Impact of Visual Distortion on Medical Images

Citation for published version:
Sun, Y & Mogos, G 2022, 'Impact of Visual Distortion on Medical Images', *IAENG International Journal of Computer Science*, vol. 49, no. 1, pp. 36-45.

Published In:
IAENG International Journal of Computer Science

Impact of Visual Distortion on Medical Images

Yuhao Sun, *Member, IAENG* and Gabriela Mogos, *Member, IAENG*

***Abstract*—The distortion is a common occurrence in the imaging area, especially medical imaging. The most common kinds of distortion in medical imaging are blurry, contrast- and noise-distorted images. The purpose of this study is to provide a four-step technique for determining if current Objective Image Quality Assessment (IQA) mathematical models function as well as human eyes. Throughout the investigation, the appropriate quality and source of X-ray CT scans were chosen. The results indicate that the Perception-Based Image Quality Evaluator (PIQE) is a moderately effective mathematical model of No-Reference IQA (NR-IQA). However, in comparison to PIQE, both the Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) and the Naturalness Image Quality Evaluator (NIQE) performed poorly when used with X-ray CT scans.**

***Index Terms*—Medical Imaging, Image Quality Assessment, BRISQUE, NIQE, PIQE.**

I. Introduction

MEDICAL imaging techniques are rapidly being employed in the medical sector, particularly in hospitals and other healthcare facilities [1]. Doctors, including clinicians and radiologists, can see the interior anatomy of a patient's specific organ or tissue in order to validate the patient's diagnosis and therapy recommendations. However, as is the case with the majority of pictures, distortion (blurry, contrast- and noise-distorted) is possible in the medical imaging field [2]. Medical pictures that are distorted may influence clinicians' clinical judgments, resulting in misdiagnosis, missed diagnosis, or other inaccuracies. The purpose of this study is to use both Subjective and Objective Image Quality Assessments (Subjective & Objective IQA) to assess if the findings of current mainstream mathematical models can perceive medical pictures as well as human eyes. Additionally, we seek to develop a mathematical model that is relatively efficient and may be applied in the future to the field of medical imaging or industry. Despite the restricted number of observers in Subjective IQA, this research demonstrated the effectiveness of a mathematical model called the Perception-based Image Quality Evaluator (PIQE) on X-Ray CT images.

II. Related Work

A. Medical Imaging

Previous research has demonstrated that since the 1890s, when X-rays were discovered, the public has paid growing attention to medical imaging methods [3]. Medical imaging has grown in strength over the last few decades, and it is now widely employed in the medical community, particularly in hospitals. From a contemporary viewpoint, traditional medical imaging methods include X-ray computed tomography (CT), magnetic resonance imaging (MRI), ultrasound imaging, and radionuclide imaging, all of which are classified as radiology procedures [4]. Cardiology, pathology, and ophthalmology are among well-known medical imaging specialities [5].

Medical pictures created by medical imaging methods can vividly depict the interior structure of a specific organ or tissue in relation to a patient [4]. By examining the pictures, experienced radiologists or doctors can identify potentially worrisome lesions. Appropriate diagnosis and therapy for the patient can be preliminarily confirmed based on appropriate judgments. In summary, medical imaging methods aid physicians in the clinical setting through two processes that occur between physicians and images: visual perception and cognition [5].

Medical imaging has become a growing area of study for computer science experts in recent years. They integrate computational intelligence techniques into the field of medical imaging to aid physicians in analysing patients' unusual situations and attempting to discover viable remedies. Neural networks, evolutionary optimization methods, and colour analysis of wound inflammation all contribute to the advancement of the medical imaging field [6]. Additionally, computer science experts' research is directed toward various medical imaging techniques and specific organs or tissues of the human body. For instance, image-guided lung biopsy, ultrasound imaging for osteoarthritis diagnosis in the knee, and virtual surgery [7]. The advancement of medical imaging technology has resulted in the resolution of several unresolved situations in the medical community.

As previously stated, two processes–visual perception and cognition–affect doctors' decision-making about diagnosis and treatment based on medical pictures provided by medical imaging technology. The procedure, however, is not always accurate and perfect. Subjective influencing variables, such as the lighting environment in the room and the picture display equipment; Objective influencing elements, such as distortions in medical pictures, both of these types of influential variables have a role in the processes [5]. Rather than that, erroneous measurements of subjective and objective influencing variables may provide diagnostic conclusions that are inconsistent with the patient's real condition. IQA must be included in this process to assure the highest possible quality of medical pictures.

B. Image Quality Assessment

In the field of computer science, picture quality assessment is a critical technique for determining the image's quality. IQA is subdivided into two components: Subjective IQA and Objective IQA [2]. The complete categorization of IQA is depicted in Figure 1.

Manuscript received October 18, 2021; revised January 3, 2022.

Yuhao Sun is a PhD student of University of Edinburgh, Scotland. E-mail: yuhao.sun@ed.ac.uk

Gabriela Mogos is an Associate Professor of School of Advanced Technology, Department of Computing, Xi'an Jiaotong-Liverpool University, China. E-mail: Gabriela.Mogos@xjtlu.edu.cn

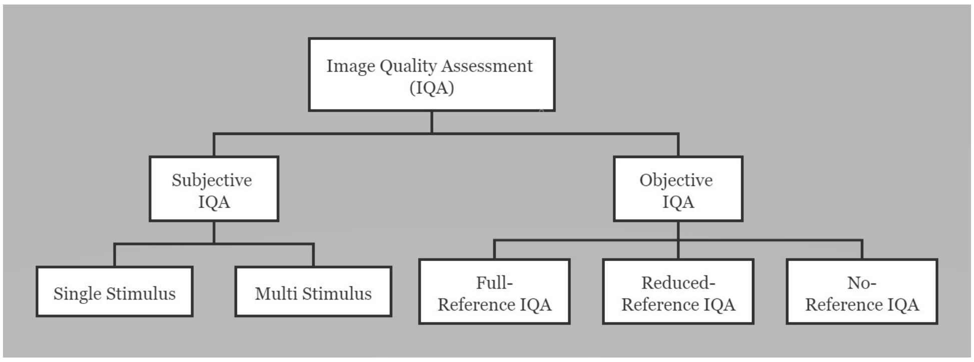

Fig. 1: The Full Classification of IQA

1) Subjective Image Quality Assessment: Prior research shows that Subjective IQA is the most trustworthy technique to assess the quality of a picture as the eventual users of most multimedia applications are human beings [8]. Specifically, in Subjective IQA, the selected observers are needed to evaluate the quality of the provided pictures, within a defined duration [8]. The previous researches have studied two primary categories are under Subjective IQA, including Single Stimulus Methods and Multi Stimulus Methods [8].

The most important difference between the two aforementioned approaches is the number of stimuli [2]. In Single Stimulus Methods, just one incentive can be supplied to observers, i.e. one test picture. In Multi Stimulus Methods, two incentives can be provided to observers, which comprise a reference picture and corresponding test image. Now please imagine a scenario, when a person sees a picture in the smartphone and is required to judge the quality of the picture, this is a Single Stimulus Methods; when a person sees two pictures and has been told one of two is a high-quality image and is required to judge the quality of another picture, this is a Multi Stimulus Methods. For all two approaches, except the aforementioned difference, the human observers are needed to classify the quality of the test image according to a 5-point Likert scale, within a set time [9].

An additional concern is how to effectively minimise the limits and downsides Subjective IQA presented by. Although we can expect from the name of this approach that the results of this approach is still highly subjective, it is exactly important to control the negative impact of influential factors including system, context, and human influential factors [10], which usually include system, context and human influential factors [11]. System important factors are the techniques we employ, in this situation, are Single Stimulus Methods. Context important elements include seeing conditions while processing the assessments. Human influencing variables are the level of human emotional moods. The key to guaranteeing the effectiveness of the evaluation is to reduce the negative impacts produced by these significant elements to the fullest extent.

Additionally, numerous researches are consistent with that Subjective IQA is a costly and time-consuming technique [2], [8]. Indeed, the human observers are necessarily required in Subjective IQA makes the approach expensive; it takes time for observers to complete the assessment causes the technique time-consuming. Even so, Subjective IQA plays a vital part in IQA research, as the final consumers of it are completely human people [2].

2) Objective Image Quality Assessment: Objective IQA is meant to deliver the numerical score values created by mathematical models, which should function similarly with human observers [8]. Objective IQA may be classified into three categories, which are Full-Reference IQA (FR-IQA), Reduced-Reference IQA (RR-IQA), and No-Reference IQA (NR-IQA), depending on the availability of reference pictures [12]. Like their names, FR-IQA indicates reference pictures are available; RR-IQA means reference images are partially accessible, and NR-IQA means there are no reference images [12], [13].

FR-IQA indicates reference pictures are accessible fully, which can be considered as ”high-quality” or ”distortion-free”. Under FR-IQA, the algorithms firstly perceive the reference pictures and then assess the test images. Famous mathematical metrics of FR-IQA include Mean Squared Error (MSE) and Peak Signal to Noise Ratio (PSNR), in which PSNR is a reduced complexity version transformed by MSE [14]. Different from the previous two models, the classic model Structural Similarity (SSIM) creates the structure of the picture and therefore compare the similarity of the reference image and matching test image [15]. SSIM is commonly utilised in the study since the findings of it are have a more substantial likeness to the results created by human eyes, compared to MSE and PSNR [14], [15]. In addition to MSE and SSIM, [16], [17], [18] and [19] are other FR-IQA models.

RR-IQA implies reference pictures are partially accessible, midway between FR-IQA and NR-IQA. For example, there are some watermarks on the photos [20]. In the 1990s’, without current sophisticated technology, it constantly happened that the entire picture could not be retrieved from the videos on multimedia communication networks; instead, certain aspects of the image, which is where RR-IQA was created from [21]. RR-IQA is less often utilised than FR-IQA and NR-IQA. There are three types of RR-IQA techniques, including the models of source pictures, the models of distortion of recorded images and the models of Human Visual System [20]. [22], [23], [24], [25] and [26] are some famous mathematical models of RR-IQA covering above three types.

NR-IQA indicates reference pictures are absent totally, meaning the image quality solely may be assessed by the corresponding test images. NR-IQA is matching most of the situations that transpired in actuality. Now imagine a scenario, you have a photo on your phone and are asked to assess the quality of it - this is precisely NR-IQA. The key characteristic of NR-IQA is there are no reference pictures but test images. The purpose of NR-IQA models seeks to develop mathematical models which can sense the quality of pictures automatically and are similar to the results obtained by human eyes to the fullest extent [27]. NR-IQA is more difficult than FR-IQA and RR-IQA, as the models need to consider numerous unexpected distortion types [27]. The human can assess the quality of the image without the reference image is because the brains are sufficiently informed to store a lot of information which informs them how a good-quality picture should seem like [8], [12]. Most commonly adopted NR-IQA models include Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [28] and Naturalness Image Quality Evaluator (NIQE) [29]. In this work, we also utilised [30] another mathematical model , Perception based Image Quality Evaluator (PIQE) [31]. [32]–[36] are alternative mathematical models of NR-IQA.

C. Method Choices

There are two approaches to Subjective IQA (Single Stimulus Methods and Multi Stimulus Methods). There are three approaches used in Objective IQA (FR-IQA, RR-IQA, and NR-IQA). However, doing Subjective IQA and Objective IQA does not need performing all of the techniques associated with each IQA.

The research items in this study are medical pictures. In Subjective IQA, we seek qualified physician observers to assess the quality of medical pictures in a manner comparable to what they encounter on the job. Single Stimulus Methods in Subjective IQA should be chosen carefully in this situation, as there are no reference pictures for physicians to assess. To maintain the same quality in Objective IQA, which is devoid of reference pictures, we must choose NR-IQA. Eventually, Single Stimulus Methods and NR-IQA were chosen for Subjective IQA and Objective IQA, respectively.

III. Methodology

A. Dataset Construction

The approach begins with the selection of an appropriate dataset. It is important to consider the features that should be included in this project-specific dataset. In summary, before generating the dataset, we should properly evaluate the following factors:

1) Image files: "Image files" refers to two-dimensional representations of visual experience, particularly PNG/JPEG files. A direct explanation for this is that certain picture collections are presented as numerical data in the context of computer science. However, using solely statistical data is unacceptable since it is difficult for human observers to judge the picture quality in Subjective IQA by reading a sequence of numbers.

2) Distortion-free: The medical images used should be of the highest quality. The image is distortion-free. These images are then considered reference images and can be processed further to create various types of distortion images.

3) Currency: The medical images used should be recent, which implies they should not have been generated "a long time" before. While this seems imprecise, medical images generated within the last five years (i.e., since 2015) can be classified as a useful currency. A significant reason for using the time period "within five years" is because Google Scholar only displays the amount of "All Citations" and "Since 2015 Citations" for each personal profile, to showcase the author's total number of paper citations and current active citations.

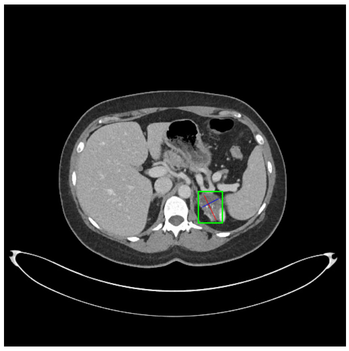

Fig. 2: An Annotated Medical Image

4) Annotations: When a lesion arises in a medical image, we must at the very least know its precise position and size. Medical specialists' instruction on how to assess lesions is critical. The chosen medical images should have the necessary comments. Refer to Figure 2 for an example of an annotated image. A lesion is defined as the area contained by the green frame.

After amassing the medical images, we decide to transform them into various sorts of distorted images. Several comparable images can be analysed in subsequent assessments as a group. We intended to transform 20 original ("high-quality") images into 30 deformed images. The distortion kinds we process are determined by the frequency with which certain types occurred during doctors' real practice. Figure 3 contains a sample set of processed images. As a result, the final collection will have 50 medical images in total.

B. Subjective Image Quality Assessment (Stage I)

Before the official assessment, we would want to examine the research subject from the standpoint of medical personnel [37]. As this issue is centred on medical and clinical contexts, the concept from the medical domains other than computer science seems to play a shared vital function. Therefore, a questionnaire is essential. The questionnaire comprises two components surrounding "the distortion" topic: personal background information and medical imaging related questions. We firstly ask numerous questions regarding position and experience, then ask them questions concerning the distortion according to their perspectives. The number of items should be regulated between eight and twelve, and most of them should be MCQ instead of sentence responses. Most significantly, the questionnaire cannot be prepared for replying longer than five minutes since failing to do so may result in a bad mental effect on takers, such as pressure.

The respondents should be either clinical doctors or radiologists and currently are working at a valid hospital or healthcare institution, which is because radiologists and clinicians are supporting each other in the clinical area [38], and the eventual decisions from clinicians are possibly different after the investigations and discussions from radiologists [39].

To summarise, the questionnaire should be designed as follows:

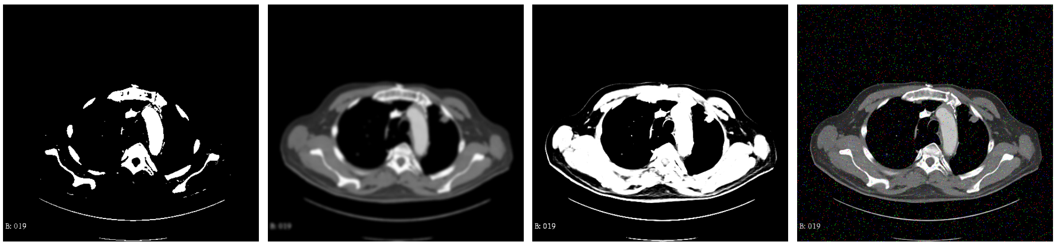

Fig. 3: Sample Set of Processed Images (from left to right: original, blurry, contrast-distorted, and noisy)

- Create and gather data using Survey@XJTLU (a questionnaire platform);
- Concentrate solely on clinical physicians' and radiologists' responses;
- Respond to the questions on medical image distortion;
- Design a restricted number of questions, the majority of which are MCQs, so that they can be completed in a short amount of time;
- Respondents clearly understand where their data will be used and have the ability to remove their data once it has been finished.

For detailed method and assessment information, please refer to another article we have published [37].

C. Subjective Image Quality Assessment (Stage II)

The observers shall evaluate the quality of the medical image in the dataset during the formal evaluation of Subjective IQA. The observers are expected to view the given medical images and then respond to four questions on the image's quality, including the following:

1) Can you get useful information from this image?
2) Do you think there is at least a lesion in this image?
3) To what extent do you think the quality of this image is good enough for you to get the above answer?
4) To what extent do you think the quality of this image is similar to the images you actually encountered during your work?

Questions 1 and 2 are Yes/No type questions. Questions 3 and 4 are scoring type questions based on a 5-point Likert scale. On each page, the observers are allowed to view one unique medical image in detail and then answer four questions aforementioned linked. They are permitted to click "Next Page" and view the next medical image once they have made judgments for the current image. Before the commencement, all observers have been informed of the terms of the Subjective IQA and then are needed to sign the consent papers. They also need to answer numerous inquiries on the preferences of the research.

We visited roughly ten doctors and expect three participants involve in Subjective IQA, two clinicians and one radiologist. We take the findings of one clinician and one radiologist into consideration in Subjective IQA as well as the data from another clinician are utilised in testing and evaluation. Considering the potentially detrimental consequences produced by influencing variables, we made the following actions: 1) to choose a sunny day as the assessment day and ask several questions related to personal emotions (prevent lousy mood caused by the weather or other reasons); 2) to set all participants to have the assessment in the same day (ensure that most environmental factors are consistent); 3) to adopt the places and computer displays which the participants usually use for daily work, in the assessment.

To demonstrate the third action in the above further, this was meant to acquire the findings as much comparable as when the physicians and radiologists operate in a regular scenario, instead of holding an assessment in a conventional method, such as utilising the same facilities and under the same hall. Specifically, most of the time, clinicians are in the clinic rooms or departmental office when they are seeing the medical images presented on the computer displays; nevertheless, radiologists are generally utilising the more high-quality and customised displays when they process the medical images in the office. Our Subjective IQA was conducted in such a method. The techniques linked to minimising the negative impacts of influencing elements in Subjective IQA followed the guidelines offered by International Telecommunication Union [40].

D. Objective Image Quality Assessment

In Objective IQA, MATLAB Image Processing Toolbox [41] has been chosen. There are seven sections in the toolbox. Correctly, we employ functions in the four components, including "Import, Export, and Conversion", "Display and Exploration", "Image Filtering and Enhancement" and notably "Image Segmentation and Analysis". For each mathematical model, three stages are processed and presented on a step-by-step basis below:

1) Image Quality Score: All medical images in FYP-Dataset have been scored using the relevant mathematical model, i.e., BRISQUE and NIQE, in MATLAB. In every No-Reference IQA, the lower score indicates the greater quality. The score can vary from 0 (the greatest quality) to 100 (the least quality).

2) Initial Check: For the scores provided by the model aforementioned, a first check is necessary. The goal of the initial check is to make sure the findings are logical. In the initial check, the quality of the original image should be better than deformed images. One of the probably acceptable conditions is the score of the original image is lower than the score of the distorted image for the identical original image set. Outliers should be deleted cautiously. As stated before, there are 50 medical images in FYPDataset, including 20 original medical images, which means we need to do the first checks for all 20 groups.

3) Confirm the Ultimate Result of Objective IQA: We examine the outcomes across all mathematical models utilised and then pick the optimal performance model. The matching assessment scores are considered as the Objective IQA findings, which are employed in the last stage, analysis.

IV. Results and Analysis

A. Subjective Image Quality Assessment I

Here are some key findings from complete statistics. In all, 59 questionnaires were returned, and 51 of those returned were declared genuine. All responders to these 51 valid surveys are now employed in a hospital or healthcare facility, with 84% being clinicians and 16% being radiologists. 90% of respondents believe their degree of professional experience is comparable to or greater than the average for similar or identical professionals.

92% of respondents claim to have experienced the problem of medical image distortion, and 98% of these respondents indicate that this is not a common occurrence, with 78% selecting the lowest frequency level. According to 76% of respondents, the most often occurring distortion is fuzzy. Certain negative consequences may result from the distortion, including "misdiagnosis" (63%) and "inaccurate assessments of illness severity" (63%). 86% of all respondents believe that medical image distortion is a serious concern for the medical community.

Radiologists have far fewer responders than clinicians, which is not surprising given that the number of radiologists in most hospitals/healthcare facilities is significantly less than the number of clinicians. In light of the questionnaire responses, we decided to place a greater emphasis on this sort of distortion, blurred. The findings of the most common forms of distortion aided in the construction of the FYPDataset. Additionally, their (the respondents') high assessment of their level of expertise demonstrates that their replies have a high degree of trustworthiness.

The data have provided us with some first insight into how medical images are distorted from the perspective of medical professionals. In the section Comparative Analysis of this chapter, we analyse some of the questionnaire's data in further detail. After that, the formal assessment takes place. Candidates for formal evaluation were chosen from responders who provided their contact information in the questionnaire's final question.

B. Subjective Image Quality Assessment II

As seen in the phase Dataset Construction, the FYPDataset contains 50 medical images. Observers should make judgments about the quality of each of the 50 medical images based on their own experience. Additionally, we established a 50-minute time restriction for the evaluation, which was created in accordance with the principles of Single Stimulus Methods. There are four questions about the image quality of each medical image. There are a total of 200 questions.

Three doctors, two of whom are clinicians and one of whom is a radiologist, were finally involved in the official assessment of Subjective IQA. The participants were thoroughly briefed on the topics of Subjective IQA and the specific activities they would do during the assessment. They then signed the consent papers if they had no more queries. Additionally, an electronic copy of the permission form was provided to each participant's mailbox upon successful completion of the evaluation. More significantly, individuals have all been notified that they have the right to withdraw all of their own experimental data prior to May 31, 2020, if they are dissatisfied with the research or have other personal reasons, as mentioned explicitly on the permission form.

TABLE I: Subjective IQA Observers

Observer Index	Position	Experience (/yrs)	Time Cost (/mins)	Data Usage
001	Clinician	>13	31	Subjective
002	Clinician	3	40	Testing
003	Radiologist	>13	23	Subjective

The full details of three observers in the formal assessment of Subjective IQA please refer to Table I.

C. Objective Image Quality Assessment

Throughout the experiment, Objective IQA has played a critical role. The entire procedure of Objective IQA is carried out within the MATLAB programme. The Image Processing Toolbox in MATLAB contains an extremely comprehensive and powerful collection of mathematical models and algorithms for image processing. It is divided into several subsections, including but not limited to Geometric Transformation and Image Registration, Image Filtering and Enhancement, Image Segmentation and Analysis, and Deep Learning for Image Processing. In the subheading Image Segmentation and Analysis, we focused on the Image Quality portion.

As previously said, three phases have been processed and presented step by step for each mathematical model, including "Result Image Quality Score," "Initial Check," and "Confirm the Ultimate Result of Objective IQA."

1) BRISQUE: The mathematical model BRISQUE has been conducted firstly. In the publication [28], Mittal and colleagues have suggested and given a software release of BRISQUE in MATLAB. In addition to this, as indicated previously, we also use the functions (particularly "brisque()") in the Image Processing Toolbox supplied by MATLAB.

For the following phase, the initial check might provide us with a sense of whether these findings are desired or not. In the initial check for BRISQUE, only the BRISQUE scores in 15 medical images of 6 groups are within our expectations, i.e., 30%. In this condition, the IQA scores generally vary from 40 to 55, therefore there is not much difference although the quality of one image is lower or higher. However, on another hand, the quality of most original images is assessed as being worse than the quality of equivalent distorted images, according to the IQA ratings given by BRISQUE. Obviously, it is not rational.

Taking a collection of results as an example, the original image labelled as "F51_030", two other images are distorted images. The types of distortion respectively are "contrast-issue with a grayscale range [0.1 0.4]", and "blurry with Gaussian filter with a standard deviation of 0.5". However, the scores were 50.6976 (original), 49.9166 (contrast-distorted), 50.1219 (blurry) (blurry). Three scores are quite near yet it still can be noticed that the contrast-distorted image is the best quality from the model's standpoint. This happens frequently in the results of BRISQUE and, it may indirectly show that BRISQUE is not trustworthy in this investigation.

Less than one-third of the findings qualified shows a serious fault that occurred by practising BRISQUE in the medical imaging field. Therefore, practising various models is highly essential.

2) NIQE: Following BRISQUE, the mathematical model NIQE was run. As with BRISQUE, Mittal and colleagues suggested and released NIQE software in November 2012 [29]. The experiment ran well because of the functions in MATLAB's Image Processing Toolbox, especially ”niqe().”

As with BRISQUE, the initial check might indicate whether or not these outcomes are desired. Regrettably, only 28% of the NIQE scores in 14 medical images from six groups are within our expectations, which is slightly lower than the findings of BRISQUE. Around one-quarter of qualifying findings indicate a severe defect caused by the use of another mathematical model, NIQE, in the medical imaging profession.

Consider the initial check findings for BRISQUE and NIQE, which were 30% and 28%, respectively; both are significantly different from our initial assumptions. We decide to do a thorough review of the whole Objective IQA procedure before proceeding to the next level. Two procedures are required: 1) thoroughly inspect the MATLAB files and debug the associated algorithms in case of any problems; and 2) independently test two mathematical models, BRISQUE and NIQE, using 20 natural images and another 20 unneeded medical images (not included in FYPDataset).

We reviewed the manual numerous times and conducted thorough debugging before concluding that there were no probable mistakes in the scripts given here. This demonstrated a possible mistake caused by the image type. By comparing the results of 20 natural images and 20 other medical images in BRISQUE and NIQE, we discovered that the results of the majority of natural images were logical (i.e., the quality of the original image should be higher than the quality of distorted images, and the IQA score of the original image should be lower than the scores of distorted images); however, whether in BRISQUE or NIQE, the results of medical images remained consistent with the previous experiments. In summary, we may expect that BRISQUE and NIQE would likely underperform, particularly in the medical imaging field, or at least in our instance, medical images obtained by X-ray CT.

Due to the fact that both BRISQUE and NIQE produced findings that were considerably lower than our expectations, another alternative mathematical model, dubbed PIQE, should be examined, as we stated in related work.

3) PIQE: Finally, the mathematical model NIQE was conducted. This time, we utilised the same toolbox in MATLAB as previously and the method ”piqe().”

The initial check has been completed using the scores given by the alternative model PIQE. Surprisingly, 41 out of 50 medical images in 16 images are within our expectations or 82%. This finding is markedly different from prior models' predictions. In comparison to the findings in BRISQUE (30%) and NIQE (28%), this demonstrates the PIQE model's effectiveness in the medical imaging discipline. Observe the findings further; we can already detect an apparent difference in scores between the original and distorted images, demonstrating the PIQE model's rather high efficiency. The PIQE results are regarded as the final experiment results for Objective IQA.

TABLE II: An Example of Subjective and Objective IQA Matching

	Scores	
	Subjective IQA	**Objective IQA**
Original	53	5
Blurry	52	4
Contrast	46	3

D. Data Analysis

The final stage is to conduct a comparative analysis of the experiment data collected before. As a result, this section summarises the major and minor findings. Figure 4 illustrates the numerical results of Subjective IQA as well as all Objective IQA. Please keep in mind that a higher score indicates higher image quality in Subjective IQA, but a lower score indicates higher image quality in all Objective IQA (BRISQUE, NIQE, and PIQE).

The findings of Subjective IQA (containing the Pre-Assessment Questionnaire) and Objective IQA were compared. According to the analysis's properties, the results are divided into three subsections:

- Main Results
  In this section, the main and significant findings are demonstrated.
- Clinical Results
  In this section, the finding relevant to the clinical area (e.g. clinicians and radiologists) are demonstrated.
- Computational Results
  In this section, the findings relevant to the mathematical models and computing discipline are demonstrated.

Clinical and computational outcomes are equally essential, as are the results that address the research issues.

Subjective IQA and Objective IQA findings are analysed in groups, which implies that comparable medical images should be analysed together. The original ”high-quality” image is included in a set of comparable medical images, as are additional distorted images processed by the original. We use the Subjective IQA findings as a guide. The two IQA are deemed identical/similar when the Subjective IQA findings for image quality are ”original image” (best), ”blurry image” (middle), and ”contrast-distorted image” (worst); and Objective IQA can also provide the same connection between images. Refer to Table II for an example of subjective and objective IQA matching, and Table III for an example of mismatching. Please keep in mind that higher Subjective IQA scores correspond to greater image quality, but lower Objective IQA scores correspond to higher image quality.

1) Main Results:

- The experiment results of Subjective IQA and Objective IQA were similar, with 80% similarity.
- PIQE (Perception-based Image Quality Evaluator), a mathematical model, performed admirably in our experiments, which were conducted in the context of medical imaging.

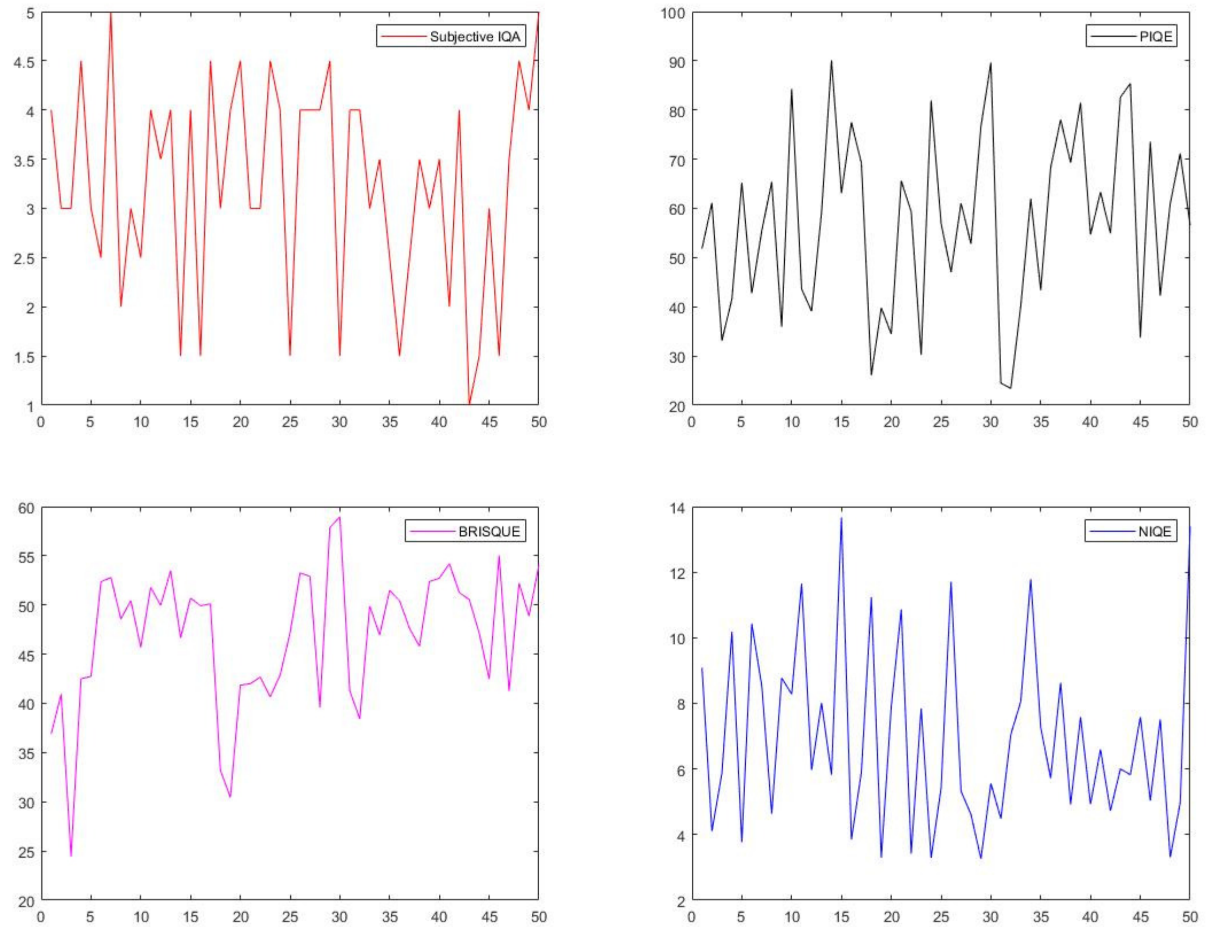

Fig. 4: Subjective and Objective IQA Results (including PIQE, BRISQUE, and NIQE). X-axis is image index, Y-axis is assessment scores.

TABLE III: An Example of Subjective and Objective IQA Mismatching

	Scores	
	Subjective IQA	**Objective IQA**
Original	46	5
Blurry	52	4
Contrast	53	3

2) Clinical Results:

- For the most part, clinicians and radiologists share the same or comparable judgments about the quality of medical images.
- Both clinicians and radiologists were highly sensitive to the changes of contrast value of medical images in the experiments.
- Within a certain range of standard deviation for Gaussian filter, the distortion of blurry in medical images possibly might not cause the negative effects to doctors.
- The experiment results in Subjective IQA were similar to the results in Pre-Assessment Questionnaire.

3) Computational Results:

- In medical imaging discipline, BRISQUE and NIQE models (no custom) did not have good performances when assessed the quality of X-ray Computed Tomography (CT) scans.
- In the range (0.5, 1.0) for Gaussian filter with a standard deviation, the blurry distortion may not result in a negative effect to doctors.
- In a certain range of grayscale for contrast values, the quality of the medical images may possibly improve to some extent. Specifically, the range for "low_in" in the grayscale = (0.1, 0.3) and the range for "high_in" in the grayscale = (0.6, 0.8).

The results indicate that the mathematical model employed in Objective IQA is capable of automatically predicting perceived image quality to the same degree as human judgements in Subjective IQA, which is consistent with our predictions. Additionally, two mathematical models in NR-IQA, BRISQUE and NIQE, do not perform well in X-ray CT scans; yet, these two models have traditionally been considered as two of the finest NR-IQA models during the last decade.

V. Discussion and Evaluation

In order to analyse the findings above that have been generated, testing is required to undertake. The technique

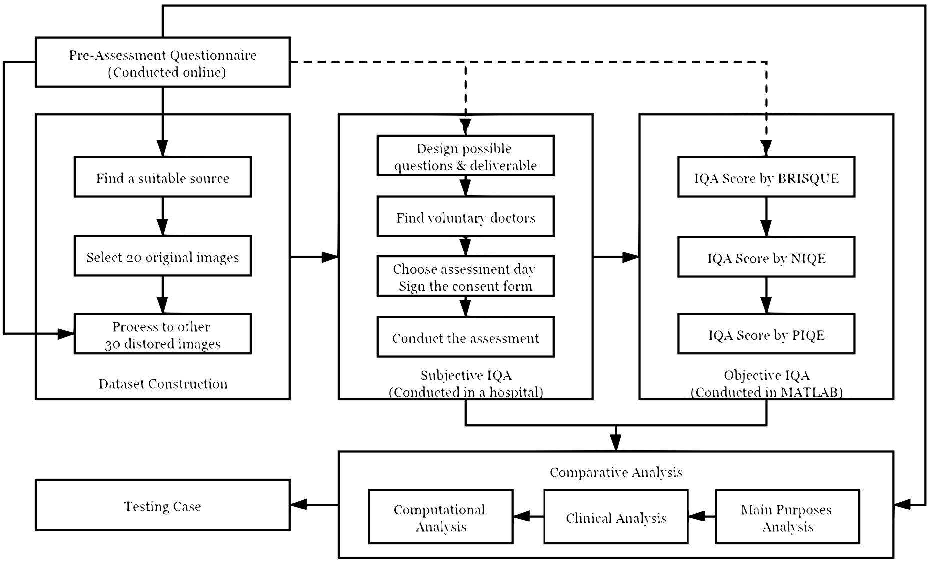

Fig. 5: Research Flowchart

may be retained as same as previously, for the reason that this has been assessed by the supervisor and us in the Specification and Design Report [42]. However, the testing performed using fresh data samples which have never been utilised in the previous phases.

For the testing sample of Subjective IQA and Objective IQA, we first build a new testing dataset of 20 medical images, and then practise IQA using the medical images in this dataset. This new dataset comprises 8 original medical images, 7 medical images as blur distorted, 4 medical images as contrast distorted, and 1 medical image as noise distorted, which matches the proportion of distortion kinds in the FYPDataset. In Subjective IQA, we accept the outcome statistics of the second observers (index is 002), as shown in Table I. In Objective IQA, the same three mathematical models are all examined, notably PIQE.

We have examined the outcomes of Subjective IQA and Objective IQA comparably, as to how we have done earlier. Unsurprisingly, the same outcomes spanning the primary findings, clinical results, and computational results may be obtained. Broadly speaking, the testing findings indicate that there is no substantial error, defect, or coincidence in our prior trials. Our results given in the section Comparative Analysis are in a generally fair approach. Additionally, a comprehensive flowchart of the investigation may be referenced in Figure 5.

The next work will mainly concentrate on natural aberrations and unique situations. The FYPDataset in the research comprises actual medical images and the associated distorted images. Please notice that distorted images are entirely artificial created by MATLAB, in this case, as illustrated in the part of the technique, Dataset Construction. This method of introducing common distortion artificially is used by most of the prior relevant work. As a matter of fact, the formats of picture data are different. Meanwhile, the various layers of distortion conceivable can occur throughout the process of generation, transmission, and handling, which is hard to replicate all by simply a computer and particular software [43]. Therefore, the deformed image simply processed by MATLAB cannot reflect the circumstances that transpired in real applications accurately. Due to the constraint of the source dataset, the extensibility of the mathematical models has been constrained in the same way [43]. Consequently, it is substantially necessary to focus on the study about natural distortions for IQA, in the following phase.

Besides, the results indicate that acknowledged mathematical models for IQA may ”crash” given the specific situation. It also indicates more experimental testing and investigations are needed to identify these unique scenarios, so as to improve the performances of existent mathematical models. We are fully aware that it is incredibly challenging to develop a mathematical model of IQA for research usage; nonetheless, we are optimistic that our study will serve as a basis for future studies on the image distortion problem under diverse scenarios. For example, the visual deterioration induced by tone-mapping of HDR images [44]. Future studies should concentrate on increasing the capacity to manage the image distortion difficulties under diverse conditions, as well as building a more sophisticated mathematical model for No-Reference IQA for the full-aspects situations, which is also a crucial topic for future research on IQA.

VI. Conclusion

Our research has led us to believe that both subjective and objective image quality assessments (IQA) can yield comparable findings at this point. Additionally, the collected findings indicate that the Perception-Based Image Quality Evaluator (PIQE) is a rather acceptable mathematical model for assessing the quality of non-referenced images (NR-IQA). We discovered an innovative result: both the Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) and the Naturalness Image Quality Evaluator (NIQE) performed poorly in comparison to PIQE in the context of medical imaging, specifically X-ray CT scans in our case; however, BRISQUE and NIQE are considered to be two of the best NR-IQA models. We used numerous assumptions to get at this finding, and future research will examine the underlying causes of this phenomenon in greater detail. There are also a few modest results in the clinical and computational sciences. Despite the constraints inherent in Subjective IQA owing to the small number of observers, we employ appropriate techniques to assure the elimination of coincidences, mistakes, and outliers. This article lays the groundwork for future medical imaging technology equipment to address the issue of distortion, which might potentially result in substantial advancements in the medical imaging field. We anticipate that more testing and research will corroborate our findings and enhance our understanding of NR-IQA in the context of medical imaging.

REFERENCES

[1] Y. Sun and G. Mogos, "Data Analysis of Medical Images", *International Journal of Design, Analysis & Tools for Integrated Circuits & Systems*, vol.9, no.1, pp. 37-40, 2020.

[2] L. Lévêque et all, "On the subjective assessment of the perceived quality of medical images and videos", *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1-6, 2018.

[3] J.K. Beutel, H.L. Van Metter and L. Richard, "Handbook of medical imaging", *Spie Press*, vol.1, 2000.

[4] K.K. Shung, M. Tsui and M.W. Benjamin, "Principles of medical imaging", *Academic Press*, 2012.

[5] A.E. Krupinski, "'Current perspectives in medical image perception", *Attention, Perception, & Psychophysics*, vol.72, no.5, pp. 1205-1217, Springer, 2010.

[6] G.H. Schaefer and A.J. Jianmin, "Computational intelligence in medical imaging: techniques and applications", *CRC Press*, 2009.

[7] K.W. Lai and D.E.O. Dewi, "Medical Imaging Technology: Reviews and Computational Applications", *Springer*, 2015.

[8] E-M.P. Mohammadi and A.S. Shahram, "Subjective and objective quality assessment of image: A survey", *arXiv preprint arXiv:1406.7799*, 2014.

[9] Series BT, "Methodology for the subjective assessment of the quality of television pictures", *Recommendation ITU-R BT*, pp.500-13, 2012.

[10] U.B. Reiter and K.K. De Moor, "Quality of Experience: advanced concepts applications and methods", *T-Lab Series in Telecommunication Services*, pp. 55-72, 2014.

[11] P. Le Callet, S.P. Möller et all, "Qualinet white paper on definitions of quality of experience", *European network on quality of experience in multimedia systems and services (COST Action IC 1003)*, vol.3, no. 2012, version 1.1.

[12] Z. Wang and A.C. Bovik, "Modern image quality assessment", *Synthesis Lectures on Image, Video, and Multimedia Processing*, Morgan & Claypool Publishers, vol.2, no.1, pp. 1-156, 2006.

[13] V. Azad and P. Sharma, "A review on objective image quality assessment techniques", *International Journal of Emerging Engineering Research and Technology*, vol.2, no.5, pp.188-192, 2014.

[14] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM", *2010 20th International Conference on Pattern Recognition*, pp.2366-2369. 2010.

[15] Z. Wang et all, "Image quality assessment: from error visibility to structural similarity", *IEEE Transactions on Image Processing*, vol.13, no.4, pp. 600-612, 2004.

[16] Z. Wang, E.P. Simoncelli and A.C Bovik, "Multiscale structural similarity for image quality assessment", *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers*, vol.2, pp.1398-1402, 2003.

[17] R.H. Sheikh and A.C. Bovik, "Image information and visual quality", *IEEE Transactions on Image Processing*, vol.15, no.2, pp.430-444, 2006.

[18] E.C. Larson and D.M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy", *International Society for Optics and Photonics*, vol.19, no.1, 2010.

[19] L. Zhang, L.M. Zhang and Z.D. Xuanqin, "FSIM: A feature similarity index for image quality assessment", *IEEE Transactions on Image Processing*, vol.20, no.8, pp. 2378-2386, 2011.

[20] Z. Wang and A.C Bovik, "Reduced-and no-reference image quality assessment", *IEEE Signal Processing Magazine*, vol.28, no.6, pp. 29-40, 2011.

[21] A.A. Webster et all, "Objective video quality assessment system based on human perception, *Human Vision, Visual Processing, and Digital Display IV - International Society for Optics and Photonics*, vol. 1913, pp. 15-26, 1993.

[22] A. Rehman and Z. Wang, "Reduced-reference image quality assessment by structural similarity estimation", *IEEE Transactions on Image Processing*, vol.21, no.8, pp. 3378-3389, 2012.

[23] Z. Wang et all, "Quality-aware images", *IEEE Transactions on Image Processing*, vol.15, no.6, pp. 1680-1689, 2006.

[24] S. Wolf and M.H. Pinson, "Spatial-temporal distortion metric for in-service quality monitoring of any digital video system", *Multimedia Systems and Applications II, International Society for Optics and Photonics*, vol.3845, pp. 266-277, 1999.

[25] K. Chono, L. Keiichi, V. Yao-Chung, M. David, G. Yoshihiro, "Reduced-reference image quality assessment using distributed source coding", *2008 IEEE International Conference on Multimedia and Expo*, pp. 609–612, 2008.

[26] M.LC. Carnec and B.D. Patrick, "Visual features for image quality assessment with reduced reference", *IEEE International Conference on Image Processing*, vol.1, pp. I-421, 2005

[27] P. Ye and D. Doermann, "No-reference image quality assessment using visual codebooks", *IEEE Transactions on Image Processing*, vol.21, no.7, pp.3129-3128, 2012.

[28] A. Mittal, A.K. Moorthy and A.C. Bovik, "No-reference image quality assessment in the spatial domain", *IEEE Transactions on Image Processing*, vol.21, no.12, pp. 4695-4708, 2012.

[29] A. Mittal, R. Soundararajan and A.C. Bovik, "Making a "completely blind" image quality analyzer", *IEEE Signal Processing Letters*, vol.20, no.3, pp. 209-212, 2012.

[30] "Train and Use No-Reference Quality Assessment Model - MATLAB & Simulink - MathWorks United Kingdom", https://uk.mathworks.com/help/images/train-and-use-a-no-reference-quality-assessment-model.html, accessed on 04/05/2021.

[31] N.P. Venkatanath, D.Bh.M. Chandrasekhar and S.S.M. Channappayya, "Blind image quality evaluation using perception based features", *2015 Twenty First National Conference on Communications (NCC)*, pp.1-6, 2015.

[32] H.R. Sheikh, A.C. Bovik and L. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000", *IEEE Transactions on Image Processing*, vol.4, no.11, pp. 1918-1927, 2005.

[33] L.W. Liang et all, "No-reference perceptual image quality metric using gradient profiles for JPEG2000", *Signal Processing: Image Communication*, vol.25, no.7, pp.502-516, 2010.

[34] T. Brandão, Tomás and M.P. Queluz, "No-reference image quality assessment based on DCT domain statistics", *Signal Processing*, vol.88, no.4, pp. 822-833, 2008.

[35] Z. Wang, H.R. Sheikh and A.C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images", *Proceedings. International Conference on Image Processing*, vol.1, pp. I-I, 2002.

[36] R. Ferzli and Lina J. Karam, "A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB)", *IEEE transactions on image processing*, vol.18, no.4, pp. 717-728, 2009.

[37] Y. Sun, Y. Zhao and J. Sun, "Subjective Image Quality Assessment: A Pre-Assessment on Visual Distortion of Medical Images by Clinicians and Radiologists", *2020 7th International Conference on Information Science and Control Engineering (ICISCE)*, pp. 1378-1381, 2020.

[38] L.S. Dalla Palma, F. Meduri, S.G. J Te, "Relationships between radiologists and clinicians: results from three surveys, *Clinical radiology, Elsevier*, vol.55, no.8, pp. 602-605, 2000.

[39] J.ML. Bosmans et all, "How do referring clinicians want radiologists to report? Suggestions from the COVER survey", *Insights into imaging, Springer*, vol.2, no.5, pp. 577-584, 2011.

[40] Series BT, "Specifications and alignment procedures for setting of brightness and contrast of displays", *Recomandation ITU-R*.

[41] "Image Processing Toolbox™ Reference", https://www.mathworks.com/help/releases/R2019b/, MathWorks, 2019.

[42] S.Yuhao, "Predict the Impact of Visual Distortion on Medical Images", *Final Year Project (CSE305): Specification and Design Report, Xi'an Jiaotong-Liverpool University*, no.1611049, China, 2020.

[43] D. Ghadiyaram and A.C. Bovik, "Massive online crowdsourced study of subjective and objective picture quality", *IEEE Transactions on Image Processing*, vol.25, no.1, pp. 372-387, 2015.

[44] D.G. Kundu et all, "Large-scale crowdsourced study for tone-mapped HDR pictures", *IEEE Transactions on Image Processing*, vol.26, no.10, pp. 4725-4740, 2017.

Yuhao Sun (M'21) Yuhao Sun is PhD student of University of Edinburgh, Scotland. He received his Master's degree in Medical Robotics from the Imperial College London, UK and Bachelor's degrees (Honours) in Information and Computing Science from the University of Liverpool (UK) and Xi'an Jiaotong-Liverpool University (China). His research interests include Human-Computer Interaction (HCI), Human-Machine Interaction, Cognitive Science, Medical Imaging, Computer Graphics, Computer Vision and some of the interdisciplinary fields joint together with techniques of computer science. He is also an active student member of IEEE.

Gabriela Mogos (M'09) Gabriela Mogos is Associate Professor of School of Advanced Technology, Department of Computing, Xi'an Jiaotong-Liverpool University, China. She received her PhD in Computer Science from the Alexandru Ioan Cuza University of Iasi, Romania. Her research interests are in the areas of Image Processing, Data Science and Information Security. She is affiliated with IEEE as senior member. In Springer, Elsevier, IEEE journals, and other scientific publications, she has served as invited reviewer.