

Article

Local Indicators of Spatial Autocorrelation (LISA): Application to Blind Noise-Based Perceptual Quality Metric Index for Magnetic Resonance Images

Michael Osadebey ^{1,*}, Marius Pedersen ^{1,†}, Douglas Arnold ^{2,†} and Katrina Wendel-Mitoraj ^{3,†}

¹ Department of Computer Science, Norwegian University of Science and Technology, Teknologivegen 22, N-2815 Gjøvik, Norway; marius.pedersen@ntnu.no

² Montreal Neurological Institute and Hospital, McGill University, 3801 University St, Montreal, QC H3A 2B4, Canada; douglas.arnold@mcgill.ca

³ BrainCare Oy, Finn-Medi 1, PL 2000, 33521 Tampere, Finland; katrina@braincare.fi

* Correspondence: michael.osadebey@ntnu.no; Tel.: +47-913-09-771

† These authors contributed equally to this work.

Received: 23 November 2018; Accepted: 2 January 2019; Published: 15 January 2019



Abstract: Noise-based quality evaluation of MRI images is highly desired in noise-dominant environments. Current noise-based MRI quality evaluation methods have drawbacks which limit their effective performance. Traditional full-reference methods such as SNR and most of the model-based techniques cannot provide perceptual quality metrics required for accurate diagnosis, treatment and monitoring of diseases. Although techniques based on the Moran coefficients are perceptual quality metrics, they are full-reference methods and will be ineffective in applications where the reference image is not available. Furthermore, the predicted quality scores are difficult to interpret because their quality indices are not standardized. In this paper, we propose a new no-reference perceptual quality evaluation method for grayscale images such as MRI images. Our approach is formulated to mimic how humans perceive an image. It transforms noise level into a standardized perceptual quality score. Global Moran statistics is combined with local indicators of spatial autocorrelation in the form of local Moran statistics. Quality score is predicted from perceptually weighted combination of clustered and random pixels. Performance evaluation, comparative performance evaluation and validation by human observers, shows that the proposed method will be a useful tool in the evaluation of retrospectively acquired MRI images and the evaluation of noise reduction algorithms.

Keywords: magnetic resonance imaging (MRI); image quality; noise; local indicators of spatial autocorrelation (LISA); local moran statistics (LMS); global moran statistics (GMS); perceptual quality; contrast; sharpness

1. Introduction

Noise is a type of distortion which is observed as the random variation of pixel intensity levels within an image. Its presence is undesirable because it can obscure useful information and degrade perceived visual quality required for the diagnosis, treatment and monitoring of diseases [1]. Sources of noise in magnetic resonance imaging (MRI) system images can be broadly classified into physiological noise and electronic noise [2,3]. Physiological noise arise from the imaged subject while electronic noise originates from system components such as the radio frequency (RF) coils.

There are several reasons to justify the need for noise-based quality evaluation methods. Image quality evaluation is a nontrivial task. Since noise is one of the several types of distortion that can degrade an image, noise-based quality evaluation method can be efficiently combined with other

distortion-specific quality evaluation methods to evaluate the quality of an image. Some operating conditions, applications and acquisition techniques encourage noise to dominate over other types of distortion. During image acquisition for the elderly or patients in trauma, noise can dominate because concern for patient comfort requires trade-off between length of scan time and image resolution. Parallel imaging (PI) acquisition technique is vulnerable to noise because its primary purpose is to reduce acquisition time through acquiring and averaging fewer data points [4]. In image processing application, the regularization parameters incorporated in most denoising algorithms, to allow a balance between noisiness and blurriness, can be easily optimized using noise-based quality evaluation methods.

Current contributions on noise-based MRI quality evaluation can be classified as traditional, model-based, Moran statistics and window-based Moran coefficient methods. The traditional methods include signal-to-noise ratio (SNR), peak signal-to-noise ratio (PSNR), mean square error (MSE) and root mean square error (RMSE). Model-based methods assess the quality of a noisy MRI image by estimating the noise variance. Model-based methods include the report in [5] which adapt the median absolute deviation (MAD), originally proposed to estimate Gaussian noise in the wavelet domain, to estimate noise in MRI images. In [6], the level of noise was estimated from the distribution of local moments. The approach in [7] estimate noise from a Rician distribution. Knowledge of the Rician distribution is obtained by double acquisition of same image. The statistics of the background signal in MRI images is exploited in [8] for the estimation of noise. The report in [9] estimates noise level from a mathematical model that describe the relationship between the clique potential of Markov random fields and noise variance. The report in [10] uses maximum likelihood principle to estimate noise in local regions of the image. Moran statistics method include the contribution by [11] which evaluates a noisy image by performing joint-count statistics and Moran test on a residual image. The residual image is formed by subtracting the original image from its Gaussian smoothed version. Window-based Moran coefficient methods slide a fixed-size window through an image. Thereafter, Moran coefficients computed from local regions are combined to measure the sharpness quality of the image. Techniques in this category include the contributions by [12–15]. The report in [16] applied machine learning technique to perceptual quality assessment of MRI images. It modifies the popular Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [17] by training regression model of MRI image features and subjective mean opinion scores.

The appropriateness of traditional full-reference methods as indicators of average error has raised concerns among researchers [18–21]. The concerns arose from multiple definitions associated with each metric, resulting in different interpretations. Multiple definitions of a quality index makes it difficult to compare results from different imaging systems, modalities and researchers [21]. Furthermore, the traditional methods are not formulated from structural and spatial information, an important requirement for perceptual image quality. Traditional methods and window-based Moran coefficient methods operate only in applications where a reference image is available. No-reference technique is the most practical and realistic approach to quality evaluation because in the real-world there is no perfect image [22]. Model-based techniques evaluate image quality by estimating noise variance. Noise variance, a physical attribute of an image, does not directly translate to perceptual image quality. Model-based quality methods such as [8] will be unreliable where there is a limited or corrupted background signal. In most current techniques, noise in the foreground is assumed to follow Gaussian and Rician distributions at high and low SNR, respectively. Since there is no clearly defined threshold which demarcates low and high SNR, this assumption can introduce modeling error which compromises accurate estimation of noise. Additional resources such as pre-correction [23] and post-correction [24] techniques are required to retrieve correct noise estimates.

Noise is one of many attributes which combine to give an image its perceptual quality. Even in noise-dominant environment, it is very difficult to estimate noise. Since noise contributes to visual perception, we are of the opinion that it is more practical to evaluate the perceptual quality of a noisy image. In this paper, we propose to combine global Moran statistics and local indicators of spatial

autocorrelation (LISA) [25] for a no-reference perceptual quality evaluation of noisy grayscale images such as MRI images. LISA, introduced by Luc Anseli in 1995 to the geographical analysis community, is an extension of Moran statistics [26] developed by Patrick Moran in 1948. LISA is applied for the identification of local patterns of spatial association in mapped data and for the decomposition of global measures of autocorrelation such as Moran’s statistics [26], Geary’s statistics [27] and Getis-Ord G statistics [28]. LISA can be regarded as a variant of window-based Moran coefficient proposed by [11] for pixels within a local region in an image. Furthermore, LISA extracts structural information and provide spatial variation of image quality, thus, satisfying the requirement of a good perceptual quality indicator outlined in [12].

This paper is organized as follows. Sections 2 and 3 describe our method and experiment on performance evaluation for the proposed quality assessment. Performance evaluation results are displayed in Section 4. Results from the experiment are discussed in Section 5. Section 6 concludes this report.

2. Methods

2.1. Problem Formulation

2.1.1. Noise, Sharpness and Contrast in Grayscale Images

Clustered pixels in grayscale images allow human observers and man-made devices to distinguish between the different anatomical structures and also distinguish between normal and abnormal structures under different pathological conditions. The degree of clustering measures the level of contrast and sharpness quality attributes of an image. Contrast is the perceived visual differences between the different structures within the image [29]. Sharpness is the visibility of small structures as measured by the deviation and difference of grey levels in the image [30].

Noise influences the contrast and sharpness quality attributes of medical images [31]. Rician noise in grayscale images such as MRI images induces randomness which causes break-up of clustered pixels and erode edges. Erosion of edges reduces contrast between different structures thereby making it difficult to utilize images for disease diagnosis. A global measure of spatial autocorrelation cannot provide information on the variations in locally perceived visibility of noise because it assumes uniform cluster patterns throughout the image. Therefore, quantification of local spatial autocorrelation will be a useful parameter for noise-based quality evaluation.

2.1.2. Local Indicators of Spatial Autocorrelation (LISA) Statistics

The most popular LISA statistics are the local Moran statistics (LMS). Henceforth LISA statistics will be used interchangeably with LMS throughout this paper. The feature of interest is pixel intensity level x_i at locations $i \in \{1, 2, \dots, N\}$ in an image. Each pixel location i is the center of a neighbourhood defined by a fixed-size $r \times c$ window. In each neighbourhood, x_j are the neighbouring pixels to x_i . The LMS I_i in each neighbourhood is expressed as:

$$I_i = z_i \sum_j w_{ij} z_j \tag{1}$$

where z_i, z_j are the deviations of x_i, x_j , respectively from the mean pixel intensity level \bar{X} within the neighbourhood:

$$\begin{aligned} z_i &= \frac{(x_i - \bar{X})}{S_i^2} \\ z_j &= (x_j - \bar{X}) \end{aligned} \tag{2}$$

where w_{ij} is a spatial weight matrix which define the spatial interaction between x_i and its neighbouring pixels x_j , and S_i^2 is the standard deviation of pixels x_j that are neighbours to x_i :

$$S_i^2 = \sum_{j=1}^N \frac{(x_j - \bar{X})^2}{N} \tag{3}$$

The followings characterize the information that can be derived from LISA statistics:

1. Average of I_i across the image gives the global Moran Statistics (GMS) I .

$$I = \frac{1}{N} \sum_i I_i \tag{4}$$

The range of values for Global Moran statistics is $\{+1 \leq I \leq -1\}$. For gray level images, GMS of 1 indicates the highest degree of clustering, GMS of 0 indicates randomness and GMS of -1 indicates highest degree of randomness and dispersion:

$$I = \begin{cases} 1 & \text{Highest Degree of Clustering} \\ 0 & \text{Randomness} \\ -1 & \text{Highest Degree of Randomness and Dispersion} \end{cases} \tag{5}$$

The range of values for LMS is $\{+k_1 \leq I \leq -k_2\}$, where k_1, k_2 , the upper and lower limits of LMS is determined by the type of grayscale image and the distribution of pixel intensity levels.

2. LISA statistics can identify the presence of outliers and the degree of spatial clustering at specific location.
3. The magnitude of positive and negative values of LISA statistics measures the degree of pixel clusters and pixel dispersion, respectively, at specific location.
4. Positive values of LISA statistics are an indication of clustering; there is relatively low margin between the pixel intensity level at specified location and corresponding intensity levels of neighbouring pixels. On the other hand negative values of LISA statistics indicates the presence of an outlier; there is a relatively wide margin between the intensity of pixel at specified location and corresponding intensity levels of neighbouring pixels.
5. Clustered pixels can be classified as clusters of high pixel intensity levels HH and clusters of low pixel intensity levels LL . Outliers can be classified as pixels with high intensity values surrounded primarily by pixels of low intensity values HL and pixels of low intensity values surrounded primarily by pixels of high intensity values LH .

2.2. Implementation

The flowchart in Figure 1 and the images in Figure 2 describe how the proposed method can be implemented. Figure 2 is a slice from a T2 MRI volume data provided by NeuroRx Research Inc. (Montreal, QC, Canada). The four successive steps to implement the algorithm are outlined below.

1. Step 1: Foreground Extraction

Foreground extraction is the extraction of the regions of interest in the test image from the background region. Foreground image I_f was extracted using the threshold method. There are three steps to extract the foreground. First is global threshold. The threshold was set at the mean intensity level of the image. The next step is a morphological filling operation followed by area threshold where small regions within the image are eliminated. Knowledge of the foreground shown in Figure 2g allows the determination of the indices of pixels as well as the total number of spatial locations in the foreground region. The number of spatial locations is required in the later implementation steps.

2. Step 2: Feature Extraction

The local Moran feature image is derived by computing the local Moran statistics of the test image according to Equation (1). The spatial weight w_{ij} which define the interaction of pixels is determined by the kernel dimension. In this research, the spatial weight was implemented using a 3×3 kernel. The local Moran statistics is averaged according to Equation (4) to obtain the global Moran statistics.

3. Step 3: Feature Classification

Using global threshold, the local Moran feature image I_{LMF} is classified into two classes. The first class consist of random and dispersed pixels I_{LMFA} . The second class I_{LMFB} consist of clustered pixels:

$$I_{LMF} = \begin{cases} I_{LMFA} & \text{if } I_i \leq 0 \\ I_{LMFB} & \text{otherwise} \end{cases} \quad (6)$$

The two classes of pixels are calculated over the foreground region. Figure 2a,b are the images resulting from the addition of 8 percent and 16 percent Rician noise levels to the image in Figure 2a. Random and dispersed feature images corresponding to noise level of 0 percent (Figure 2a), 8 percent (Figure 2b) and 16 percent (Figure 2c) are displayed in Figure 2d–f, respectively.

4. Step 4: Quality Prediction

Quality prediction is based on two concepts. First, the GMS is considered a perceptual weight which modulates the LMS. Second, the test image is a real grayscale image having heterogeneous features, that is, images in which pixels can be assigned to at least two different classes. In contrast, sharpness and total quality scores shown in Figure 2h are predicted from the perceptually weighted sum of the clustered and dispersed pixels within a grayscale image.

The contrast quality score Q_1 is defined as:

$$Q_1 = I \left(1 - \frac{N_{CA}}{N_{If}} \right) + (1 - I) \left(\frac{N_{CB}}{N_{If}} \right) \quad (7)$$

where N_{CA} , N_{CB} , N_{If} are the number of dispersed, clustered and foreground pixels, respectively in the image.

The sharpness quality score Q_2 is defined as:

$$Q_2 = 1 - \left(\frac{N_{CA}}{N_{If}} \right) + (1 - I) \left(\frac{N_{CB}}{N_{If}} \right) \quad (8)$$

The total quality score Q_T is the average of the contrast Q_1 and sharpness Q_2 quality scores:

$$Q_T = \frac{(Q_1 + Q_2)}{2} \quad (9)$$

Here, we show how the quality scores defined in Equations (7) and (8) can predict the contrast and sharpness quality scores of an ideal, extremely degraded and real MRI slices.

(a) Ideal MRI Slice

In an ideal MRI slice, the pixels tend towards the highest degree of clustering. According to Equation (5),

$$I \approx 1 \quad (10)$$

Since random and dispersed pixels are sparse in an ideal MRI image,

$$N_{CA} \approx 0 \quad (11)$$

Inserting Equations (10) and (11) into Equations (7) and (8), the contrast and sharpness quality scores are both equal and optimized towards a value of 1,

$$Q_1 = Q_2 \approx 1 \tag{12}$$

(b) Extremely Degraded MRI Slice

For an extremely degraded MRI slice, the pixels tend towards the highest degree of randomness and dispersion. According to Equation (5),

$$I \approx -1 \tag{13}$$

Random and dispersed pixels are dominant and contained within the foreground region,

$$N_{C_A} \approx N_{I_f} \tag{14}$$

Clustered pixels are sparse, thus

$$N_{C_B} \approx 0 \tag{15}$$

Inserting Equations (13)–(15) into Equations (7) and (8), the contrast and sharpness quality scores are both equal and minimized towards a value of 0,

$$Q_1 = Q_2 \approx 0 \tag{16}$$

(c) Real MRI Slice

For a real MRI slice, the contrast, sharpness and total quality scores in Figure 2h are defined to lie in the range of values between ideal and extremely degraded MRI slices:

$$\begin{aligned} \{Q_1 : 0 \leq Q_1 \leq 1\} \\ \{Q_2 : 0 \leq Q_2 \leq 1\} \\ \{Q_T : 0 \leq Q_T \leq 1\} \end{aligned} \tag{17}$$

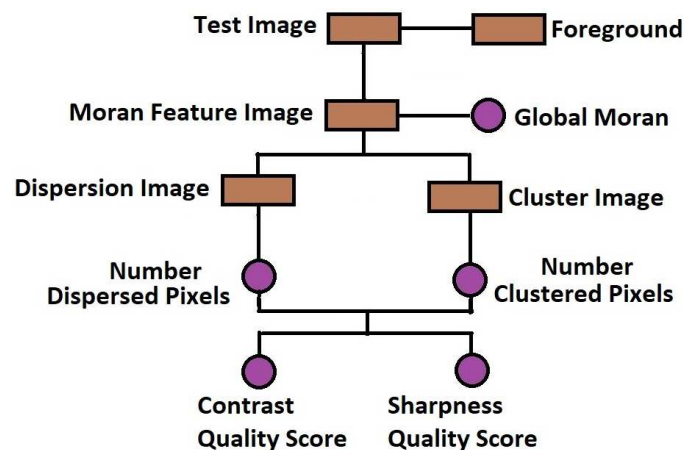


Figure 1. Flow chart of the proposed method for blind noise-based quality assessment in MRI images.

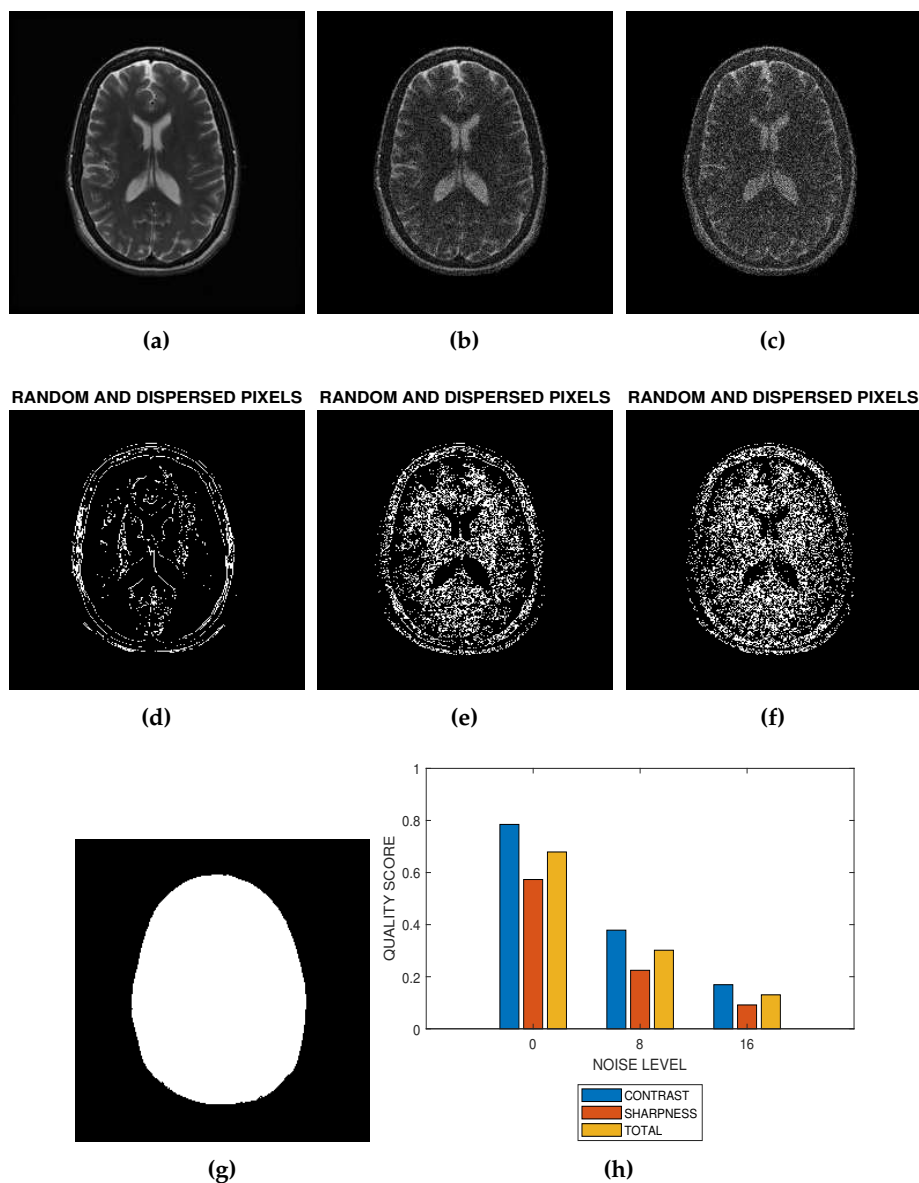


Figure 2. The implementation of blind noise-based quality assessment in MRI images. (a) The test image provided by NeuroRx Research Inc. (b) Degraded version of the test image in (a) at 8 percent Rician noise. (c) Degraded version of the test image in (a) at 16 percent Rician noise. (d) Random and dispersed pixels in the test image in the absence of noise distortion. (e) Random and dispersed pixels in the test image at 8 percent Rician noise. (f) Random and dispersed pixels in the test image at 16 percent Rician noise. (g) The Foreground extracted from the test image in (a). (h) Variation of contrast, sharpness and total quality scores at Rician noise levels of 0 percent, 8 percent and 16 percent.

3. Experiment

3.1. Sources and Description of Test Data

The dataset for the experiment consist of 18 brain, ten cardiac and ten breast MRI volume data. The brain data consist of 15 real and three simulated datasets. In all the MRI data file information, spatial resolution was not explicitly expressed by the field of view (FOV). The available metric closest to the FOV is slice thickness.

3.1.1. Real Brain MRI Data

The real brain MRI volume data were acquired without perceived distortion. It was provided by NeuroRx Research Inc. (<https://www.neurorx.com>), BrainCare Oy (<http://braincare.fi/>) and The Alzheimer's Disease Neuroimaging Initiative ADNI (www.adni.loni.usc.edu). They consist of 10 T2 images and 5 T1 images. Each slice in the T2 and T1 MRI volume data from NeuroRx has dimension of 256×256 and 2.4 mm thickness. Data from ADNI has dimension 192×166 and 1.2 mm thickness. The T1 MRI volume data from BrainCare has dimension 448×512 pixels and 5 mm thickness.

3.1.2. Cardiac MRI Data

The cardiac MRI were short axis MRI volume data from the Department of Diagnostic Imaging of the Hospital for Sick Children in Toronto, Canada (<http://www.sickkids.ca/DiagnosticImaging/index.html>). The images were acquired using the Fast Imaging Employing Steady State Acquisition (FIESTA) sequence protocol. The data were among the experimental data in the report [32]. The report describes the framework for the analysis of short axis cardiac MRI using statistical models of shape and appearance. Each volume data has four dimensions and contains 20 frames. The number of slices in each frame varies from 8 to 15. The dimension of each slice is 256×256 along the long axis.

3.1.3. Breast MRI Data

The breast MRI was obtained from the Reference Image Database to Evaluate Therapy Response (RIDER) [33] in the Cancer Imaging Archive (TCIA) (<https://wiki.cancerimagingarchive.net/display/Public/RIDER+Breast+MRI#>) database. Each volume data contains 60 slices with dimension 288×288 and thickness of 2.5 mm.

3.1.4. Simulated MRI Data

Three simulated T1, T2 and PD MRI volume data were downloaded from McGill University BrainWeb [34]. There are 60 slices in each volume data. Each slice has dimension 217×181 and thickness of 3 mm.

3.2. Generation of Noise Distortion

Noise was added to a real MRI image according to the procedure outlined in [5]. Two separate and identical realizations of white Gaussian noise $\mathcal{N}(0, \sigma)$ were generated. The standard deviation σ of the Gaussian noise is expressed as a percentage τ of the maximum intensity level ν in the image. Thereafter, we simulate the complex plane of MRI acquisition process. The real component I_r was simulated by adding a realization of the Gaussian noise to a real MRI image A :

$$I_r = A + \mathcal{N}(0, \sigma) \quad (18)$$

The imaginary component I_i is the second realization of the Gaussian noise:

$$I_i = \mathcal{N}(0, \sigma) \quad (19)$$

Added Rician noise m is the magnitude of the complex data.:

$$m = \sqrt{I_r^2 + I_i^2} \quad (20)$$

White Gaussian noise equivalent of Rician noise level is defined as in [5]:

$$\sigma = \frac{\nu\tau}{100} \quad (21)$$

The Rician noise level was scaled from 1 to 15 in unit step.

3.3. Experiment Category

Objective and subjective evaluation are the two main categories of the experiment.

3.3.1. Objective Evaluation

The objective evaluation has three categories; retrospection, noise reduction and comparative performance evaluation.

1. Retrospection

This category utilize real MRI data that were retrospectively acquired without degradation. It was further divided into T2, T1, breast and cardiac MRI data. 250 slices were utilized for each category.

2. Noise Reduction

Two hundred slices were selected from the retrospectively acquired MRI data for performance evaluation of the bilateral filter proposed in [35]. The bilateral filter is a non-linear filter which became popular because of its edge-preserving feature. We chose to evaluate only one state-of-the-art noise reduction algorithm because the goal of this research is not comparative performance evaluation. The MATLAB implementation code was downloaded from (people.csail.mit.edu/jiawen/software/). Our proposed method assessed the noise reduction algorithm at mild (4 percent), moderate (8 percent) and severe (12 percent) levels of Rician noise. The parameters of the bilateral filter are as follows; smoothing parameter in the spatial dimension $\sigma_S = 5$, smoothing parameter in the range dimension $\sigma_R = 15$, amount of downsampling in the spatial dimension $\gamma_S = 5$ and amount of downsampling in the range dimension $\gamma_R = 15$.

3. Comparative Performance Evaluation

Comparative performance evaluation was carried out on simulated and real MRI data. The simulated data were 15 slices from T1, T2 and PD MRI volume data. The real data were retrospectively acquired from 15 T2, 15 T1, 15 breast and ten cardiac MRI images. The proposed method was compared to SSIM, PSNR and BRISQUE. Rician noise, from level 0 to level 15 was added to each slice in a MRI volume. For each level of noise, quality prediction from each quality metric is the average quality scores from all slices in the MRI volume.

The SSIM and our proposed method have the same lower and upper limit quality indices. This is not the case for PSNR and BRISQUE. The quality indices from the PSNR and BRISQUE were modified to have same lower and upper limit quality indices as our proposed method. Since PSNR will give very large number, it was computed in the decibel scale. The decibel value was further divided by 100. BRISQUE quality index is such that the image with best quality is 0 while that image with worst quality is 100. To make the quality index comparable to our proposed method, output from BRISQUE was subtracted from 100. The difference is further divided by 100.

3.3.2. Subjective Validation

The objective experiment was validated by four observers; two radiologists and two medical imaging professionals. The validation experiment was facilitated by **QuickEval** [36], a web-based tool for psychometric image evaluation provided by the Norwegian Colour and Visual Computing Laboratory (www.colourlab.no/quickeval) at the Norwegian University of Science and Technology, Gjøvik, Norway. The experiment was classified according to the objective experiment category, but we report results for only comparative performance evaluation.

The observer assigns a score between 0 and 100, in unit step, to each slice. Each score assigned by the observer is divided by 100 to ensure that the subjective and objective scales are in the same range. An observer was first presented with an undistorted version of an MRI slice, followed by increasing distortion levels in the original slice. The distorted levels are 5, 10 and 15. Spearman's rank correlation coefficient ρ [37] is the metric we use to measure the relationship between our objective results and the score assigned by human observers.

4. Results

Noise, like other classical quality attributes contributes to the visual quality of an image. This research exploits the principle of local indicators of spatial autocorrelation for blind assessment of the perceptual quality of magnitude MRI images in an environment where Rician noise is the dominant type of distortion. Contrast and sharpness quality of an image is expressed by the degree of clustering within the image. The degree of clustering is quantitatively expressed by local Morgan statistics.

The proposed noise-based quality evaluation method was implemented in the MATLAB computing environment. The demonstration and implementation codes are attached as supplementary file to this paper. Upon acceptance, the codes will be available for download from (www.colourlab.no/software). In this paper, nine figures are used to explain the performance evaluation of the proposed method.

The first slice images (Figures 3a, 4a, 5a and 6a) in Figures 3–6 are the original images without degradation. The following five slice images are Rician noise degraded versions of the original images at Rician noise levels of 3 percent, 6 percent, 9 percent, 12 percent and 15 percent, respectively. For each level of noise, the contrast, sharpness and total quality scores are displayed in Figures 3g, 4g, 5g and 6g. Expectedly, the quality scores decrease with increasing level of noise.

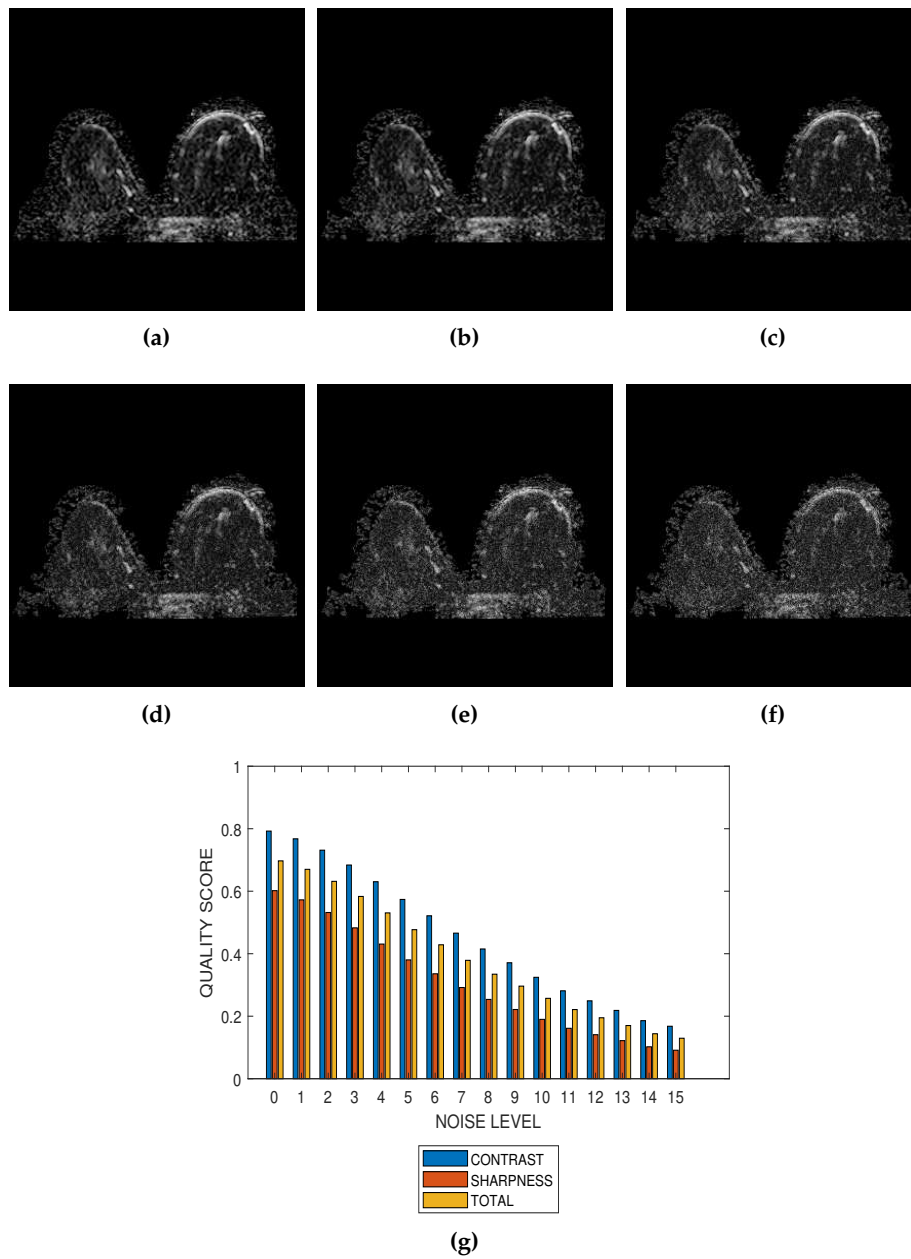


Figure 3. (a) A breast MRI slice provided by the RIDER project of The Cancer Imaging Archive (TCIA) and its degraded versions at Rician noise levels (b) 3 percent, (c) 6 percent, (d) 9 percent, (e) 12 percent and (f) 15 percent. (g) Variation of contrast, sharpness and total quality scores for Rician noise levels that vary from 0 percent to 15 percent.

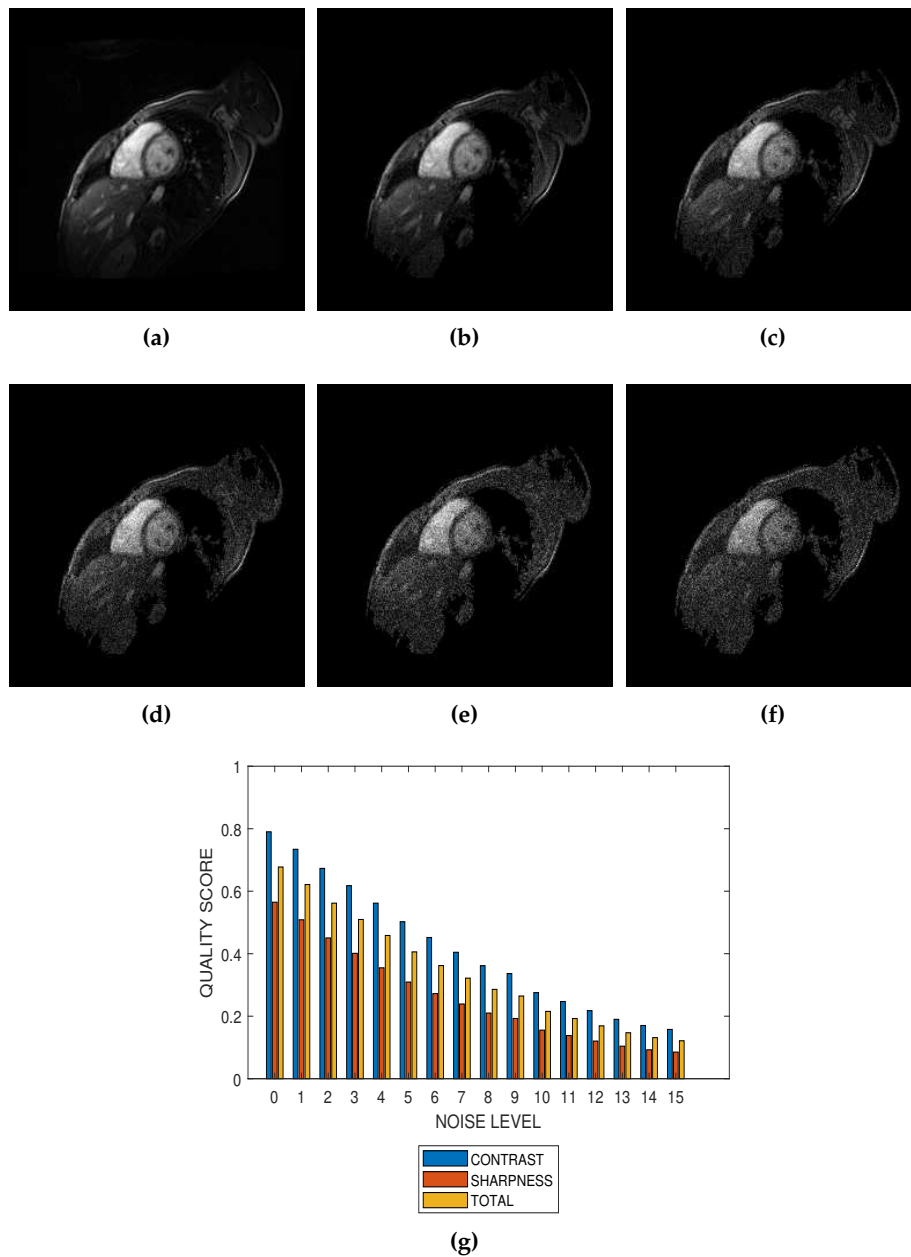


Figure 4. (a) A short axis cardiac MRI slice provided by the Department of Diagnostic Imaging of the Hospital for Sick Children in Toronto, Canada, and its degraded versions at Rician noise levels (b) 3 percent, (c) 6 percent, (d) 9 percent, (e) 12 percent and (f) 15 percent. (g) Variation of contrast, sharpness and total quality scores for Rician noise levels that vary from 0 percent to 15 percent.

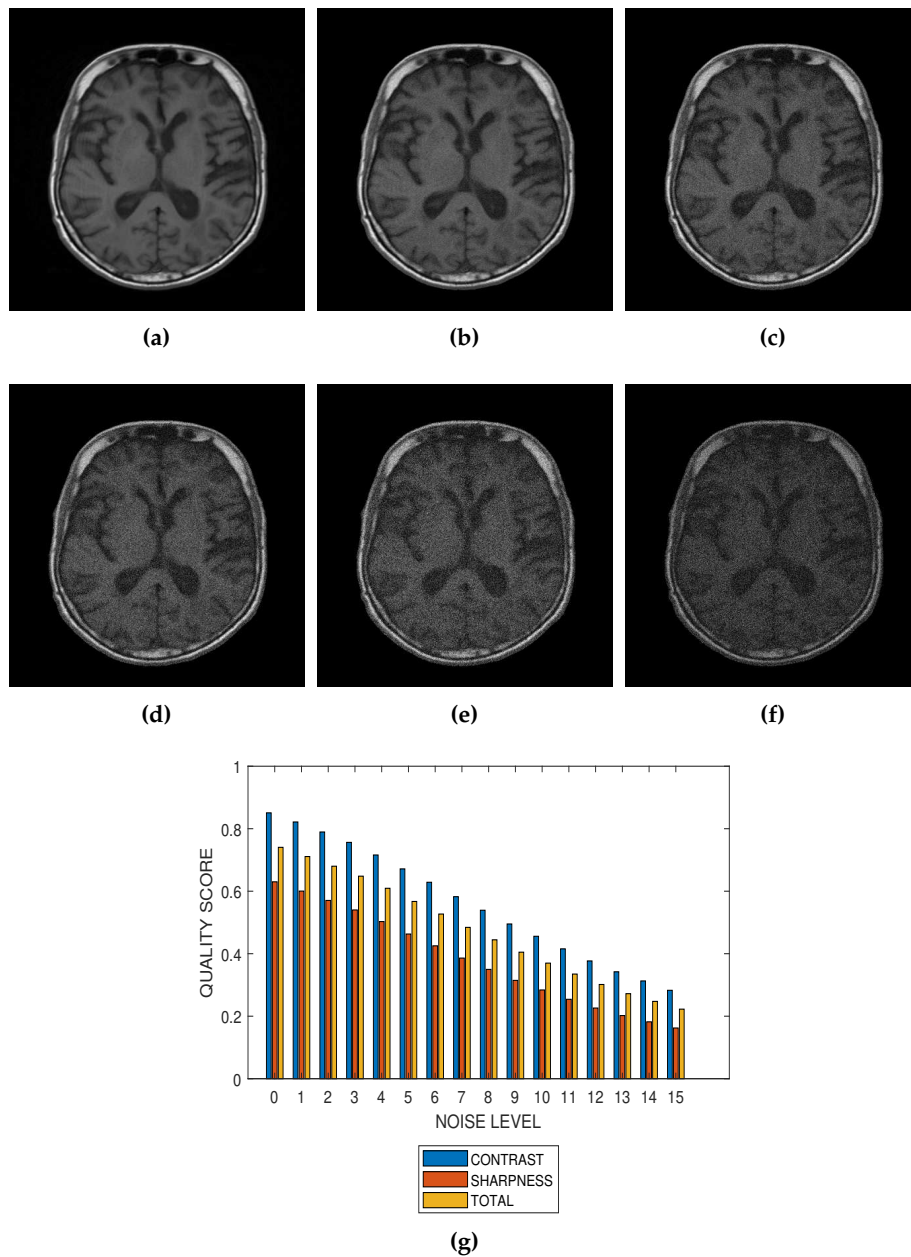


Figure 5. (a) A conventional T1 weighted MRI slice from BrainCare Oy., and its degraded versions at Rician noise levels (b) 3 percent, (c) 6 percent, (d) 9 percent, (e) 12 percent and (f) 15 percent. (g) Variation of contrast, sharpness and total quality scores for Rician noise levels that vary from 0 percent to 15 percent.

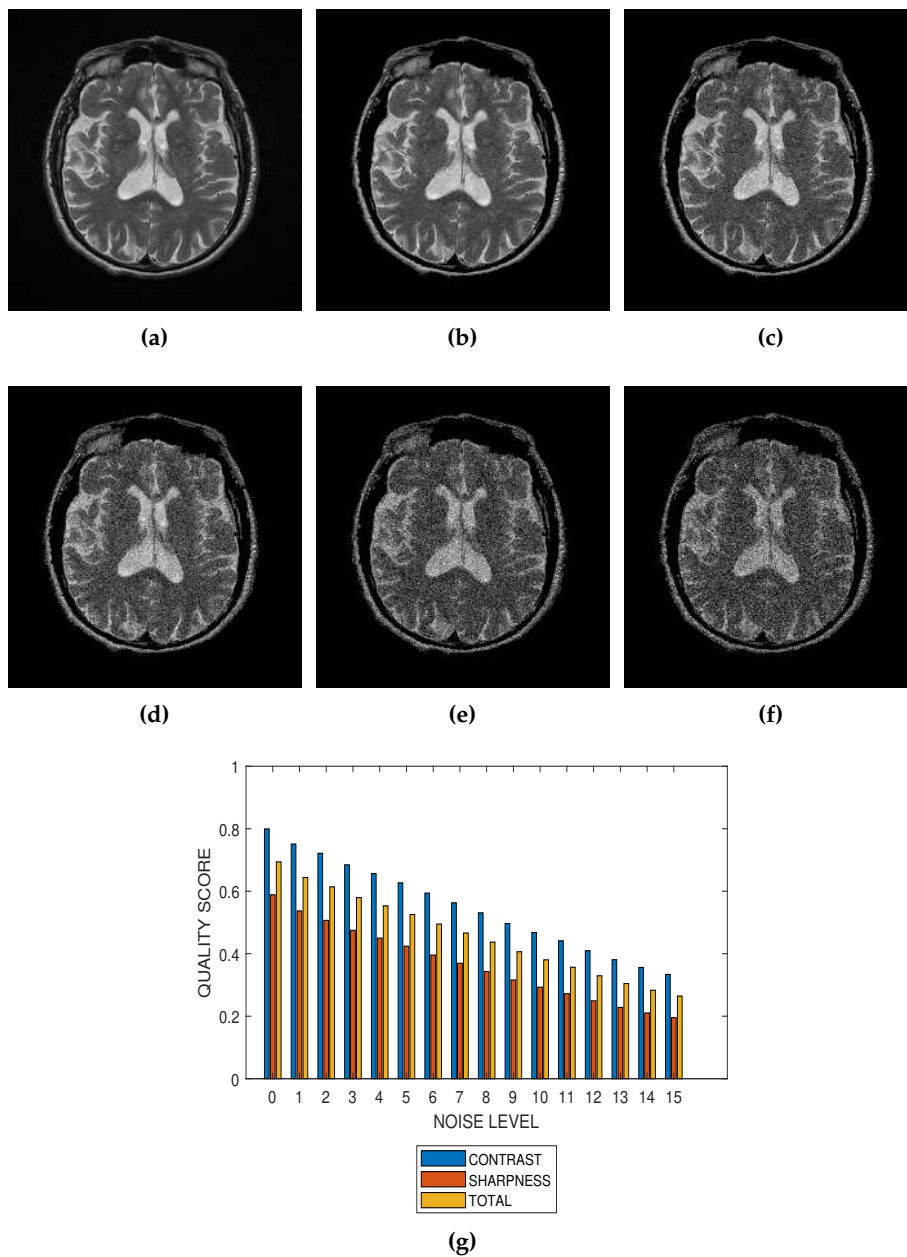


Figure 6. (a) A T2 weighted MRI slice from ADNI and its degraded versions at Rician noise levels (b) 3 percent, (c) 6 percent, (d) 9 percent, (e) 12 percent and (f) 15 percent. (g) Variation of contrast, sharpness and total quality scores for Rician noise levels that vary from 0 percent to 15 percent.

Performance of the noise reduction algorithm at 4 percent, 8 percent and 12 percent noise levels are displayed in Figures 7–9, respectively. In the three levels of Rician noise, the post processing quality scores are higher than corresponding preprocessing quality scores.

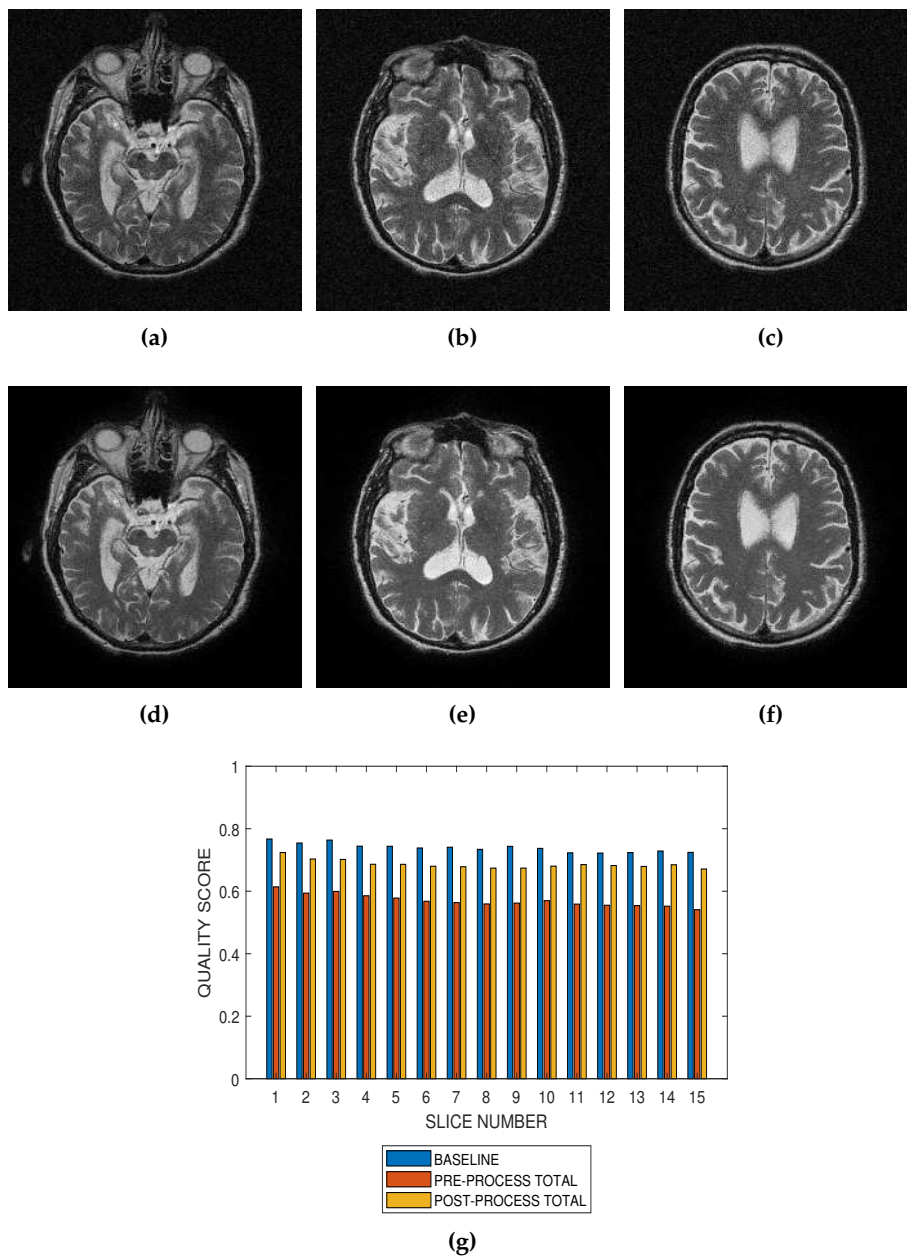


Figure 7. Application of the proposed method to the assessment of an image enhancement algorithm. (a–c) are slices numbers 20, 27 and 34, respectively in a T2 MRI volume degraded with 4 percent Rician noise. (d–f) are the slices displayed in (a–c) but their quality have been enhanced using a bilateral filter. (g) The baseline quality score of the original image, the predicted total quality indices of 12 successive T2 MRI slices in a volume data before and after enhancement with a bilateral filter.

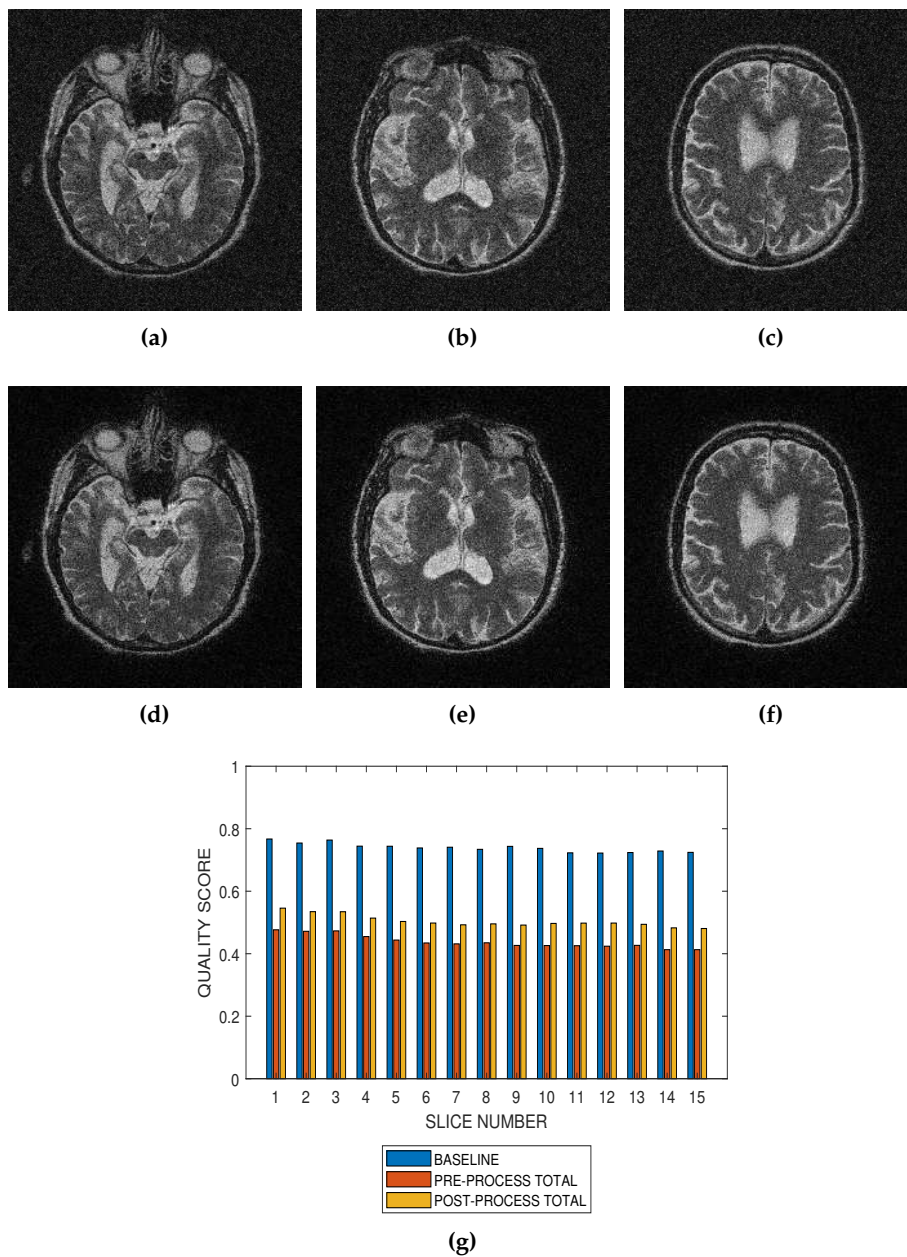


Figure 8. Application of the proposed method to the assessment of an image enhancement algorithm. (a–c) are slices numbers 20, 27 and 34, respectively in a T2 MRI volume degraded with 8 percent Rician noise. (d–f) are the slices displayed in (a–c) but their quality have been enhanced using a bilateral filter. (g) The baseline quality score of the original image, the predicted total quality indices of 12 successive T2 MRI slices in a volume data before and after enhancement with a bilateral filter.

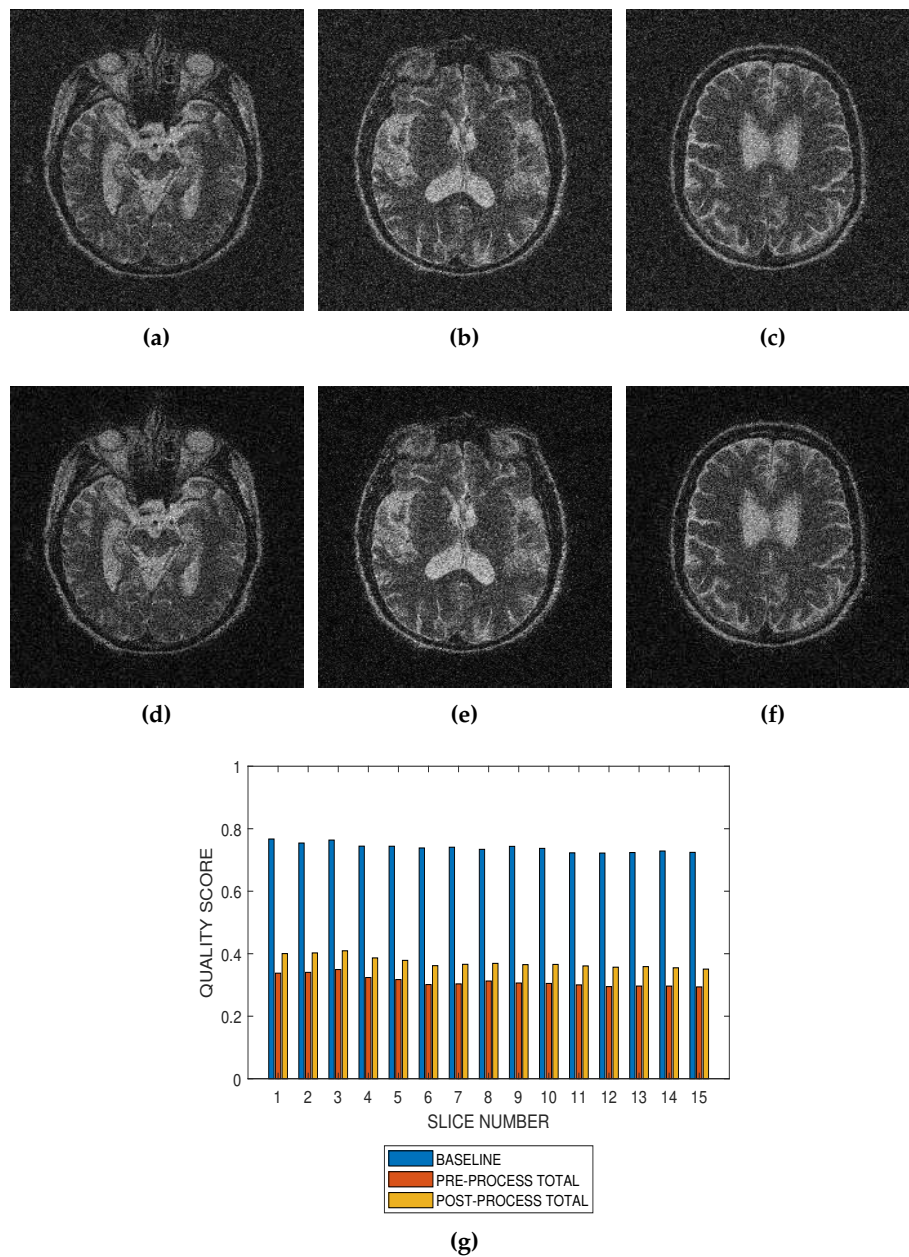


Figure 9. Application of the proposed method to the assessment of an image enhancement algorithm. (a–c) are slices numbers 20, 27 and 34, respectively in a T2 MRI volume degraded with 12 percent Rician noise. (d–f) are the slices displayed in (a–c) but their quality have been enhanced using a bilateral filter. (g) The baseline quality score of the original image, the predicted total quality indices of 12 successive T2 MRI slices in a volume data before and after enhancement with a bilateral filter.

Results from comparative performance evaluation of the proposed method with SSIM, PSNR and BRISQUE are displayed in Figures 10 and 11, for real and simulated MRI data, respectively. Figure 10a–d are quality predictions on T2, T1, breast and cardiac MRI slices for Rician noise levels from 0 to 15. The images in the first, second and third columns of Figure 11 are sample MRI images from simulated T1, T2 and PD MRI volume data. Quality predictions corresponding to each MRI volume data, for Rician noise levels from 0 to 15, are the plots in the last row of Figure 11. For all the quality metrics, there is a general trend of decreasing quality score with increasing noise level. Tables 1–4 are the subjective evaluation results from the comparative performance of our proposed method with SSIM, PSNR and BRISQUE quality metrics.

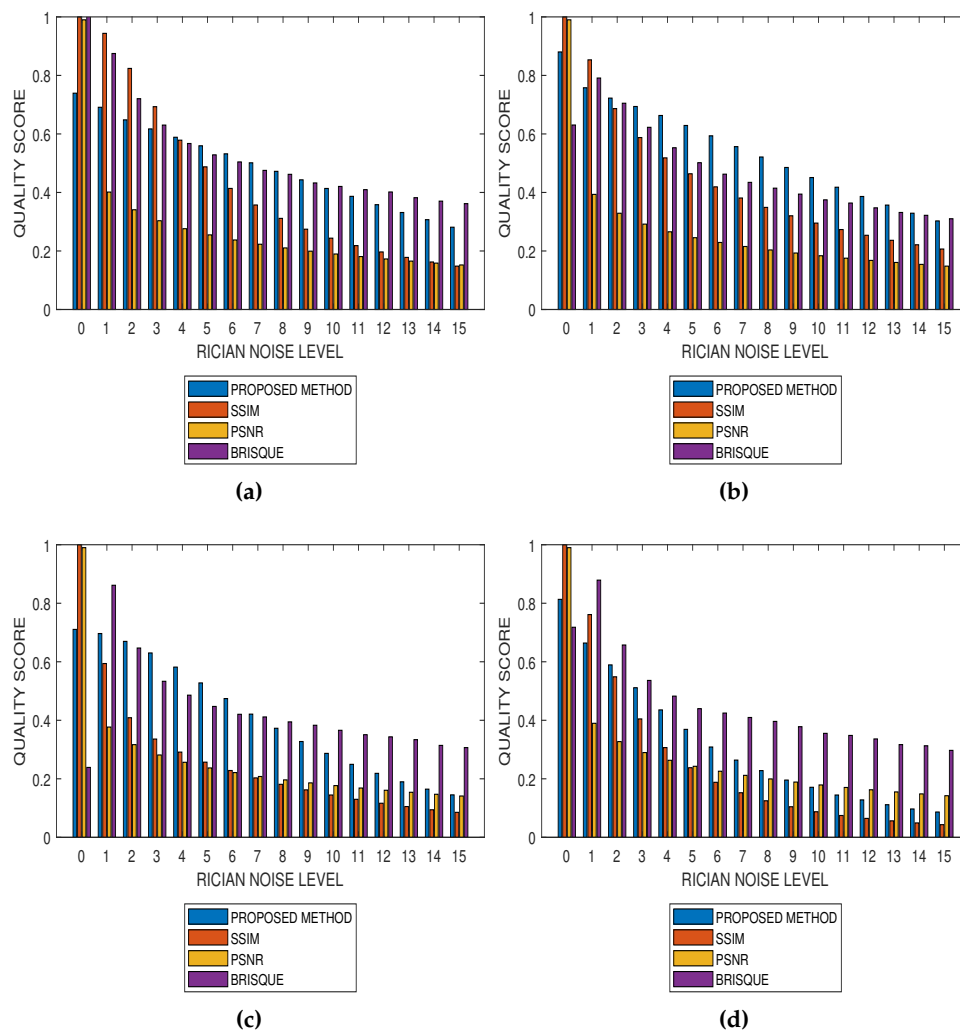


Figure 10. Comparative performance evaluation of the proposed method with SSIM, PSNR and BRISQUE quality metrics using (a) 15 slices in a T2 MRI volume data, (b) 15 slices in a T1 MRI volume data, (c) 15 slices in a breast MRI volume data and (d) 10 slices in a short axis cardiac MRI volume data.

Table 1. Comparative Performance Evaluation of the Proposed Method with SSIM, PSNR and BRISQUE Quality Metrics using T2 MRI Sequence.

Noise Level	Number of Slices	Average Subjective Score	Average Objective Score			
			Proposed Method	SSIM	PSNR	BRISQUE
0	15	0.70	0.75	1.00	0.95	1.00
5	15	0.65	0.60	0.60	0.30	0.60
10	15	0.55	0.50	0.30	0.25	0.50
15	15	0.40	0.35	0.20	0.20	0.40

Table 2. Comparative Performance Evaluation of the Proposed Method with SSIM, PSNR and BRISQUE Quality Metrics using T1 MRI Sequence.

Noise Level	Number of Slices	Average Subjective Score	Average Objective Score			
			Proposed Method	SSIM	PSNR	BRISQUE
0	15	0.81	0.85	1.00	0.90	0.62
5	15	0.73	0.70	0.52	0.30	0.60
10	15	0.50	0.55	0.35	0.20	0.45
15	15	0.35	0.40	0.30	0.18	0.40

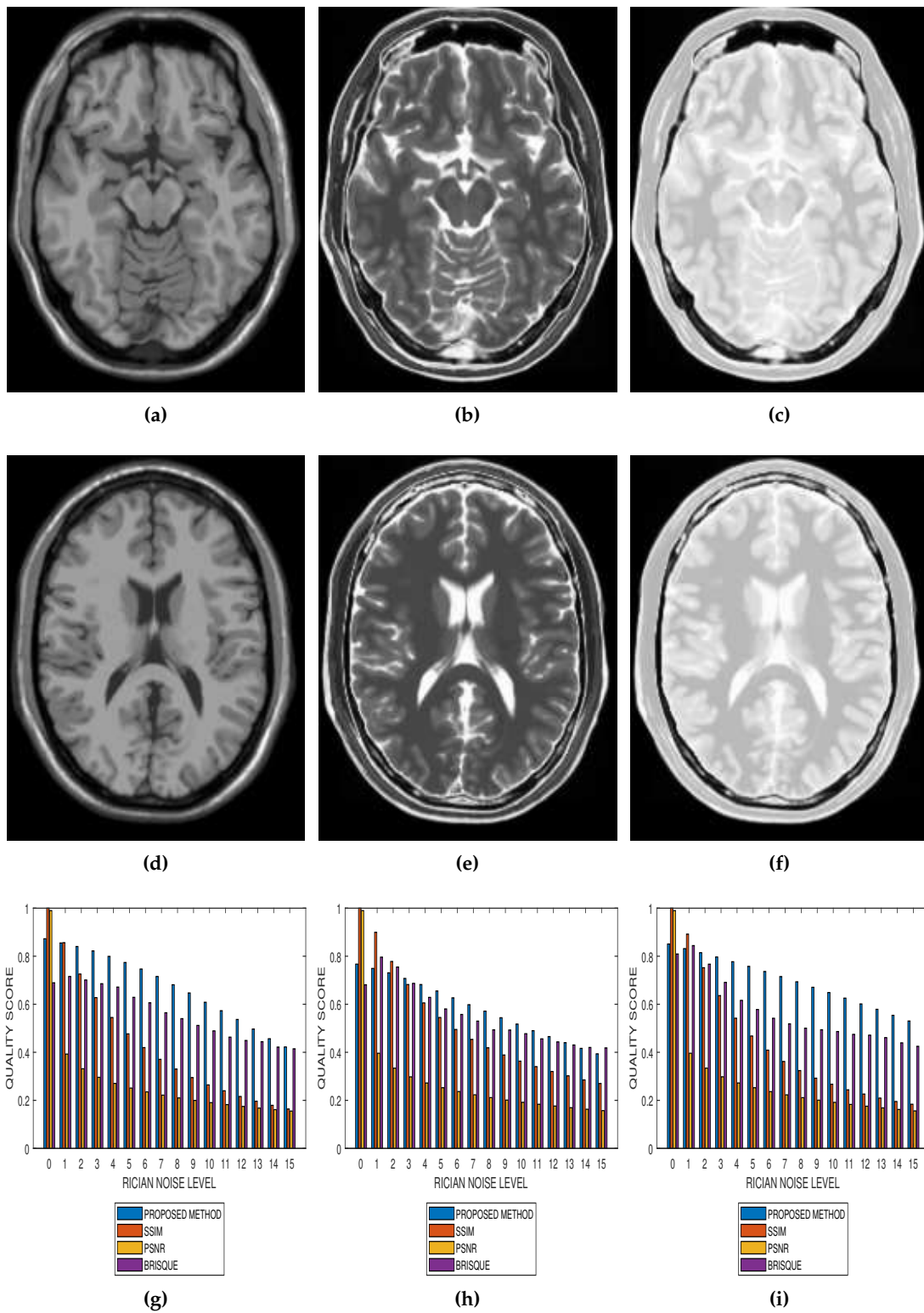


Figure 11. Comparative performance evaluation of the proposed method with SSIM, PSNR and BRISQUE quality metrics using 15 slices in (a,d) T1, (b,e) T2 and (c,f) PD simulated MRI volume data from McGill University BrainWeb (g–i).

Table 3. Comparative Performance Evaluation of the Proposed Method with SSIM, PSNR and BRISQUE Quality Metrics using Breast MRI Images.

Noise Level	Number of Slices	Average Subjective Score	Average Objective Score			
			Proposed Method	SSIM	PSNR	BRISQUE
0	15	0.75	0.70	1.00	0.95	0.23
5	15	0.61	0.55	0.30	0.25	0.50
10	15	0.40	0.35	0.20	0.22	0.40
15	15	0.20	0.15	0.15	0.15	0.35

Table 4. Comparative Performance Evaluation of the Proposed Method with SSIM, PSNR and BRISQUE Quality Metrics using Cardiac MRI Images.

Noise Level	Number of Slices	Average Subjective Score	Average Objective Score			
			Proposed Method	SSIM	PSNR	BRISQUE
0	10	0.76	0.80	1.00	0.95	0.70
5	10	0.47	0.45	0.32	0.27	0.50
10	10	0.25	0.20	0.10	0.10	0.40
15	10	0.15	0.10	0.05	0.03	0.35

5. Discussion

The goal of image quality evaluation systems is to predict the quality of images with the same level of efficiency as human visual system (HVS). In this research, we propose to evaluate the quality of MRI images in a noise-dominant environment. The formulation of the proposed method incorporates four steps which mimic HVS characteristics. First, the magnitude of local Moran statistics at each location can be regarded as a measure of contrast sensitivity threshold. Second, the spatial weight which define the spatial interaction in a local neighbourhood accounts for the visibility of locally perceived noise. Third, averaging the local Moran statistics to give the global Moran statistics is the equivalent of integrating perceptually weighted local noise. Fourth, perceptual weight, expressed as a function of GMS is assigned to contrast and sharpness quality attributes in the computation of the total quality score. The proposed quality evaluation system transforms noise level into an easy-to-interpret and standardized quality metric. The lower and upper limits of the quality index are 0 and 1, respectively. Below, we discuss the suitability of our proposed method to different applications.

The proposed method is a practical and realistic approach to quality evaluation. The images in Figures 3–6 were evaluated without the need of a reference image. The images used for the performance evaluation were acquired with different levels of details and, hence, possess different perceptual quality. Variation in the profiles of quality scores shown in Figures 3–6 shows that the proposed quality evaluation system can provide fairly good perceptual quality assessment for different types of MRI images.

Performance evaluation of noise reduction algorithms will be a useful application of the proposed method. Evaluation of the proposed methods before and after denoising MRI images (See Figures 7–9) show that the performance of noise reduction algorithms decreases with the severity of noise.

There is no doubt that SSIM, PSNR and BRISQUE are popular and efficient quality evaluation techniques. The comparative performance evaluation results displayed in Tables 1–4 as well as Figures 10 and 11 reveal their limitations and add to the growing calls for application-specific quality metrics. At lower noise levels, current techniques provide an inaccurate estimate of perceptual quality. This is expected for SSIM and PSNR as their predicted quality score is based on noise level and with reference to an assumed ideal image. The no-reference BRISQUE performed better than SSIM and PSNR. However, inaccurate estimation of noise, particularly at lower noise levels in T2, T1 and Breast MRI, was evident in BRISQUE. This can be attributed to training data extracted from images corrupted by Gaussian noise rather than the Rician noise present in MRI images. The proposed method has the

lowest margin between a predicted quality score and the corresponding score assigned by human observers. Furthermore, it has the highest perceptual quality discrimination for different noise levels. For these reasons, we can say that the proposed method demonstrates superior performance over current techniques.

A good image quality evaluation model should not only deliver high quality prediction accuracy but also be computationally efficient [38]. The proposed method meets these requirements. Unlike some current techniques such as BRISQUE, there is no need for additional resources such as complex feature extraction and training of data for quality prediction. The simple and efficient feature extraction coupled with processing in the binary domain makes the proposed method useful in applications where large volumes of MRI data are processed.

Evaluation of the proposed method on real MRI images was complemented with anatomically realistic phantoms. Since spatially invariant Rician noise was artificially added to these images, the performance of the proposed method on spatially variant noise in multiple-coil acquired MRI images, remains an open question. However, the performance evaluation results demonstrate the promise of a new state-of-the-art algorithm.

6. Conclusions

Noise limits the utility of medical images for the diagnosis of diseases. This paper propose a new noise-based quality evaluation for MRI images. The LMS estimates the visibility of locally perceived noise from the clustered and dispersed pixels. Noisiness at each local area is perceptually weighted using the spatial weight matrix. The perceptually weighted local noise is integrated to form the GMS. Perceptual weight assigned to the contrast and sharpness quality scores is expressed as a function of the GMS. Quality prediction is based on the perceptually weighted sum of the clustered and dispersed pixels in the image. The proposed method is computationally efficient and performance evaluation shows good correlation with subjective evaluation by human observers. Potential applications of the proposed method include the evaluation of retrospectively acquired MRI images, performance evaluation of noise reduction algorithms, parameter optimization for denoising algorithms and the selection of an appropriate acceleration factor in parallel imaging acquisition techniques.

Author Contributions: M.O. carried out the design and implementation of the proposed metric system. M.P. contributed to the technical development, analysis and interpretation of the results. D.A. and K.W.-M. were involved in data analysis as well as interpretation of the experimental results. All authors have been involved in drafting and revising the manuscript and approved the final version to be published. All authors read and approved the final manuscript.

Funding: Marius Pedersen have been supported by the Research Council of Norway, project no. 247689 'IQMED: Image Quality enhancement in MEDical diagnosis, monitoring and treatment'. Data collection and sharing for this project was, in part, funded by the Alzheimer's Disease Neuroimaging Initiative(ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; MesoScale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Conflicts of Interest: The authors declare no conflict of interests.

Abbreviations

MRI	Magnetic Resonance Imaging
SNR	Signal-to-Noise Ratio
PSNR	Peak Signal-to-Noise Ratio
MSE	Mean Square Error
RMSE	Root Mean Square Error
ADNI	The Alzheimer’s Disease Neuroimaging Initiative
RIDER	Reference Image Database to Evaluate Therapy Response
TCIA	The Cancer Imaging Archive
HVS	Human Visual System
LMS	Local Moran Statistics
GMS	Global Moran Statistics
P1	Parallel Imaging
T1	Longitudinal Relaxation
T2	Transverse Relaxation

References

- Mandić, I.; Peić, H.; Lerga, J.; Štajduhar, I. Denoising of x-ray images using the adaptive algorithm based on the LPA-RICI algorithm. *J. Imaging* **2018**, *4*, 34. [[CrossRef](#)]
- Triantafyllou, C.; Hoge, R.; Krueger, G.; Wiggins, C.; Potthast, A.; Wiggins, G.; Wald, L. Comparison of physiological noise at 1.5 T, 3 T and 7 T and optimization of fMRI acquisition parameters. *Neuroimage* **2005**, *26*, 243–250. [[CrossRef](#)] [[PubMed](#)]
- Aja-Fernández, S.; Pie, T.; Vegas-Sánchez-Ferrero, G. Spatially variant noise estimation in MRI: A homomorphic approach. *Med. Image Anal.* **2015**, *20*, 184–197. [[CrossRef](#)]
- Aja-Fernández, S.; Vegas-Sánchez-Ferrero, G.; Tristán-Vega, A. Noise estimation in parallel MRI: GRAPPA and SENSE. *Magn. Reson. Imaging* **2014**, *32*, 281–290. [[CrossRef](#)] [[PubMed](#)]
- Coupé, P.; Manjón, J.V.; Gedamu, E.; Arnold, D.; Robles, M.; Collins, D.L. Robust Rician noise estimation for MR images. *Med. Image Anal.* **2010**, *14*, 483–493. [[CrossRef](#)] [[PubMed](#)]
- Aja-Fernández, S.; Tristán-Vega, A.; Alberola-López, C. Noise estimation in single-and multiple-coil magnetic resonance data based on statistical models. *Magn. Reson. Imaging* **2009**, *27*, 1397–1409. [[CrossRef](#)] [[PubMed](#)]
- Sijbers, J.; Den Dekker, A.; Van Audekerke, J.; Verhoye, M.; Van Dyck, D. Estimation of the noise in magnitude MR images. *Magn. Reson. Imaging* **1998**, *16*, 87–90. [[CrossRef](#)]
- Sijbers, J.; Poot, D.; den Dekker, A.J.; Pintjens, W. Automatic estimation of the noise variance from the histogram of a magnetic resonance image. *Phys. Med. Biol.* **2007**, *52*, 1335. [[CrossRef](#)]
- Osadebey, M.; Bouguila, N.; Arnold, D. The clique potential of Markov random field in a random experiment for estimation of noise levels in 2D brain MRI. *Int. J. Imaging Syst. Technol.* **2013**, *23*, 304–313. [[CrossRef](#)]
- Tabelow, K.; Voss, H.U.; Polzehl, J. Local estimation of the noise level in MRI using structural adaptation. *Med. Image Anal.* **2015**, *20*, 76–86. [[CrossRef](#)]
- Chuang, K.-S.; Huang, H. Assessment of noise in a digital image using the join-count statistic and the Moran test. *Phys. Med. Biol.* **1992**, *37*, 357. [[CrossRef](#)] [[PubMed](#)]
- Chen, T.-J.; Chuang, K.-S.; Wu, J.; Chen, S.C.; Hwang, M.; Jan, M.-L. A novel image quality index using Moran I statistics. *Phys. Med. Biol.* **2003**, *48*, 131. [[CrossRef](#)]
- Shiao, Y.-H.; Chen, T.-J.; Chuang, K.-S.; Lin, C.-H.; Chuang, C.-C. Quality of compressed medical images. *J. Digit. Imaging* **2007**, *20*, 149. [[CrossRef](#)]
- Chen, T.-J.; Chuang, K.-S.; Chang, J.-H.; Shiao, Y.-H.; Chuang, C.-C. A blurring index for medical images. *J. Digit. Imaging* **2006**, *19*, 118. [[CrossRef](#)] [[PubMed](#)]
- Chen, T.-J.; Chuang, K.-S.; Wu, J.; Chen, S.C.; Hwang, M.; Jan, M.-L. Quality degradation in lossy wavelet image compression. *J. Digit. Imaging* **2003**, *16*, 210–215. [[CrossRef](#)] [[PubMed](#)]
- Chow, L.S.; Rajagopal, H. Modified-BRISQUE as no reference image quality assessment for structural MR images. *Magn. Reson. Imaging* **2017**, *43*, 74–87. [[CrossRef](#)] [[PubMed](#)]
- Mittal, A.; Moorthy, A.K.; Bovik, A.C. No-reference image quality assessment in the spatial domain. *IEEE Trans. Image Process.* **2012**, *21*, 4695–4708. [[CrossRef](#)] [[PubMed](#)]

18. Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding rmse in the literature. *Geosci. Model Dev.* **2014**, *7*, 1247–1250. [[CrossRef](#)]
19. Willmott, C.J.; Matsuura, K.; Robeson, S.M. Ambiguities inherent in sums-of-squares-based error statistics. *Atmos. Environ.* **2009**, *43*, 749–752. [[CrossRef](#)]
20. Willmott, C.J.; Matsuura, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim. Res.* **2005**, *30*, 79–82. [[CrossRef](#)]
21. Kupinski, M.A.; Clarkson, E. Objective assessment of image quality. In *Small-Animal Spect Imaging*; Springer Science & Business Media: Berlin, Germany, 2005; pp. 101–114.
22. Chow, L.S.; Paramesran, R. Review of medical image quality assessment. *Biomed. Signal Process. Control* **2016**, *27*, 145–154. [[CrossRef](#)]
23. Pieciak, T.; Vegas-Sánchez-Ferrero, G.; Aja-Fernández, S. Variance stabilization of noncentral-chi data: Application to noise estimation in MRI. In Proceedings of the 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), Prague, Czech Republic, 13–16 April 2016; pp. 1376–1379.
24. Manjón, J.V.; Coupé, P.; Buades, A. MRI noise estimation and denoising using non-local PCA. *Med. Image Anal.* **2015**, *22*, 35–47. [[CrossRef](#)] [[PubMed](#)]
25. Timischl, F. The contrast-to-noise ratio for image quality evaluation in scanning electron microscopy. *Scanning* **2015**, *37*, 54–62. [[CrossRef](#)]
26. Moran, P.A. The interpretation of statistical maps. *J. R. Stat. Soc. Ser. B (Methodol.)* **1948**, *10*, 243–251. [[CrossRef](#)]
27. Geary, R.C. The contiguity ratio and statistical mapping. *Incorporated Statist.* **1954**, *5*, 115–146. [[CrossRef](#)]
28. Getis, A.; Ord, J.K. The analysis of spatial association by use of distance statistics. *Geogr. Anal.* **1992**, *24*, 189–206. [[CrossRef](#)]
29. Pedersen, M.; Bonnier, N.; Hardeberg, J.Y.; Albrechtsen, F. Attributes of image quality for color prints. *J. Electron. Imaging* **2010**, *19*, 011016.
30. Kim, H.; Cho, K.; Kim, J.; Jin, K.; Kim, S. Robust parameter design of derivative optimization methods for image acquisition using a color mixer. *J. Imaging* **2017**, *3*, 31. [[CrossRef](#)]
31. Singh, P.; Mukundan, R.; De Ryke, R. Texture based quality analysis of simulated synthetic ultrasound images using local binary patterns. *J. Imaging* **2017**, *4*, 3. [[CrossRef](#)]
32. Andreopoulos, A.; Tsotsos, J.K. Efficient and generalizable statistical models of shape and appearance for analysis of cardiac MRI. *Med. Image Anal.* **2008**, *12*, 335–357. [[CrossRef](#)]
33. Clark, K.; Vendt, B.; Smith, K.; Freymann, J.; Kirby, J.; Koppel, P.; Moore, S.; Phillips, S.; Maffitt, D.; Pringle, M.; et al. The cancer imaging archive (TCIA): Maintaining and operating a public information repository. *J. Digit. Imaging* **2013**, *26*, 1045–1057. [[CrossRef](#)] [[PubMed](#)]
34. Collins, D.L.; Zijdenbos, A.P.; Kollokian, V.; Sled, J.G.; Kabani, N.J.; Holmes, C.J.; Evans, A.C. Design and construction of a realistic digital brain phantom. *IEEE Trans. Med. Imaging* **1998**, *17*, 463–468. [[CrossRef](#)] [[PubMed](#)]
35. Tomasi, C.; Manduchi, R. Bilateral filtering for gray and color images. In Proceedings of the Sixth International Conference On Computer Vision, Bombay, India, 7 January 1998; pp. 839–846.
36. Van Ngo, K.; Storvik, J.J.; Dokkeberg, C.A.; Farup, I.; Pedersen, M. QuickEval: A web application for psychometric scaling experiments. In *SPIE/IS&T Electronic Imaging*; International Society for Optics and Photonics: Bellingham, WA, USA, 2015; p. 93960.
37. Corder, G.W.; Foreman, D.I. *Nonparametric Statistics: A Step-by-Step Approach*; John Wiley & Sons: Hoboken, NJ, USA, 2014.
38. Xue, W.; Zhang, L.; Mou, X.; Bovik, A.C. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE Trans. Image Process.* **2014**, *23*, 684–695. [[CrossRef](#)] [[PubMed](#)]

