A novel algorithm for comprehensive quality assessment of clinical magnetic resonance images based on natural scene statistics in spatial domain

幾嶋, 洋一郎

Magnetic Resonance Imaging 92 (2022) 203–211

Contents lists available at ScienceDirect

Magnetic Resonance Imaging

journal homepage: www.elsevier.com/locate/mri

A novel algorithm for comprehensive quality assessment of clinical magnetic resonance images based on natural scene statistics in spatial domain

Yoichiro Ikushima [a,b,*], Shogo Tokurei [b], Hiroyuki Tarewaki [c], Junji Morishita [d], Hidetake Yabuuchi [d]

[a] Department of Health Sciences, Graduate School of Medical Sciences, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, Japan
[b] Department of Radiological Science, Faculty of Health Sciences, Junshin Gakuen University, 1-1-1 Chikushigaoka, Minami-ku, Fukuoka 815-8510, Japan
[c] Division of Radiology, Department of Medical Technology, Osaka University Hospital, 2-15 Yamadaoka, Suita, Osaka 565-0871, Japan
[d] Department of Health Sciences, Faculty of Medical Sciences, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, Japan

ARTICLE INFO

Keywords:
Image-quality assessment
Feature statistics
Brain

ABSTRACT

Background: A magnetic resonance imaging (MRI)-specific objective image quality assessment (IQA) algorithm, the quality evaluation using multidirectional filters for MRI (QEMDIM), was previously reported. QEMDIM requires a set of reference images to calculate the quality score (S_Q) for an assessed image. S_Q may be affected by the quality of the reference set owing to the calculation procedure.

Purpose: To propose a modified version of the IQA algorithm and compare the IQA performance of the original and modified algorithms.

Assessment: Brain axial T_1- and T_2-weighted spin-echo images of varying quality levels (noise and blurring) were acquired from seven healthy men. Subjective IQA (paired comparisons) was conducted on the images, and subjective quality scores were obtained. With reference sets of various quality levels, QEMDIM and modified IQA were applied to the same images that underwent the subjective IQA. The correlation of each S_Q and modified score (S_{mod}) with the subjective scores was evaluated for content-related subsets of assessed images and for each reference set. The effect of the reference-set quality on the distribution of the correlation coefficients (CCs) was statistically evaluated for S_Q and S_{mod} using a one-way analysis of variance test with a significance level of 0.05. We also evaluated the variation in S_{mod} for images with almost the same qualities using the standard deviation (SD).

Results: The CCs of S_Q varied significantly with the quality of the reference set, whereas that of S_{mod} did not. The SD of S_{mod} for almost-same-quality images was less than that corresponding to the confidence interval of the subjective scores.

Conclusion: Our modified algorithm was superior to QEMDIM in terms of IQA performance in clinical practice, especially in terms of accuracy, robustness, and reproducibility.

1. Introduction

Many researchers of magnetic resonance imaging (MRI) have made efforts to develop pulse sequences and/or image reconstruction to shorten the scan time without severe artifacts [1–7]. Such technical developments for shortening the scan time have often sacrificed the image quality in terms of low signal-to-noise ratio (SNR) and spatial resolution, which has led to lower diagnostic accuracy or detectability of some types of disease. Quality degradation usually becomes severe as time passes. To optimize the length of the scan time and/or any parameters regarding pulse sequence and reconstruction, it is important to quantitatively evaluate the quality of clinical magnetic resonance (MR) images.

There are many methods for the image quality assessment (IQA) of medical images. One of the most reliable and frequently used methods is subjective IQA by human observers, such as receiver operating

* Corresponding author at: Department of Health Sciences, Graduate School of Medical Sciences, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, Japan.
E-mail address: yoichiro.ikushima@gmail.com (Y. Ikushima).

https://doi.org/10.1016/j.mri.2022.07.010
Received 24 February 2022; Received in revised form 7 June 2022; Accepted 11 July 2022
Available online 14 July 2022

characteristic analysis [8], paired comparisons [9,10], and alternative forced choices [11,12]. However, subjective IQA has the disadvantages of requiring a long time and large inter- and/or intra-observer variability [13]. Therefore, many automated and objective IQA methodologies without observers have been developed [14–18]. Jang et al. proposed an MRI-specific objective IQA algorithm, quality evaluation using multi-directional filters for MRI (QEMDIM) [18]. QEMDIM requires an assessed image and a set of high-quality reference images to compute the quality score of the assessed image. The quality score is based on the absolute value of the difference in the quality between them, and QEMDIM may be unable to determine whether the assessed image has a higher or lower quality than the reference set because of the absolute score, possibly leading to misinterpretation of the QEMDIM scores by clinicians. To prevent this misinterpretation, the quality of the reference set should be higher than that of the assessed image. However, the preparation of high-quality images is difficult in clinical practice. If the reference set consists of images with a quality similar to or lower than of the assessed image, QEMDIM may not correctly assess the image quality. Therefore, a novel IQA algorithm is required to compensate for the drawbacks of the QEMDIM.

In a report by Jang et al. [18], the QEMDIM algorithm was applied to images that had any one of the various types of distortions, including noise, blurring, and some types of artifacts, which were artificially added to the original high-quality images using image processing. These artificial distortions may differ from those encountered in clinical practice. Moreover, only one type of distortion changes may be rarely observed because most scan and reconstruction parameters in MRI can cause two or more types of distortions [19]. Therefore, it is necessary to validate whether the QEMDIM and our proposed algorithm can correctly assess the quality of images distorted by MRI scan parameters.

This study aimed to propose a novel MRI-specific IQA algorithm by modifying the QEMDIM algorithm and to compare QEMDIM with our modified algorithm in terms of their IQA performance in clinical use. The IQA performance mentioned above includes the correlation with subjective IQA (*i.e.*, accuracy) and variations in the quality score with the quality of the reference set (robustness across the reference qualities) and the quality type of the assessed image (robustness across the quality types assessed). We also evaluated the distribution of the modified quality scores of images with consistent qualities to validate our modified algorithm in terms of the quantity and reproducibility (average and standard deviation [SD] of the distribution).

The remainder of this paper is organized as below. The algorithms of QEMDIM and our modified method are described and compared in Section 2. Section 3 consists of four subsections that explain the procedures for acquiring MR images, calculating subjective and objective IQ scores, and analyzing these scores. The analytical results are shown in Section 4. We describe our novel findings, the reasons underlying our results, and the limitations of this study in Section 5. Finally, we present some concluding remarks in Section 6.

2. Theory

The QEMDIM algorithm is based on natural scene statistics in the spatial domain [16,18]. In the algorithm (Fig. 1), an MR image is transformed into a map of the mean-subtracted contrast-normalized (MSCN) coefficients (MSCN map) [Eq. (1)] to extract quality-related features. Specifically, a given pixel value at the x-th column and y-th row, $I(x,y)$, on an MR image with H rows and W columns is translated to the MSCN coefficient, $\widehat{I}(x,y)$, using Eq. (1):

$$\widehat{I}(x,y) = \frac{I(x,y) - \mu(x,y)}{\sigma(x,y) + C} \tag{1}$$

where C is the stabilizing factor of 0.2, and $\mu(x,y)$ and $\sigma(x,y)$ are the mean and SD, respectively, calculated from the pixel values around $I(x, y)$ as follows:

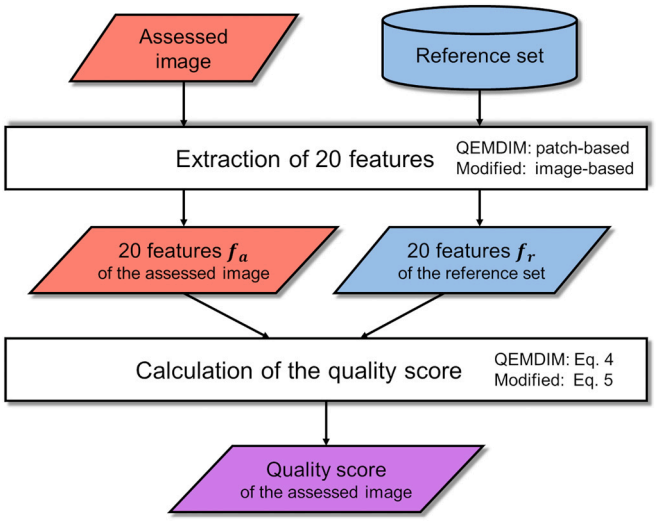

Fig. 1. Flowchart of the previously-reported algorithm (QEMDIM) and our modified IQA algorithm. The differences between the QEMDIM and modified algorithms were the "Extraction of 20 features" and "Calculation of the quality score," which are denoted as "QEMDIM:" and "Modified:."
QEMDIM: Quality Evaluation using Multi-Directional Filters for MRI.

$$\mu(x,y) = \sum_{k=-K}^{K}\sum_{l=-L}^{L} w(k,l)\, I(x+k, y+l) \tag{2}$$

$$\sigma(x,y) = \sqrt{\sum_{k=-K}^{K}\sum_{l=-L}^{L} w(k,l)\, [I(x+k, y+l) - \mu(x,y)]^2} \tag{3}$$

where w is a two-dimensional Gaussian-weighting function with a kernel size of $(2K+1) \times (2L+1)$ and SD of 1.167 pixels. K and L were set to 3. The QEMDIM algorithm applies four types of multidirectional filters to an MSCN map. Each of the original and four filtered MSCN maps were divided into 16 square-shaped image patches with a size of $H/4 \times W/4$ without overlapping. From each patch, a histogram of the MSCN coefficients (MSCN histogram) was drawn. The shape of an MSCN histogram varies depending on image quality [16,18]. Each of the MSCN histograms was fitted to the generalized Gaussian distribution (GGD) [20] to extract two features, α and β, representing the shape of the histogram. The above processes are also performed for the scaled MR image that is generated by applying low-pass filtering and downsampling by a factor of two to the original MR image. Twenty features (two GGD-related features × five MSCN maps × two scaled images) were extracted from each patch of the MR image. For an assessed image, each of the 20 types of features was averaged over the 16 patches, which were used as the "20 features $\boldsymbol{f}_\mathbf{a}$ of the assessed image," as shown in in Fig. 1. For the reference set, each of the 20 types of features was averaged over all patches of all the images included in the set after 10% outliers at the top and bottom of the feature distribution were excluded, which were the "20 features $\boldsymbol{f}_\mathbf{r}$ of the reference set," as shown in Fig. 1. Finally, the QEMDIM score S_Q representing the quality of the assessed image was calculated as follows:

$$S_Q = \sqrt{(\boldsymbol{f}_\mathbf{r} - \boldsymbol{f}_\mathbf{a})^T\, (\boldsymbol{f}_\mathbf{r} - \boldsymbol{f}_\mathbf{a})} \tag{4}$$

where $\boldsymbol{f}_\mathbf{r}$ and $\boldsymbol{f}_\mathbf{a}$ are the 20-dimensional (20D) vectors of the features of the reference set and assessed image, respectively. According to Eq. (4), the S_Q means the Euclidean distance between the $\boldsymbol{f}_\mathbf{r}$ and $\boldsymbol{f}_\mathbf{a}$ vectors in the 20D space. In Fig. 2, the S_Q can be depicted as the distance between the gray square and open circle for the high-quality assessed image and the distance between the gray square and open triangle for the low-quality assessed image. The two S_Qs in Fig. 2 may have equal values, despite their different image qualities.

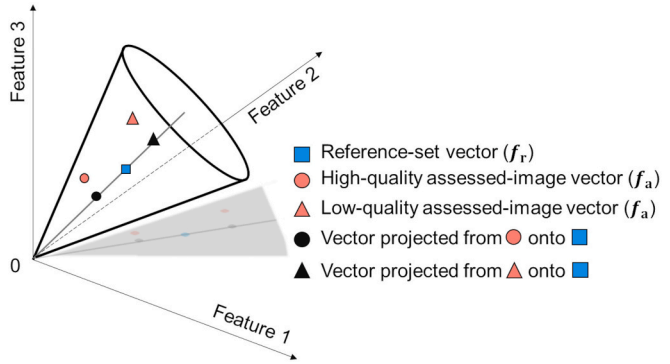

Fig. 2. Schematic illustration of the reference set vectors (a blue square), high- and low-quality assessed images (a red circle and triangle), and their projections onto the reference-set vector (a black circle and triangle). The three-dimensional coordinate is depicted here for simplicity, although the actual vectors have 20 dimensions. The cone depicts the range where a vector is likely to exist. The gray line inside the cone is perpendicular to the base circle of the cone and parallel to the reference-set vector. The gray triangle and the dots and line inside it represent the shadows of the cone and plots. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Based on the report of Jang et al. [18], we assumed that the norm or many elements of a 20-feature vector of an image would reach zero as the image quality increased and would reach a high or stable value when the image quality decreased. This implies that the data point of a 20-feature vector of a high-quality image would be close to the origin of the coordinate in 20D space, and that of a low-quality image would be far from the origin within a range whose shape was similar to that of a relatively narrow cone with its tip at the origin (Fig. 2). Therefore, under the hypothesis that the vector of a reference set would be parallel to a vector normal to the base circle of the cone, we believe that a useful feature for IQA may be the distance from the origin to a data point projected from the assessed image vector onto the reference vector (closed circle and closed triangle in Fig. 2). Under these assumptions, to compensate for the drawback of QEMDIM, which may not correctly assess the image quality, as described in the previous paragraph, we proposed a new quality score S_{mod} as follows:

$$S_{mod} = \|\boldsymbol{f}_r\| - \frac{\boldsymbol{f}_a \cdot \boldsymbol{f}_r}{\|\boldsymbol{f}_r\|} \tag{5}$$

where $\|\boldsymbol{f}_r\|$ is the norm of $\boldsymbol{f}_r$ and "·" is the inner product operator. The $\|\boldsymbol{f}_r\|$ is depicted in Fig. 2 as the distance from the origin to the gray square, and the $\frac{\boldsymbol{f}_a \cdot \boldsymbol{f}_r}{\|\boldsymbol{f}_r\|}$ is depicted as the distance from the origin to the closed circle for the high-quality assessed image and distance from the origin to the closed triangle for the low-quality assessed image (Fig. 2). When the projected vector of the assessed image (closed circle or closed triangle in Fig. 2) is closer to the origin than the reference set vector, the S_{mod} shows a positive number, which means that the assessed image has a higher quality than the reference set (and *vice versa*). The S_{mod} of zero represents the quality of the assessed image, which is the same as that of the reference set.

We made two modifications to the QEMDIM algorithm (Fig. 1). First, the algorithm calculates the quality score S_{mod} based on Eq. (5). Second, our algorithm does not divide the MR images into patches. In the QEMDIM algorithm, 20 types of features were obtained by averaging each feature over 16 patches. In contrast, in our modified algorithm, 20 features are obtained directly from an image because drawing of the MSCN histograms and extraction of the α and β values are performed on the entire region of the image. The calculation efficiency is expected to improve by removing the process of dividing the patches and averaging the features. The patch-based algorithm in QEMDIM is intended to overestimate the quality of the reference set by excluding outliers because, as mentioned before, the QEMDIM algorithm requires a high-quality reference set. Therefore, image-based calculations will work well as our algorithm does not require high-quality reference images.

3. Materials and methods

This study comprised MR image acquisition and image observation sessions that required cooperation from the participants. This study was approved by the institutional review board, and written informed consent was obtained from all participants.

3.1. MR image acquisition and reconstruction

A 3.0-T MR scanner (SIGNA™ Architect, GE Healthcare, Milwaukee, WI, USA) and a 48-channel head coil were used in this study. A total of seven healthy volunteers (labeled Vol_1–Vol_7) participated in the MR acquisition experiment. The average and SD of their ages were 30.6 ± 3.8 (range, 25–38), and all of them were males. Five (Vol_1–Vol_5) of the seven volunteers underwent eight sets of T_1-weighted (T_1W) brain axial two-dimensional (2D) imaging, and five volunteers (Vol_1–Vol_3, Vol_6, and Vol_7) underwent two sets of T_2-weighted (T_2W) brain axial 2D imaging. The number of the sets differing between the T_1W and T_2W imaging was determined not to exceed the scan duration of 1 h. The imaging coverage was set to include the centrum semiovale and the upper edge of the eyeballs. To minimize the intra-volunteer variability in image quality among the multiple acquisitions, all k-spaces were concurrently filled with the MR signals. This was achieved by modifying the vendor-supplied pulse sequence using the environment for pulse programming in C (EPIC version DV28, GE Healthcare, Milwaukee, WI, USA). The scan parameters which differed between the T_1W and T_2W images were repetition time (TR) and echo time (TE): TR/TE = 545 ms/12 ms for T_1W imaging and 3500 ms/80 ms for T_2W imaging. The other similar parameters for T_1W and T_2W imaging were as follows: spin-echo pulse sequence (a single echo per TR); Cartesian trajectory; number of excitations (NEX), 1; field of view (FOV) in the frequency-encoding (FE) direction, 66 cm; FOV in the phase-encoding (PE) direction, 22 cm; slice thickness, 2 mm; spacing, 2 mm; FE direction, anterior-posterior; number of slices, 12; number of FE steps (Nx), 768; number of PE steps (Ny), 256; and receive bandwidth (BW), ±250 kHz. We did not utilize the parallel imaging reconstruction [3–5] and fast spin echo sequence [1] for simplicity, although these techniques have been frequently used to shorten scan duration in clinical practice. It is thought to be difficult to precisely predict the quality of images acquired using these techniques [21,22]. Aiming to make the image quality predictable, these techniques were not adopted in this study.

From each of the original k-spaces with a rectangular matrix size (Nx × Ny = 768 × 256) and NEX = 1, three k-spaces with a square shape (256 × 256) were generated so that the p-th square-shaped k-space consisted of the signals stored in $(3x + p)$-th column of the rectangular k-space, where $p = \{1, 2, 3\}$ and $x = \{0, 1, 2, \cdots, 255\}$. We can regard the three square-shaped k-spaces as those that were acquired at different temporal phases whose time intervals were zero and that had the same image quality and image content [23]. Then, square-shaped k-spaces of NEX = n were generated by averaging the n k-spaces that had the same content (*i.e.*, the same volunteer and cross-section), where n = {1, 2, 4, 8} for T_1W k-spaces and n = {1, 2} for T_2W k-spaces. In addition, the filling rate of the square-shaped k-spaces in the PE direction (FR_{Ky}) varied at four levels: 50% (Ny = 128), 60% (Ny = 154), 75% (Ny = 192), and 100% (Ny = 256), which were implemented by padding zeros to the high-frequency regions of the k-spaces. This process simulates under-sampling in the PE direction and zero-fill interpolation processing. A low FR_{Ky} indicates low spatial resolution in the PE direction. After applying the elliptic Fermi filter, each square-shaped k-space was inversely Fourier-transformed to obtain magnitude images. All reconstructions were performed using Orchestra toolbox (GE Healthcare, Milwaukee, WI, USA). Finally, T_1W images were categorized into 16 types of image

quality, and T_2W images were categorized into eight types of quality (columns in Fig. 3); each of the quality types consisted of 180 images (Fig. 3).

3.2. Subjective IQA

Thirteen radiologic technologists (Obs_1–Obs_{13}) participated in this observation session. The average and SD of their ages were 34.5 ± 6.3 (range, 28–48), and those of their years of experience as radiologic technologists were 9.4 ± 5.2 (range, 4–23). Out of 13, one was female, and the others were male. A total of 16 T_1W and eight T_2W images at the level of the basal ganglia of Vol_1 were used in the observation (Fig. 4). These images were of the same phase, and each had a different quality than the others (Figs. 3 and 4). The Ura-modified Schéffe method [24], which is a paired comparison method, was used in this study to comprehensively evaluate the subjective image quality of each MR image. The observation session consisted of T_1W and T_2W sub-sessions, and the 13 observers participated in either or both of the sub-sessions so that ten observers were assigned to each sub-session (Obs_1–Obs_{10} in the T_1W sub-session and Obs_1–Obs_7 and Obs_{11}–Obs_{13} in the T_2W sub-session). In the observation session, two images of different qualities were simultaneously displayed on a display monitor in a horizontal orientation, and the observers were asked to evaluate the relative and comprehensive image quality on a three-point scale (the left image is better, they are similar, or the right one is better). The value resulting from a pair of left image l and right image r evaluated by an observer o was defined as the preference $\alpha_{lro} \in \{1, 0, -1\}$, where $l \in \{1, 2, \cdots, N_{Im} \mid l \neq r\}$, $r \in \{1, 2, \cdots, N_{Im} \mid r \neq l\}$, $o \in \{1, 2, \cdots, N_{Ob}\}$, N_{Im} and N_{Ob} denote the numbers of images and observers in the sub-session, respectively. The $\alpha_{lro} = 1$ means the "left one is better." The three-point scale evaluations were repeated for 240 image pairs ($_{16}P_2$) in the T_1W sub-session and 56 image pairs ($_8P_2$) in the T_2W sub-session. The order of display of the image pairs was randomized among the observers to remove order effects. The observers were not allowed to adjust the luminance, contrast, or magnification factor of the displayed images. The viewing distance and time were arbitrary. The observational procedure was instructed before the observation session using several pairs of images extracted from the images used in the observation session. Finally, the preference for the i-th image $\widehat{\alpha}_i$ was calculated as follows [24]:

$$\widehat{\alpha}_i = \frac{\sum_{o=1}^{N_{Ob}} \sum_{r=1}^{N_{Im}} \alpha_{iro} - \sum_{o=1}^{N_{Ob}} \sum_{l=1}^{N_{Im}} \alpha_{lio}}{2N_{Im}N_{Ob}} \tag{6}$$

The $\widehat{\alpha}_i$ ranging within [−1, 1] was regarded as the subjective image quality score in this study, and a high score indicated a good quality of the i-th image. One-way analysis of variance (ANOVA) was conducted to evaluate the effect of image quality on subjective quality scores. Moreover, as an index to determine whether there is a significant difference in the subjective quality score between any two images with a significance level ϕ, we calculated the yardstick value Y_ϕ as follows:

$$Y_\phi = q_\phi \sqrt{\widehat{\sigma}^2 / (2N_{Ob}N_{Im})} \tag{7}$$

where q_ϕ is the studentized range and $\widehat{\sigma}^2$ is the unbiased variance of errors computed in the ANOVA [25]. Two quality scores ($\widehat{\alpha}_i$ and $\widehat{\alpha}_j$) are regarded as significantly different if $|\widehat{\alpha}_i - \widehat{\alpha}_j| \rangle Y_\phi$ is satisfied. The statistical significance levels of the ANOVA and yardstick analyses were set at 0.05.

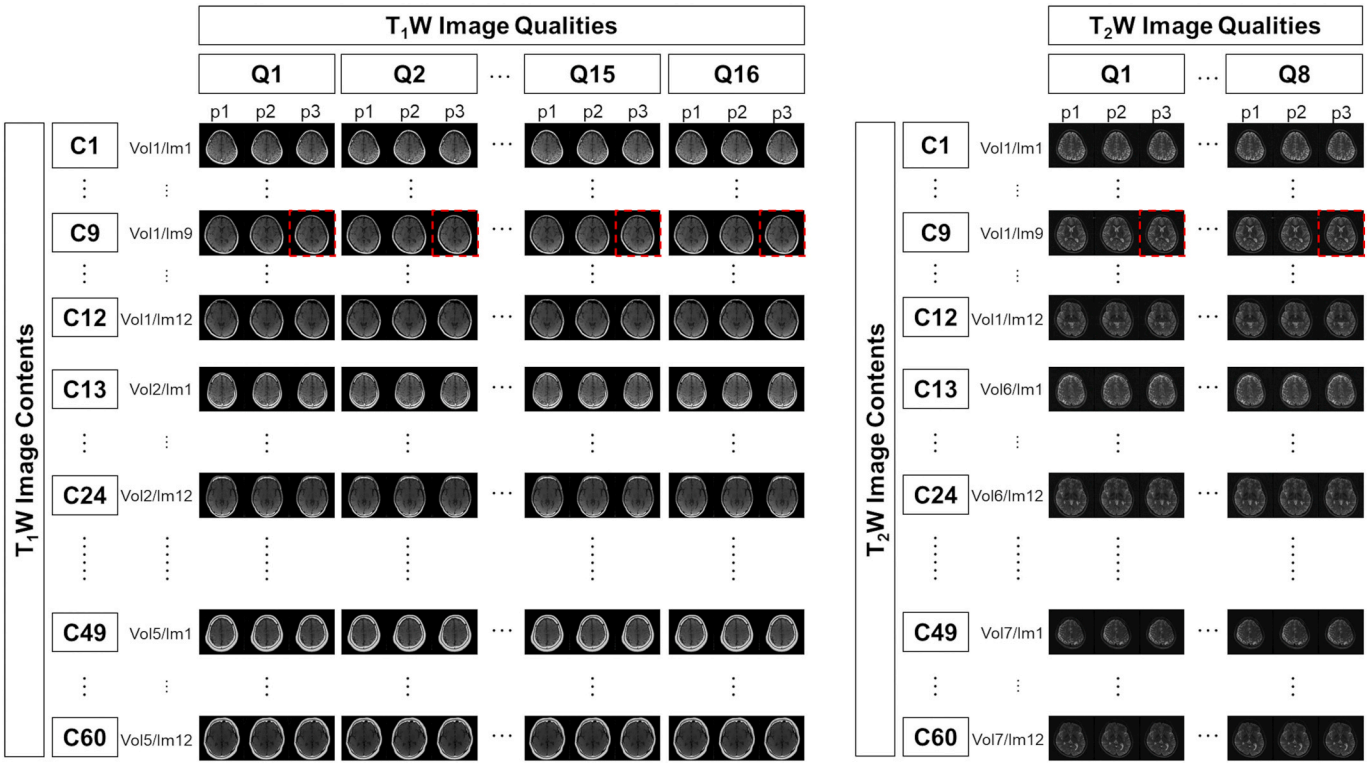

Fig. 3. Types of images reconstructed in this study. The left-hand side is a set of T_1W images and the right-hand side is a set of T_2W images. The 180 images on the same column have the same quality (NEX and FR_{Ky}), and 16 and eight types of qualities were prepared for T_1W and T_2W images, respectively (denoted by "Q" on the top of the figure). Each row consists of images with the same content (subject and cross section), and 60 types of content were acquired (denoted by "C" on the left of the figure). The "p" just above the "C1" images represents the temporal phase. The "Vol" and "Im" immediate left of the "Q1" images denote the volunteer (subject) and image (cross section). The red dashed rectangles denote the images used in the observation, as shown in Fig. 4. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

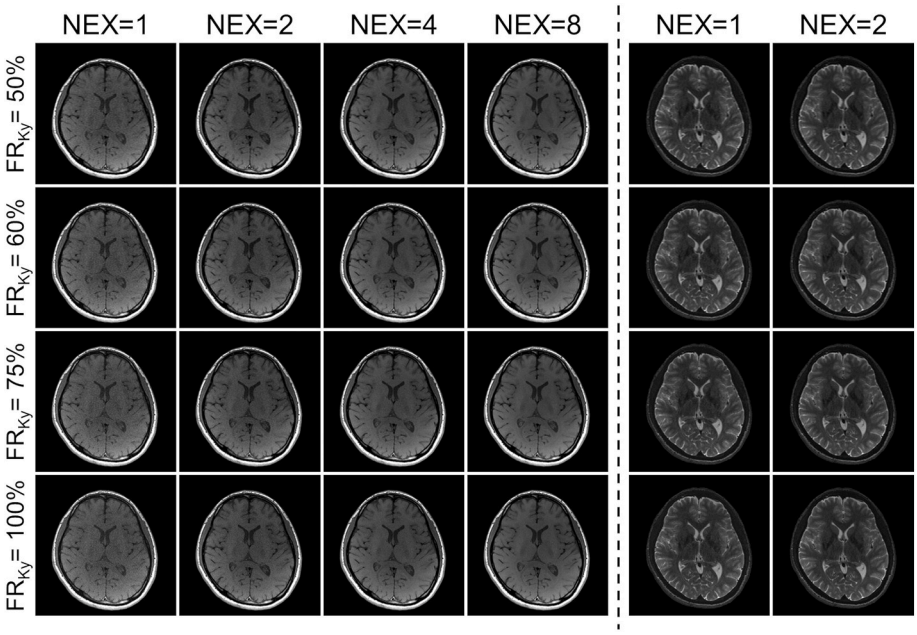

Fig. 4. Images used in the observation. The images in the left of the vertical dashed line are T_1W images, and those of the right are T_2W images. Each column contains images with the same NEX, and each row contains images with the same FR_{Ky}.
NEX: number of excitations.
FR_{Ky}: filling rate of k-space in phase-encoding direction.

3.3. Comparison of IQA performance between the previous and our algorithms

The reconstructed T_1W and T_2W images were divided into 16 T_1W and eight T_2W reference sets based on their quality (columns in Fig. 3). Feature vector $\boldsymbol{f_r}$ (Fig. 1) was extracted from each reference set using an image-based modified algorithm. We defined $\boldsymbol{f_{r,Q1}}$, $\boldsymbol{f_{r,Q2}}$, ⋯, and $\boldsymbol{f_{r,Q16}}$ for the T_1W reference sets and $\boldsymbol{f_{r,Q1}}$, ⋯, and $\boldsymbol{f_{r,Q8}}$ for the T_2W reference sets, each corresponding to the qualities of Q1, Q2, … (Fig. 3).

The feature vector $\boldsymbol{f_a}$ of an assessed image was extracted from each of all the 4320 reconstructed images (Fig. 3) using the image-based modified algorithm (Fig. 1). Every $\boldsymbol{f_a}$ was used to calculate the 16 or eight quality scores S_{mod}s (Fig. 1 and Eq. (5)), each corresponding to the score calculated using the $\boldsymbol{f_{r,Q1}}$, $\boldsymbol{f_{r,Q2}}$, ⋯. Note that in the calculation of the scores, the contrast (T_1W or T_2W) of the reference sets was kept similar to that of the assessed image. Eventually, 16 and eight S_{mod}s were assigned to each T_1W and T_2W assessed images, respectively.

The reconstructed T_1W and T_2W assessed images, each including 16 and eight S_{mod}s, respectively, were divided into 60 assessed subsets according to the image contents (rows in Fig. 3). Each assessed subset consisted of 48 T_1W images (16 quality × three phases) or 24 T_2W images (eight quality × three phases). Among images belonging to the same assessed subset, the variation in image quality was based only on NEX and FR_{Ky}. We assumed that the order and difference in the true quality (subjective quality score $\widehat{\alpha}_i$) among the images of the same subset should be the same as that of the images used in the observation (Fig. 4). Under this assumption, the correlation coefficient (CC) was calculated between "the $\widehat{\alpha}_i$s obtained for the images in Fig. 4" and "the S_{mod}s of images in an assessed subset" for each reference set and each assessed subset. The Pearson linear correlation coefficient (PLCC), Spearman rank-ordered correlation coefficient (SROCC), and Kendall rank-ordered correlation coefficient (KROCC) were adopted as CCs [26–28]. Furthermore, the QEMDIM score S_Qs [Eq. (4)] and their CCs with $\widehat{\alpha}_i$s were calculated in the same manner as described above. The distribution of CCs between S_Qs and $\widehat{\alpha}_i$s ($CC_{Q,\alpha}$ including $PLCC_{Q,\alpha}$, $SROCC_{Q,\alpha}$, and $KROCC_{Q,\alpha}$) was compared to that of CCs between S_{mod}s and $\widehat{\alpha}_i$s ($CC_{mod,\alpha}$ including $PLCC_{mod,\alpha}$, $SROCC_{mod,\alpha}$, and $KROCC_{mod,\alpha}$) for each reference set with various qualities. The Kruskal–Wallis test was used to analyze the effect of the quality of the reference set on the $PLCC_{Q,\alpha}$, $SROCC_{Q,\alpha}$, $KROCC_{Q,\alpha}$, $PLCC_{mod,\alpha}$, $SROCC_{mod,\alpha}$, and $KROCC_{mod,\alpha}$. The statistical significance level was set at a p-value of 0.05.

3.4. Quantity and reproducibility of our algorithms

All the reconstructed images were divided into 60 × 16 T_1W and 60 × 8 T_2W reference sets according to their qualities and image content; each of which consisted of three images with the same quality and content but different phases (Fig. 3). S_{mod}s were calculated for all the reconstructed images with the reference set to which the assessed image belonged. Since each assessed image theoretically had the same quality and content as the corresponding reference set, all S_{mod}s would ideally indicate zeros. We evaluated the normality, average, and SD of the distribution constructed using all S_{mod}s. The Shapiro-Wilk test was used for normality evaluation.

4. Results

ANOVA demonstrated that the subjective score $\widehat{\alpha}$ significantly varied with the image quality ($p < 0.001$ for among the T_1W images and among the T_2W images in Fig. 4). The yardstick values $Y_{0.05}$ were 0.130 for the T_1W images and 0.184 for the T_2W images.

4.1. Comparison of IQA performance between the previous and our algorithms

The distributions of 60 $PLCC_{Q,\alpha}$s and 60 $PLCC_{mod,\alpha}$s for each of the reference sets with different qualities are shown in Fig. 5. The distributions of SROCCs and KROCCs is not shown in this study because of their similarity to those of PLCCs. Tables 1 and 2 show the medians and interquartile ranges (IQRs) of the distributions of $CC_{mod,\alpha}$ and $CC_{Q,\alpha}$ for each quality of the reference set. For $CC_{mod,\alpha}$s, including $PLCC_{mod,\alpha}$, $SROCC_{mod,\alpha}$, and $KROCC_{mod,\alpha}$ (Fig. 5b and d and Table 1), the variation in the medians and IQRs caused by the qualities of the reference set was no more than 0.02. These variations with various qualities were not statistically significant ($p = 1.00$ for all $CC_{mod,\alpha}$s). For $CC_{Q,\alpha}$s (Fig. 5a and c and Table 2), the medians and IQRs significantly varied in up to 1.85 and 0.63, respectively ($p = 0.00$ for all $CC_{Q,\alpha}$s). In particular, the IQRs were wide, and the medians were close to zero when the reference set had an intermediate quality (NEX/FR_{Ky} = 2/100% or 4/60% for T_1W

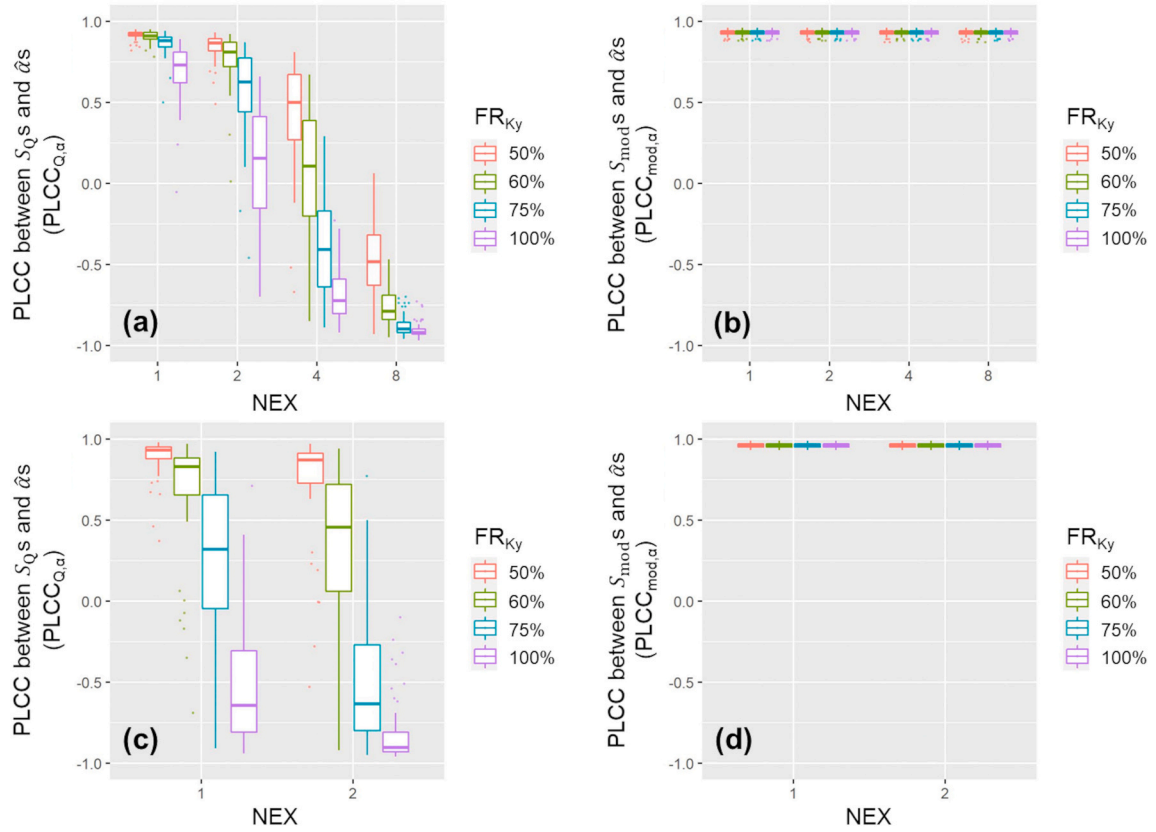

Fig. 5. Boxplots representing the variation in the distribution of PLCC with the quality of the reference sets. The figures on the top row (**a** and **b**) and bottom row (**c** and **d**) denote the results of T_1W and T_2W images, respectively. The Y-axes are $PLCC_{Q,\alpha}$ for **(a)** and **(c)** and $PLCC_{mod,\alpha}$ for **(b)** and **(d)**. The X-axes depict the NEX of the reference set, and the colors of the plots depict the FR_{Ky} of the reference set. Each box plot consists of 60 data points that are the same as the number of the image contents (rows in Fig. 3). The height of each box indicates the interquartile range (IQR), and the horizontal line in each box represents the median of the distribution. The dot plots indicate the outliers defined as data points higher than the 75th percentile +1.5 times the IQR or lower than the 25th percentile – 1.5 times the IQR.
NEX: number of excitations.
FR_{Ky}: filling rate of a k-space in phase-encoding direction.
PLCC: Pearson linear correlation coefficient.
S_Q: objective quality score obtained from the QEMDIM algorithm.
S_{mod}: objective quality score obtained from the modified algorithm.
$\hat{\alpha}$: subjective quality score.

images and 1/75% or 2/60% for T_2W images). It is visually apparent that the variation in the CC distributions with the various qualities of the reference set was much smaller for the modified scores than for the QEMDIM scores (Fig. 5).

Representative scatter plots depicting the correlations between the modified scores (S_{mod}s) and the subjective scores ($\hat{\alpha}_i$s) and between the QEMDIM scores S_Qs and $\hat{\alpha}_i$ are shown in Fig. 6. Each of the 48 S_{mod}s was equally and parallelly shifted by the quality of the reference set (Fig. 6d–f). In other words, the difference in S_{mod} between any two images was consistent irrespective of the quality of the reference set. The slope of the linear regression equation was 2.17 for Fig. 6d and 2.18 for Fig. 6e and f, respectively, and the average ± SD of the slopes of all the combinations of the assessed subset and reference set, that is, all plots in Fig. 5, were 2.38 ± 0.403 for T_1W images and 3.28 ± 0.314 for T_2W images. Fig. 6a–c show that despite the same assessed images as in Fig. 6d–f, the difference in S_Q between any two images largely varied with the qualities of the reference set. There was a positive correlation for the low-quality reference set (Fig. 6a), negative correlation for the high-quality reference set (Fig. 6c), and almost no correlation for the intermediate-quality reference set (Fig. 6b).

4.2. Quantity and reproducibility of our algorithms

We obtained a histogram of the 4320 S_{mod}s, which was calculated for the 4320 assessed images with the corresponding reference set that consisted of three images with the same quality and content as the assessed image. The histogram shows a Gaussian distribution (p = 0.478). The average ± SD was $-1.67 \times 10^{-17} \pm 0.0262$.

5. Discussion

The QEMDIM scores (S_Qs) significantly varied with the quality of the reference set, even if the assessed images used were same among the different reference sets (Fig. 6a–c). The median and IQR of the $CC_{Q,\alpha}$ also varied significantly because of the quality of the reference set (Fig. 5a and c and Table 2). In particular, when the quality of the reference set was intermediate, the medians were close to zero, and the IQRs were wide. These results imply that the QEMDIM algorithm is sensitive to the quality of the reference set, which is in line with the hypothesis described in the Theory section (Fig. 2 and Eq. (4)). The reason why the S_Qs showed a positive correlation with $\hat{\alpha}$ for the low-quality reference set (Fig. 6a) is that all 48 assessed images had higher quality than the reference set. Therefore, even the lowest-quality assessed image had a higher quality than the reference set and a small distance from the reference set in the 20D feature space, which led to a lower S_Q than that of the high-quality images. Similarly, a high- or intermediate-quality assessed image had a higher quality than the reference set and a large or intermediate distance from the reference set, which led to a high or

Table 1
The medians of correlation coefficients (PLCC/SROCC/KROCC) between the subjective and modified scores calculated using the reference sets with various qualities (different NEX in the rows and FR_{Ky} in the columns). The parenthesized values next to the medians are the interquartile ranges.

Image Contrast	NEX	FR_{Ky} [%]			
		50	60	75	100
T_1W	1	.93 (.02)/ .94 (.02)/ .82 (.03)	.93 (.02)/ .95 (.01)/ .82 (.03)	.93 (.02)/ .95 (.01)/ .82 (.03)	.93 (.02)/ .95 (.01)/ .83 (.03)
	2	.93 (.02)/ .94 (.02)/ .82 (.04)	.93 (.02)/ .95 (.01)/ .82 (.03)	.93 (.02)/ .95 (.01)/ .82 (.03)	.93 (.02)/ .95 (.01)/ .83 (.03)
	4	.93 (.02)/ .94 (.02)/ .82 (.03)	.93 (.02)/ .94 (.02)/ .82 (.03)	.93 (.02)/ .95 (.01)/ .82 (.03)	.93 (.02)/ .95 (.01)/ .82 (.03)
	8	.93 (.02)/ .94 (.02)/ .82 (.03)	.93 (.02)/ .94 (.02)/ .82 (.03)	.93 (.02)/ .95 (.01)/ .82 (.03)	.93 (.02)/ .95 (.01)/ .82 (.03)
T_2W	1	.96 (.02)/ .96 (.01)/ .86 (.05)	.96 (.02)/ .96 (.01)/ .86 (.05)	.96 (.02)/ .96 (.02)/ .86 (.05)	.96 (.02)/ .96 (.02)/ .86 (.05)
	2	.96 (.02)/ .96 (.01)/ .86 (.05)	.96 (.02)/ .96 (.01)/ .86 (.05)	.96 (.02)/ .96 (.02)/ .86 (.05)	.96 (.02)/ .96 (.02)/ .86 (.05)

PLCC: Pearson linear correlation coefficient.
SROCC: Spearman rank-order correlation coefficient.
KROCC: Kendall rank-order correlation coefficient.
FR_{Ky}: filling rate of square-shaped k-space data in phase-encoding direction.
NEX: the number of excitations.

Table 2
The medians of correlation coefficients (PLCC/SROCC/KROCC) between the subjective and QEMDIM scores calculated using the reference sets with various qualities (different NEX in the rows and FR_{Ky} in the columns). The parenthesized values next to the medians are the interquartile ranges.

Image Contrast	NEX	FR_{Ky} [%]			
		50	60	75	100
T_1W	1	.92 (.02)/ .94 (.02)/ .80 (.04)	.91 (.04)/ .93 (.04)/ .78 (.06)	.88 (.06)/ .89 (.08)/ .72 (.09)	.73 (.19)/ .72 (.22)/ .54 (.19)
	2	.87 (.07)/ .88 (.09)/ .72 (.10)	.81 (.15)/ .81 (.17)/ .63 (.19)	.63 (.33)/ .60 (.34)/ .44 (.28)	.16 (.57)/ .16 (.50)/ .11 (.37)
	4	.50 (.40)/ .44 (.43)/ .32 (.33)	.10 (.59)/ .08 (.54)/ .06 (.38)	-.41 (.47)/ -.33 (.45)/ -.26 (.33)	-.73 (.22)/ -.69 (.30)/ -.52 (.25)
	8	-.49 (.31)/ -.46 (.35)/ -.32 (.25)	-.79 (.15)/ -.76 (.19)/ -.59 (.18)	-.90 (.06)/ -.91 (.06)/ -.75 (.09)	-.92 (.03)/ -.94 (.03)/ -.80 (.05)
T_2W	1	.93 (0.07)/ .93 (0.05)/ .80 (0.09)	.83 (.23)/ .84 (.28)/ .69 (.28)	.32 (.70)/ .32 (.67)/ .22 (.52)	-.65 (.50)/ -.54 (.54)/ -.40 (.49)
	2	.87 (.18)/ .89 (.18)/ .73 (.23)	.46 (.66)/ .44 (.66)/ .33 (.52)	-.64 (.53)/ -.51 (.55)/ -.36 (.47)	-.91 (.12)/ -.89 (.16)/ -.74 (.21)

PLCC: Pearson linear correlation coefficient.
SROCC: Spearman rank-order correlation coefficient.
KROCC: Kendall rank-order correlation coefficient.
FR_{Ky}: filling rate of square-shaped k-space data in phase-encoding direction.
NEX: the number of excitations.

intermediate S_Q. For the high-quality reference set (Fig. 6c), all the assessed images had lower quality than the reference set, and the difference in quality between the assessed image and the reference set increased as the quality of the assessed image decreased, which resulted in a negative correlation between the S_Qs and $\hat{\alpha}$s. For the intermediate-quality reference set (Fig. 6b), the difference in quality between the assessed image and the reference set increased as the quality of the assessed image differed greatly from that of the reference set, which resulted in almost no correlation. On the other hand, the modified scores (S_{mod}s) of our modified IQA algorithm showed a consistently high correlation with $\hat{\alpha}$s, irrespective of the quality of the reference set (Fig. 5b and d and Table 1), which implies that our modified IQA algorithm had high accuracy and robustness across the reference quality in terms of IQA performance. In clinical practice, collecting many high-quality images for a reference set is difficult, and an assessed image is often superior to the reference set when a newly introduced machine or technology is used for the assessed images. Therefore, our modified IQA algorithm is more suitable for clinical use than the QEMDIM algorithm owing to its excellent IQA performance.

When an assessed image has almost the same quality as the reference set, S_{mod} is assumed close to zero (Eq. (5)). We evaluated the distribution of S_{mod} under such conditions, and the average ± SD of S_{mod} was $-1.67 \times 10^{-17} \pm 0.0262$ in this study. The average value was very close to zero compared to the range in S_{mod} of approximately 0.8 (Fig. 6), which was observed among the T_1W images used in the observational experiment. Therefore, our modified IQA algorithm is considered highly quantitative. SD is thought to rely on both, the intra-measurement variability of the MR signal and the detection limit of the IQA algorithm, which is inevitable. The SD value can be a useful index for determining if the difference in quality between two assessed images when using the same reference set is significant. Owing to both, the SD of 0.0262 and the normality of the S_{mod} distribution, a S_{mod} difference of up to 0.0514 (1.96 × SD) can occur with a 95% probability, even if the two assessed images have almost the same quality. A S_{mod} difference of approximately 0.05 was observed between the images with NEX = 8 and FR_{Ky} = 50% and those with NEX = 2 and FR_{Ky} = 100% (Fig. 4), and the $\hat{\alpha}$ difference between these images was 0.156. The slope of the linear regression equation can predict $\hat{\alpha}$ from S_{mod}. Its value was 2.38 ± 0.403 for T_1W images and 3.28 ± 0.314 for T_2W images on average ± SD; therefore, a S_{mod} difference of 0.0514 that can occur for two almost same-quality images should correspond to an $\hat{\alpha}$ difference of 0.122 ± 0.0207 for T_1W images and 0.169 ± 0.016 for T_2W images. These $\hat{\alpha}$ differences that can occur for two almost same-quality images are equivalent to and less than the yardstick value $Y_{0.05}$ of 0.130 for T_1W images and 0.184 for T_2W images. This implies that our modified algorithm can have IQA reproducibility and sensitivity equivalent to those of subjective IQA.

In this study, the two types of image quality were concurrently changed using the scan and reconstruction parameters. This quality is assumed to be similar to that seen in the clinical situation. Our modified quality scores showed a high correlation with subjective quality scores in this situation. Thus, our modified algorithm has high robustness across the quality types of assessed images. Such a high correlation was observed in both T_1W and T_2W images, implying that our modified algorithm is insensitive to the contrast of the assessed images. As the high IQA performance proved the above, our modified IQA algorithm has the potential to be used as an alternative to subjective IQA.

Our modified IQA algorithm computes the S_{mod} of an assessed image based on comparison with a reference set (Eq. (5)), and S_{mod}s exhibited a high correlation with the $\hat{\alpha}$, irrespective of the quality of the reference set (Fig. 5 and Table 1). Furthermore, our modified IQA has equivalent sensitivity and reproducibility to subjective IQA in terms of the subtle difference in quality between the two images. Therefore, in clinical practice, S_{mod}s could be a criterion that MR operators use to determine if a retake of the MR image is needed without the subjective IQA before the examination has finished. In this case, the reference set needs to be composed of the images previously taken for the same purpose or patient, which will cause S_{mod} to be lower than zero if the retake is needed, that is, the quality of the assessed image is lower than that of the reference set. Furthermore, it may be useful for MR researchers to use our modified algorithm to compare their developing methodology with the conventional one with respect to image quality without human observers. In this case, the reference set should be composed of conventional images. In addition, our modified IQA algorithm could be helpful

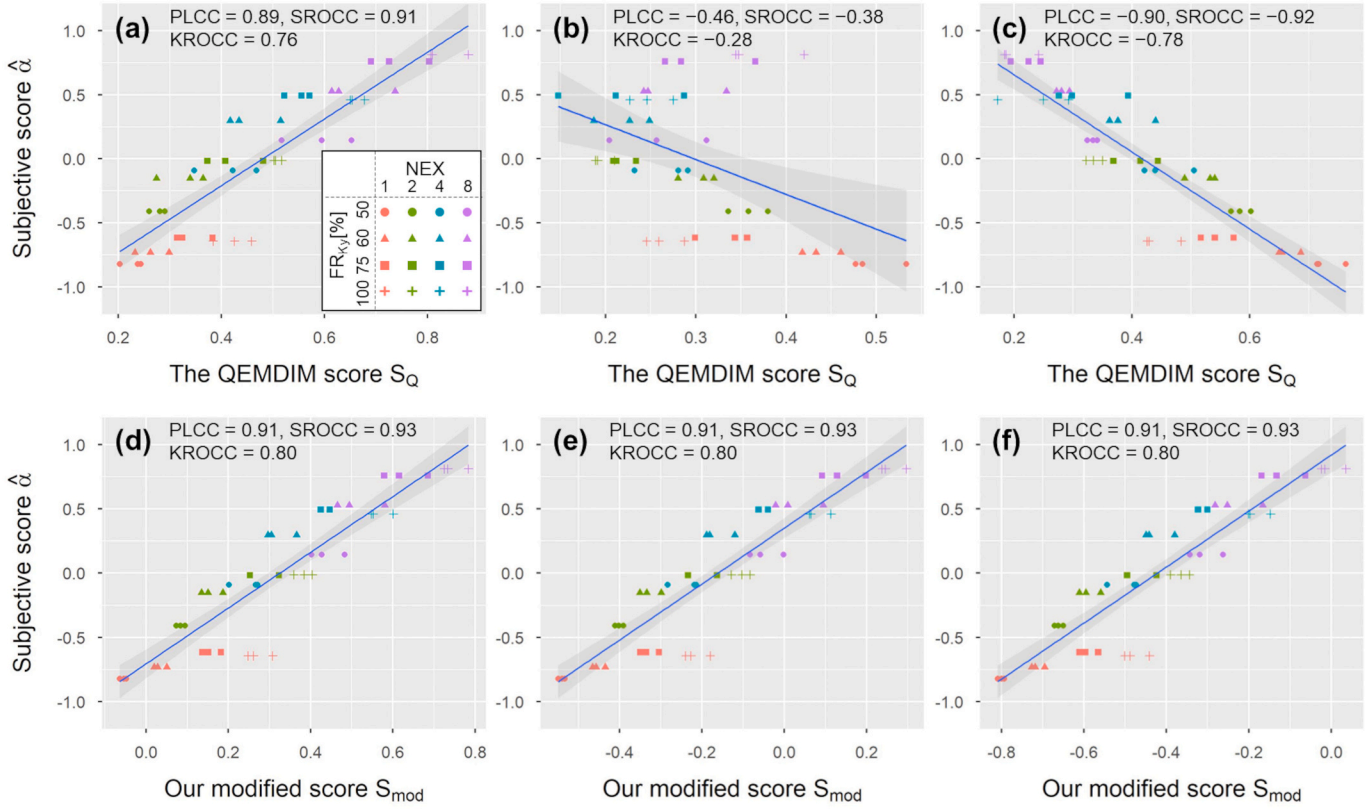

Fig. 6. Representative scatter plots representing the correlation between the subjective scores $\hat{\alpha}$ and QEMDIM scores S_Q (a–c) and between the $\hat{\alpha}$ and the objective modified scores S_{mod} (d–f). The figures on the different columns are different in the quality of the reference set each other: the left, low quality (NEX/FR_{Ky} = 1/50%); center, intermediate quality (4/75%); and right, high quality (8/100%). Each of the 48 plots in all figures denotes a measure of a T_1W assessed image capturing the 9th cross-section of the subject Vol_1 (images included in the "C9" row in Fig. 3). The colors and shapes of the plots denote the NEX and FR_{Ky}, respectively, of the assessed images (a). The blue straight line and the surrounding gray area depict the linear regression line and the confidence interval, respectively. The PLCC, SROCC, and KROCC are shown above each figure. Note that the scales of the X-axes are not fixed among the figures.
NEX: number of excitations.
FR_{Ky}: filling rate of a k-space in phase-encoding direction.
PLCC: Pearson linear correlation coefficient.
SROCC: Spearman rank-ordered correlation coefficient.
KROCC: Kendall rank-ordered correlation coefficient.
S_Q: objective quality score obtained from the QEMDIM algorithm.
S_{mod}: objective quality score obtained from the modified algorithm.
$\hat{\alpha}$: subjective quality score. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

in a variety of conditions, such as quantification of the misdiagnosis rate due to the image quality and standardization of distorted image quality among different medical facilities.

We investigated the relationship between IQA quality scores and subjective scores. However, the effect of the quality score on the detectability and diagnostic accuracy of any disease on MR images has not been determined. Clarifying this relationship would make this modified algorithm highly useful in clinical practice. For example, based on this relationship, we can determine the cutoff value or acceptable range of the quality scores of the acquired image. To indicate the influence of the quality score on diagnostic accuracy, further studies are needed to investigate the relationships between our modified scores, subjective IQA, and diagnostic accuracy for various types of diseases.

This study had two possible limitations. First, we applied our IQA algorithm to images acquired under limited conditions, including the scan parameters (spin-echo axial T_1W and T_2W images), participants (five healthy men with relatively young ages), and organs (brain). For example, fast spin-echo and fluid-attenuated inversion recovery pulse sequences, diffusion-weighted imaging, and MR angiography have been used in clinical brain MRI along with the spin-echo sequence. Most of the brain MRI examinations have been performed for elderly patients with abnormalities, including cerebral infarction, tumors, or hemorrhage. Additionally, MRI examinations have been applied to any part of the whole body, not only the brain. These factors might have affected the image quality and/or appearance, which was not evaluated in this study. Further studies are required to investigate their effects on the performance of the modified IQA. Furthermore, we used an MR scanner produced by a single vendor. There may be differences in the reconstruction, image processing, and pulse sequences among scanners produced by different vendors, which can affect the modified scores. Hence, it is necessary to compare the scores among scanners produced by different vendors. Nevertheless, we believe that our results are clinically useful because we verified our algorithm with fewer biases that could influence our scores. Second, we assessed the quality of images without artifacts, such as motion, susceptibility, wraparound, heterogeneous signal intensity, and geometrical distortion, although the presence and/or severity of artifacts is one of the most important factors affecting subjective quality and diagnostic accuracy. Jang et al. reported that the QEMDIM algorithm has the potential to assess the severity of artifacts, including motion, aliasing caused by Cartesian undersampling, streak, and partial volume, along with Gaussian and Rician noise [18]. It may be clinically useful to develop a methodology based on our

modified algorithm to separately assess the severity of noise, blurring, and several types of artifacts, in addition to the comprehensive image quality assessed in this study.

6. Conclusion

In conclusion, the quality scores of the QEMDIM algorithm are significantly dependent on the quality of the reference set. Our proposal of the modified IQA algorithm exhibited a high correlation with the subjective IQA scores irrespective of the quality of the reference set, and achieved quality scores of 0.00 ± 0.026 for images with the same quality and content. We demonstrated that the modified IQA algorithm can be applied to image distortion caused by changing MR scan parameters. Therefore, our modified algorithm is thought to have a clinically acceptable IQA performance (that is, accuracy, robustness, quantity, and reproducibility) for brain MR images, which may imply that our algorithm has the potential to be an alternative to subjective IQA.

Funding

This study was supported by JSPS KAKENHI (grant number JP19K17249).

CRediT authorship contribution statement

Yoichiro Ikushima: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Funding acquisition. **Shogo Tokurei:** Conceptualization, Investigation, Writing – original draft, Writing – review & editing. **Hiroyuki Tarewaki:** Investigation, Resources. **Junji Morishita:** Writing – review & editing, Supervision. **Hidetake Yabuuchi:** Writing – review & editing, Supervision.

Acknowledgements

The authors thank the radiological technologists at Osaka University Hospital and Junshin Gakuen University for participating in our observational experiments and the volunteers for participating in our scanning experiments.

References

[1] Henning J, Nauerth A, Friedburg H. RARE imaging: a fast imaging method for clinical MR. Magn. Reson. Med. 1986;3:823–33. https://doi.org/10.1002/mrm.1910030602.
[2] Noll DC, Nishimura DG, Macovski A. Homodyne detection in magnetic resonance imaging. IEEE Trans. Med. Imaging 1991;10:154–63. https://doi.org/10.1109/42.79473.
[3] Sodickson DK, Manning WJ. Simultaneous acquisition of spatial harmonics (SMASH): fast imaging with radiofrequency coil arrays. Magn. Reson. Med. 1997;38:591–603. https://doi.org/10.1002/mrm.1910380414.
[4] Pruessmann KP, Weiger M, Scheidegger MB. SENSE: sensitivity encoding for fast MRI. Magn. Reson. Med. 1999;42:952–62. https://doi.org/10.1002/(SICI)1522-2594(199911)42:5<952::AID-MRM16>3.0.CO;2-S.
[5] Griswold MA, Jakob PM, Heidemann RM, Nittka M, Jellus V, Wang J, et al. Generalized autocalibrating partially parallel acquisitions (GRAPPA). Magn. Reson. Med. 2002;47:1202–10. https://doi.org/10.1002/mrm.10171.
[6] Lustig M, Donoho DL, Santos JM, Pauly JM. Compressed sensing MRI. IEEE Signal Process Magaz. 2008;25:72–82. https://doi.org/10.1109/MSP.2007.914728.
[7] Lustig M, Donoho D, Pauly JM. Sparse MRI: the application of compressed sensing for rapid MR imaging. Magn. Reson. Med. 2007;58:1182–95. https://doi.org/10.1002/mrm.21391.
[8] Swets JA. ROC analysis applied to the evaluation of medical imaging techniques. Investig. Radiol. 1979;14:109–21. https://doi.org/10.1097/00004424-197903000-00002.
[9] Franken EA, Smith WL, Berbaum KS, Kao SC, Sato Y. Comparison of a PACS workstation with conventional film for interpretation of neonatal examinations: a paired comparison study. Pediatr. Radiol. 1991;21:336–40. https://doi.org/10.1007/BF02011481.
[10] Nishiyama Y, Tada K, Nishiyama Y, Mori H, Maruyama M, Katsube T, et al. Effect of the forward-projected model-based iterative reconstruction solution algorithm on image quality and radiation dose in pediatric cardiac computed tomography. Pediatr. Radiol. 2016;46:1663–70. https://doi.org/10.1007/s00247-016-3676-x.
[11] Burgess AE. Comparison of receiver operating characteristic and forced choice observer performance measurement methods. Med. Phys. 1995;22:643–55. https://doi.org/10.1118/1.597576.
[12] Ikejimba LC, Salad J, Graff CG, Ghammraoui B, Cheng WC, Lo JY, et al. A four-alternative forced choice (4AFC) methodology for evaluating microcalcification detection in clinical full-field digital mammography (FFDM) and digital breast tomosynthesis (DBT) systems using an inkjet-printed anthropomorphic phantom. Med. Phys. 2019;46:3883–92. https://doi.org/10.1002/mp.13629.
[13] Ciccone G, Vineis P, Frigerio A, Segnan N. Inter-observer and intra-observer variability of mammogram interpretation: a field study. Eur. J. Cancer 1992;28:1054–8. https://doi.org/10.1016/0959-8049(92)90455-B.
[14] Mortamet B, Bernstein MA, Jack Jr CR, Gunter JL, Ward C, Britson PJ, et al. Automatic quality assessment in structural brain magnetic resonance imaging. Magn. Reson. Med. 2009;62:365–72. https://doi.org/10.1002/mrm.21992.
[15] Mittal A, Moorthy AK, Bovik AC. No-reference image quality assessment in the spatial domain. IEEE Trans. Image Process. 2012;21:4695–708. https://doi.org/10.1109/TIP.2012.2214050.
[16] Mittal A, Soundararajan R, Bovik AC. Making a "completely blind" image quality analyzer. IEEE Signal Process Lett. 2013;20:209–12. https://doi.org/10.1109/LSP.2012.2227726.
[17] Osadebey M, Pedersen M, Arnold D, Wendel-Mitoraj K. No-reference quality measure in brain MRI images using binary operations, texture and set analysis. IET Image Process. 2017;11:672–84. https://doi.org/10.1049/iet-ipr.2016.0560.
[18] Jang J, Bang K, Jang H, Hwang D. Quality evaluation of no-reference MR images using multidirectional filters and image statistics. Magn. Reson. Med. 2018;80:914–24. https://doi.org/10.1002/mrm.27084.
[19] Edelstein WA, Glover GH, Hardy CJ, Redington RW. The intrinsic signal-to-noise ratio in NMR imaging. Magn. Reson. Med. 1986;3:604–18. https://doi.org/10.1002/mrm.1910030413.
[20] Sharifi K, Leon-Garcia A. Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video. IEEE Trans Circuits Syst Video Technol 1995;5:52–6. https://doi.org/10.1109/76.350779.
[21] Aja-Fernández S, Vegas-Sánchez-Ferrero G, Tristán-Vega A. Noise estimation in parallel MRI: GRAPPA and SENSE. Magn. Reson. Imaging 2014;32:281–90. https://doi.org/10.1016/j.mri.2013.12.001.
[22] Miyati T, Fujita H, Kasuga T, Koshida K, Sanada S, Banno T, et al. Measurements of MTF and SNR(f) using a subtraction method in MRI. Phys. Med. Biol. 2002;47:2961–72. https://doi.org/10.1088/0031-9155/47/16/311.
[23] Machida Y, Kusahara H, Kassai Y. A novel SNR estimation technique applicable to clinical parallel MR images: Triple band-width single acquisition method (TriSAM). In: Proceedings of the 16th Annual Meeting of ISMRM, Toronto; 2008 (abstract 3089).
[24] Nagasawa S. Improvement of the Schéffe's method for paired comparisons. Kansei Eng Int 2002;3:47–56. https://doi.org/10.5057/kei.3.3_47.
[25] Schéffe H. An analysis of variance for paired comparisons. J. Am. Stat. Assoc. 1952;47:381–400. https://doi.org/10.1080/01621459.1952.10501179.
[26] Gu K, Zhai G, Lin W, Yang X, Zhang W. No-reference image sharpness assessment in autoregressive parameter space. IEEE Trans. Image Process. 2015;24:3218–31. https://doi.org/10.1109/TIP.2015.2439035.
[27] Gu K, Zhou J, Qiao J, Zhai G, Lin W, Bovik AC. No-reference quality assessment of screen content pictures. IEEE Trans. Image Process. 2017;26:4005–18. https://doi.org/10.1109/TIP.2017.2711279.
[28] Gu K, Jakhetiya V, Qiao J, Li X, Lin W, Thalmann D. Model-based referenceless quality metric of 3D synthesized images using local image description. IEEE Trans. Image Process. 2018;27:394–405. https://doi.org/10.1109/TIP.2017.2733164.