

Assessment of Perceptual Distortion Boundary Through Applying Reversible Watermarking to Brain MR Images

Asaad F. Qasim^{1,3}, Rob Aspin¹, Farid Meziane¹ and Peter Hogg²

A.Qasim@edu.salford.ac.uk, R.Aspin@salford.ac.uk, F.Meziane@salford.ac.uk, P.Hogg@salford.ac.uk

¹ School of Computing, Science and Engineering, University of Salford, Greater Manchester, M5 4WT, UK.

² School of Health Sciences, University of Salford, Greater Manchester, M6 6PU, UK.

³ Ministry of Higher Education and Scientific Research, Baghdad, Iraq.

Abstract

The digital medical workflow faces many circumstances in which the images can be manipulated during viewing, extracting and exchanging. Reversible and imperceptible watermarking approaches have the potential to enhance trust within the medical imaging pipeline through ensuring the authenticity and integrity of the images to confirm that the changes can be detected and tracked. This study concentrates on the imperceptibility issue. Unlike reversibility, for which an objective assessment can be easily made, imperceptibility is a factor of human cognition that needs to be evaluated within the human context. By defining a perceptual boundary of detecting the modification, this study enables the formation of objective guidelines for the method of data encoding and level of image/pixel modification that translates to a specific watermark magnitude.

This study implements a relative Visual Grading Analysis (VGA) evaluation of 117 brain MR images (8 original and 109 watermarked), modified by varying techniques and magnitude of image/pixel modification to determine where this perceptual boundary exists and relate the point at which change becomes noticeable to the objective measures of the image fidelity evaluation.

The outcomes of the visual assessment were linked to the images Peak Signal to Noise Ratio (PSNR) values, thereby identifying the visual degradation threshold. The results suggest that, for watermarking applications, if a watermark is applied to the 512x512 pixel (16 bpp grayscale) images used in the study, a subsequent assessment of PSNR=82dB or greater would mean that there would be no reason to suspect that the watermark would be visually detectable.

Keywords: Medical imaging; DICOM; Reversible Watermarking; Imperceptibility; Image Quality; Visual Grading Analysis.

1. Introduction

In most medical imaging systems, the conventional file-based diagnosis has migrated to e-diagnosis workflows. Hospital Information Systems (HIS) and medical imaging platforms generate and manage digital images across many modalities comprising X-ray, Ultrasound, Magnetic Resonance Imaging (MRI), Computerized Tomography (CT), etc. Typically, the images are managed within a digital workflow based on the Digital Imaging and Communications in Medicine (DICOM) standard [1]. Images captured in a hospital are inserted into the Picture Archiving and Communication Systems (PACS) and then transferred to the hierarchically upper PACS systems until they reach the top-PACS. In the top-PACS, the data is permanently archived in tapes, physical drives, or optical supports to become available for the diagnostic workflow services through drawdown within the PACS system [2].

During the production and transmission, the integrity of these medical images, and wider data sets may not be strictly preserved [2]. The exchange of these images through, and across, hospitals, locations and administrative organizations, has become a common practice for many purposes, such as diagnosis, treatment, training, distance learning and medical discussions between clinicians and radiologists [3]. In most cases, this will be within the defined workflows of the PACS systems, but there are also many cases in which images and data are withdrawn from one system to be transmitted to other institutions or people. The capacity to maintain the authenticity and integrity confirmation of these images has become crucial, both within the internal systems and during transferring them to other systems [4].

Digital watermarking has been shown to be a robust approach to ensure the integrity and authenticity of digital data. Digital watermarking is the hiding of information within the digital object. The embedded data can then be detected/extracted to confirm the validity of the object [5]. In critical applications, such as healthcare, there are rigorous controls on data reliability that prevent any deformation of the data as a side-effect of the watermarking operation. Therefore, any robust watermarking technique implemented on the medical image should consider special requirements including imperceptibility, reversibility, and reliability [6].

1.1. Imperceptibility

Usually referred to as invisibility or fidelity, it represents the highest requirement of watermarking systems. A digital watermark is called imperceptible if the original and watermarked images are perceptually indistinguishable. It might be fulfilled by sacrificing either robustness, capacity or both [7]. Robustness indicates the ability of the watermarking scheme to resist to different image processing operations. Capacity refers to the number of bits that can be concealed into the cover image without impacting the image quality. Therefore, a suitable trade-off might be found depending on the desired application [8].

1.2. Reversibility

In the medical domain, if an image is modified during the workflow process a collapse in trust regarding the validity of the images is formed. Any small change to the image could lead to misdiagnosis with possible life threatening consequences, or legal implications. Therefore, fully reversible watermarking techniques have been developed which can completely recover both the original unmodified image and the embedded watermark [9].

Reversible watermarking approaches can be categorized into four groups: compression based [10, 11], histogram modification based [12-15], quantization based [16-19], and Difference Expansion (DE) based [20-22]. Recently, reversible watermarking based on the DE technique has been suggested in many kinds of studies, and they typically exceed the other types of reversible methods in that they offer higher payload capacity and lower complexity compared to the other methods [23-26].

1.3. Reliability

This may be decomposed into two aspects: [8]

- Integrity: the ability to confirm that the data has not been changed without authorization.
- Authentication: the ability to identify data source and verifying that the information relates to the right patient.

Unfortunately, there is no standard approach for automatically assessing the amount of noticeable distortion within watermarked images. Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM) indices are often cited in the literature; however, they do not reflect the characteristics of the human visual system and perceptual process [27]. In exploring the use of digital watermark within medical imaging, the question of how much data could be encoded within the image became an important one to explore and establish trust in the medical environments. This research investigates this issue. Specifically, it seeks to answer two questions; (i) is there a reliable technique to measure the degradation of images that have been watermarked? (ii) is there a threshold of imperceptibility which can be employed to calibrate an automated image quality measure? The aim of this investigation is to determine a set of guidelines for embedding the watermark, in terms of technique and level of modification/data encoding that ensure that the watermarked image has no perceivable difference to the original. This seeks to define an assessment approach based on a clinical trial that can be used to validate the watermarked images, before they are inserted into the PACS system, to ensure their integrity and authenticity within the digital medical workflow. This can be achieved by asking experts in reading medical images to detect the noticeable differences of the anatomical structure of images modified by varying techniques and magnitudes. To the best of our knowledge, there is no similar study conducted before to clinically evaluate the watermarked MR images by using standard quality criteria dealing with the visibility of the anatomical details of the brain.

The rest of this paper is organized as follows. Section 2 explains the techniques used to assess the image quality. Section 3 highlights several studies conducted to evaluate the quality of medical images in terms of the applicability of using them in medical practices. The whole process of generating the watermarked images and conducting the visual assessment is demonstrated in section 4. In section 5, the experimental results including comparison with previously reported studies are presented. Section 6 concludes this work.

2. Assessment of Image Quality

The measurement of image quality is vital for various image processing purposes. In general, image quality scales fulfil three kinds of applications [28].

1. To examine and monitor the image quality in quality control systems.

2. To improve the algorithms and the parameter setting of image processing systems.
3. As an indicator for selecting the applicable image processing algorithms.

Image quality can be evaluated either directly (e.g. physical measurements) or indirectly (e.g. visual/clinical approaches). Physical metrics are easy and commonly used in assessing image quality. However, their efficacy in achieving a measurement which is relevant to the observer judgment is not yet confirmed as they do not consider all the clinical characteristics that are related to medical investigations [29]. Therefore, they should be accompanied by observers' attitudes to ensure their efficiency and validity [30]. Visual assessments are complicated, expensive and time-consuming, which makes them ineffective for real-world applications. They also require specific equipment and conditions to be conducted [31].

2.1. Physical Assessment

The goal of this approach is to design mathematical models that are able to autonomously evaluate the quality of a modified image, against its unmodified version. The similarity between the reference and watermarked images can be measured using the following most commonly adopted metrics [32]. In all of the used equations, $N \times M$ is the images dimension, and I_{ref} and I_{tst} represent the reference and test images respectively.

2.1.1. Peak Signal to Noise Ratio

It is a basic measure used to estimate the distortion amount between the reference and test images (Eq.1). A higher PSNR value indicates lower distortion [8].

$$PSNR = (I_{ref}, I_{tst}) = 10 * \log_{10} \frac{MAX_I^2}{MSE} \quad (Eq.1)$$

Where MAX_I represents the highest possible pixel value of the input images, MSE is the Mean Squared Error between the tested images (Eq. 2).

$$MSE = \frac{1}{MN} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} (I_{ref}(i, j) - I_{tst}(i, j))^2 \quad (Eq.2)$$

2.1.2. Structural Similarity Index

SSIM measures the degradation in the structural information between two images. This metric compares the similarity of three factors: luminance, contrast, and structure (Eq.3). It takes a value between -1 and 1 where the value of 1 indicates that the two tested images are equal [8].

$$SSIM(I_{ref}, I_{tst}) = \frac{(2\mu_{ref}\mu_{tst}+c1)(2cov+c2)}{(\mu_{ref}^2+\mu_{ref}^2+c1)(\sigma_{ref}^2+\sigma_{tst}^2+c2)} \quad (Eq.3)$$

$$\begin{cases} c1 = (k_1L)^2 & k_1 = 0.01 \\ c2 = (k_2L)^2 & k_2 = 0.03 \end{cases}$$

Where $\mu_{I_{ref}}$ and $\mu_{I_{tst}}$ are the average of I_{ref} and I_{tst} , respectively, $\sigma_{I_{ref}}^2$ and $\sigma_{I_{tst}}^2$ are the variances of I_{ref} and I_{tst} , respectively. Cov is the covariance of I_{tst} , $c1$ and $c2$ are variables to stabilize the division with weak denominator, and L is the dynamic range of pixel values ($L=2^{\wedge}(\text{number of bits per pixels}) - 1$).

2.2. Visual Assessment

Visual testing methods represent the most clinically related approach for evaluating the quality of images since human observers are the definitive users in most multimedia applications. In this measurement, a group of experts are required to give their subjective response regarding the quality of each image [31]. When adopting this approach, the average of the observers scores of different observers are calculated to analyze the assessment results [33]. Two main visual techniques are employed to evaluate the quality of the images and the observer's performance.

2.2.1. Receiver Operating Characteristic

The main task of an observer in medical imaging is to identify whether a displayed patient's image presents a proof of pathology, or not. Therefore, a system to measure the observers' performance about the diagnosis quality is necessary [29]. Receiver Operating Characteristic (ROC) approach is often employed in radiology to evaluate the observers' performance against known diagnostic images. This method, constructed from the Signal Detection Theory (SDT), assesses whether an observer can identify a low contrast signal (artefact) in a noisy environment (digital image). The clinical equivalent to this is the distinguishing of the irregular case, from a series of regular cases [33]. Accordingly, an observer is required to identify features within the image and the performance of the observer's group can then be measured by counting the number of right responses [29].

ROC analysis has a severe limitation in that it is strongly reliant on the ubiquity of the disease. Moreover, the images must be classified into two categories (normal and abnormal), indicating that a significant number of images with subtle pathology are needed. The ROC approach does not serve adequately for many lesions within the same image, and the localization of lesions is not considered, therefore an image may be diagnosed as abnormal for the incorrect reason [29, 34]. To overcome these weaknesses in the ROC methodology, several measures have been developed to enhance its efficiency. These measures involved the development of ROC related approaches to improve its statistical strength while utilizing a low number of images [33].

2.2.2. Visual Grading Analysis

Visual grading of the visibility and reproduction of the anatomical structures is a popular, simple and valid scheme to visually assessing the quality of the clinical images [35]. Its implementation is based on the visualization of the anatomical structures by asking a viewer to estimate the clarity of some details in the medical images. This method, based on the human decision, offers a clinically favored method for evaluating the image quality [36]. The significance of the Visual Grading Analysis (VGA) in the detection of diseases has been studied and confirmed as defining a robust relationship between the anatomical clarity of normal anatomy and the ability to detect the pathological structures [33, 37]. The reasons for using visual grading as a preferred technique are reported as [29]:

- The validity of VGA investigations can be considered high when the anatomical structures are chosen based on their clinical relevance and the observers are experts in radiography.
- In special cases, VGA methods have been proved to coincide with both detection investigations using human observers [38, 39] and utilising physical assessment for image quality [40, 41].

- In comparison to ROC methods, VGA studies are comparatively easy to implement, especially when optimising equipment at the local level. This is because, with VGA method, a lesser number of images are needed, and fewer evaluators may be sufficient than that of ROC approach.
- The time required to perform VGA assessment is comparatively short, at least for the observers, which means that it can be conducted in any dispensary or hospital.
- Special preparations are needed to conduct ROC analysis, for example, half of the images should contain pathologies and particular software is needed to conduct the test; these issues are not required for VGA investigations.

Two common ways can be employed to conduct VGA trial to assess the image quality [34]:

Absolute VGA

In this method, each image is viewed individually and the observer is asked to give his/her opinion about the visibility of the anatomical structures in the image. The absolute VGA score ($VGAS_{abs}$) can be calculated from the collected ratings (Eq. 4) [37].

$$VGAS_{abs} = \frac{\sum_{i=1}^I \sum_{c=1}^C \sum_{o=1}^O G_{(abs)i,c,o}}{I \times C \times O} \quad (\text{Eq. 4})$$

Where $G_{(abs)i,c,o}$ is the absolute rating for a given image (i), criterion (c), and observer (o). I, C and O represent the total number of images, criteria, and observers, respectively.

Relative VGA

In relative VGA, the observer compares and rates the visibility of anatomical structures of a test image against the same structures of a reference image. A range of scores is used to define the observers' judgment. The relative VGA score ($VGAS_{rel}$) can be computed from the collected ratings (Eq.5) [35]. It is recommended that when implementing this method, the reference image should always be displayed side by side on a screen similar to the screen used to display the test image to guarantee that these images are presented with the identical monitor brightness and contrast [33-35].

$$VGAS_{rel} = \frac{\sum_{i=1}^I \sum_{c=1}^C \sum_{o=1}^O G_{(rel)i,c,o}}{I \times C \times O} \quad (\text{Eq. 5})$$

Where $G_{(rel)i,c,o}$ is the relative grading for a given image (i), criterion (c) and observer (o). I, C and O indicate the total number of images, criteria, and observers, respectively.

In this research, it became apparent that utilizing the visual approaches to evaluate digital image quality would make the outcomes more appropriate to clinical environments since these measures concentrate on how obviously an observer can visualize the anatomical structure of a given image. Two key shortcomings are identified; VGA reveals the observer's view and hence can be sensitive to inter-observer variability [37], and the anatomical details, required to be assessed, must be determined previously. No official and validated guidelines on this are available and there is a difference of opinion in the published literature; hence, performing comparisons is difficult [42, 43].

3. Related Work

Investigation of the transparency of the watermarked medical images is a critical issue prior to release them in the clinical workflow. Therefore, several subjective and clinical evaluations have been conducted to inspect the imperceptibility of watermarked images from a quality perspective, and also in terms of the applicability of using them in medical practices.

Dowling, et al. [27] implemented subjective testing to determine the visual threshold of perceptibility in which the observers cannot detect any notable differences between the original images and images that have been distorted. A set of 15 medical images were each degraded with four level of Gaussian noise to produce a set of 60 image pairs. A group of 20 volunteers from the CSIRO ICT e-Health Centre and National ICT Australia were asked to identify the original image from each pair of displayed images. They were also directed to choose images randomly if they cannot reveal any difference between the two images. The opportunity of discovering differences between the original and distorted images varied extremely; even within the same noise level. Therefore greater sample size and further calibration have been suggested to be used in the future research. Maeder, et al. [44] conducted a subjective evaluation based on Two Alternative Force Choice (2AFC) technique. A set of 32 of mammograms was watermarked by three different embedding strengths utilizing two watermarking techniques based on Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT). A group of 12 participants, who were all medical imaging researchers and were familiar with the artefacts that might appear in the distorted images, were recruited to evaluate the images. The watermarked images were presented in random order, beside the original images and the observers were asked to point out which image is watermarked. Further refinement by increasing the image sample size and the number of observers has been suggested due to the significant variability in the observers' scores.

Zain, et al. [45] conducted a clinical trial to assess the impact of digital watermarking on medical diagnoses. A set of 75 images were watermarked using a dual layer technique. A group of 3 consultant radiologists were asked to perform a clinical investigation on a random collection of original and watermarked images, which was then compared to the ground truth diagnosis. Giakoumaki, et al. [46] performed a visual evaluation of 120 medical images from 6 different modalities modified by encoding 4 different sizes of the watermark. The original and watermarked images were presented as a pair at the same time to 2 radiologists with no declaration of the original one. In each step, the radiologists were required to announce any differences between the images which may lead to a wrong diagnosis.

In Das and Kundu [47] approach, an expert clinician was required to assess the performance of the proposed watermarking method. A set of 430 medical images of different modalities, sizes and file formats were utilized to test the proposed technique. During the subjective evaluation, the expert clinician was requested to classify the viewed images into different categories: original, watermarked and watermark extracted. Zear, et al. [48] proposed an approach for subjectively evaluating the quality of watermarked medical images obtained by varying the gain factor and encoding different magnitudes of watermarks into various image modalities. A group of 6 persons were asked to evaluate the acceptability of the visual quality of the watermarked images for diagnosis at different gain factors.

Although these studies highlighted the ability to recognize the watermarked images and evaluate the acceptability of using them for diagnosis, however, they did not take into consideration the anatomical structures of the organs during the evaluation. In many cases, the embedded watermark may not affect the diagnosis, although it is visible to human eyes. This is a significant issue in watermarking techniques where the transparency of the hidden data is an essential requirement. Therefore, we conducted a clinical trial to assess the visualization of the anatomical details of brain MR images distorted by various payload to define the perceptual boundary of detecting the modifications.

4. Study Design

The literature reviewed demonstrated that the wide majority of published studies used physical/mathematical metrics to reach their proposed objectives. In this research, both approaches were adopted, but a special attention is given to the visual method since it is more suitable for image assessment within the clinical environment [49]. However, physical metrics (e.g. PSNR and SSIM) were utilized to support the visual assessment and validate the evaluation scale. These tests help to determine the amount of information that can be inserted into the images as a watermark and specify the acceptable level of distortion. *Fig 1* summarizes the whole process adopted for evaluating the watermarked images quality.

We conducted a visual assessment trial based on relative VGA method to evaluate the images. This approach was selected because it is very sensitive to the slight changes between the images and also it can aid to decrease bias in decision-making [50]. In the relative VGA implementation, all images (watermarked) are compared to a reference image (the un-watermarked image). The reference and modified images are shown to the observer together at the same time on two separated and identical screens. A particular criteria items were utilized to visually rate the images and then determine the differences between the images. A Likert scale (scored from 1 to 5) was used to rate the observers' scores; where a score of 1 indicates "strongly disagree", 2 "disagree", 3 "neither agree nor disagree", 4 "agree", and 5 "strongly agree". A five-point Likert scale was adopted because it offers a more reliable measure of the observer's attitude [51]. A bespoke, Java-based application was utilized to show the criteria items and the images in a random order on twin monitors [52]. This software displays the original image on the same screen throughout the assessment process.

4.1. Data Collection

This research uses a dataset provided by the MRI unit of Al Kadhimiya Teaching Hospital (Iraq), from patients' records for use in this research conducted at the University of Salford (UK) [53, 54]. The medical images dataset contains 165 brain MRI scans, in DICOM format, taken during the regular diagnostic process. These images have been independently diagnosed and categorized clinically into normal and abnormal pathologies by clinicians of this unit.

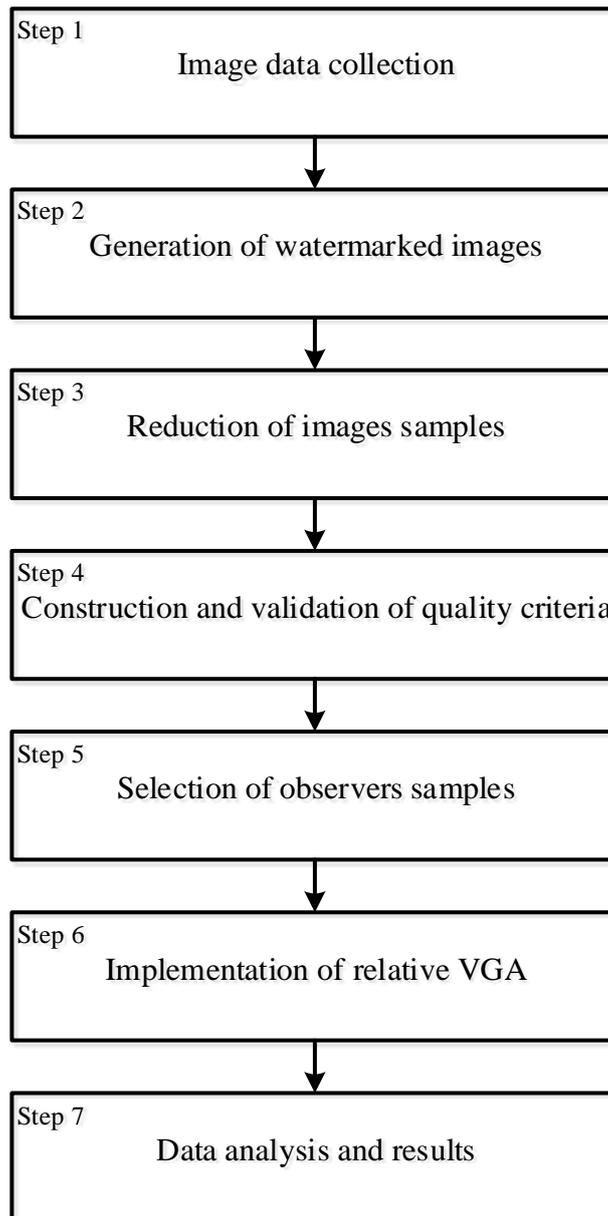


Fig 1. Key steps adopted for visually assessing the imperceptibility of the watermarked images

4.2. Generation of Watermarked Images Samples

To produce a set of watermarked images, three reversible watermarking based on Difference Expansion (DE) technique have been applied. These approaches were chosen due to they offer high capacity and low computational complexity compared to the other methods and were, therefore, suitable as potential techniques for the wider research project [26].

1. Tian [21] (embeds 1-bit per 2-pixels) method, adapted to operate within a 16 bpp (signed) color space.
2. Alattar [20] (embeds 3-bits per quad-pixels) method, adapted to operate within a 16 bpp (signed) color space.
3. Extended (within this research) Tian [21] method (by embedding 2-bits per 2-pixels) and adapted to operate within a 16 bpp (signed) color space.

The objective of these algorithms is to controllably hide information within a defined subset of the image pixels to generate a set of images with various distortion levels, defined by the quality of information encoded and the number of pixels modified. Each image was then assessed against the original, with specific assessment criteria relating to the clarity of features within the images to determine the level of modification at which the perceptual difference became noticeable. These algorithms allow to exactly recovering the complete original image after extracting the watermark successfully, thereby additionally meeting the requirement for a fully reversible process. All the encoding techniques have been applied to eight different brain MR images in 16bpp DICOM format using MATLAB (*Fig 2*). The size of all images is 512×512 pixels. These images were chosen on the basis of the following:

- They contain all the characteristics of the anatomical structure of the brain.
- They have been independently diagnosed and categorized into normal and abnormal pathologies by the clinicians.
- They contain different sizes of a tumors/lesions.
- They have different sizes of Region of Interest (ROI) and Region of Non Interest (RONI). ROI region comprises the informative part of the image which is utilized for diagnostic. However, RONI includes the non-critical part of the image (e.g. background). Occasionally this region may contain grey level parts of slight interest [8].

The embedding process was performed in ten incremental steps. In each step, an additional ten percent of the image matrix has been used to embed the watermark bits, with the entire matrix modified in the final step. After modification, standard PSNR (*Figs. 3-5*) and SSIM metrics (*Tables 1-3*) have been utilized to measure the distortion level and the structural similarity between the original images and their corresponding watermarked versions. Higher PSNR value indicates lower distortion, while SSIM value of 1 denotes that both images are structurally similar. SSIM values for all the executed techniques are either 1 or very close to 1 which denotes that the change in structural information between the original and watermarked images is unworthy.

In some of these figures (*Figs. 3-5*), there is a slight discontinuity in the PSNR reduction in the 40-60% region of the image pixel modification. PSNR values depend on which part of the image has been selected to hide the watermark, and this region marks a threshold region in the proportion of pixels within the image ROI and RONI. This difference does not impact on the aim of these algorithms, which is to determine the acceptable distortion level in each algorithm PSNR values decrease by increasing the amount of pixels modification (capacity). PSNR also depends on the nature of the image when using the same algorithm and encoding the same watermark.

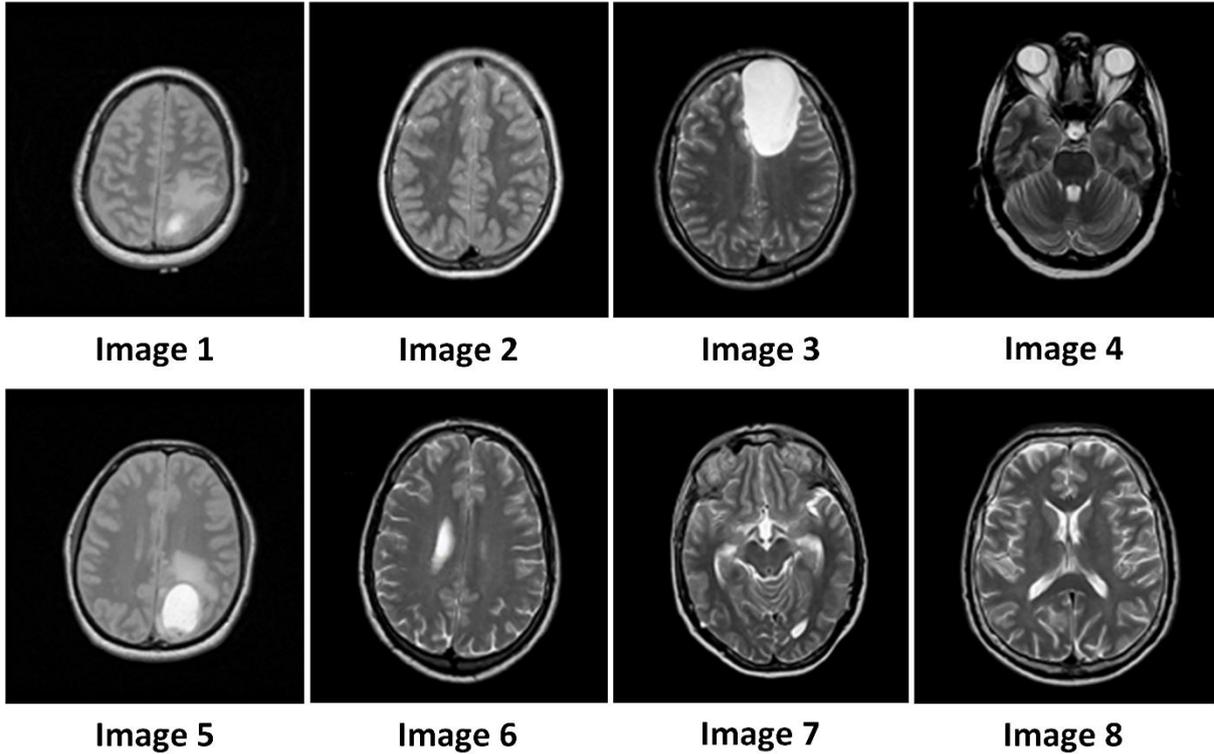


Fig 2. The eight Brain MR images in DICOM format (16bpp, 512x512 pixels) used in the implemented reversible watermarking techniques

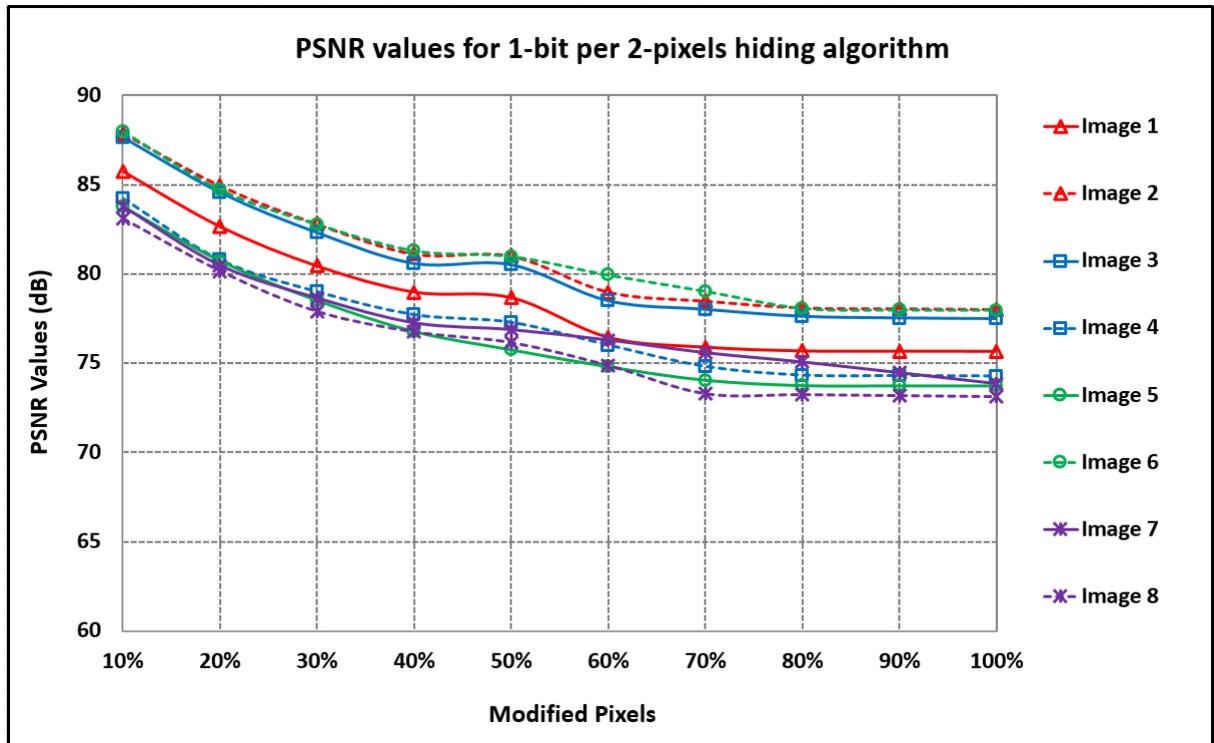


Fig 3. Distortion level (PSNR) between the original eight images and their corresponding watermarked versions by implementing technique 1 (1-bit per 2-pixels) to hide the watermark in ten steps (10-100%)

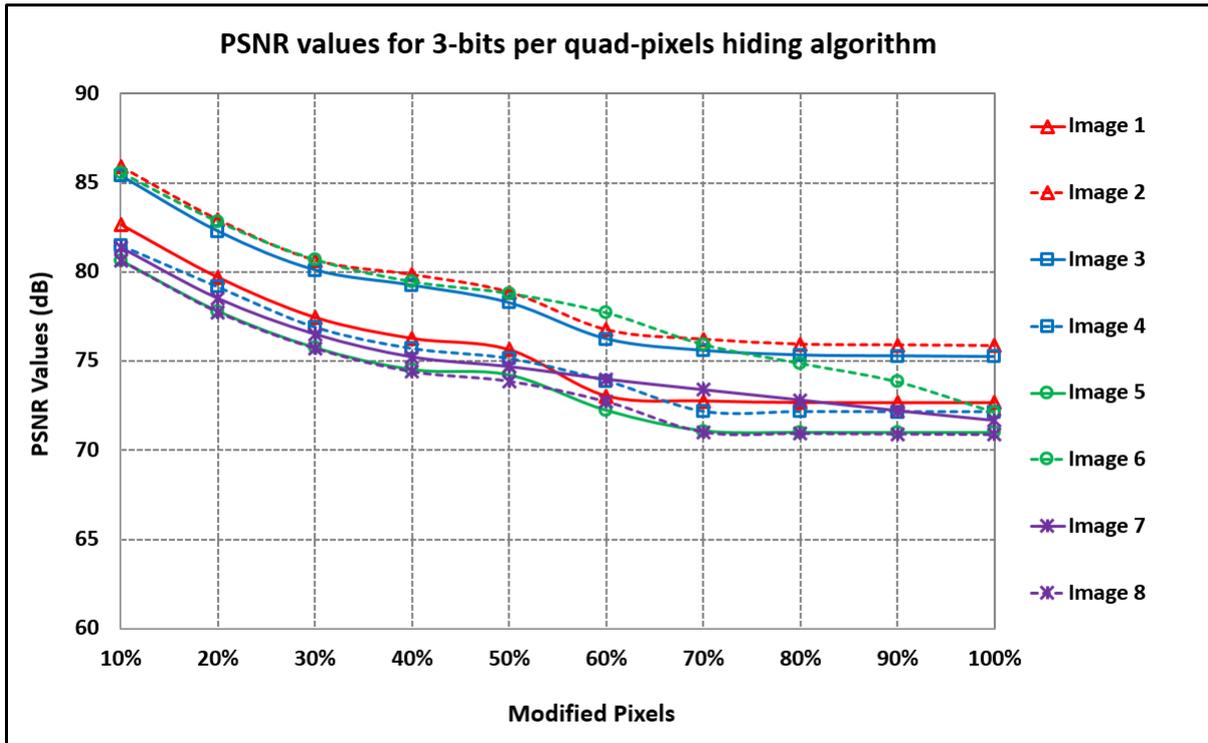


Fig 4. Distortion level (PSNR) between the original eight images and their corresponding watermarked versions by implementing technique 2 (3-bits per quad-pixels) to hide the watermark in ten steps (10-100%)

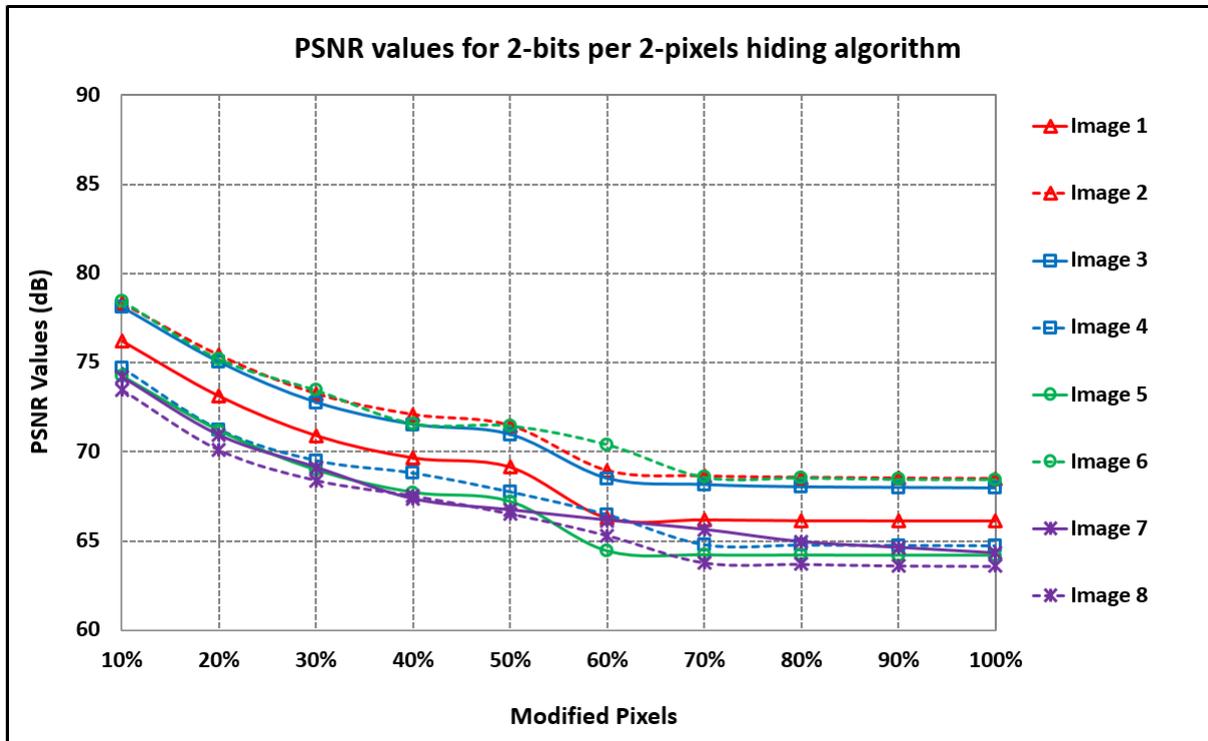


Fig 5. Distortion level (PSNR) between the original eight images and their corresponding watermarked versions by implementing technique 3 (2-bits per 2-pixels) to hide the watermark in ten steps (10-100%)

Table 1. Distortion level (SSIM) between the original eight images and their corresponding watermarked versions by implementing technique 1 (1-bit per 2-pixels) to hide the watermark in ten steps (10-100%)

Modified pixels	Image 1	Image 2	Image 3	Image 4	Image 5	Image 6	Image 7	Image 8
10%	1	1	1	1	1	1	1	1
20%	1	1	1	1	1	1	1	1
30%	1	1	1	1	1	1	1	1
40%	1	1	1	1	1	1	1	1
50%	1	1	1	1	1	1	1	1
60%	1	1	1	1	1	1	1	1
70%	1	1	1	1	1	1	1	0.9999
80%	1	1	1	1	1	1	1	0.9999
90%	1	1	1	1	1	1	1	0.9999
100%	1	1	1	1	1	1	1	0.9999

Table 2. Distortion level (SSIM) between the original eight images and their corresponding watermarked versions by implementing technique 2 (3-bit per quad-pixels) to hide the watermark in ten steps (10-100%)

Modified pixels	Image 1	Image 2	Image 3	Image 4	Image 5	Image 6	Image 7	Image 8
10%	1	1	1	1	1	1	1	1
20%	1	1	1	1	1	1	1	1
30%	1	1	1	1	1	1	1	1
40%	1	1	1	1	1	1	1	1
50%	1	1	1	1	1	1	1	1
60%	0.9999	1	1	1	0.9999	1	1	0.9999
70%	0.9999	1	1	0.9999	0.9999	1	0.9999	0.9999
80%	0.9999	1	1	0.9999	0.9999	1	0.9999	0.9999
90%	0.9999	1	1	0.9999	0.9999	1	0.9999	0.9999
100%	0.9999	1	1	0.9999	0.9999	0.9999	0.9999	0.9999

Table 3. Distortion level (SSIM) between the original eight images and their corresponding watermarked versions by implementing technique 3 (2-bit per 2-pixels) to hide the watermark in ten steps (10-100%)

Modified pixels	Image 1	Image 2	Image 3	Image 4	Image 5	Image 6	Image 7	Image 8
10%	1	1	1	1	1	1	1	1
20%	0.9999	1	1	0.9999	0.9999	1	0.9999	0.9999
30%	0.9999	0.9999	0.9999	0.9999	0.9999	1	0.9999	0.9998
40%	0.9999	0.9999	0.9999	0.9999	0.9998	0.9999	0.9998	0.9998
50%	0.9999	0.9999	0.9999	0.9998	0.9998	0.9999	0.9998	0.9998
60%	0.9997	0.9999	0.9998	0.9998	0.9996	0.9999	0.9996	0.9997
70%	0.9997	0.9999	0.9998	0.9996	0.9996	0.9998	0.9996	0.9995
80%	0.9997	0.9998	0.9998	0.9996	0.9996	0.9998	0.9996	0.9995
90%	0.9997	0.9998	0.9998	0.9996	0.9996	0.9998	0.9996	0.9995
100%	0.9997	0.9998	0.9998	0.9996	0.9996	0.9998	0.9996	0.9995

4.3. Reduction of Images Samples

The total number of the generated image set is 248 images (8 original and 240 modified images) where each original image has been modified ten times using each of the three algorithms. This presents a significant challenge for the observers, in terms of time and effect, which may also impact on the outcome of the evaluation as tiredness and constancy could become an issue. The images set size was, therefore, reduced to create a subset that covers both extreme cases and presented a wider range of images spanning the anticipated perceptual boundary as defined by the evaluated PSNR values for the image set. In the reduction process, the images were categorized into three groups according to their distortion level in terms of PSNR values:

- Group 1 contains the images that have $PSNR \geq 80dB$.
- Group 2 contains the images that have $70dB \leq PSNR < 80dB$.
- Group 3 contains the images that have $PSNR < 70dB$.

For each group, a different number of images was selected by excluding the images that have convergent PSNR values taking into account the inclusion of all ranges of PSNR (*Table 4*). The new sample size after applying the reduction steps includes 117 images (8 original and 109 modified images).

Table 4. Selected images after applying the reduction strategy

Image set	Total number of modified images			Selected images		
	Group 1 PSNR \geq 80dB	Group 2 PSNR[70-80)dB	Group 3 PSNR<70dB	Group 1 PSNR \geq 80dB	Group 2 PSNR[70-80)dB	Group 3 PSNR<70dB
Image 1	4	19	7	4	7	3
Image 2	8	17	5	4	7	2
Image 3	8	17	5	4	8	2
Image 4	3	19	8	3	7	4
Image 5	3	19	8	3	7	4
Image 6	8	18	4	4	6	2
Image 7	3	19	8	3	7	4
Image 8	3	19	8	3	6	5
Total	240			109		

4.4. Construction and Validation of Quality Criteria Items

Content validity indicates the adequacy of the selected criteria items to cover the subject and then to achieve the purposes of the investigation; items that are not relevant to the concept being evaluated could drive a wrong in the analysis, and therefore wrong conclusions may be drawn [55]. Two major recommendations have been suggested to ensure the investigation validity; utilizing large numbers of items and employing items created from previous studies [49]. Unfortunately, no standard criteria for MR images can be found in the literature and adopted for this investigation. Therefore, the criteria items used in this research, which have been identified as fundamental to evaluate the quality of brain scans, were taken from various sources dealing with CT images. These criteria have been selected to fit the anatomical structure details of brain MR images. European guidelines on quality

criteria for CT images [56] have been recognized as one of the essential sources for medical images. These guidelines concentrate on the visibility of anatomical structures within the clinical image and how this helps in getting a correct diagnosis. Moreover, the level of clarity of anatomical structures was classified into three main definitions [35]:

- *Visualization*, which means that the distinctive characteristics are discoverable but details are not entirely reproduced; *only features are clear*;
- *Reproduction*, which indicates that the details of anatomical structures are noticeable but not indeed obviously identified; *detail is appearing*;
- *Visually sharp reproduction*, which refers to the clear representation of the anatomical structure details; *details are clear*.

Additional criteria were drawn from a published study that has utilized brain image as an area for study was the second source for generating the quality measures [57]. In addition, several items have been created to examine some cases that may appear as a result of image processing operations (e.g. encoding the watermark data). Within this research, eight items have been constructed to assess the image quality and measure the distortion level between the experimental images (*Table 5*), where items 1 to 7 refer to the reproduction of the structure, and item 8 estimates the overall image quality. These scale items were revised by an expert (professor of radiography) alongside researchers to ensure their validity and applicability.

Table 5. Image quality criteria adopted within this research

Criterion no.	Description
1	There is a visually sharp reproduction of the border between white and grey matter
2	There is a visually sharp reproduction of the mesencephalon (midbrain)
3	There is a visually sharp reproduction of the cerebrospinal fluid space over the brain
4	The superior sagittal sinus is clearly distinguishable
5	The presence or absence of the tumor is clearly identifiable
6	There are no noticeable regular/periodic intensity patterns in the image
7	There are no noticeable irregular/non-periodic intensity artefacts in the image
8	The image quality is adequate for diagnosis

4.5. Selection of Observers

The number of participants is a significant issue in scale validation as it is directly related to the number of random errors that may appear. Reliability scale and factor analysis utilized for content validation need a small number of participants [49]. According to the European guidelines on quality criteria, at least two observers should examine the assent of each image with the quality criteria individually [56, 58]. Rubin [59] stated that five (or even three in some cases) of such observers are sufficient in many situations. Some recommended a rule of five members per item [60]. Consequently, five qualified radiographers from the University of Salford were invited to assess the images. This is considered to be adequate due to this investigation being concerned with the differences in the anatomical structures of these images and their quality, not for diagnostic purposes.

All observers (three males and two females) are experienced in radiographers and their age range from 30 to 40 years. Two observers have PhD in diagnostic radiography while the other three have a Master's degree in diagnostic radiography. At the time of the assessment, three of the observers had more than eight years' experience as radiographers while the other two had three years. To confirm that all the observers have a normal visual function, they were asked whether their eyesight was a typical vision (20/20), the date of their latest eyesight test and if their eyesight was corrected with glasses or contact lenses. All observers had checked their eyesight within the last 12 months, and they had a typical vision (20/20), two of the observers used glasses, and the rest (three) did not require any eyesight correction. The participated radiographers held qualifications in image reading and reported that they have substantial experience of visually assessing medical images quality for research purposes.

4.6. Implementation of Visual Assessment

Under the visual assessment approach, expert medical image readers (radiographers) were asked to visually compare the images and evaluate the differences through an objective questions set (criteria). This seeks to determine the human perceptual boundary and identify where that coincides with the context of the PSNR. The relative VGA trial was conducted with five qualified radiographers on an image set comprising 117 (8 original and 109 modified) images. Observers were required to evaluate each original image against its modified variants by giving their opinion about eight criteria items for each image. This trial was conducted in a room with PCs and computer screens devoted to medical image analysis at the University of Salford. A five-point Likert scale was utilized to rank criteria items, ranging from 1= strongly disagree to 5= strongly agree, producing in a digital form for individual scores. Before starting any evaluation process, it was considered necessary to fulfil the following steps:

- All the criteria were explained to the observers.
- Two 23.2 inches Liquid Crystal Display (LCD) flat monitors were utilized in this trial to view the images. Both screens were calibrated to DICOM Grayscale Standard Display Function (GSDF) to imitate the clinical requirements and optimise the displaying mode that is recommended for obtaining reliable detection and analysis [61].
- The surrounding light was kept dimmed at 20-38 Lux throughout the evaluation operation.
- No time restrictions were imposed on the observers during images assessing.
- No restrictions were imposed on the distance between the observers and monitors.
- Observers were blinded to image acquiring factors and watermarking techniques.
- No image manipulation was allowed.
- During the evaluation process, the images were randomised to minimise observers' bias.

After reading the information sheet, the participants in this investigation were asked to sign a consent form. The whole experiment for each observer took approximately three hours to complete the assessment. Four thousand, six hundred and eighty (4680) scores were gathered from the participants, involving their ratings on the eight criteria items for all experimental images.

5. Experimental Results and Discussion

5.1. Approach Reliability

After the data have been collected, it is now essential to test the internal reliability of each experimental image to identify the scores that are inconsistent with the measurement. These items can then be excluded to improve assessment validity and reliability [62]. Cronbach's alpha is the most common statistic method utilized to measure the internal consistency. A lenient cut off point for the Alpha coefficient is 0.6 [63]. However, an acceptable reliability value has been recommended to be 0.7 and greater [64]. Calculating the internal reliability of each experimental image is superfluous due to many images have approximately the same distortion level and scores. Therefore, the Alpha coefficient values for the images located within the same range of PSNR have been measured (*Table 6*). The relative VGA approach compares the original images with each other. This is necessary to provide a clear impression of the validity of the assessment process, especially on images that are slightly distorted. Therefore, the PSNR values of the original images is infinity (Inf) which denotes that there is no numerical difference between these images. All Cronbach's alpha values for the observers' scores are above 0.7. This ensures the reliability of the conducted trial.

Table 6. Cronbach's alpha values for the observers' scores on all experimental images.

Images PSNR	Alpha coefficient
Inf	0.928
[86-88)	0.900
[84-86)	0.934
[82-84)	0.903
[80-82)	0.906
[78-80)	0.798
[76-78)	0.776
[74-76)	0.815
[72-74)	0.799
[70-72)	0.776
[68-70)	0.838
[66-68)	0.732
[64-66)	0.794
[63-64)	0.874

5.2. Data Analysis and Results

The outcomes of both assessment approaches (visual and physical) have been connected to identify the visual degradation boundary in which the observers can identify the noticeable differences between the tested images. The observers' ratings have been only linked to the modified images PSNR values due to the SSIM values of all modified images are either one or very close to one. This seeks to determine to what level of modification the distortion is invisible to the observers. The overall observers' scores for the eight criteria items have been combined and categorized according to the PSNR values of the tested images (*Fig 6*) to define a collective assessment of the perceptual degradation boundary that applied to the generalized case for all images. In addition, the utilized five-point Likert scale was reduced to three-point by gathering the 'strongly disagree' and 'disagree' scales to one scale (disagrees) and gathering the 'strongly agree' and 'agree' scales to one scale (agrees) (*Fig 7*). This contributes to formulating the final conclusion for identifying the imperceptibility threshold in which the observers cannot recognize any differences between the original and watermarked images.

In the five-point (*Fig 6*) and three-point (*Fig 7*) plots of the Likert assessment for image quality, the range in which there is no uncertainty over the perception of no difference between the source and modified images extends down to PSNR=82dB. Uncertainty over whether a difference is noticeable starts at around PSNR=80dB (there are no reports of a perceived difference, but some observers (less than 3% of the overall scores) report they are uncertain of whether there is a difference or not). Considering the mean scores for the criteria, there is also strong evidence indicating that there is no opportunity of detecting any discernible difference for images that have PSNR \geq 82dB. This suggests, for brain MR images watermarking applications, that if a watermark is applied to the 16bpp DICOM image, a subsequent assessment of PSNR=82dB or greater would mean that there would be no reason to suspect that the watermark would be visually detectable.

By considering the results of relative VGA trial against the actual PSNR measured for the image set (*Table 7*) this would suggest that technique 3, which modifies 2-bits for every 2-pixels, will be visually detectable in every case. The others implemented techniques performed better, with technique 1 (1-bit per 2-pixels) being undetectable visually, when 10% of the pixels is modified. This equates to hiding 1.6KB of payload into the image. The size of the DICOM header data is highly variable and depends on the imaging modality, capture device and institutional practice for the composition of the data encoded [65]. Disconnection of the image from this header, or obliteration of the header renders the image useless for medical purposes, so encoding this information as the watermark is highly advantageous. While there are few studies on the typical size of the header, one does suggest that data in the range 0.5-4KB (per image) is normal, depending on the encoding scheme and Application Programming Interface (API) used [66]. Even the best case for these encoding techniques (technique 1 at 20% of pixels modified - 3.2KB of payload, technique 2 at 10% of pixels modified - 2.4KB of payload) is insufficient for the maximum full header to be used as the watermark. However, careful selection of the metadata fields and compression of the raw metadata could bring this down to an achievable descriptor of the patient data, sufficient to connect image and metadata, for the watermark payload.

Mean and Standard Deviation (SD) were computed to measure the observer evaluation of each score and to assess confidence in statistical conclusions (*Fig 8 and Fig 9*).

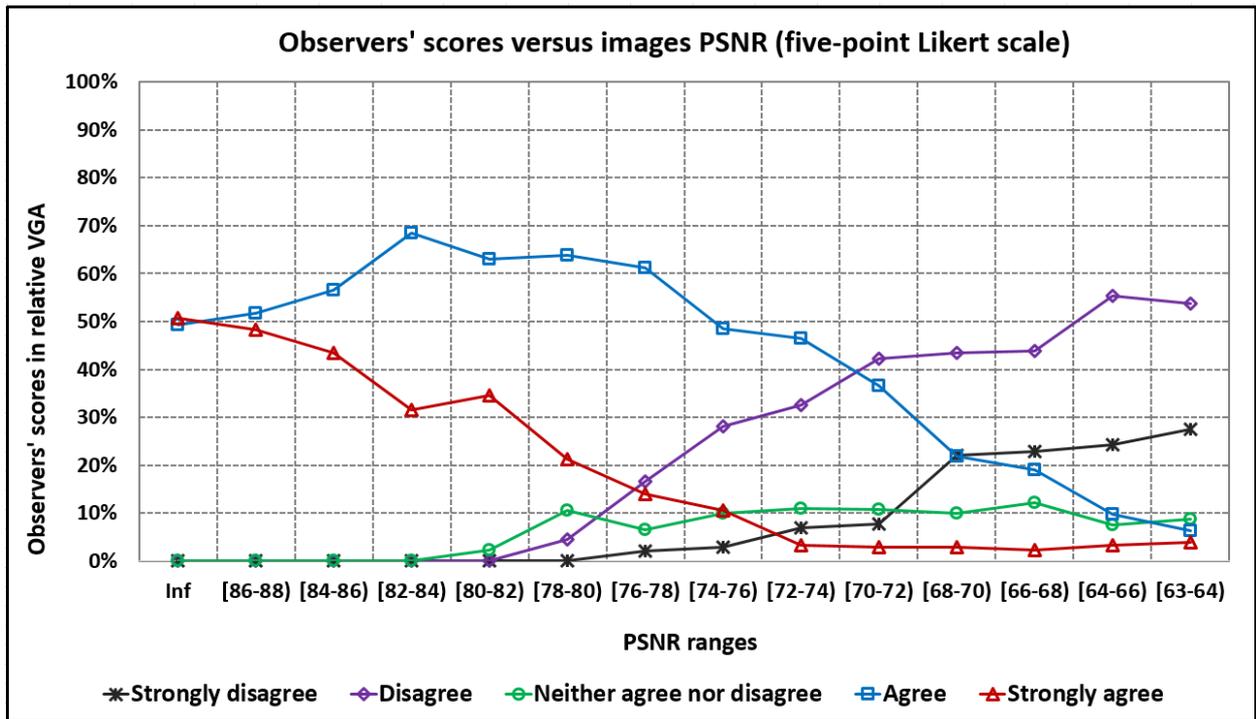


Fig 6. The overall observers' scores (*five-point Likert scale*) for the eight criteria items against PSNR values for all experimental images.

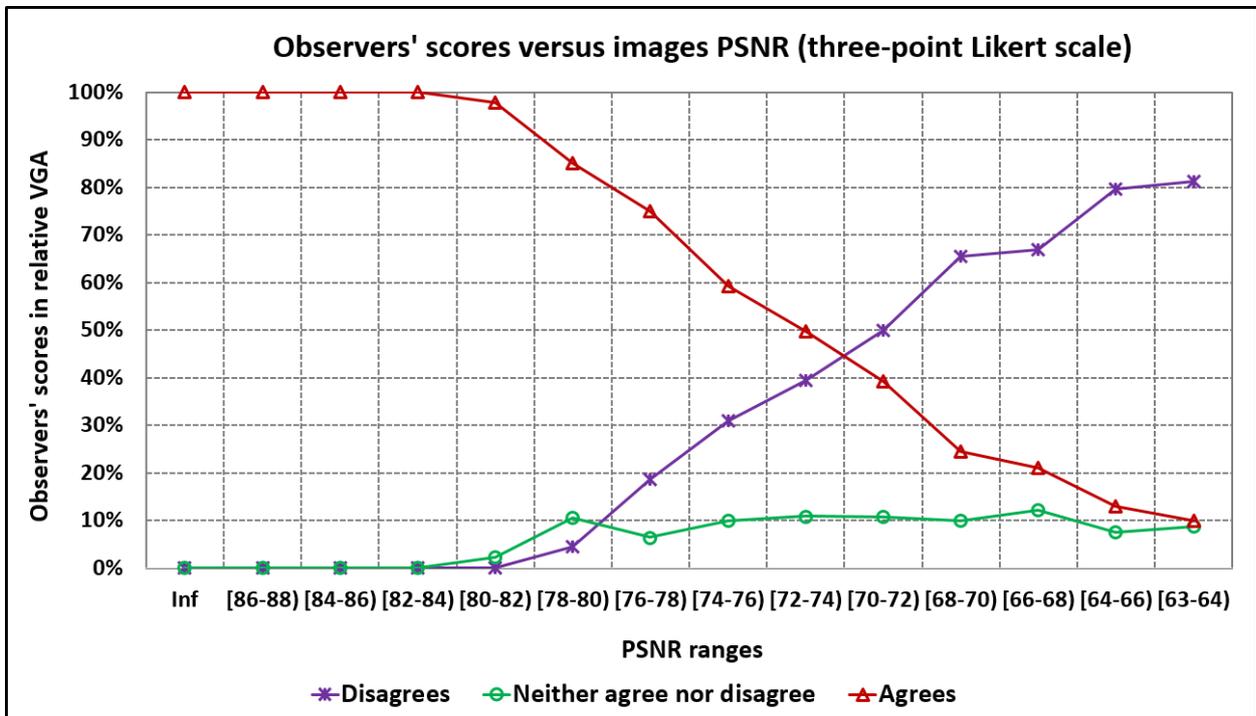


Fig 7. The overall observers' scores (*three-point Likert scale*) for the eight criteria items against PSNR values for all experimental images.

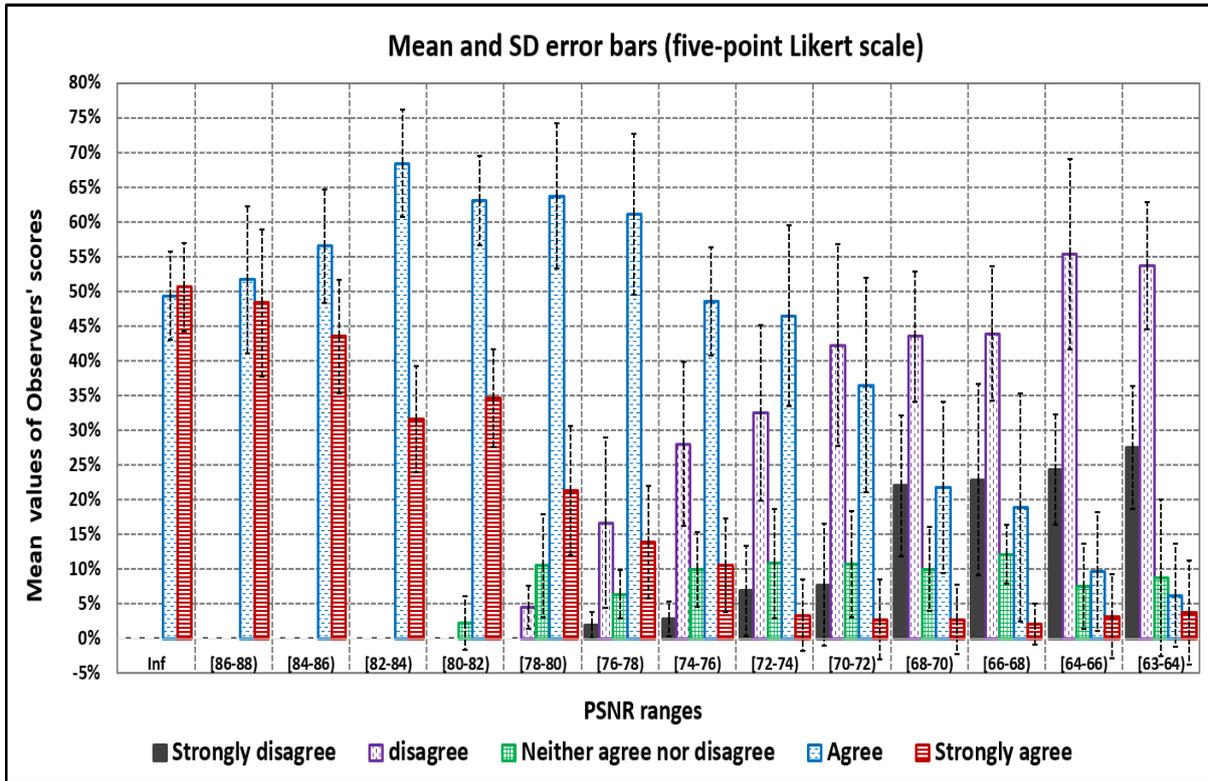


Fig 8. The mean and SD error bars for the overall observers' scores (*five-point Likert scale*) for the eight criteria items against PSNR values for all experimental images

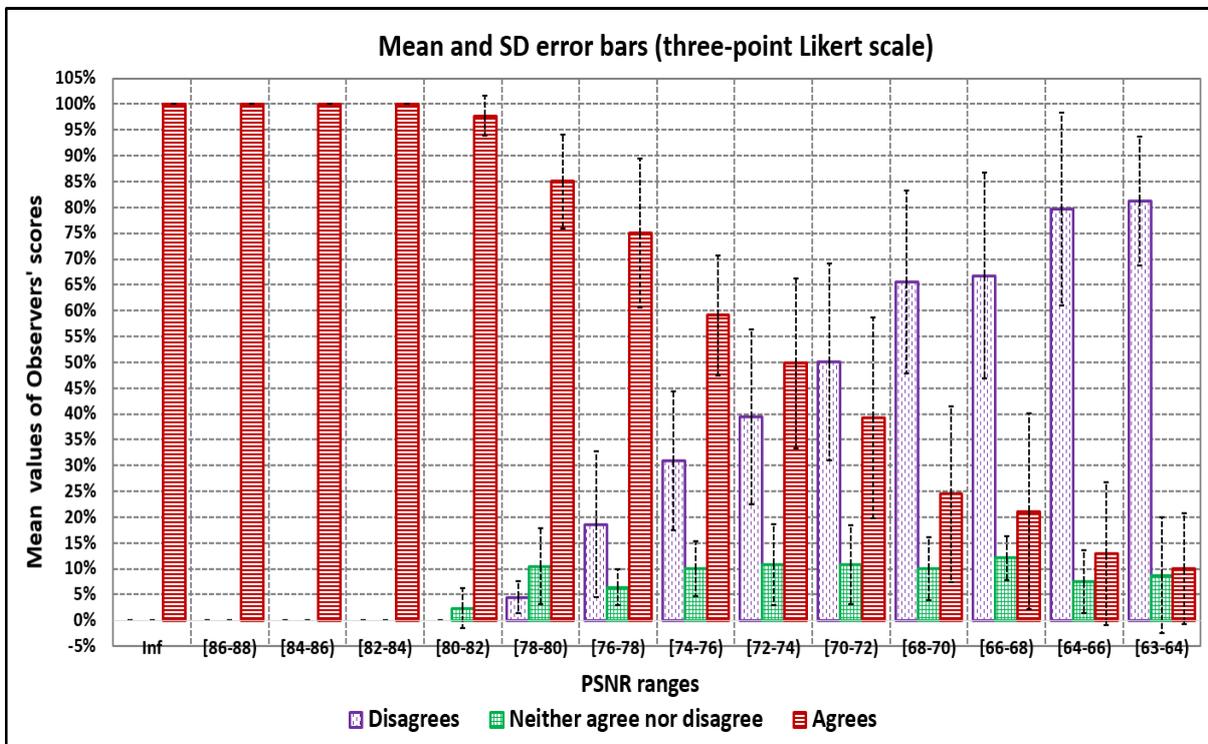


Fig 9. The mean and SD error bars for the overall observers' scores (*three-point Likert scale*) for the eight criteria items against PSNR values for all experimental images

Table 7: Aggregated (mean) PSNR values for all experimental images with the SD considered. Green cells denote the region in which no perceivable difference in the images was noticed, orange, where some uncertainty exists

Modified pixels	Technique 1 1-bit per 2-pixels		Technique 2 3-bits per 4-pixels		Technique 3 2-bits per 2-pixels	
	Mean+SD	Mean-SD	Mean+SD	Mean-SD	Mean+SD	Mean-SD
10%	87.58	83.43	85.24	80.66	78.04	73.86
20%	84.47	80.27	82.33	77.89	74.95	70.58
30%	82.36	78.22	80.13	75.80	72.88	68.70
40%	80.76	76.85	79.13	74.52	71.50	67.55
50%	80.59	76.20	78.29	74.07	71.04	66.76
60%	78.90	75.05	76.61	72.51	69.07	65.04
70%	78.28	74.00	75.65	71.37	68.22	64.26
80%	77.70	73.75	75.15	71.27	68.09	63.91
90%	77.60	73.62	74.84	71.12	68.05	64.00
100%	77.56	73.46	74.55	70.82	68.03	63.92

5.3. Comparison with Other Approaches

Although comparing the performance of the proposed approach is difficult due to the lack of investigations that used standard criteria to evaluate the visualization of the anatomical detail of brain MR images, we compared our approach to other studies stated in the literature (*Table 8*).

In Zain, et al. [45] approach, the radiologists diagnosed a random collection of original and watermarked images, which was then compared to the ground truth diagnosis. The study did not take into account the visual distortions of the anatomical details of the images that can appear without impacting the diagnosis. The aim of Giakoumaki, et al. [46] and Das and Kundu [47] approaches is to define whether there is a difference between original and watermarked images not to determine the level of visual perception of distortion. Furthermore, the number of assessors is small which may affect the evaluation outcomes and therefore leads to wrong conclusions. In Dowling, et al. [27], Maeder, et al. [44], and Zear, et al. [48] approaches, the perception of distortion boundaries have been determined through identifying the differences between the original and modified images. A significant difference in the values of imperceptibility threshold can be observed due to the large variability in the observers' scores. This happened due to the observers do not have experiences to conduct similar investigations. Moreover, the sample size of images used in these studies to determine the perception threshold is small. Therefore, increasing the images sample size and using further calibration have been suggested for future research.

Our approach has identified the threshold at which the observers can detect the slight differences between the anatomical details of the brain. Qualified radiographers have evaluated the differences in the anatomical structure between the original and manipulated images based on universal criteria. The result of imperceptibility threshold is much higher than the approaches under comparison, and no variability has been observed in the observers' scores. This is due to adopting standard criteria for evaluating the anatomical details of the brain and involving participants who have experiences in conducting related investigations.

Table 8: Performance comparison of the proposed approach against approaches identified in the literature

Approach	Year	No. of images	Images modalities	Images format	No./Experience of observers	Standard criteria?	Objective assessment	Subjective assessment
Dowling, et al. [27]	2007	60	MRI CT	DICOM	20 volunteers	No	PSNR (30-75dB)	PSNR threshold (57dB)
Maeder, et al. [44]	2008	32	Mammogram	-	12 semi-skilled researchers	No	PSNR (44.59-64.92dB)	PSNR threshold (45.5dB)
Zain, et al. [45]	2009	225	X-rays Ultrasound CT	-	3 radiologists (each evaluated 75 images)	No	Average PSNR (54.15dB)	No effect on medical diagnosis
Giakoumaki, et al. [46]	2010	120	CT MRI MRA Ultrasound Dermatological Radiological	JPEG BMP TIF	2 radiologists	No	PSNR (52.78±0.08-72.64±0.09dB)	No variation detected
Das and Kundu [47]	2013	430	CT MRI USG X-ray Mammogram	BMP TIF GIF DICOM	1 clinician	No	Average PSNR (42.16-44.8dB)	No noticeable difference found
Zear, et al. [48]	2018	6	CT	-	6 persons	No	PSNR (27.29-43.88dB)	PSNR threshold (27.29dB)
Proposed	2018	117	MRI	DICOM	5 radiographers	Yes	PSNR (63.58 -87.99dB)	PSNR threshold (82dB)

6. Conclusion

This study has conducted a relative VGA trial to determine the range of modification, for brain MR images, within which changes to the image data (pixels) are unperceivable to the observer. This seeks to define a perceptual boundary, below which change is noticeable, to determine heuristic guidelines for the method of watermarking and the level of modification that can be applied to encode a known magnitude of payload data in an imperceptible manner. Relating this to objective measures for image fidelity (PSNR) is then undertaken to define quantitative criteria to guide the selection of watermark encoding technique and enable an objective post modification assessment of the watermarked image to ensure the condition of imperceptibility is met. The outcomes propose that, when applying digital watermarking to medical images, the modification of the images to a level of PSNR=82dB or greater, between the reference and watermarked images, is undetectable to all observers, and modification level to a PSNR=80dB should not be noticeable in the vast majority of cases. This translates to a watermark payload of 1.6Kb (approx.) in the 512x512 pixel (16 bpp grayscale) images used in the study. While this is insufficient to encode a typical DICOM header collection of metadata into these images, careful selection of the metadata components and compression should enable sufficient information to be encoded to ensure the image pixel data can be re-connected to the patient record, if required, and enable the authenticity and integrity evaluation that the wider research is seeking. These images are relatively small, by modern standards, and are a specific requirement of the research, but more typical 1024x1024 images should enable a potential 4x increase in payload, which is close to the typical magnitude of a single image DICOM header. Further research will need to be undertaken to confirm this.

Providing a reliable and dependable method for digital watermarking of images within the medical imaging workflow is intended to enhance the security of data within the complex document management pipeline, thereby reducing the risk of data being compromised through intentional or unintentional changes, and enhancing trust in the medical imaging system. The definition of a reversible and unperceivable watermark, which can be evaluated by objective measures before the image is released into the clinical process, ensures that security can be achieved and, importantly, the original (raw) image data can always be recovered when required for critical activities such as diagnosis.

For future work, we recommend increasing the images samples utilizing different modalities commonly used in medical practices. Furthermore, involving expert radiologists to evaluate the images that have distortion amount close to the threshold of imperceptibility. This includes using the ROC approach to determine if the modification applied to medical images impacts diagnosis.

Acknowledgement

We would like to thank the ministry of higher education and scientific research (Iraq), for providing scholarship support to the first author of this research paper. We also want to thank all participants who attended our visual assessment trial.

Declarations of Interest

None.

References

- [1] O. S. Pianykh, *Digital imaging and communications in medicine (DICOM): a practical introduction and survival guide*: Springer Science & Business Media, 2009.
- [2] M. Fontani, A. De Rosa, R. Caldelli, F. Filippini, A. Piva, M. Consalvo, *et al.*, "Reversible watermarking for image integrity verification in hierarchical pacs," in *Proceedings of the 12th ACM workshop on Multimedia and security*, 2010, pp. 161-168.
- [3] O. M. Al-Qershi and B. E. Khoo, "Authentication and data hiding using a hybrid ROI-based watermarking scheme for DICOM images," *Journal of digital imaging*, vol. 24, pp. 114-125, 2011.
- [4] G. Coatrieux, L. Lecornu, B. Sankur, and C. Roux, "A review of image watermarking applications in healthcare," in *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*, 2006, pp. 4691-4694.
- [5] H. Nyeem, W. Boles, and C. Boyd, "A review of medical image watermarking requirements for teleradiology," *Journal of digital imaging*, vol. 26, pp. 326-343, 2013.
- [6] S. M. Mousavi, A. Naghsh, and S. Abu-Bakar, "Watermarking techniques used in medical images: a survey," *Journal of digital imaging*, vol. 27, pp. 714-729, 2014.
- [7] A. H. Ali, L. E. George, A. Zaidan, and M. R. Mokhtar, "High capacity, transparent and secure audio steganography model based on fractal coding and chaotic map in temporal domain," *Multimedia Tools and Applications*, pp. 1-30, 2018.
- [8] A. F. Qasim, F. Meziane, and R. Aspin, "Digital watermarking: Applicability for developing trust in medical imaging workflows state of the art review," *Computer Science Review*, vol. 27, pp. 45-60, 2018.
- [9] C. Qin, Z. He, H. Yao, F. Cao, and L. Gao, "Visible watermark removal scheme based on reversible data hiding and image inpainting," *Signal Processing: Image Communication*, vol. 60, pp. 160-172, 2018.
- [10] M. Arsalan, S. A. Malik, and A. Khan, "Intelligent reversible watermarking in integer wavelet domain for medical images," *Journal of Systems and Software*, vol. 85, pp. 883-894, 2012.
- [11] M. U. Celik, G. Sharma, A. M. Tekalp, and E. Saber, "Lossless generalized-LSB data embedding," *IEEE transactions on image processing*, vol. 14, pp. 253-266, 2005.
- [12] P. Tsai, Y. C. Hu, and H. L. Yeh, "Reversible image hiding scheme using predictive coding and histogram shifting," *Signal Processing*, vol. 89, pp. 1129-1143, 2009.
- [13] C. C. Lin, W. L. Tai, and C. C. Chang, "Multilevel reversible data hiding based on histogram modification of difference images," *Pattern Recognition*, vol. 41, pp. 3582-3591, 2008.
- [14] Z. Ni, Y. Q. Shi, N. Ansari, and W. Su, "Reversible data hiding," *IEEE Transactions on circuits and systems for video technology*, vol. 16, pp. 354-362, 2006.
- [15] C. De Vleeschouwer, J. F. Delaigle, and B. Macq, "Circular interpretation of bijective transformations in lossless watermarking for media asset management," *IEEE Transactions on Multimedia*, vol. 5, pp. 97-105, 2003.
- [16] L. T. Ko, J. E. Chen, Y. S. Shieh, M. Scalia, and T. Y. Sung, "A novel fractional-discrete-cosine-transform-based reversible watermarking for healthcare information management systems," *Mathematical Problems in Engineering*, vol. 2012, 2012.
- [17] L. T. Ko, J. E. Chen, Y. S. Shieh, H. C. Hsin, and T. Y. Sung, "Nested quantization index modulation for reversible watermarking and its application to healthcare information management systems," *Computational and mathematical methods in medicine*, vol. 2012, 2012.
- [18] M. J. Saberian, M. A. Akhaee, and F. Marvasti, "An invertible quantization based watermarking approach," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2008, pp. 1677-1680.
- [19] Y. M. Cheung and H. T. Wu, "A sequential quantization strategy for data embedding and integrity verification," *IEEE transactions on circuits and systems for video technology*, vol. 17, pp. 1007-1016, 2007.
- [20] A. M. Alattar, "Reversible watermark using difference expansion of quads," in *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'04)*, 2004, pp. 377-380.
- [21] J. Tian, "Reversible data embedding using a difference expansion," *IEEE transactions on circuits and systems for video technology*, vol. 13, pp. 890-896, 2003.
- [22] A. M. Alattar, "Reversible watermark using difference expansion of triplets," in *Proceedings IEEE International Conference on Image Processing. Barcelona, Spain*, 2003, pp. 501-504.
- [23] A. Roček, K. Slavíček, O. Dostál, and M. Javorník, "A new approach to fully-reversible watermarking in medical imaging with breakthrough visibility parameters," *Biomedical Signal Processing and Control*, vol. 29, pp. 44-52, 2016.
- [24] V. Kumar and V. Natarajan, "Hybrid local prediction error-based difference expansion reversible watermarking for medical images," *Computers & Electrical Engineering*, vol. 53, pp. 333-345, 2016.
- [25] B. Lei, E. L. Tan, S. Chen, D. Ni, T. Wang, and H. Lei, "Reversible watermarking scheme for medical image based on differential evolution," *Expert Syst. Appl.*, vol. 41, pp. 3178-3188, 2014.

- [26] A. Khan, A. Siddiqa, S. Munib, and S. A. Malik, "A recent survey of reversible watermarking techniques," *Information sciences*, vol. 279, pp. 251-272, 2014.
- [27] J. A. Dowling, B. M. Planitz, A. J. Maeder, J. Du, B. Pham, C. Boyd, *et al.*, "Visual quality assessment of watermarked medical images," in *Society of Photo-Optical Instrumentation Engineers (SPIE)*. 2007.
- [28] Z. Wang, A. C. Bovik, and L. Lu, "Why is image quality assessment so difficult?," in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, 2002, pp. IV-3313-IV-3316.
- [29] M. Båth, "Evaluating imaging systems: practical applications," *Radiation protection dosimetry*, vol. 139, pp. 26-36, 2010.
- [30] C. H. McCollough, M. R. Bruesewitz, and J. M. Kofler Jr, "CT dose reduction and dose management tools: overview of available options 1," *Radiographics*, vol. 26, pp. 503-512, 2006.
- [31] P. Mohammadi, A. Ebrahimi-Moghadam, and S. Shirani, "Subjective and objective quality assessment of image: A survey," *arXiv preprint arXiv:1406.7799*, 2014.
- [32] K. M. Nasr and M. G. Martini, "A visual quality evaluation method for telemedicine applications," *Signal Processing: Image Communication*, vol. 57, pp. 211-218, 2017.
- [33] L. Månsson, "Methods for the evaluation of image quality: a review," *Radiation protection dosimetry*, vol. 90(1-2), pp. 89-99, 2000.
- [34] F. Zarb, L. Rainford, and M. F. McEntee, "Image quality assessment tools for optimization of CT images," *Radiography*, vol. 16, pp. 147-153, 2010.
- [35] E. Seeram, R. Davidson, S. Bushong, and H. Swan, "Image quality assessment tools for radiation dose optimization in digital radiography: An overview," *Radiologic technology*, vol. 85, pp. 555-562, 2014.
- [36] Ö. Smedby and M. Fredrikson, "Visual grading regression: analysing data from visual grading experiments with regression models," *The British journal of radiology*, vol. 83, pp. 767-775, 2010.
- [37] P. Sund, M. Båth, S. Kheddache, and L. G. Månsson, "Comparison of visual grading analysis and determination of detective quantum efficiency for evaluating system performance in digital chest radiography," *European radiology*, vol. 14, pp. 48-58, 2004.
- [38] A. Tingberg, C. Herrmann, B. Lanhede, A. Alm, J. Besjakov, S. Mattsson, *et al.*, "Comparison of two methods for evaluation of the image quality of lumbar spine radiographs," *Radiation protection dosimetry*, vol. 90, pp. 165-168, 2000.
- [39] C. Herrmann, P. Sund, A. Tingberg, S. Keddache, L. G. Mansson, A. Almen, *et al.*, "Comparison of two methods for evaluating image quality of chest radiographs," in *Proc. SPIE 2000*, pp. 251-257.
- [40] M. Sandborg, A. Tingberg, D. Dance, B. Lanhede, A. Almén, G. McVey, *et al.*, "Demonstration of correlations between clinical and physical image quality measures in chest and lumbar spine screen-film radiography," *The British journal of radiology*, vol. 74, pp. 520-528, 2001.
- [41] M. Sandborg, G. McVey, D. Dance, and G. A. Carlsson, "Comparison of model predictions of image quality with results of clinical trials in chest and lumbar spine screen-film imaging," *Radiation protection dosimetry*, vol. 90, pp. 173-176, 2000.
- [42] N. Shet, J. Chen, and E. L. Siegel, "Continuing challenges in defining image quality," *Pediatric radiology*, vol. 41, pp. 582-587, 2011.
- [43] Y. Li, A. Poulos, D. McLean, and M. Rickard, "A review of methods of clinical image quality evaluation in mammography," *European journal of radiology*, vol. 74, pp. e122-e131, 2010.
- [44] A. Maeder, J. Dowling, A. Nguyen, E. Brunton, and P. Nguyen, "Assuring authenticity of digital mammograms by image watermarking," in *International Workshop on Digital Mammography*, 2008, pp. 204-211.
- [45] J. M. Zain, A. R. Fauzi, and A. A. Aziz, "Clinical assessment of watermarked medical images," *Journal of Computer Science*, vol. 5, pp. 857-863, 2009.
- [46] A. Giakoumaki, K. Perakis, K. Banitsas, K. Giokas, S. Tachakra, and D. Koutsouris, "Using digital watermarking to enhance security in wireless medical image transmission," *Telemedicine and e-Health*, vol. 16, pp. 306-313, 2010.
- [47] S. Das and M. K. Kundu, "Effective management of medical information through ROI-lossless fragile image watermarking technique," *Computer methods and programs in biomedicine*, vol. 111, pp. 662-675, 2013.
- [48] A. Zear, A. K. Singh, and P. Kumar, "A proposed secure multiple watermarking technique based on DWT, DCT and SVD for application in medicine," *Multimedia Tools and Applications*, vol. 77, pp. 4863-4882, 2018.
- [49] H. Mraity, A. England, and P. Hogg, "Developing and validating a psychometric scale for image quality assessment," *Radiography*, vol. 20, pp. 306-311, 2014.
- [50] D. G. Pelli and B. Farell, "Psychophysical methods," in *Handbook of Optics: Fundamentals, techniques, and design*, Second ed New York: McGraw-Hill, 1995.
- [51] R. Likert, "A technique for the measurement of attitudes," *Archives of psychology*, vol. 140, pp. 1-55, 1932.

- [52] P. Hogg and P. Blindell, "Software for image quality evaluation using a forced choice comparison method," in *E-poster in UKRC*, Manchester, UK, 2012.
- [53] A. M. Hasan and F. Meziane, "Automated screening of MRI brain scanning using grey level statistics," *Computers & Electrical Engineering*, 2016.
- [54] A. M. Hasan, F. Meziane, R. Aspin, and H. A. Jalab, "Segmentation of brain tumors in MRI images using three-dimensional active contour without edge," *Symmetry*, vol. 8, p. 132, 2016.
- [55] I. McDowell, *Measuring health: a guide to rating scales and questionnaires*: Oxford university press, 2006.
- [56] H. Menzel, H. Schibilla, and D. Teunen, "European guidelines on quality criteria for computed tomography," *Luxembourg: European Commission publication*, vol. EUR 16262 EN, 2000.
- [57] K. Ledenius, E. Svensson, F. Stålhammar, L. Wiklund, and A. Thilander-Klang, "A method to analyse observer disagreement in visual grading studies: example of assessed image quality in paediatric cerebral multidetector CT images," *The British journal of radiology*, 2014.
- [58] EC., *European guidelines on quality criteria for diagnostic radiographic images: UR 16260 EN*: Office for Official Publications of the European Communities. Brussels, Luxembourg, 1996.
- [59] D. B. Rubin, "Multiple imputation after 18+ years," *Journal of the American statistical Association*, vol. 91, pp. 473-489, 1996.
- [60] B. G. Tabachnick and L. S. Fidell, *Using multivariate statistics (5 ed.)*, 5 ed. ed. Boston: Pearson Education, 2013.
- [61] J. T. Norweck, J. A. Seibert, K. P. Andriole, D. A. Clunie, B. H. Curran, M. J. Flynn, *et al.*, "ACR–AAPM–SIIM technical standard for electronic practice of medical imaging," *Journal of digital imaging*, vol. 26, pp. 38-52, 2013.
- [62] R. Ho, *Handbook of univariate and multivariate data analysis and interpretation with SPSS*: Boac Raton: Taylor & Francis Group, 2006.
- [63] L. J. Cronbach, "Coefficient alpha and the internal structure of tests," *psychometrika*, vol. 16, pp. 297-334, 1951.
- [64] D. Streiner and G. Norman, "Health measurement scales: a practical guide to their development and use 4 edition Oxford University Press," *New York*, 2008.
- [65] D. R. Varma, "Managing DICOM images: Tips and tricks for the radiologist," *The Indian journal of radiology & imaging*, vol. 22, p. 4, 2012.
- [66] M. Ismail, Y. Ning, and J. Philbin, "Separation of metadata and bulkdata to speed DICOM tag morphing," in *SPIE Medical Imaging*, 2014, pp. 903905-903905-6.