



HAL
open science

Towards Recommendations and Guidelines for Subjective Medical Image and Video Quality Assessment

Meriem Outtas, Lu Zhang, Maria Martini

► To cite this version:

Meriem Outtas, Lu Zhang, Maria Martini. Towards Recommendations and Guidelines for Subjective Medical Image and Video Quality Assessment. 2024 12th European Workshop on Visual Information Processing (EUVIP), Sep 2024, Geneve, Switzerland. pp.1-6, <10.1109/EUVIP61797.2024.10772875>. <hal-04884095>

HAL Id: hal-04884095

<https://hal.science/hal-04884095v1>

Submitted on 13 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Copyright - All rights reserved

Towards Recommendations and Guidelines for Subjective Medical Image and Video Quality Assessment

Meriem OUTTAS

Univ. Rennes, INSA Rennes
CNRS, IETR - UMR 6164 F-35000
Rennes, France
meriem.outtas@insa-rennes.fr

Lu ZHANG

Univ. Rennes, INSA Rennes
CNRS, IETR - UMR 6164 F-35000
Rennes, France
lu.zhang@insa-rennes.fr

Maria MARTINI

School of Computer Science and Mathematics
Faculty of Engineering, Computing
and the Environment, Kingston University
m.martini@kingston.ac.uk

Abstract—Evaluating the quality of medical images is a critical step in developing precise, trustworthy, and clinically applicable image processing algorithms and machine learning models for medical imaging applications. This evaluation serves to both benchmark and optimize algorithms. However, there is a lack of standard or guidelines for conducting subjective Medical Image and Video Quality Assessment (MIVQA) tests. Although there are existing standards for natural image and video quality assessment that provide information on selecting images and video sequences, assessor types and numbers, subjective testing procedures (test environment, participant selection, methodology, etc.) and model performance evaluation, they are not tailored to medical images. Although this study does not aim to propose a comprehensive method for MIVQA subjective tests, it addresses several aspects that may be worth considering when a MIVQA subjective test is conducted.

Index Terms—Medical imaging, Quality assessment, Subjective tests, recommendations.

I. INTRODUCTION

Image processing has gained significant popularity across all sectors, particularly with the advent and advancement of deep learning and artificial intelligence. The significance of image processing is particularly pronounced in the medical field, where it serves a multitude of critical purposes. These include early disease detection, aiding healthcare professionals in making accurate diagnoses and enhancing their overall efficiency. The versatility and impact of image processing in this field cannot be overstated. This raises the important issue of medical image quality. In fact, the effectiveness, accuracy and reliability of an analysis based on an image, whether performed by a human or an algorithm, is largely determined by the image quality. Assessing this quality plays thus an important role in benchmarking and optimising image processing algorithms involved in a diagnostic process or in a surgical operation, ultimately leading to improved patient treatment, care and safety. For humans, high-quality images can help identify subtle patterns and anomalies that might be missed in lower-quality images. Similarly, algorithms require clear and detailed visual data to perform a certain task, such as making accurate predictions and classifications.

Today, Artificial Intelligence (AI) has become a dominant force in the field of medical image analysis, with a growing number of models relying on deep learning algorithms. The role of image quality is also not ignorable in the training and testing of these models. Different models may have a different robustness to the image quality, thus image quality can have a dramatic impact on the accuracy and reliability of these models, ultimately on the tasks performed by the models (e.g., segmentation, diagnosis, surgery guide).

Since the work of Cosman *et al.* [1], efforts have been made to establish a “gold standard” for Medical Image Quality Assessment (MIQA). However, due to unique characteristics of medical images, it is difficult to define an objective and universally accepted reference standard that could achieve consensus in the medical community. In this context, subjective tests are often considered more valuable than objective assessments. This is because subjective testing involves human observers who provide valuable insight into how well medical images convey information to the human visual system. Ultimately, medical images are interpreted by clinicians, and their ability to accurately perceive and diagnose conditions from these images is essential. In this paper, we will focus on subjective MIVQA, which involves human observers who can provide insight into how well medical images deliver information to the human visual system. The aim is not to propose a comprehensive standard, but to provide guidance on important issues that arise when conducting subjective tests, which may pave a way for a future recommendation/guideline in this domain. The proposed approach aims to initially present a preliminary version and subsequently engage in discussions with the Video Quality Experts Group (VQEG), which holds significant role in the validation of objective and subjective video quality methods. This collaboration can potentially pave the way for the development of international standards (e.g., within ITU) for subjective assessment of the quality of medical contents.

II. RELATED/PRIOR WORKS

Several efforts have been made in order to discuss and review Medical Image and Video Quality Assessment (MIVQA). For example, Razaak and Martini [2] discussed the importance of quality evaluation in medical images, focusing their work on e-health applications. They provide an overview of previous works and offer a classification of MIVQA methods by mainly proposing three categories: objective quality metrics, subjective quality metrics and hybrid approach. Chow and Paramesran [3] divided MIVQA into subjective and objective assessments, with the latter further categorised into Full Reference Image Quality Assessment (FR-IQA), Reduced Reference IQA (RR-IQA) and No Reference IQA (NR-IQA) methods. They conclude that NR-IQA methods are particularly relevant to medical images due to the lack of pristine reference images. A review by Leveque *et al.* [4] provides a summary of existing studies on subjective MIVQA, including details on the used protocols, objective metrics, correlation measures, and employed statistical analysis methods. Barman *et al.* [5] proposed a process for documenting the various aspects of a subjective quality assessment dataset, such as details about the source sequences, encoding settings, number of test subjects, test methodology, etc.

More efforts have been made for objective MIVQA. For example, Zhang *et al.* [6] focused on task-based approaches and summarized some key issues and specificities including diagnostic task, appropriate pathology for different diagnostic task, figure of merit, expertise and subjective experimental protocol. A comprehensive survey [7] reviews the methodologies employed for objective quality assessment of MIVQA and categorizes them into two main approaches: visual quality-based and task-based. The visual quality-based methods calculate a quality index directly from visual attributes, while the task-based methods, which are gaining increasing attention, evaluate the impact of quality impairments on the performance of a particular task.

The rapid development of artificial intelligence (AI) applications for medical images has obviously led to many contributions in MIVQA. Baum *et al.* [8] raised the key issue of MIQA in the development of machine learning models for lung anomaly detection. A quality assessment module is trained to label images as of sufficient or insufficient quality prior to the diagnosis assistance module. In [9], the authors proposed a neural network-based tool that guides reviewers to focus on potentially problematic images in distributed imaging studies, minimizing the human effort and time required for quality control. [10] proposed a model based on convolution self-attention to predict medical image quality without a reference with limited amount of annotated data. Focusing on eye fundus images, in [11] the authors classified medical images, via AI, in three levels: good, bad, and an additional level where the quality is adequate only for some specific tasks. Morgan *et al.* [12] propose a Checklist for AI in Medical Imaging (CLAIM). Our work presented here is directly related to three sections of CLAIM and can meet the expected requirements: data,

ground truth and evaluation. Indeed, ensuring the quality of these elements could lead to more rigorous, reproducible and generalisable methods, especially those based on AI.

A large number of subjective and objective quality assessment tests have been performed on medical images to validate methods such as enhancement, compression, lesion detection [6][13][14][15][16][17][18], and different methods have been used to perform these tests [7] [4]. Most of them follow or are inspired by existing standards mainly dedicated to natural images and videos such as ITU-R BT.500 [19], ITU-T P.910 [20], ITU-T P.913 [21]. For instance, in [14] ITU-R BT.500 was adopted, in particular with double-stimulus continuous quality scale (DSCQS). These standards provide comprehensive guidelines and recommendations on various aspects such as the selection of images and video sequences, the type and number of assessors, procedures for subjective testing (test environment, participant selection, methodology, etc.), and the evaluation of model performance. However, it would be important to have a specific procedure to be followed for performing subjective quality assessment tests in the medical context, to ensure that the medical image quality is suitable for accurate diagnosis, analysis, etc. To the best of our knowledge, there is currently no established standard for subjective quality assessment of medical images and videos.

III. SPECIFIC ASPECTS AND GUIDELINES FOR THE SUBJECTIVE QUALITY ASSESSMENT OF MEDICAL IMAGES AND VIDEOS

We discuss in the following specific aspects related to the subjective quality assessment of medical imaging, in particular the importance of the purpose of the subjective study and how this can drive the methodology for subjective assessment, the different methods to establish the ground truth, the selection of viewing subjects, and the analysis of the subjective results.

A. Study purpose

1) *Research question and objectives of the study:* As mentioned above, assessing the quality of medical images serves mainly two goals: first benchmark and provide high-quality medical images to healthcare professionals, image processing algorithms or deep learning models to ensure accurate diagnosis, intervention, and effective treatment planning. The second is to optimise a specific procedure, technique, protocol or device to improve the overall imaging process. By identifying the research goal, researchers can define the remaining elements, mostly the pathology and modality, and the type of quality assessment method to be used.

2) *Define the pathology / medical aim and modality:* This step must be carried out meticulously to ensure that the research questions are answered with significant and relevant results. In the case of benchmarking, the pathology is likely been defined beforehand. The modality will be defined according to its reliability and diagnostic efficiency. In the case of optimisation, the selection and specification of the pathology and modality could allow researchers to identify the particular disease or visual condition that highlights the

important features to be studied. While the selection of the most suitable imaging modalities for a specific pathology or intervention is selected by the medical specialist, the consequent quality assessment protocol should take this information into consideration. In particular the number of points in the Likert scale/the number of quality classes should be selected considering the objective of the study [22].

3) *Quality assessment method:* Medical quality assessment is typically task-related since the aim of medical imaging is, as discussed above, to perform a specific task such as obtaining a diagnosis or support intervention (e.g., guided ultrasound, remote surgery).

An important way to measure the quality of medical imaging is measuring the **quality in the performance of the medical task** based on the available image or video, for instance measuring how the medical specialist is able, based on the images, to detect a certain type of lesion or identify a particular anatomical structure, or to perform a specific operative task (e.g., considering suturing time and quality in the case of remote surgery). This approach is useful for evaluating the effectiveness of a specific imaging technique or algorithm for a particular clinical application. It should be noted that the task-based approach is considered in subjective assessment here because it implies the involvement of a human.

On the other hand, quality assessment based on **opinion scores** relies on subjective ratings from human observers to evaluate the quality of an image or video. In this case, two options are possible: ask the subjects to rate the general quality of the image or ask the subjects to assess the confidence on a diagnosis / task performance based on that image or video. Obviously, for the second case expert medical viewers are required for the subjective tests, while for the first case it is definitely preferred but not strictly required. More details on this are provided in the relevant subsection below. The instructions provided to the subjects are critical and it is important that any medical image/video quality assessment study reports clearly the question asked / instructions provided to the subjects. The quality is typically estimated from the scores of multiple subjects via the calculation of the mean, to obtain the mean opinion score (MOS). Other methods to estimate the quality based on subjective results considering subjects bias and content ambiguity could be considered (e.g., [23]). In the medical case, the opinion of different medical specialists with different experience may be weighted differently based on the level of experience; also, some types of medical content or modality may result in a higher variability in the score provided by different subjects/experts.

B. Types of impairment

Impairments in medical images and video may be due to the acquisition device (e.g., speckle noise in ultrasound video) or to processing pipeline following acquisition (e.g., compression or delivery via an error prone channel). In the latter case, a reference image / video is often available and the subjective study can be performed using a hidden reference (the original image can be presented in the tests to subjects

who are unaware that this is the original one). Pair comparison tests are possible. In the first case, single stimulus subjective tests are the norm, due to a lack of reference image or video.

C. Ground Truth

As mentioned above, in several cases there is no pristine reference in MIVQA. In the case the image / video quality is measured based on the quality in the performance of a diagnosis, a Ground Truth (GT) can be established or accepted based on the objective of the study. This GT is considered the true or actual diagnosis associated to an image and can be used as a benchmark to compare and evaluate the performance of different subjective MIVQA methods. Here are a few selected examples.

- **Diagnostic tests:** these are widely regarded as the closest thing to a gold standard in medical diagnosis, as they have been extensively validated by the medical community and are considered to be the most accurate and reliable methods. These include, for example, anatomic pathology tests that involve the examination of tissue samples under a microscope to identify abnormalities and are widely used in the diagnosis of adenomas and cancer. Antigen tests detect the presence of specific proteins on the surface of viruses or bacteria.
- **Clinical evaluation or clinical diagnosis:** This is based on the expertise of healthcare professionals, which may include physical examination, history, symptoms, etc. While this approach is commonly used in medical practice, and while clinical intuition is seen as beneficial for advancing diagnoses [24], it can be difficult to apply in the context of MIVQA due to its subjectivity and reliance on the judgement of the healthcare professional.
- **Physical and digital phantoms :** These are used to mimic the properties of human tissues and structures; however, it is difficult for these phantoms to fully reflect a clinical image.
- **Synthetic images:** These images, mimicking the characteristics of genuine medical images, can be used to generate ground truth data with specific properties and features. It is worth mentioning the use of generative models for medical image synthesis, as proposed by Meharban *et al.* [25] and Osuala *et al.* [26]; such methods are able to generate synthetic images with a certain degree of realism and accuracy.

The latter two kind of GT allow the creation of a controlled environment and reproducible method. Therefore, they are more suitable for the evaluation of some impairments such as image contrast, reconstruction algorithm, compression/transmission artefacts or viewing systems.

D. Observers, viewers and assessors

1) Participants/Assessors:

- **Expert:** It is essential that assessors involved in subjective MIVQA have relevant expertise in the interpretation of medical images and videos, in particular in the specific imaging task being assessed. These assessors

may be radiologists, clinicians or medical physicists. The inclusion of observers with different levels of expertise and specialisation in anatomy, pathology and imaging techniques can also provide valuable insights into the variability of image interpretation. However, it is not always possible to involve such experts due to the limited numbers of specialists in each discipline and their busy work schedules. An interesting method to cope with the limited availability of expert viewers can be to perform repeated experiments with a small pool of subjects, leveraging the observation that apparent differences among subjects can, for the most part, be explained by random factors [27].

- **Naïve** observers are individuals who have little or no training or experience in interpreting medical images. They have not necessarily already been involved in research studies to assess the perceptual quality of medical images or to investigate the effects of image degradation on visual perception. When naive observers are employed, in general the question asked is not associated to a diagnosis or task, but only to the perceived quality. However, it was shown in a proof of concept that naive observers can be trained using deep learning methodologies to reliably perform some tasks of medical image perception assessment and visual pattern recognition of certain cancers [28]. Regardless of the promising first results, it is still recommended that naive participants are involved as little as possible due to their lack of expertise and potential for misinterpretation of medical images.
- are mathematical or computational models that simulate human performance in visual tasks, taking into account the statistical properties of images and the characteristics and cognition of the human visual system. They can provide quantitative measures of image quality or perform specific tasks (e.g., lesion detection, segmentation). However, it is important to consider the ethical implications of using model observers and to use them as a complementary tool rather than to replace human participants.

2) *Number of participants:* Including a larger number of observers can increase the statistical power of a study. However, this is not always feasible. According to a study conducted in [4], it is advisable to include at least four (4) participants, but it is important to note that this is the absolute minimum number, especially in the case of expert observers. As highlighted above, the limited number of participants in a test can be compensated via repeated observations, i.e., presenting multiple times the same content randomly in the test sequence to the same subject [27].

E. Test methodology and protocol

The design of the subjective tests protocol is a key element to ensure that it is appropriate to the research question and that the results are valid and reliable. At this stage, reference should be made to the methods and procedures defined in various ITU recommendations [19] [20] [21] relating to images and video, and [29] which sets the standard for recognition tasks

and could be adapted to task-based MIVQA. One could cite the two most used protocols in existing subjective MIVQA:

- **Double Stimulus Continuous Quality Scale (DSCQS)**, where the sequences are presented in pairs (reference and impaired) and subjects are asked to assess the quality of each of them, being unaware of which one is the reference. This protocol can only be used when the impaired image or video is available.
- **SAMVIQ**, where video sequences are shown in multi-stimulus form, so that the user can choose the order of tests and correct their votes, as appropriate [19].

In the absence of an established standard for subjective MIVQA, the chosen protocol can be adapted and modified according to the needs of the research.

F. Analysis of the results of subjective tests: figure of merit, metrics and statistical analysis

The choice of **figures of merit**, metrics, and statistical analysis depends on all aforementioned items. Mean Opinion Score (MOS) and Receiver Operating Characteristic (ROC) analysis are the most widely used to evaluate the task/diagnostic accuracy or perceived diagnostic quality. **Statistical analysis** evaluates the significance of the results obtained, for instance via Analysis of Variance (ANOVA), t-test, and correlation analysis. In addition, **Confidence intervals** and **Inter/Intra-observer variability** could be used to estimate precision and reliability of the assessment method. As mentioned above, methods different than MOS can be used to estimate the quality based on the results of subjective tests, for instance via generalised linear models, as in [23]. It is recommended that when results of subjective tests are published, complete information on opinion scores of all subjects on all contents is provided, to enable detailed statistical analysis and modelling.

G. Ethical consideration

Ethical considerations are crucial when dealing with human information in medical studies. While **prospective studies** require approval from dedicated authorities responsible for human protection, **retrospective studies** involve the use of pre-existing data, thus ethical considerations are typically limited to obtaining patient authorization and ensuring data anonymisation.

IV. CONCLUSION AND FUTURE WORK

In this paper, we present a first attempt to overcome the lack of standardised recommendations for performing subjective tests to assess the quality of medical images and videos. It is our aim to gather interest and contributions towards this work as a long-term initiative that we aim to launch and continuously improve over time. To achieve this, we plan to engage in ongoing discussions with other experts, seek their feedback and suggestions, and incorporate new work and research findings.

REFERENCES

- [1] P. C. Cosman, R. M. Gray, and R. A. Olshen, "Evaluating Quality of Compressed Medical Images: SNR, Subjective Rating, and Diagnostic Accuracy," *Proceedings of the IEEE*, vol. 82, no. 6, pp. 919–932, 1994.
- [2] M. Razaak and M. G. Martini, "Medical image and video quality assessment in e-health applications and services," in *2013 IEEE 15th International Conference on e-Health Networking, Applications and Services (Healthcom 2013)*, 2013, pp. 6–10.
- [3] L. S. Chow and R. Paramesran, "Review of medical image quality assessment," *Biomedical Signal Processing and Control*, vol. 27, pp. 145–154, 2016.
- [4] L. L ev eque, M. Outtas, H. Liu, and L. Zhang, "Comparative study of the methodologies used for subjective medical image quality assessment," *Physics in Medicine & Biology*, vol. 66, no. 15, p. 15TR02, 2021.
- [5] N. Barman, Y. Reznik, and M. Martini, "Datasheet for subjective and objective quality assessment datasets," in *2023 15th International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2023, pp. 111–114.
- [6] L. Zhang, C. Cavaro-M enard, and P. Le Callet, "Key issues and specificities for the objective medical image quality assessment," in *Sixth International Workshop on Video Processing and Quality Metrics for Consumer Electronics VPQM 2012*, 2012, pp. 1–6.
- [7] R. Rodrigues, L. L ev eque, J. Guti errez, H. Jebbari, M. Outtas, L. Zhang, A. Chetouani, S. Al-Juboori, M. Martini, and A. M. Pinheiro, "Objective quality assessment of medical images and videos: Review and challenges," *arXiv preprint arXiv:2212.07396*, 2022.
- [8] Z. M. Baum, E. Bonmati, L. Cristoni, A. Walden, F. Prados, B. Kanber, D. C. Barratt, D. J. Hawkes, G. J. Parker, C. A. G. Wheeler-Kingshott *et al.*, "Image quality assessment for closed-loop computer-assisted lung ultrasound," in *Medical Imaging 2021: Image-Guided Procedures, Robotic Interventions, and Modeling*, vol. 11598. SPIE, 2021, pp. 183–189.
- [9] D. Zuki c, A. Haley, C. Lisle, J. Klo, H. J. Johnson, and A. Chaudhary, "Medical image quality assurance using deep learning," in *Medical Imaging with Deep Learning*, 2022.
- [10] M. Tliba, A. Sekhri, M. A. Kerkouri, and A. Chetouani, "Deep-based quality assessment of medical images through domain adaptation," *2022 IEEE International Conference on Image Processing (ICIP)*, pp. 3692–3696, 2022.
- [11] A. Raj, N. A. Shah, A. K. Tiwari, and M. G. Martini, "Multivariate regression-based convolutional neural network model for fundus image quality assessment," *IEEE Access*, vol. 8, pp. 57 810–57 821, 2020.
- [12] J. Mongan, L. Moy, and C. E. Kahn, "Checklist for Artificial Intelligence and Medical Imaging (Claim)," *Radiology: Artificial Intelligence*, 2020.
- [13] M. Gray, H. Morchel, and V. Hazelwood, "Evaluating the effect of bit rate on the quality of portable ultrasound video," in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, 2015, pp. 247–250.
- [14] A. Chaabouni, Y. Gaudeau, J. Lambert, J. M. Moureaux, and P. Gallet, "Subjective and objective quality assessment for H.264 compressed medical video sequences," in *2014 4th International Conference on Image Processing Theory, Tools and Applications (IPTA)*, 2014, pp. 1–5.
- [15] A. Chaabouni, Y. Gaudeau, J. Lambert, J.-M. Moureaux, and P. Gallet, "H.264 medical video compression for telemedicine: A performance analysis," *IRBM*, vol. 37, no. 1, pp. 40–48, 2016, Medical Image Analysis for Computer Aided Diagnosis.
- [16] L. L ev eque, W. Zhang, P. Parker, and H. Liu, "The impact of specialty settings on the perceived quality of medical ultrasound video," *IEEE Access*, vol. 5, pp. 16 998–17 005, 2017.
- [17] I. A. Urbaniak and M. W. Wolter, "The fidelity of compressed and interpolated medical images," *Technical Transactions*, 2020.
- [18] M. Outtas, L. Zhang, O. Deforges, A. Serir, W. Hamidouche, and Y. Chen, "Subjective and objective evaluations of feature selected multi output filter for speckle reduction on ultrasound images," *Physics in Medicine and Biology*, vol. 63, no. 18, p. aadbc9, 2018.
- [19] ITU-R, "Methodologies for the subjective assessment of the quality of television images," *REC. ITU-R BT.500-15 (05/2023)*.
- [20] ITU-T, "Subjective video quality assessment methods for multimedia applications," *Recommendation ITU-T P.910*, October 2023.
- [21] ITU-T, "Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment," *Recommendation ITU-T P.913*, 2021.
- [22] A. Raj, A. K. Tiwari, and M. G. Martini, "Fundus image quality assessment: survey, challenges, and future scope," *IET Image Processing*, vol. 13, no. 8, pp. 1211–1224, 2019.
- [23] S. Pezzulli, M. G. Martini, and N. Barman, "Estimation of quality scores from subjective tests-beyond subjects' mos," *IEEE Transactions on Multimedia*, vol. 23, pp. 2505–2519, 2020.
- [24] M. Vanstone, S. Monteiro, E. Colvin, G. Norman, J. Sherbino, M. Sibbald, K. L. Dore, and A. J. Peters, "Experienced physician descriptions of intuition in clinical reasoning: a typology," *Diagnosis*, vol. 6, pp. 259 – 268, 2019.
- [25] M. S. Meharban, M. K. Sabu, and S. Krishnan, "Introduction to medical image synthesis using deep learning:a review," *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, vol. 1, pp. 414–419, 2021.
- [26] R. Osuala, G. Skorupko, N. Lazrak, L. Garrucho, E. Garc a, S. Joshi, S. Jouide, M. Rutherford, F. Prior,

- K. Kushibar, O. Díaz, and K. Lekadir, “medigan: a Python library of pretrained generative models for medical image synthesis,” *Journal of Medical Imaging*, vol. 10, no. 06, 2023.
- [27] P. Pérez, L. Janowski, N. García, and M. Pinson, “Subjective assessment experiments that recruit few observers with repetitions (FOWR),” *IEEE Transactions on Multimedia*, vol. 24, pp. 3442–3454, 2021.
- [28] J. Hegdé, “Deep learning can be used to train naïve, nonprofessional observers to detect diagnostic visual patterns of certain cancers in mammograms: a proof-of-principle study,” *Journal of Medical Imaging*, vol. 7, no. 02, p. 1, 2020.
- [29] ITU-T, “Subjective video quality assessment methods for recognition tasks, ITU Recommendation P.912 (03/2016),” 2016.