user experience of cloud gaming is satisfactory and identify any potential issues that might degrade the gaming experience. Secondly, it is important for ensuring the performance of cloud gaming platforms is reliable and allowing developers to address potential issues. Then, it is critical for ensuring user satisfaction with cloud gaming services and identifying areas for improvement to increase user retention. Furthermore, objective quality metric assessment can provide a competitive advantage to cloud gaming providers and increase market share. Finally, it can help developers to determine areas for improvement and ensure the cloud gaming platform continues to evolve.

4. CONCLUSION

Cloud gaming is a new way to provide a high-quality gaming experience to users anytime and anywhere. The future of cloud gaming seems to be promising, but there are still technical issues that need to be solved to improve gaming experience. Machine learning algorithms can be used in this context to assess the quality based on analysis specific characteristics of network data.

With advancement in the field of machine learning, the domain of assessment has witnessed numerous models and metrics based on these algorithms. This paper contributes to this trend by proposing models for objective quality assessment of two metrics, i.e., blockiness and blurriness.

Three different machine learning techniques (i.e., Random Forest, Random Three and J-48) were employed to create models for objective quality assessment of these two metrics. The results indicate that Random Forest has the best performance in objective quality assessment of both metrics with an accuracy of 78.33% for blockiness and 67.33% for blurriness. These results can serve as a starting point for further research activities.

The future work could therefore consider the introduction of additional features and combination of different objective quality metrics assessed by more precise machine learning algorithms leading to multimodal quality, experience, or perception of cloud gaming.

REFERENCES

- [1] M. Johansson, "Mobile cloud gaming: Network slicing early commercial use cases series," Ericsson, 2022. [Online]. Available: <https://www.ericsson.com/en/blog/2022/6/mobile-cloud-gaming-an-early-use-case-for-network-slicing>
- [2] RootMetrics, "Mobile cloud gaming: the real-world cloud gaming experience in Los Angeles," 2020. [Online]. Available: <https://rootmetrics.com/en-US/content/us-LA-gaming-report-2020>
- [3] Huawei Cloud, "Cloud Gaming Experience Model (Cloud gMOS)," 2019. [Online]. Available: https://www-file.huawei.com/-/media/corporate/pdf/x-lab/2019/cloud_gmos_white_paper_en.pdf?la=en
- [4] A. Wahab, N. Ahmad, M. G. Martini, J. Schormans, "Subjective Quality Assessment for Cloud Gaming," *Multidisciplinary Scientific Journal*, vol. 4, no. 3, pp. 404–419, 2021.
- [5] J. Baraković Husić, S. Baraković, "Multidimensional modelling of quality of experience for video streaming," *Computers in Human Behaviour*, vol. 129, 10715, 2022.
- [6] M. I. Jordan, T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, pp. 255–60, 2015.
- [7] S. Ray, "A Quick Review of Machine Learning Algorithms," in *Proc. 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (Com-IT-Con)*, India, 2019.
- [8] A. Leontaris, A. R. Reibman, "Comparison of blocking and blurring metrics for video compression," in *Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, Philadelphia, PA, USA, 2005.
- [9] M. Shahid, A. Rossholm, B. Löfström, H.-J. Zepernick, "No-reference image and video quality assessment: a classification and review of recent approaches," *EURASIP Journal on Image and Video Processing*, 2014.
- [10] H. Choi, C. Lee, "No-reference image quality metric based on image classification," *EURASIP Journal on Advances in Signal Processing*, 2011.
- [11] Z. Akhtar, T. H. Falk, "Audio-Visual Multimedia Quality Assessment: A Comprehensive Survey," *IEEE Access*, vol. 5, pp. 21090 – 21117, 2017.
- [12] X. Yu, Z. Ying, N. Birkbeck, Y. Wang, B. Adsumilli, A. C. Bovik, "Subjective and objective analysis of streamed gaming videos," 2022, arXiv:2203.12824.
- [13] N. Barman, S. Schmidt, S. Zadtootaghaj, M. G. Martini, S. Möller, "An evaluation of video quality

- assessment metrics for passive gaming video streaming,” in Proc. 23rd Packet Video Workshop (PV), The Netherlands, 2018.
- [14] N. Barman, S. Zadtootaghaj, M. G. Martini, S. Möller, S. Lee, “A comparative quality assessment study for gaming and non-gaming videos,” in Proc. 10th International Conference Quality Multimedia Experience (QoMEX), Italy, 2018.
- [15] N. Barman, M. G. Martini, “H.264/MPEG-AVC, H.265/MPEG-HEVC and VP9 codec comparison for live gaming video streaming,” in Proc. 9th International Conference Quality Multimedia Experience (QoMEX), Germany, 2017.
- [16] N. Barman, S. Zadtootaghaj, S. Schmidt, M. G. Martini, S. Möller, “GamingVideoSET: A dataset for gaming video streaming applications,” in Proc. 16th Annual Workshop Network and System Support Games (NetGames), The Netherlands, 2018.
- [17] N. Barman, E. Jammeh, S. A. Ghorashi, M. Martini, “No-reference video quality estimation based on machine learning for passive gaming video streaming applications,” IEEE Access, vol. 7, pp. 74511-74527, 2019.
- [18] S. Zadtootaghaj, N. Barman, S. Schmidt, M. Martini, S. Möller, “NR-GVQM: A No Reference Gaming Video Quality Metric,” in Proc. International Symposium on Multimedia (ISM), Taiwan, 2018.
- [19] S. Zadtootaghaj, N. Barman, R. R. Ramachandra Rao, S. Göring, M. Martini, A. Raake, S. Möller, “DEMI: Deep video quality estimation model using perceptual video quality dimensions,” in Proc. 22nd International Workshop on Multimedia Signal Processing (MMSP), Finland, 2020.
- [20] M. Utke, S. Zadtootaghaj, S. Schmidt, S. Bosse, S. Möller, “NDNetGaming - development of a no-reference deep CNN for gaming video quality prediction,” Multimedia Tools and Applications, vol. 81, pp. 3181–3203, 2020.
- [21] S. Van Damme, M. Torres Vega, J. Heyse, F. De Backere F., De Turck, “A low-complexity psychometric curve-fitting approach for the objective quality assessment of streamed game videos,” Signal Processing: Image Communication, vol. 88, 115954, 2020.
- [22] S. Göring, R. R. Ramachandra Rao, A. Raake, “nofu -a lightweight no-reference pixel-based video quality model for gaming content,” in Proc. 11th International Conference on Quality of Multimedia Experience (QoMEX), Germany, 2019.
- [23] Steam. (2023). [Online]. Available: <https://store.steampowered.com/about/>
- [24] Steam Link. (2023). [Online]. Available: https://store.steampowered.com/app/353380/Steam_Link/
- [25] Wireshark. (2023). [Online]. Available: <https://www.wireshark.org/>
- [26] Aiseesoft Screen Recorder. (2023). [Online]. Available: <https://www.aiseesoft.com/screen-recorder/>
- [27] MSU Video Group: Video filtering and compression. (2023). [Online]. Available: <https://www.compression.ru/video/>
- [28] MATLAB: Math. Graphics. Programming. (2023). [Online]. Available: <https://www.mathworks.com/products/matlab.html>
- [29] WEKA: The Data Platform for the cloud & AI Era. (2023). [Online]. Available: <https://www.weka.io/>
- [30] M. Gong, “A Novel Performance Measure for Machine Learning Classification,” International Journal of Managing Information Technology (IJMIT), vol. 13, no. 1, 2021.
- [31] G. Kouglioumtzidis, V. Poulkov, Z. D. Zaharis, P. I. Lazaridis, “A Survey on Multimedia Services QoE Assessment and Machine Learning-Based Prediction,” IEEE Access, vol. 10, pp. 19507-19538, 2022.
- [32] D. Vučić, L. Skorin-Kapov, “QoE assessment of mobile multiparty audiovisual telemeetings,” IEEE Access, vol. 8, pp. 107669-107684, 2020.
- [33] C. Baena, O. S. Peñaherrera-Pulla, R. Barco, S. Fortes, “Measuring and estimating Key Quality Indicators in Cloud Gaming services,” Computer Networks, 109808, 2023.
- [34] O. Izima, R. de Fréin, A. Malik, “A Survey of Machine Learning Techniques for Video Quality Prediction from Quality of Delivery Metrics,” Electronics, vol. 10, no. 22, 2851, 2021.
- [35] P. Casas, S. Wassermann, “Improving QoE prediction in mobile video through machine learning,” in Proc. of 8th Int. Conf. Netw. Future (NOF), London, UK, 2017.
- [36] S. S. Sabet, M. R. Hashemi, S. Shirmohammadi, M. Ghanbari, “A Novel Objective Quality Assessment Method for Perceptually-Coded Cloud Gaming Video,” in Proc. IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), Miami, FL, USA, 2018.

BIOGRAPHY

Jasmina Baraković Husić is a Full Professor at the University of Sarajevo. She also works as a specialist in the Department for Planning and Development of Technologies and Services, at BH Telecom, Joint Stock Company, Sarajevo. She is co-founder of a research laboratory Little Mama Labs. She is/was principal investigator on many national and international projects (NATO, Erasmus+, Ministry of Education, RAI, etc.) and has co-operators throughout the world (e.g., EU member countries, SEE region countries, etc.). Her research interests include a variety of topics in networked multimedia services, QoS/QoE/QoL management, 5G-IoT communications.

Sara Kozić completed her master's degree at the Faculty of Electrical Engineering, University of Sarajevo,

Department of Telecommunications. She works in an IT company, System Verification, as a quality assurance engineer. Her research focuses around creating efficient quality assurance methodologies and strategies for software engineering, with a specific focus on enhancing software quality through meticulous testing and analysis. In addition, she is engaged in research on IoT systems to address real-world problems, as well as assessment of objective quality metrics for cloud gaming.

Sabina Baraković is an Associate Professor at the University of Sarajevo and Expert Adviser at the Ministry of

Security of Bosnia and Herzegovina. She is a Co-founder of a research laboratory Little Mama Labs. She is/was principal investigator on many national and international projects (Horizon2020, NATO, ITU, Erasmus+, CoE, OSCE, etc.) and has co-operators throughout the world (e.g., EU member countries, SEE region countries, Qatar, etc.). Her research interests include multidimensional QoE modelling of multimedia, mobile Web-based applications, and unified communications, cybersecurity in smart environments and IoT/WoT, QoS/QoE/QoL management, etc.