

Article

# Magnetic Resonance Image Quality Assessment by Using Non-Maximum Suppression and Entropy Analysis

Rafał Obuchowicz <sup>1</sup>, Mariusz Oszust <sup>2</sup>, Marzena Bielecka <sup>3,\*</sup>, Andrzej Bielecki <sup>4</sup>  
and Adam Piórkowski <sup>5</sup>

<sup>1</sup> Department of Diagnostic Imaging, Jagiellonian University Medical College, 19 Kopernika Street, 31-501 Cracow, Poland; rafalobuchowicz@su.krakow.pl

<sup>2</sup> Department of Computer and Control Engineering, Rzeszow University of Technology, W. Pola 2, 35-959 Rzeszow, Poland; marosz@kia.prz.edu.pl

<sup>3</sup> Faculty of Geology, Geophysics and Environmental Protection, AGH University of Science and Technology, al. Mickiewicza 30, 30-059 Cracow, Poland

<sup>4</sup> Faculty of Electrical Engineering, Automation, Computer Science and Biomedical Engineering, AGH University of Science and Technology, al. Mickiewicza 30, 30-059 Cracow, Poland; azbielecki@gmail.com

<sup>5</sup> Department of Biocybernetics and Biomedical Engineering, AGH University of Science and Technology, al. Mickiewicza 30, 30-059 Cracow, Poland; pioro@agh.edu.pl

\* Correspondence: bielecka@agh.edu.pl

Received: 5 January 2020; Accepted: 13 February 2020; Published: 16 February 2020



**Abstract:** An investigation of diseases using magnetic resonance (MR) imaging requires automatic image quality assessment methods able to exclude low-quality scans. Such methods can be also employed for an optimization of parameters of imaging systems or evaluation of image processing algorithms. Therefore, in this paper, a novel blind image quality assessment (BIQA) method for the evaluation of MR images is introduced. It is observed that the result of filtering using non-maximum suppression (NMS) strongly depends on the perceptual quality of an input image. Hence, in the method, the image is first processed by the NMS with various levels of acceptable local intensity difference. Then, the quality is efficiently expressed by the entropy of a sequence of extrema numbers obtained with the thresholded NMS. The proposed BIQA approach is compared with ten state-of-the-art techniques on a dataset containing MR images and subjective scores provided by 31 experienced radiologists. The Pearson, Spearman, Kendall correlation coefficients and root mean square error for the method assessing images in the dataset were 0.6741, 0.3540, 0.2428, and 0.5375, respectively. The extensive experimental evaluation of the BIQA methods reveals that the introduced measure outperforms related techniques by a large margin as it correlates better with human scores.

**Keywords:** blind image quality assessment; magnetic resonance images; entropy; non-maximum suppression

---

## 1. Introduction

The ubiquity of advancements in imaging has brought significant attention of medical specialists due to the role of the quality of displayed content in diagnosis [1–3]. The quality of Magnetic Resonance (MR) images depends on used hardware parts, software techniques, as well as human errors involving

patient noncompliance or operator mistakes [4–8]. Therefore, the development of automatic image quality assessment (IQA) methods for MR scans is particularly important since the contamination of acquired images may compromise subsequent diagnosis and treatment. Moreover, such methods may support a selection of algorithms for image processing or parameters of imaging systems. Hopefully, a time-consuming examination of images by trained medical specialists can be avoided. Furthermore, the lack of reproducibility of subjective tests and personal quality preferences impeding scores of small groups encourages the use of automatic and repeatable IQA methods. The IQA measures are divided into three categories: Full-reference (FR), reduced-reference (RR), and no-reference or blind (BIQA) methods [9]. The full-reference methods compare input images with their non-distorted versions. However, most medical imaging systems do not produce pristine images, limiting the application range of FR methods [10]. The reduced reference techniques, in turn, require only a part of the information on the pristine image, and blind IQA methods assess images without any external information. Therefore, the development of BIQA approaches is desired.

Among the applications of FR-IQA methods to MR images, Baselice et al. [10] compared results of denoising approaches using Mean Square Error (MSE) with the Structural Similarity Index (SSIM) [11]. Jang et al. [12] employed the SSIM and the Root-Mean-Square Error (RMSE) for an evaluation of BIQA methods on synthetically distorted MR scans. In the work of Chow and Rajagopal [13], Noise Quality Measurement (NQM) with Feature SIMilarity (FSIM) were applied to evaluate a BIQA method. Recognition of a supportive role of FR measures in the assessment of medical images and the need for creating new datasets are among findings of that work [13]. The reduced-reference techniques are not used for the assessment of MR images. In the literature, several BIQA methods have been introduced. Interestingly, as the Signal-to-Noise Ratio (SNR) and Contrast-to-Noise Ratio (CNR) are frequently used for the assessment of medical images [14], they are often criticized due to the need of indication of clearly defined regions with tissue and background [14–16]. Considering MR images, Chow and Rajagopal [13] adapted Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [17] by training it on MR images instead of natural images. The method fits the Mean Subtracted Contrast Normalization (MSCN) of an image to the Generalized Gaussian Distribution (GGD). Similarly, the GGD was used by Jang et al. [12]. In that work, the characteristics of MR scans were taken into account by employing a multidirectional-filtering of images. In the work of Yu et al. [16], four BIQA methods, i.e., BRISQUE, Natural Image Evaluator (NIQE), Blind Image Integrity Notator using DCT statistics (BLIINDS-II), and Blind Image Quality Index (BIQI), were trained on the SNR scores. Their correlation with the SNR was investigated by Zhang et al. [18]. The BIQA methods for the assessment of brain MR scans were employed by Sandilya and Nirmala [19] and Osadebey et al. [20]. In the first work, the reconstructed scans were assessed with BRISQUE, while in the second approach, binary images of brain scans were evaluated considering noise, lightness, contrast, sharpness, and texture details.

The literature review reveals that the lack of BIQA approaches designed for MR scans is caused mainly by the lack of IQA databases of such images with subjective scores. Moreover, natural images differ from MR images concerning characteristics of used imaging systems for their registration, the complexity of captured structures, or noise. Considering the popularity of IQA methods designed for natural images, some of BIQA approaches for MR scans adapted or modified them. However, there exist many concepts in the IQA of natural images that are not yet utilized for MR images and they should be examined. Therefore, in this paper, apart from the novel BIQA approach designed for MR scans, a set of representative IQA methods is evaluated. Furthermore, a dataset with MR scans and subjective scores used in the evaluation is released.

The introduced method, ENTropy-based Magnetic resonance Image Quality Assessment measure (ENMIQA), takes into account thresholded local intensity differences obtained by using the non-maximum suppression (NMS) [21,22] operation and calculates the entropy of a sequence of extrema numbers.

The extrema represent a set of filtered versions of an input image. Then, entropy is used for quality prediction.

The major contributions of this work are a novel method for the quality assessment of MR images and a comprehensive evaluation of the measure against the state-of-the-art IQA techniques on a dataset of MR images assessed by a large group of experienced radiologists.

The remainder of this paper is organized as follows. In Section 2, the approach is introduced. Then, in Section 3, it is evaluated against the related BIQA methods. Finally, in Section 4, the paper is concluded.

## 2. Proposed Image Quality Measure

In the introduced method, ENMIQA, an input image  $I$  is filtered to determine pixels that represent local intensity extrema. To determine which pixels should be selected, the NMS operation [21,22] is performed. However, to provide a more thorough examination instead of selecting pixels that are of greater or lesser intensity value than its surrounding neighbors, in this work, a sequence of intensity thresholds  $\mathfrak{T} = [1, 2, \dots, S]$ ,  $S \in \mathbb{Z}^+$ , is introduced. The NMS uses the threshold  $t \in \mathfrak{T}$  to indicate the local extrema. Consequently, image  $I$  for each threshold  $t$  is represented by the number of found local extrema  $I(t)$ . This can be written as:

$$I(t) = \sum_{a=1}^M \sum_{b=1}^N T(a, b, t), \tag{1}$$

where a pair  $(a, b)$  denotes the pixel location within an image of the size  $M \times N$  and  $T(a, b, t)$  is a test in which the NMS is calculated using the proposed threshold  $t$ . The test is obtained as follows:

$$T(a, b, t) = \begin{cases} 1, & \text{if } \forall_{(i,j)} I(a, b) > I(a + i, b + j) + t, \\ 1, & \text{else if } \forall_{(i,j)} I(a, b) < I(a + i, b + j) - t, \\ 0, & \text{otherwise,} \end{cases} \tag{2}$$

where  $(i, j) \in \{(0, 1), (0, -1), (1, 0), (-1, 0)\}$ . The pair of indices  $(i, j)$  forms the neighborhood of  $3 \times 3$  pixels around the location  $(a, b)$ . Finally, a sequence of sums  $I(\mathfrak{T}) = [I(t = 1), I(t = 2), \dots, I(t = S)]$  is obtained. Then, it is divided by the image size to normalize the values. To determine the quality of the input image  $I$ , entropy of  $I(\mathfrak{T})$  is calculated.

Entropy is the fundamental concept of Shannon information theory [23,24]. It is usually considered in the framework of measure theory. Assuming that space  $X$  with a probabilistic measure  $\mu$  and a countable partition  $\mathfrak{P}$  of  $X$  are given [25], the entropy  $h$  is:

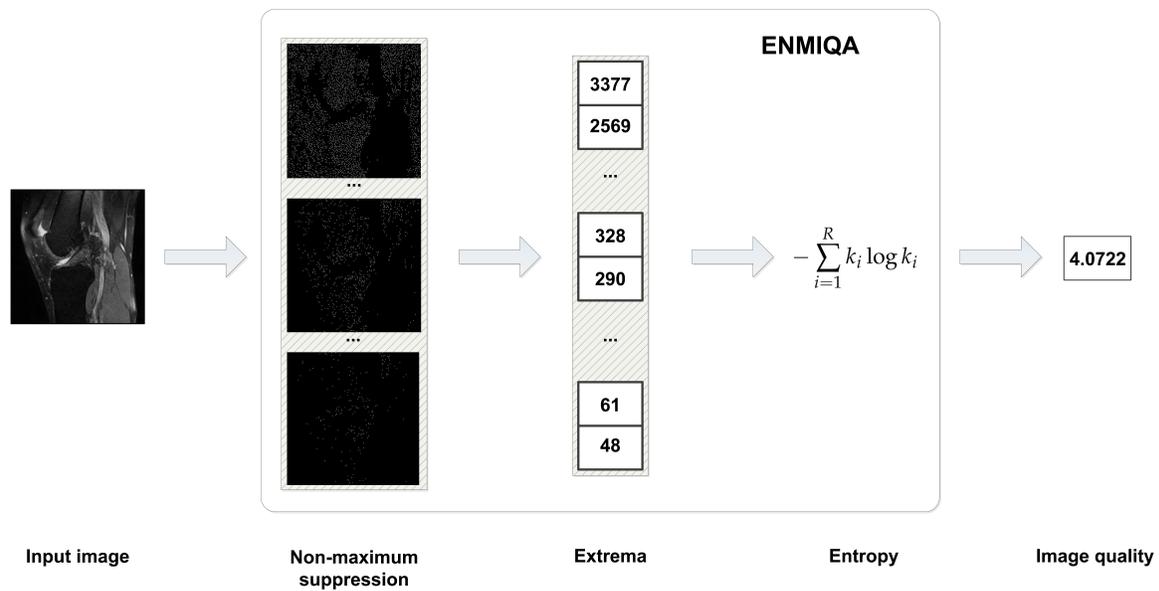
$$h(\mu, \mathfrak{P}) = \sum_{P \in \mathfrak{P}} s(\mu(P)), \tag{3}$$

where  $s: [0, 1] \rightarrow [0, \infty)$  can be expressed as  $s(x) = -x \log x$  for  $0 < x \leq 1$  and  $s(0) = 0$ . Note that entropy equals zero if and only if there exists such  $P \in \mathfrak{P}$  that  $\mu(P) = 1$ . If  $X$  contains  $R$  elements, then  $\mathfrak{P} = \{P_1, \dots, P_R\}$ . Furthermore, if  $\mu$  is based on counting measure, then Equation (3) has the following form:

$$h(\mu, \mathfrak{P}) = - \sum_{i=1}^R k_i \log k_i, \tag{4}$$

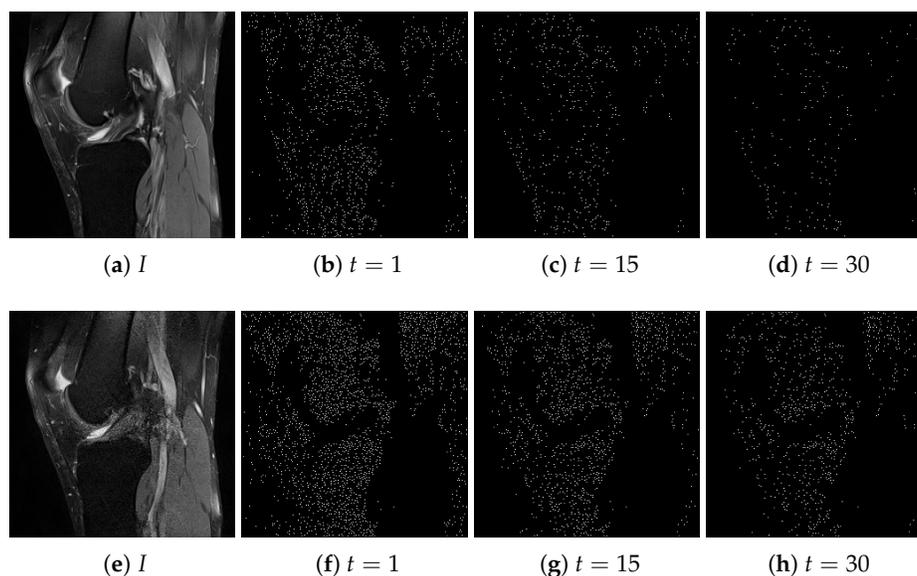
where  $k_i = \frac{m_i}{m}$ ,  $m_i$  and  $m$  are the numbers of elements in  $P_i$  and  $X$ , respectively. Entropy defined by Equation (4) reaches its maximum for the uniform distribution of the measure  $\mu$  on the family  $\mathfrak{P}$ . Such defined entropy refers to the amount of information on  $(X, \mu)$  introduced by  $\mathfrak{P}$ . Consequently, the inversely proportional relationship between entropy and information is often applied in practice.

In this paper, entropy analysis is used for the IQA of two-dimensional MR images. In such a context, it can be employed for measuring disorders. In MR scans of internal organs, single isolated impulses with higher or lower intensity concerning a local neighborhood are common in distorted images. Thus, the greater the value of the threshold  $t$  in the NMS, the greater the probability that the detected intensity irregularities are disorders that decrease the quality of an image. The observed discriminative capabilities of entropy regarding images of different qualities justify its use for the IQA of MR images. In this work, Equation (4) is directly used as a quality measure, assuming that a set  $X$  is expressed as  $\{(I, t), t \in \mathfrak{T}\}$  and  $\mathfrak{T}$  determines the partition of  $X$ . The main computational steps of the method are shown in Figure 1.



**Figure 1.** Image processing steps towards the calculation of image quality in ENtropy-based Magnetic resonance Image Quality Assessment measure (ENMIQA).

Figure 2 presents two MR images of different quality and the influence of  $t$  on the local extrema. As shown, the proposed method determines more extrema in images with more distortions.



**Figure 2.** Two magnetic resonance (MR) images of different quality and the determined local extrema for  $t = 1, 15, 30$ .

### 3. Results and Discussion

In this section, a dataset that contains MR images with associated subjective scores is introduced. Then, the performance of ENMIQA against ten state-of-the-art related methods is evaluated using a typical methodology and discussed. Finally, the influence of parameters of ENMIQA on its performance is provided.

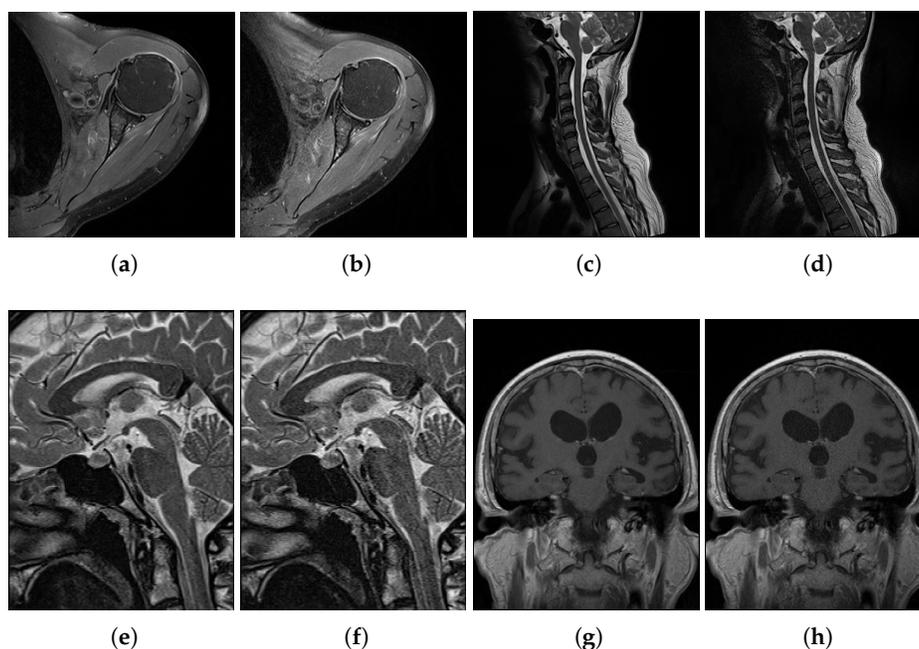
#### 3.1. Experimental Data

The introduced ENMIQA and related techniques are evaluated on a dataset that contains MR images and subjective scores collected in tests with human subjects. The dataset consists of 70 T2-weighted MR images (T2w) extracted from the lumbar and cervical spine, brain, hip, knee, and wrist sequences in axial, sagittal, and coronal planes. The sequences were obtained for a group of 51 patients of 27–41 years old (26 men and 25 women). The study protocol was designed according to the guidelines of the Declaration of Helsinki and the Good Clinical Practice Declaration Statement. The data safety was ensured by removing the personal details from images. Written acceptance for conducting the study was obtained from the Ethics Committee of Jagiellonian University (no. 1072.6120.15.2017). To produce images with different quality for the IQA purposes, shortened sequences were acquired using Process Analytical Technology (PAT) I software (Siemens) and employing the GeneRalized Autocalibrating Partially Parallel Acquisitions (GRAPPA) 3 in which 25% of the echoes were acquired with 60% signal reduction regarding the original acquisition mode [26,27]. Then, images with distortion types that were not present in all examined body parts were rejected. The obtained dataset is characterized in Table 1. There are 15, 9, and 11 image pairs captured in sagittal, axial, and coronal planes, respectively. The size of the images ranges from  $192 \times 320$  to  $512 \times 512$ . The subjective scores for images were obtained in a group of 31 experienced radiologists with more than six years of diagnostic reading residency. Each radiologist assessed two images of the same part of the body at once, spending a minute on the assessment of the pair. The images were scored from 1 to 5, with a higher score associated with better quality. The examination was repeated until all images in the dataset were assessed. Then, scores for images were averaged and the mean opinion score

(MOS) was obtained. The number of radiologists that took part in the subjective tests was large enough to ensure that personal quality preferences do not impair the MOS. However, the number of images in the database depended on the number of medical professionals and the time spent on the examination. Exemplary images from the dataset can be seen in Figure 3.

**Table 1.** Summary of images used in experiments.

Body Part	No. of Image Pairs	Axial Plane	Sagittal Plane	Coronal Plane
Lumbar and cervical spine	7	2	5	0
Knee	7	2	4	1
Shoulder	8	2	2	4
Wrist	3	0	0	3
Hip	2	1	1	0
Pelvis	2	0	0	2
Elbow	1	1	0	0
Ankle	1	0	1	0
Brain	4	1	2	1
Total pairs	35	9	15	11



**Figure 3.** Exemplary MR images used in experiments.

### 3.2. Evaluation Methodology

According to the popular protocol for the performance evaluation of IQA measures, objective scores  $Q$  for images in a database are compared with subjective scores (i.e., MOS)  $\mathcal{S}$  collected for them in tests with human subjects. Typically, the four criteria are used to characterize IQA measure [28]: Pearson correlation coefficient (PLCC), Spearman Rank order Correlation Coefficient (SRCC), Kendall Rank order Correlation Coefficient (KRCC), and Root Mean Square Error (RMSE). The PLCC and RMSE are calculated for the vector  $Q_p$  obtained via a nonlinear mapping between objective scores  $Q$  and subjective scores  $\mathcal{S}$  using fitted

parameters of the regression model  $\beta = [\beta_1, \beta_2, \dots, \beta_5]$ , i.e.,  $Q_p = \beta_1 \left( \frac{1}{2} - \frac{1}{1 + \exp(\beta_2(Q - \beta_3))} \right) + \beta_4 Q + \beta_5$ . The PLCC is obtained as:

$$PLCC(Q_p, \mathfrak{S}) = \frac{\bar{Q}_p^T \bar{\mathfrak{S}}}{\sqrt{\bar{Q}_p^T \bar{Q}_p \bar{\mathfrak{S}}^T \bar{\mathfrak{S}}}}, \tag{5}$$

where  $\bar{Q}_p$  and  $\bar{\mathfrak{S}}$  are mean-removed vectors. The SRCC is calculated as:

$$SRCC(Q, \mathfrak{S}) = 1 - \frac{6 \sum_{i=1}^m d_i^2}{m(m^2 - 1)}, \tag{6}$$

where  $d_i$  is the difference between  $i$ -th image in vectors of scores and  $m$  denotes the number of images in the dataset. The KRCC is obtained as:

$$KRCC(Q, \mathfrak{S}) = \frac{m_c - m_d}{0.5m(m - 1)}, \tag{7}$$

where  $m_c, m_d$  are the number of concordant and discordant pairs, respectively. The RMSE, in turn, is obtained as:

$$RMSE(Q_p, \mathfrak{S}) = \sqrt{\frac{(Q_p - \mathfrak{S})^T (Q_p - \mathfrak{S})}{m}}. \tag{8}$$

### 3.3. Comparative Evaluation

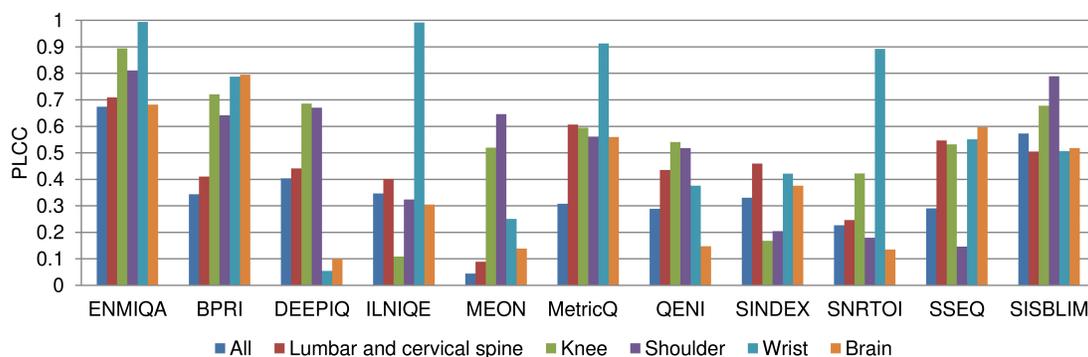
The ENMIQA is compared against the following ten related BIQA measures: SNRTOI [18], BPRI [29], ILNIQE [30], QENI [31], SISBLIM [32], metricQ [33], SSEQ [34], SINDEXT [35], MEON [36], and DEEPIQ [37]. The SNRTOI [18] was implemented by authors of this paper, while other methods were run using their publicly available Matlab implementations. All compared methods, similarly to ENMIQA, do not require training. However, MEON and DEEPIQ represent recently introduced deep learning approaches and are already trained by their authors. The ENMIQA run with  $S = 30$  in experiments and other measures used their default parameters. In cases in which a method was designed to process color images, three identical channels were used as an input. The performance of the methods and their approaches to image quality modeling and prediction are shown in Table 2.

**Table 2.** Evaluation and characteristics of compared blind image quality assessment (BIQA) measures. The best value for each performance criterion is written in bold.

Method	PLCC	SRCC	KRCC	RMSE	Approach to Image Quality Modeling and Prediction
ENMIQA	<b>0.6741</b>	<b>0.3540</b>	<b>0.2428</b>	<b>0.5375</b>	Thresholded NMS and entropy
BPRI	0.3440	0.1515	0.1120	0.6832	Distortion-specific metrics and pseudo-reference image
DEEPIQ	0.4039	0.3030	0.2037	0.6657	RankNet trained on quality-discriminable image pairs
ILNIQE	0.3465	0.1796	0.1162	0.6826	Multivariate Gaussian model of pristine images
MEON	0.0439	0.1247	0.0771	0.7272	End-to-end deep neural network with subtasks
MetricQ	0.3075	0.2300	0.1520	0.6924	Singular value decomposition of local image gradient matrix
QENI	0.2886	0.2385	0.1587	0.6967	Self-similarity of local features and saliency models
SINDEXT	0.3307	0.2802	0.1962	0.6869	Global and local phase information
SNRTOI	0.2262	0.1828	0.1245	0.7088	Signal-to-noise ratio
SSEQ	0.2903	0.0855	0.0487	0.6963	Distortion classification using local entropy
SISBLIM	0.5733	0.2885	0.1820	0.5962	Free energy theory based fusion of distortion-specific metrics

As reported, the measure introduced in this paper, ENMIQA, outperforms related techniques by a large margin in terms of all four performance indices. Depending on the considered index, it is followed

by SISBLIM (PLCC and RMSE) and DEEPIQ (SRCC and KRCC). To show the performance of the measures for images of body parts largely represented in the database, the PLCC calculated for their subsets is reported in Figure 4. Here, ENMIQA obtains greater PLCC than it can be seen for the remaining methods for images of the lumbar and cervical spine, knee, shoulder, and wrist. It is slightly worse than BPRI for brain images. Interestingly, it seems that the recently introduced BPRI is suitable for such images, despite being the second worst technique regarding the entire database and the fourth-best technique in ranking based on the individual body parts. The worse results of methods designed for the assessment of natural images, as well as by complex deep learning approaches, can be justified by the specifics of MR images in which a large portion of the area is covered by organs or tissue while the background is usually dark and may contain noise. In natural images, such empty or nearly empty spaces are seldom found. Furthermore, popular BIQA methods are often trained to recognize typical distortion types (e.g., BPRI, ILNIQE, MEON, SSEQ, or DEEPIQ). Interestingly, methods trained on images contaminated with Gaussian noise can, to some extent, correctly predict the quality of MR images since Gaussian noise manifests itself in magnitude images as a Rician distribution of pixel intensities [38]. This is confirmed by weaker performance of the SNRTOI, which, being an SNR derivative, is often used by radiologists as supporting information on the captured images. The reported results for other methods seem to justify the need for the development of measures designed for the IQA of MR images.



**Figure 4.** Pearson correlation coefficient (PLCC) performance of the BIQA methods for subsets of images of common body parts.

To evaluate the statistical significance of the obtained errors in the prediction of IQA methods, hypothesis tests based on the prediction residuals of each IQA measure after non-linear mapping were conducted using F-statistic [28]. The F-test is based on an assumption of the Gaussianity of residuals and determines whether the two compared sample sets come from the same distribution, based on the ratio of their variances. The test is often used for the comparison of IQA measures [28]. Therefore, at first, the Jarque–Bera (JB) statistic to determine whether residuals come from a normal distribution was used [39]. In the JB test, the null hypothesis is that the vector of residuals of NR measure follows a normal distribution while the alternative hypothesis is that it does not follow it. Since for all compared measures the null hypothesis was not rejected at the 5% significance level, the F-statistic could be reliably employed. In the F-test, the null hypothesis is that the vectors of residuals of two IQA measures come from the same distribution with the same variance and are statistically indistinguishable (95% confidence). The alternative hypothesis is that the vectors are statistically distinguishable and have different variances. Before the calculation of the F-statistic, a vector of residuals of a measure was used to fit a normal distribution and 1000 samples were drawn from it. The tests revealed that the residual variance of ENMIQA is statistically smaller than those of all compared IQA methods with confidence greater than 95%. This is also indicated

by the ratio in all cases. The obtained JB statistics for measures and ratios of the residual variances of algorithms to the ENMIQA are presented in Table 3.

**Table 3.** Ratios of residual variances of methods to ENMIQA and the Jarque–Bera (JB) statistics. Smaller values of JB statistics denote smaller deviations from the Gaussianity. All measures follow a normal distribution.

Method	Ratio	JB Statistic
ENMIQA	1.0000	0.8523
BPRI	0.6189	2.8999
DEEPIQ	0.6510	1.3870
ILNIQE	0.6201	3.9911
MEON	0.5462	3.8930
MetricQ	0.6032	2.8356
QENI	0.5952	2.7040
SINDEX	0.6124	3.2580
SNRTOI	0.5751	1.7389
SSEQ	0.5958	3.5343
SISBLIM	0.8128	0.1254

### 3.4. Computational Complexity

The computational complexity of ENMIQA depends on the size of processed image ( $N \times M$ ), the length of the sequence of thresholds  $S$ , and the size of the neighborhood used for the NMS ( $k = 3 \times 3$ ). Therefore, its computational complexity is of the order of  $O(NMSk^2)$ .

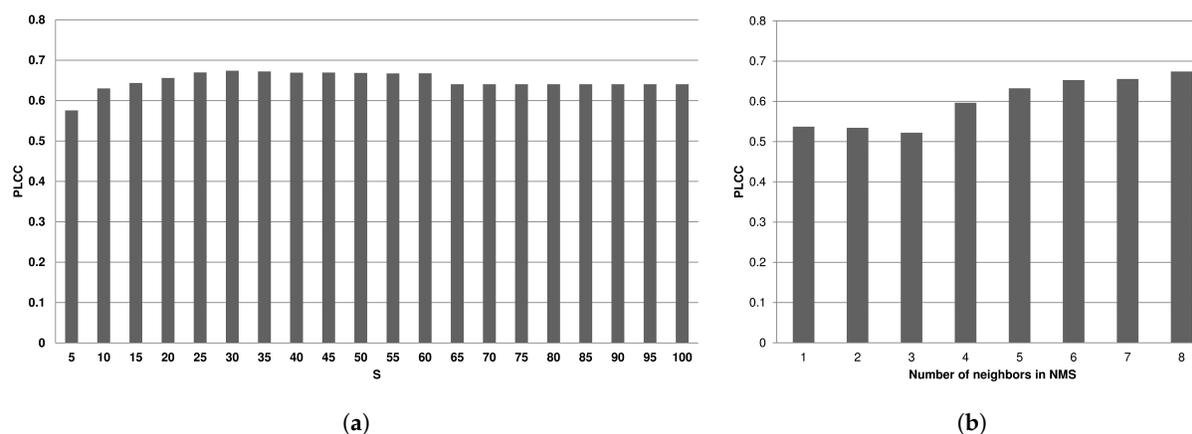
The introduced dataset was used to analyze the computational complexity of methods in terms of the average time taken to assess an image. The methods were run on a 2.2 GHz Intel Core CPU with 8 GB RAM using Matlab 2019b environment. Table 4 reports obtained timings. As shown, ENMIQA is slower than MEON, SINDEX, and SNRTOI, but it is faster than the remaining seven measures. The fastest methods (i.e., SINDEX and SNRTOI) are characterized by inferior IQA performance, and taking into account the results for more promising techniques, the introduced ENMIQA is relatively fast and provides the superior quality prediction of MR images.

**Table 4.** Time–cost comparison of BIQA measures (in seconds).

Method	ENMIQA	BPRI	DEEPIQ	ILNIQE	MEON	MetricQ	QENI	SINDEX	SNRTOI	SSEQ	SISBLIM
Runtime	0.2151	0.2524	2.439	9.299	0.1853	0.4813	1.212	0.0479	0.0069	0.9140	1.629

### 3.5. Influence of Parameters

The ENMIQA is governed by the sequence of thresholds  $\mathfrak{T} = [1, 2, \dots, S]$ ,  $S \in \mathbb{Z}^+$  used by the non-maximum suppression. Therefore, it is worth to determine how stable is its performance for various  $S$ . The  $S$  is the greatest threshold in the sequence and indicates its length. The PLCC performance of the method on the entire database, ranging  $S$  from 5 to 100 with the step of 5 is shown in Figure 5a. The previously introduced evaluation methodology was applied on the entire dataset to allow a coherent comparison with already reported results of other IQA methods (see Section 3.3). Considering the value of the threshold  $S$ , it can be set in between 20 and 60 without a visible drop in the prediction performance. Since ENMIQA exhibits a stable performance across the values of  $S$ ,  $S = 30$  used in experiments is justified.



**Figure 5.** Influence of the threshold  $S$  (a) and the number of neighboring pixels in the non-maximum suppression (NMS) (b) on the PLCC performance of ENMIQA.

The non-maximum suppression selects a pixel with the extreme value, taking into account its eight neighbors and the threshold  $t$ . Since a pixel has 8 neighbors, it is reasonable to use its full neighborhood (the size of 8). However, the suppression can be modified to accept a lesser number of neighboring pixels that are used to indicate the local extrema (see Equation (2)). Therefore, in Figure 5b, the impact of the number of neighbors on the PLCC results of ENMIQA is shown. Here, if the number of used neighbors while determining the local extrema is lower than 8, the performance of the method visibly deteriorates. Hence, the entire pixel neighborhood should be considered by ENMIQA with the NMS. Interestingly, even with a smaller neighborhood the approach still offers a promising performance.

#### 4. Conclusions

In this work, a new BIQA measure for the evaluation of MR images is proposed. The method uses the non-maximum suppression with a sequence of thresholds to detect local intensity extrema in MR images. A relationship between the number of extrema and entropy is investigated. Consequently, a new measure is introduced and experimentally validated against ten representative BIQA techniques on a database that contains MR images assessed by a large group of experienced medical professionals. The experimental comparison reveals that ENMIQA outperforms the-state-of-the-art measures by a large margin in terms of four performance criteria, confirming its suitability for the quality prediction of MR images.

To facilitate the replicability of the reported findings, as well as the applicability of the measure, the Matlab code of ENMIQA and the dataset are available at <http://marosz.kia.prz.edu.pl/ENMIQA.html>.

**Author Contributions:** Conceptualization, R.O., M.O., M.B., A.B., and A.P.; methodology, M.O.; software, M.O., M.B., A.B., and A.P.; validation, R.O., M.O., M.B., A.B., and A.P.; investigation, R.O., M.O., M.B., A.B., and A.P.; writing and editing, R.O., M.O., M.B., A.B., and A.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

#### References

1. Hameed, M.H.; Umer, F.; Khan, F.R.; Pirani, S.; Yusuf, M. Assessment of the diagnostic quality of the digital display monitors at the dental clinics of a university hospital. *Inform. Med. Unlocked* **2018**, *11*, 83–86. doi:10.1016/j.imu.2018.02.002. [CrossRef]

2. Bielecka, M. Syntactic-geometric-fuzzy hierarchical classifier of contours with application to analysis of bone contours in X-ray images. *Appl. Soft Comput.* **2018**, *69*, 368–380. doi:10.1016/j.asoc.2018.04.038. [[CrossRef](#)]
3. Bielecka, M.; Bielecki, A.; Korkosz, M.; Skomorowski, M.; Wojciechowski, W.; Zieliński, B. Application of Shape Description Methodology to Hand Radiographs Interpretation. In *Computer Vision and Graphics*; Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; pp. 11–18.
4. Kustner, T.; Liebgott, A.; Mauch, L.; Martirosian, P.; Bamberg, F.; Nikolaou, K.; Yang, B.; Schick, F.; Gatidis, S. Automated reference-free detection of motion artifacts in magnetic resonance images. *Magn. Reson. Mater. Phys. Biol. Med.* **2018**, *31*, 243–256. [[CrossRef](#)]
5. Senel, L.K.; Kilic, T.; Gungor, A.; Kopanoglu, E.; Guven, H.E.; Saritas, E.U.; Koc, A.; Cukur, T. Statistically Segregated k-Space Sampling for Accelerating Multiple-Acquisition MRI. *IEEE Trans. Med Imaging* **2019**, *38*, 1701–1714. doi:10.1109/TMI.2019.2892378. [[CrossRef](#)] [[PubMed](#)]
6. Knoll, F.; Hammernik, K.; Kobler, E.; Pock, T.; Recht, M.P.; Sodickson, D.K. Assessment of the generalization of learned image reconstruction and the potential for transfer learning. *Magn. Reson. Med.* **2019**, *81*, 116–128. doi:10.1002/mrm.27355. [[CrossRef](#)] [[PubMed](#)]
7. Chow, L.S.; Rajagopal, H.; Paramesran, R. Correlation between subjective and objective assessment of magnetic resonance (MR) images. *Magn. Reson. Imaging* **2016**, *34*, 820–831. [[CrossRef](#)] [[PubMed](#)]
8. Chen, S.; Hu, P.; Gu, Y.; Pang, L.; Zhang, Z.; Zhang, Y.; Meng, X.; Cao, T.; Liu, X.; Fan, Z.; et al. Impact of patient comfort on diagnostic image quality during PET/MR exam: A quantitative survey study for clinical workflow management. *J. Appl. Clin. Med. Phys.* **2019**, *20*, 184–192. doi:10.1002/acm2.12664. [[CrossRef](#)]
9. Chandler, D.M. Seven Challenges in Image Quality Assessment: Past, Present, and Future Research. *ISRN Signal Process.* **2013**, *2013*, 905685. doi:10.1155/2013/905685. [[CrossRef](#)]
10. Baselice, F.; Ferraioli, G.; Pascazio, V. A 3D MRI denoising algorithm based on Bayesian theory. *Biomed. Eng. OnLine* **2017**, *16*, 25. doi:10.1186/s12938-017-0319-x. [[CrossRef](#)]
11. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. doi:10.1109/tip.2003.819861. [[CrossRef](#)]
12. Jang, J.; Bang, K.; Jang, H.; Hwang, D.; Initiative, A.D.N. Quality evaluation of no-reference MR images using multidirectional filters and image statistics. *Magn. Reson. Med.* **2018**, *80*, 914–924. [[CrossRef](#)] [[PubMed](#)]
13. Chow, L.S.; Rajagopal, H. Modified-BRISQUE as no reference image quality assessment for structural MR images. *Magn. Reson. Imaging* **2017**, *43*, 74–87. doi:doi.org/10.1016/j.mri.2017.07.016. [[CrossRef](#)] [[PubMed](#)]
14. Welvaert, M.; Rosseel, Y. On the Definition of Signal-To-Noise Ratio and Contrast-To-Noise Ratio for fMRI Data. *PLOS ONE* **2013**, *8*, 1–10. doi:10.1371/journal.pone.0077089. [[CrossRef](#)] [[PubMed](#)]
15. Dietrich, O.; Raya, J.G.; Reeder, S.B.; Reiser, M.F.; Schoenberg, S.O. Measurement of signal-to-noise ratios in MR images: influence of multichannel coils, parallel imaging, and reconstruction filters. *J. Magn. Reson. Imaging* **2007**, *26*, 375–385. [[CrossRef](#)] [[PubMed](#)]
16. Yu, S.; Dai, G.; Wang, Z.; Li, L.; Wei, X.; Xie, Y. A consistency evaluation of signal-to-noise ratio in the quality assessment of human brain magnetic resonance images. *BMC Med. Imaging* **2018**, *18*, 17. doi:10.1186/s12880-018-0256-6. [[CrossRef](#)] [[PubMed](#)]
17. 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. doi:10.1109/TIP.2012.2214050. [[CrossRef](#)]
18. Zhang, Z.; Dai, G.; Liang, X.; Yu, S.; Li, L.; Xie, Y. Can Signal-to-Noise Ratio Perform as a Baseline Indicator for Medical Image Quality Assessment. *IEEE Access* **2018**, *6*, 11534–11543. doi:10.1109/ACCESS.2018.2796632. [[CrossRef](#)]
19. Sandilya, M.; Nirmala, S.R. Determination of reconstruction parameters in Compressed Sensing MRI using BRISQUE score. In Proceedings of the 2018 International Conference on Information, Communication, Engineering and Technology (ICICET), Pune, India, 29–31 August 2018; pp. 1–5. doi:10.1109/ICICET.2018.8533865. [[CrossRef](#)]

20. 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*, 672–684. doi:10.1049/iet-ipr.2016.0560. [[CrossRef](#)]
21. Neubeck, A.; Van Gool, L. Efficient Non-Maximum Suppression. In Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06), Hong Kong, China, 20–24 August 2006; pp. 850–855. doi:10.1109/ICPR.2006.479. [[CrossRef](#)]
22. Hosang, J.; Benenson, R.; Schiele, B. Learning non-maximum suppression. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4507–4515.
23. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x. [[CrossRef](#)]
24. Jiménez-García, J.; Romero-Oraá, R.; García, M.; López-Gálvez, M.I.; Hornero, R. Combination of Global Features for the Automatic Quality Assessment of Retinal Images. *Entropy* **2019**, *21*, 311. [[CrossRef](#)]
25. Śmieja, M.; Tabor, J. Entropy of the Mixture of Sources and Entropy Dimension. *IEEE Trans. Inf. Theory* **2012**, *58*, 2719–2728. doi:10.1109/TIT.2011.2181820. [[CrossRef](#)]
26. Deshmane, A.; Gulani, V.; Griswold, M.A.; Seiberlich, N. Parallel MR imaging. *J. Magn. Reson. Imaging* **2012**, *36*, 55–72. doi:10.1002/jmri.23639. [[CrossRef](#)] [[PubMed](#)]
27. Breuer, F.A.; Kellman, P.; Griswold, M.A.; Jakob, P.M. Dynamic autocalibrated parallel imaging using temporal GRAPPA (TGRAPPA). *Magn. Reson. Med.* **2005**, *53*, 981–985. doi:10.1002/mrm.20430. [[CrossRef](#)] [[PubMed](#)]
28. Sheikh, H.R.; Sabir, M.F.; Bovik, A.C. A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms. *IEEE Trans. Image Process.* **2006**, *15*, 3440–3451. doi:10.1109/tip.2006.881959. [[CrossRef](#)] [[PubMed](#)]
29. Min, X.; Gu, K.; Zhai, G.; Liu, J.; Yang, X.; Chen, C.W. Blind Quality Assessment Based on Pseudo-Reference Image. *IEEE Trans. Mult.* **2018**, *20*, 2049–2062. doi:10.1109/TMM.2017.2788206. [[CrossRef](#)]
30. Zhang, L.; Zhang, L.; Bovik, A.C. A Feature-Enriched Completely Blind Image Quality Evaluator. *IEEE Trans. Image Process.* **2015**, *24*, 2579–2591. doi:10.1109/TIP.2015.2426416. [[CrossRef](#)]
31. Oszust, M. No-Reference quality assessment of noisy images with local features and visual saliency models. *Inf. Sci.* **2019**, *482*, 334–349. doi:doi.org/10.1016/j.ins.2019.01.034. [[CrossRef](#)]
32. Gu, K.; Zhai, G.; Yang, X.; Zhang, W. Hybrid No-Reference Quality Metric for Singly and Multiply Distorted Images. *IEEE Trans. Broadcast.* **2014**, *60*, 555–567. doi:10.1109/TBC.2014.2344471. [[CrossRef](#)]
33. Zhu, X.; Milanfar, P. Automatic Parameter Selection for Denoising Algorithms Using a No-Reference Measure of Image Content. *IEEE Trans. Image Process.* **2010**, *19*, 3116–3132. doi:10.1109/TIP.2010.2052820. [[CrossRef](#)]
34. Liu, L.; Liu, B.; Huang, H.; Bovik, A.C. No-reference image quality assessment based on spatial and spectral entropies. *Signal Process. Image Commun.* **2014**, *29*, 856–863. doi:10.1016/j.image.2014.06.006. [[CrossRef](#)]
35. Leclaire, A.; Moisan, L. No-Reference Image Quality Assessment and Blind Deblurring with Sharpness Metrics Exploiting Fourier Phase Information. *J. Math. Imaging Vis.* **2015**, *52*, 145–172. doi:10.1007/s10851-015-0560-5. [[CrossRef](#)]
36. Ma, K.; Liu, W.; Zhang, K.; Duanmu, Z.; Wang, Z.; Zuo, W. End-to-End Blind Image Quality Assessment Using Deep Neural Networks. *IEEE Trans. Image Process.* **2018**, *27*, 1202–1213. doi:10.1109/TIP.2017.2774045. [[CrossRef](#)] [[PubMed](#)]
37. Ma, K.; Liu, W.; Liu, T.; Wang, Z.; Tao, D. dipIQ: Blind Image Quality Assessment by Learning-to-Rank Discriminable Image Pairs. *IEEE Trans. Image Process.* **2017**, *26*, 3951–3964. [[CrossRef](#)] [[PubMed](#)]
38. Cardenas-Blanco, A.; Tejos, C.; Irarrazaval, P.; Cameron, I. Noise in magnitude magnetic resonance images. *Concepts Magn. Reson. Part A* **2008**, *32A*, 409–416. doi:10.1002/cmr.a.20124. [[CrossRef](#)]
39. Jarque, C.M.; Bera, A.K. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Econ. Lett.* **1980**, *6*, 255–259. doi:10.1016/0165-1765(80)90024-5. [[CrossRef](#)]

