Expert Systems With Applications 135 (2019) 39–59

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Analysis of controversies in the formulation and evaluation of restoration algorithms for MR Images

Simi V.R.[a,*], Damodar Reddy Edla[a], Justin Joseph[b], Venkatanareshbabu Kuppili[a]

[a] Department of Computer Science and Engineering, National Institute of Technology, Goa-403401, India
[b] Department of Electronics and Instrumentation, St. Joseph's College of Engineering and Technology, Pala, Kerala 686579, India

ARTICLE INFO

Article history:
Received 7 April 2019
Revised 30 May 2019
Accepted 2 June 2019
Available online 4 June 2019

Keywords:
Denoising
Edge preserving filters
Magnetic resonance image
Restoration

ABSTRACT

There is no reliable guidance available in literature so far for the selection of a suitable technique for denoising Magnetic Resonance (MR) images. The performance of edge-preserving denoising schemes like Nonlocal Means, Bilateral, Total Variation, Anisotropic Diffusion, Kuwahara, wavelet denoising, Linear Minimum Mean Square Error, Smallest Univalue Segment Assimilating Nucleus and Beltrami filters on MR images are evaluated and compared in this paper. Performance evaluation is done on real-time MR Images, Shepp–Logan Phantom images and simulated MR images. Image Quality Analysis indices used for the evaluation are Structural Similarity Index Metric, Noise Quality Measure, Peak Signal to Noise Ratio, Edge Preservation Index, MetricQ, Anisotropic Quality Index, Blind Reference Image Quality Evaluator and computational time. It has been observed that the performance of each filter is completely different on Shepp–Logan, simulated MR and real-time MR images. It is critically sensitive to the strength of noise also. No filter which can offer good performance equally on Phantom, simulated MR image and real-time MR images, is available in the literature. Values of the objective indices are not in concordance with subjective quality ratings. Filter designs optimized on Phantom or simulated MR using maximum PSNR between denoised and ground-truth images as an objective function (minimum error sense in general) do not perform well on real-time MRI.

1. Introduction

Magnetic Resonance Imaging (MRI) is a modality extensively used in neuroimaging studies (Benou, Veksler, Friedman, & Raviv, 2017; Hermessi, Mourali, & Zagrouba, 2019). In neuroimaging, Magnetic Resonance (MR) images are helpful for both diagnosis and characterization of Multiple Sclerosis, Dementia, Alzheimer's disease, infectious diseases, intra-cranial lesions *etc*. MR images are extensively used as assistive tools in image-guided stereotactic surgery and Radiation Treatment (RT) planning also. Compared to other imaging modalities, MR images contain more features and structural details which help the physicians for better diagnosis. The quality of the MR images is usually hindered by random noise (Rundo et al., 2019). Even though the image acquisition techniques have undergone tremendous development in hardware engineering, extenuation of noise via hardware modifications is remaining as an unreached objective in MRI. Noise reduces the visibility of low contrast anatomical structures, especially at low signal-to-noise ratio (SNR). Presence of noise adversely affects the performance of edge-based segmentation schemes used in software packages for computerized image analysis. The presence of noise intervenes with the accurate computation of radiation dosage in RT planning. As it is not trivial to address the issue of noise in MRI through design modifications of the MR equipment, post-processing techniques have a significant role in improving MR image's quality.

The visual quality of MR images can be improved feasibly by denoising. The conventional neighbourhood averaging techniques do not preserve edges. Examples for conventional techniques are Gaussian, mean and median filters. In MR denoising, preservation of weak morphological edges while smoothing homogenous regions is important. An ideal filter should be capable to suppress noisy grey level transitions in the homogeneous regions of the image, selectively, without hindering the quality of edges. The state-of-the-art Edge Preserving Filters (EPFs) are capable of preserving the edges while smoothing the homogeneous regions.

Selection of a proper denoising technique for MR image is a challenging task as there is no reliable guidance in the literature. For real-time MR images, proper ground-truth is not available. So

* Corresponding author.
E-mail addresses: simi.vr@nitgoa.ac.in (S. V.R.), dr.reddy@nitgoa.ac.in (D.R. Edla), josephjusti@gmail.com (J. Joseph), venkatanaresh@nitgoa.ac.in (V. Kuppili).

https://doi.org/10.1016/j.eswa.2019.06.003
0957-4174/

the performance evaluation of filters designed for MRI's is often done on Phantom and simulated images with the help of image quality metrics. There is a complication that filters performing well on Phantom and simulated images may not show good performance on real-time MRI. Filters exhibiting good values of quality evaluation metrics may not produce output images with appreciable visual quality.

There is no reliable guidance available in the literature so far for the selection of a suitable technique for denoising Magnetic Resonance (MR) images. The performance evaluation of edge-preserving denoising schemes like Nonlocal Means filter (NLM) (Singh & Bala, 2019), Bilateral Filter (BF) (Akar, 2016), Total Variation (TV) (Kang, Jung, & Kang, 2018), Anisotropic Diffusion (AD) (Tong, Sun, Payet, & Ong, 2012), Kuwahara (Bartyzel, 2016; Djurović, 2017), wavelet denoising (Khatami, Khosravi, Nguyen, Lim, & Nahavandi, 2017), Linear Minimum Mean Square Error (LMMSE) (Golshan & Hasanzadeh, 2015), Smallest Univalue Segment Assimilating Nucleus (SUSAN) (Smith & Brady, 1997), and Beltrami (Fernández, & Martínez, 2010) filters are done in this study. Among the edge-preserving filters listed above, except the wavelet denoising, all other filters belong to the class of 'edge-preserving spatial filters'. Among the edge-preserving spatial filters, AD is a denoising algorithm based on Partial Differential Equations (PDEs) (Lahmiri & Boukadoum, 2016). Wavelet denoising is a 'transform domain' algorithm, outside the category of 'edge-preserving spatial filters' which is believed to have the edge-preserving capability. Wavelet denoising belongs to the broad category of 'multi-resolution denoising algorithms' (Lahmiri, 2017a). Another filter widely used in medical image denoising is 2D Wiener filter (Lahmiri, 2017b). Wiener is an inverse filtering algorithm. Even though Wiener filter is not counted within the category of edge-preserving filters, because of its wide applications, it is also included in the performance evaluation discussed in this paper. In transform domain denoising algorithms, the image is decomposed into frequency or wavelet sub-bands and the coefficients of selected sub-bands are attenuated. There are some denoising techniques analogous to transform domain algorithms. For example, in some denoising algorithms, the input image is decomposed into distinct modes with the help of techniques like Empirical Mode Decomposition (EMD) (Lahmiri, 2015; Lahmiri & Boukadoum, 2015a) or Variational Mode Decomposition (VMD) (Lahmiri & Boukadoum, 2014, 2015b) and noise-free estimate is computed from the modes after excluding the residue modes. As denoising algorithms based on EMD and VMD are outside the category of edge-preserving filters which is the prime focus of this paper, they are not included in the objective analysis presented here.

1.2. Contribution and highlights

This work is intended to analyze the issues and uncertainty in choosing adequate algorithms for denoising MR images. This analysis will help as a reliable guideline for the selection of the edge-preserving filters and their design for MR images. Performance evaluation of edge-preserving denoising schemes on Phantom, simulated and real-time MR images is done in this study. Working of nine edge-preserving filters and Wiener filter are assessed on MR Images. For performance evaluation and validation, eight different image quality metrics including various types of full-reference, partial-reference and no-reference indices are used in this study. Concordance of image quality indices with subjective fidelity ratings is also discussed. The inferences and observations made out from study are versatile as a roadmap for choosing restoration schemes, for the optimization of their operational parameters and modification.

1.3. Organization of the paper

The paper is organized as follows: In Section 2, the theory of edge-persevering filters evaluated in this study is explained. Different qualitative and quantitative methods used for the performance evaluation of denoising filters and their features are explained. Details of test images used for this experiment and system requirements are also furnished. The images denoised with different filters, Mean Opinion Scores of denoised images and objective quality evaluation results are provided in Section 3. Section 4 focuses on the inferences of the study and various issues and disputes related to the selection of filters for denoising MR images.

2. Methodology

The denoising schemes studied in this paper are exclusively edge-preserving filters. Performance of filters is assessed on different types of input images that too at different noise levels both qualitatively and quantitatively. The image sets used in this study includes real-time MR images, Shepp–Logan Phantom images and simulated brain images from BrainWeb: Simulated Brain Database.

2.1. Edge preserving filters used in this study

The popular edge preserving denoising schemes available in literature are NLM (Singh & Bala, 2019), BF (Akar, 2016), TV (Kang et al., 2018), AD (Tong et al., 2012), Kuwahara (Bartyzel, 2016; Djurović, 2017), wavelet denoising (Khatami et al., 2017), LMMSE (Golshan & Hasanzadeh, 2015), SUSAN (Smith & Brady, 1997), Beltrami filter (Fernández & Martínez, 2010) *etc.*

Kuwahara filter is an immediate extension of mean filter. In Kuwahara filter, neighbourhood region around central pixel is segregated into partially overlapping sub-areas. The variances of the sub-regions are compared and the central pixel is substituted by mean grey levels in the sub-area, showing minimum variance. BF is an extension of the conventional 2D Gaussian filter. In 2D Gaussian filter; the restored intensity is the weighted sum of the pixel intensities in a square window around the central pixel. The weight corresponding to any pixel within the window is calculated from the spatial distance between that pixel and contextual pixel. The Bilateral filter uses two Gaussian kernels. The first kernel, termed as the spatial kernel, accounts for the spatial distance of the pixels in the square window from the contextual pixel. The second kernel, termed as the radiometric kernel, accounts for the grey level distance of the pixels in the square window from the contextual pixel. SUSAN filter is closely similar to the bilateral filter, except the aspect that, the contextual pixel is not included in SUSAN filter for computing the restored intensity.

In NLM, the denoised intensity is the weighted sum of the pixels in the whole image or in a neighbourhood area of significantly large size, known as search window, of arbitrary radius. The weight corresponding to any pixel in the search window depends on the similarity of the pixels within the square window of arbitrary radius, known as similarity window around that pixel and the contextual pixel.

TV and AD filters are two edge-preserving filters outside the group of Kuwahara, SUSAN, bilateral and NLM filters. In TV filter, an approximate estimate that has smaller total variation, characterized in terms of the variance, than the original image, constrained by an objective criterion based on minimum error sense between the original image and approximation, is computed. The primal-dual method and split-Bregman method, which solves the objective constrained the partial differential equation of the TV filter are computationally complex. AD filter is a PDE based denoising algorithm. In the AD filter, the input image is denoised via the repeated

process of 'anisotropic diffusion', iteratively. To minimize the local gradient, average of the weighted sum of fraction of the gradient among contextual pixel and its connected neighbours are added to it, in each iteration.

Wavelet denoising or wavelet thresholding (WT) is a non-parametric method which does involve nonlinear shrinking of coefficients in the transform domain. It comes under the category of multi-resolution image denoising. The three main steps involved in wavelet denoising are wavelet decomposition of the image, nonlinear shrinkage denoising and computation of inverse wavelet transform. Beltrami filter is based on Beltrami flow, which aims at minimising the area of the image manifold. In the LMMSE filter, LMMSE estimator is used to predicting the noise-free signal from noisy MR acquisition. The technique has the capability to adaptively compute the strength of noise from the background. Estimation of noise is based on the assumption that the distribution of noise is Rician, which holds good for single coil MR acquisitions.

In Wiener, the principle of 'inverse filtering' is for image denoising. In Wiener filter, the strength of denoising is self-adaptive to the local image variance. When the local variance is large, the filter offers less smoothing and when the variance is small, it offers more smoothing. This is the reason behind the better denoising and edge preservation of Wiener filter than other linear filters. The Wiener filter offers the best tradeoff between noise attenuation and protection of edges. It removes the additive noise and inverts the blurring simultaneously.

All edge-preserving filter designs discussed above are based on the assumption that noise is additive white Gaussian. In real-time MRI, noise is proven to be additive Gaussian (Kuppusamy, Joseph, & Jayaraman, 2019).

2.2. Qualitative analysis

For qualitative analysis output images of NLM, Bilateral, TV, AD, Beltrami, Kuwahara, WT, LMMSE, SUSAN and Wiener filters are presented before a group of ten spectators on a Liquid Crystal Display computer screen with a resolution of 96 dpi. The spectators are tutored to evaluate the quality of the denoised images in terms of three criteria, residual noise, edge-blur and structural loss. Assessment is fully left to the individual discretion of the spectator. Spectators are asked to give scores between 1 and 5. Mean of the scores values corresponding to each image, given by the spectators are taken into regard as Mean Opinion Score (MOS) or the subjective quality evaluation rating.

2.3. Quantitative analysis

Quantitative analysis is performed in terms of full reference, no-reference and partial reference quality evaluation metrics. In the case of full reference metrics, performance is evaluated between ground-truth and denoised image. However, for MR images, noise-free ground-truth images are not available. Only partial reference or no reference metrics are possible in those cases where the ground-truth is not available. Full reference metrics used here in this paper include Peak Signal to Noise Ratio (PSNR) (Joseph & Periyasamy, 2018a), Structural Similarity Index (SSIM) (Kuppusamy, Joseph, & Sivaraman, 2017), Edge Preservation Index (EPI) (Zhang, Feng, Wang, & Xue, 2013) and Noise Quality Measure (NQM) (Damera-Venkata, Kite, Geisler, Evans, & Bovik, 2000). MetricQ (Gabarda, Cristóbal, & Goel, 2018) is a partial reference metric. Anisotropic Quality Index (AQI) (Gabarda & Cristóbal, 2007; Zhu & Milanfar, 2010) and Blind Reference Image Quality Evaluator (BRISQUE) (Chow & Rajagopal, 2017) are the no-reference metrics. In addition, the computational time of the denoising schemes also measured.

The full-reference metric, PSNR is based on the mean squared error, which quantitatively measures the similarity of the restored image to noise-free ground-truth. Output images of good denoising filters are expected to be more similar to the noise-free ground-truth and they produce comparatively larger PSNR values. The higher the PSNR value, the better the quality of the denoised image. Full-reference metric SSIM measures the structural similarity between ground-truth and denoised images. It is a bounded statistic with ideal value equal to one. The SSIM quantifies the degradation of geometric features of the MR images. The bounded statistics EPI measures the extent to which edges are preserved. High EPI values ideally equal to one are appreciable. NQM is a full reference image quality metric which reflects the quality of the denoised images in terms of the amount of residual noise. Consequently, NQM reflects how far image quality is degraded by noise. High NQM values are expected.

AQI, BRISQUE and MetricQ are three indices used to quantify the overall quality of the denoised images by taking into account the extent of residual noise and strength of edges. Out of them, BRISQUE and AQI are no-reference indices. MetricQ is a partial reference index as noisy input is used as a reference during its computation. Ideally, the values of MetricQ and AQI are supposed to be very high. Whereas, the value of BRISQUE comes down in response to the increase in the quality of the image.

2.4. Test images

The test images used for the experiments belong to three different classes, real-time MR Images, Shepp–Logan Phantom images and simulated brain images. MR images are acquired with the help of a 1.5 Tesla 2D MRI scanner, Model: Signa HDxt, manufactured by GE Medical Systems, available at Hind Labs, Government Medical College Kottayam, Kerala, India. Series of acquisition is MR Spectroscopy. The inter-slice gap and slice thickness set during the image acquisition are 1.5 mm and 5 mm, respectively. Images from T1 Fast Spin-Echo Contrast Enhanced (FS-ECE), Diffusion Weighted Imaging (DWI), T2 Fluid Attenuation Inversion Recovery (FLAIR), Gradient Recalled Echo (GRE) and 1000b Array Spatial Sensitivity Encoding Technique (ASSET) pulse sequences are used (Joseph & Periyasamy, 2018b; Joseph, Anoop, & Williams, 2019; Simi, Edla, & Joseph, 2018). Shepp–Logan Phantom image's (Jain, 1989) ground-truth is generated by Matlab®. Additive white Gaussian noise with four different variances ($\sigma = 0.001$, 0.0025, 0.005, and 0.01) is added to the ground-truth image to simulate noisy images at four different noise levels. The benchmark simulated brain data with readily available noisy free ground-truth (0% noise level) is availed from the BrainWeb database (Cocosco, Kollokian, Kwan, Pike, & Evans, 1997; Yang et al., 2015). The MR slices in this database belong to T1, T2 and Proton Density (PD) pulse sequences with 5 mm slice thickness at different noise levels, 1%, 3% and 5%.

2.5. System requirements

All experimental analysis is performed via. Matlab® on a 64-bit Personal Computer with Intel (R) Core i7 processor, 3.60 GHz CPU, 4 GB RAM memory and Microsoft Windows 7 as Operating System.

3. Results

3.1. Shepp–Logan Phantom images

Shepp–Logan Phantom images contaminated with additive Gaussian noise at four different noise levels ($\sigma = 0.001$, 0.0025, 0.005 and 0.01) denoised with different edge-preserving filters

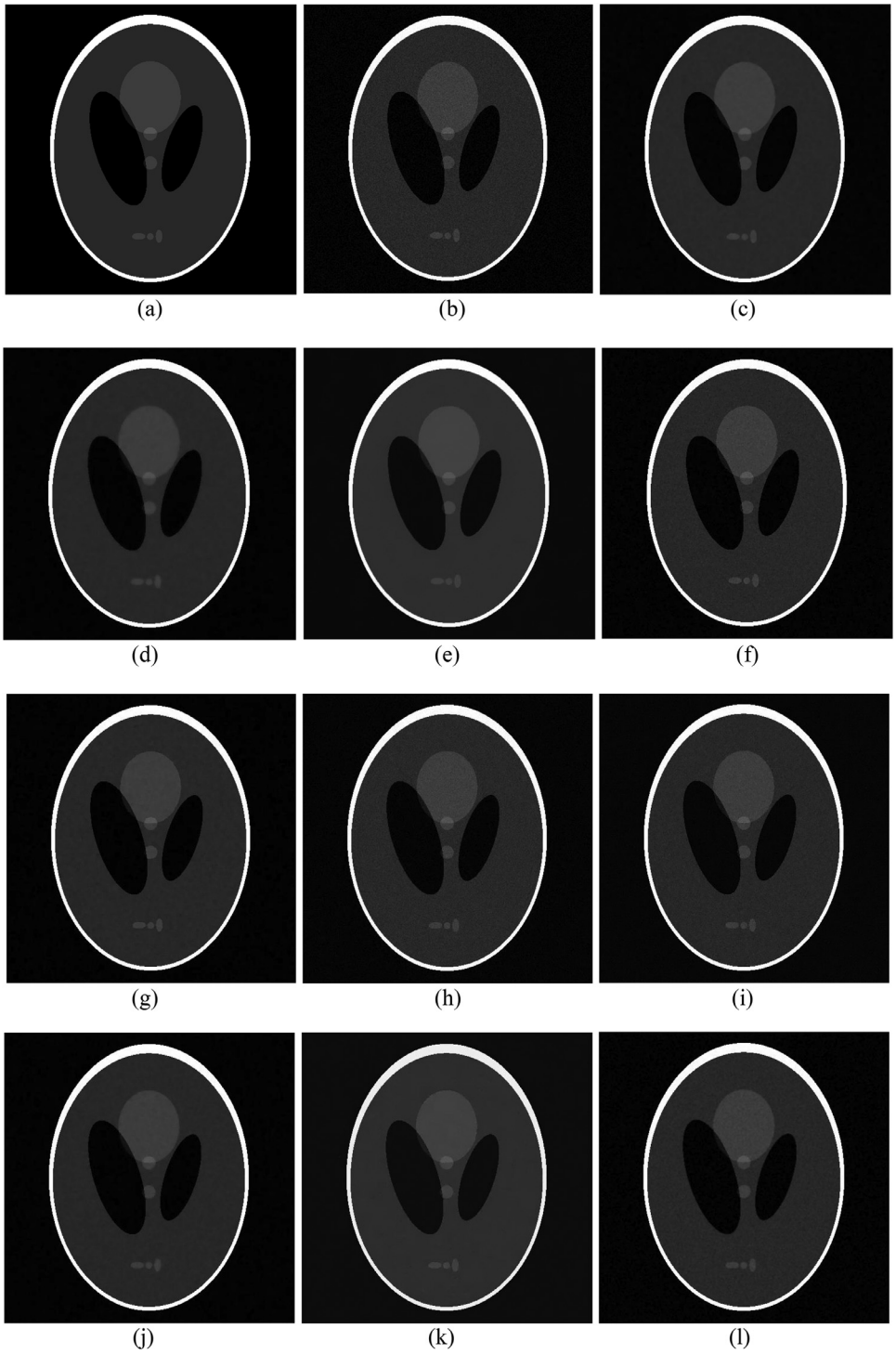


Fig. 1. Shepp–Logan Phantom images denoised with different filters (a) Ground-truth (b) Noisy image ($\sigma = 0.001$) (c) NLM (d) Bilateral filter (e) TV (f) AD (g) Kuwahara (h) Wavelet thresholding (i) LMMSE (j) SUSAN (k) Beltrami (l) Wiener.

NLM, Bilateral, TV, AD, Kuwahara, wavelet thresholding, LMMSE, SUSAN, Beltrami and Wiener are shown in Figs. 1–4. Visual inspection of denoised images is done based upon three criteria; they are residual noise, edge-blur and structural loss or similarity with ground-truth.

While visual inspecting the denoised images in Fig. 1, it can be seen that the filters NLM, TV and Beltrami (Fig. 1(c), 1(e) and 1(k)) have noise-free images without blur and structural loss. Output image of BF (Fig. 1(d)) has slight blur, but it is free from noise and structural loss. Image denoised with TV filter (Fig. 1(e)) is completely free from noise, structural loss and blur. The denoised image of AD filter (Fig. 1(f)) has residual noise, but edge-blur and structural loss are absent. Fig. 1(g), of Kuwahara filter, retains a negligible amount noise. Edge-blur and structural loss are absent. AD, WT and LMMSE (Fig. 1(f), 1(h) and 1(i)) filters retains noise in denoised images, but edge-blur and structural loss are absent. A negligible amount of noise is present in Fig. 1(j) denoised by SUSAN filter. However, it has no edge-blur and structural loss. Image

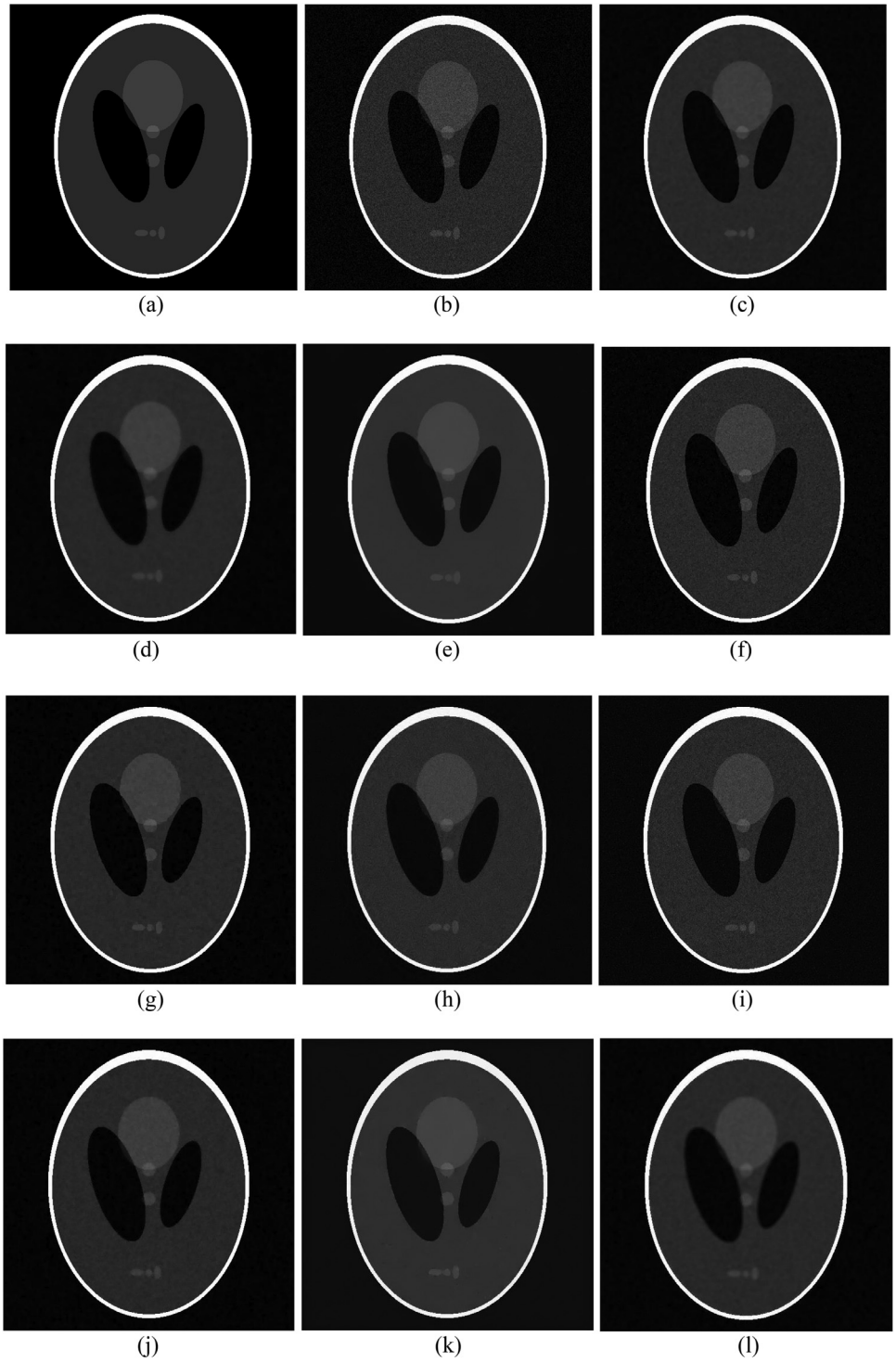


Fig. 2. Shepp–Logan Phantom images denoised with different filters (a) Ground-truth (b) Noisy image ($\sigma = 0.0025$) (c) NLM (d) Bilateral filter (e) TV (f) AD (g) Kuwahara (h) Wavelet thresholding (i) LMMSE (j) SUSAN (k) Beltrami (l) Wiener.

restored with the Wiener filter (Fig. 1(l)) shows edge-preservation, but a negligible amount of noise is present in it. Edge-blur is absent in it.

Visual analysis of denoised images in Fig. 2, shows that TV and Beltrami filters (Fig. 2(e) and 2(k)), produce noise-free images without blur and have ground-truth similarity. NLM retains slight noise in its denoised image (Fig. 2(c)) without edge-blur and structural loss. Output image of BF (Fig. 2(d)) shows edge-blur and it retains noise. Fig. 2(e), the output image of TV filter, is noise-free and blur-free without structural loss. Output images of AD and LMMSE filters (Fig. 2(f) and 2(i)) contain a high amount of noise, but edge-blur and structural loss are absent. A moderate amount of noise is present in the output images of Kuwahara and WT based denoising (Fig. 2(g) and 2(h)). These images do not show blur and structural loss. A negligible amount of noise is present in Fig. 2(j) without blur and structural loss, which is the output of SUSAN filter. Slight amount of noise and blur is present in the denoised image of Wiener filter in Fig. 2(l).

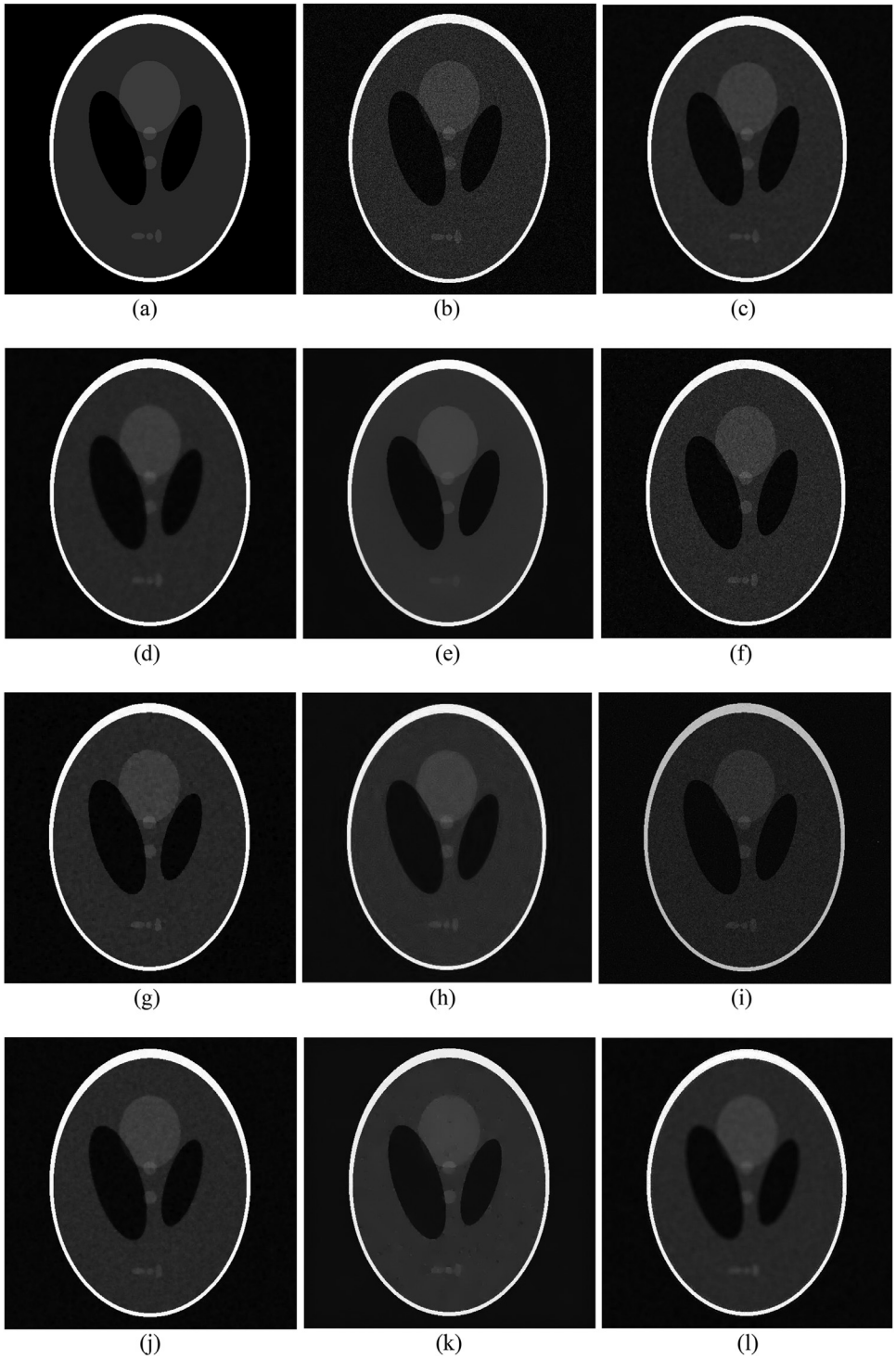


Fig. 3. Shepp–Logan Phantom images denoised with different filters (a) Ground-truth (b) Noisy image ($\sigma = 0.005$) (c) NLM (d) Bilateral filter (e) TV (f) AD (g) Kuwahara (h) Wavelet thresholding (i) LMMSE (j) SUSAN (k) Beltrami (l) Wiener.

Considering the denoised images ($\sigma = 0.005$) in Fig. 3, output images of TV and Beltrami filters (Fig. 3(e) and 3(k)) are noise-free with negligible structural loss. These images are free from edge-blur also. Denoised images of NLM and WT (Fig. 3(c) and 3(h)) contain slight residual noise but free from blur and structural loss. The restored image of BF in Fig. 3(d), have a slight blur and residual noise but free from structural loss. Denoised images of AD and LMMSE (Fig. 3(f) and 3(i)) contain severe noise. Structural loss and edge-blur are not visible on those images. Average amount of noise is noticeable in Kuwahara and SUSAN filtered images (Fig. 3(g) and 3(j)). These images show slight structural loss also, but no blur. In the Fig. 3(l) of Wiener, noise and structural loss is present.

During the visual analysis of denoised images at noise level $\sigma = 0.01$ (Fig. 4), the following observations are noted. Restored image of Kuwahara filter (Fig. 4(c)) contains considerably low noise and has slight structural change with no of edge-blur. BF blurs the image (Fig. 4(d)) and it has a slight structural loss, residual noise. TV filter completely removes noise from image (Fig. 4(e)) and is

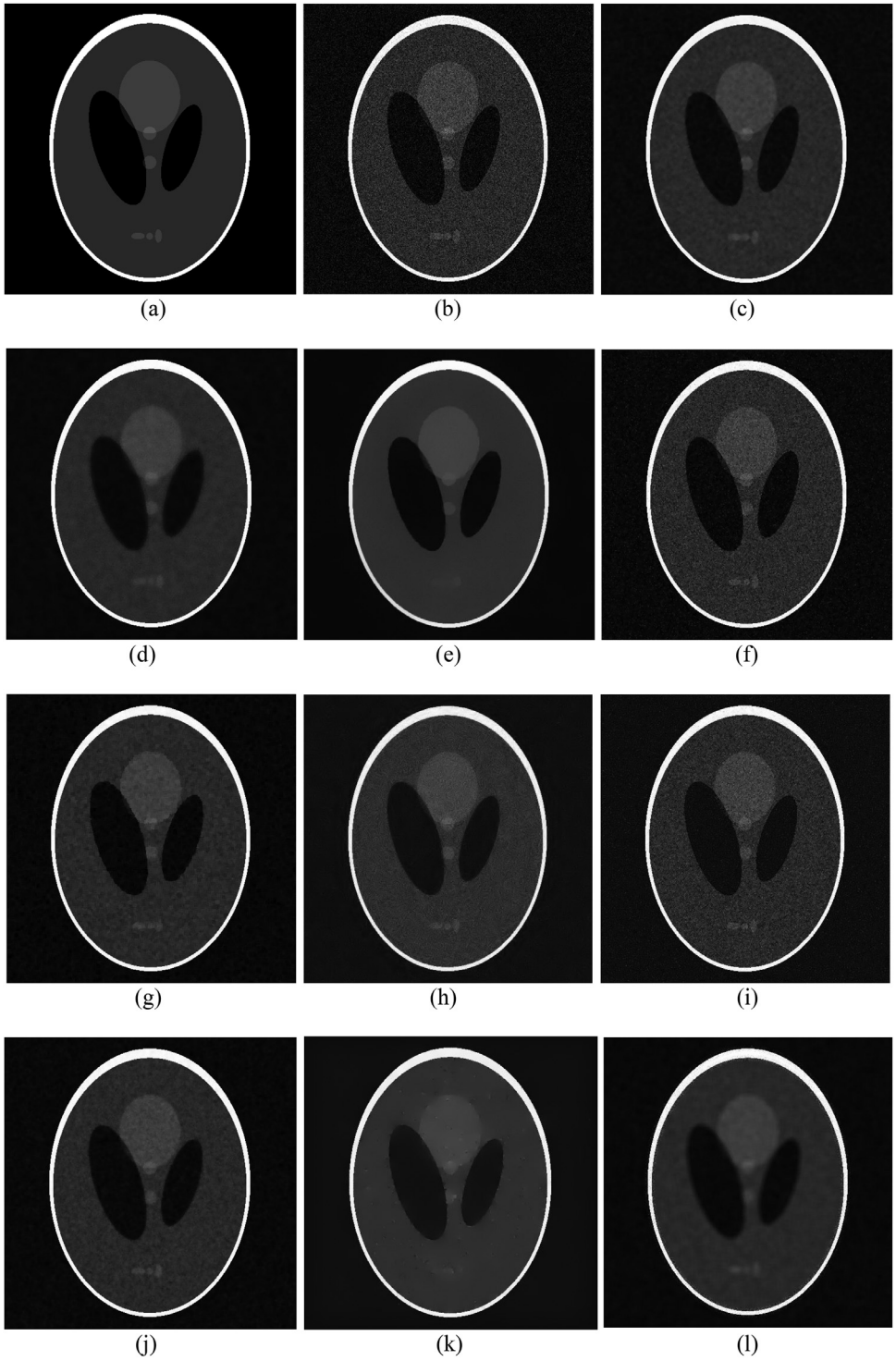


Fig. 4. Shepp–Logan Phantom images denoised with different filters (a) Ground-truth (b) Noisy image ($\sigma = 0.01$) (c) NLM (d) Bilateral filter (e) TV (f) AD (g) Kuwahara (h) Wavelet thresholding (i) LMMSE (j) SUSAN (k) Beltrami (l) Wiener.

blur-free. However, the denoised image suffers from structural loss. AD and LMMSE filters retain a high amount of noise in restored images (Fig. 4(f) and 4(i)) with structural loss but blur not present. An average amount of noise is present in both WT and LMMSE denoised images (Fig. 4(h) and 4(i)) with severe structural loss. However, edge-blur is absent in both the images. SUSAN filter preserves a slight amount of noise (Fig. 4(j)) with slight structural loss. Beltrami filter's denoised image (Fig. 4(k)) has structural loss with negligible residual noise and shows no edge-blur. Blur and noise are present in the output of the Wiener filter in Fig. 4(l). Structural loss is also visible in it.

MOS values of Phantom images denoised by different filters (Figs. 1– 4) are given in Table 1. At low noise level $\sigma = 0.001$, NLM, TV and Beltrami denoised images have high MOS values. Wiener, BF, Kuwahara and SUSAN filters also have shown good MOS values. AD, WT and LMMSE filters have very low MOSs. At noise level, $\sigma = 0.0025$, TV and Beltrami filters have high MOS values. NLM and Wiener filters also have good scores. All other filters SUSAN,

Table 1
Mean Opinion Score of Phantom images restored with various filters.

FILTERS	Noise Level				Summary
	$\sigma = 0.001$	$\sigma = 0.0025$	$\sigma = 0.005$	$\sigma = 0.01$	
NLM	**4.5**	4	3	2.5	3.50±0.91
BF	4	3	2.5	3	3.13±0.63
TV	**4.5**	**4.5**	**4**	**3.5**	**4.13±0.48**
AD	2.5	2	2	2	2.13±0.25
KUWAHARA	4	2.5	2.5	2.5	2.89±0.75
WT	2.5	2.5	3	2.5	2.63±0.25
LMMSE	2.5	2	2	2	2.13±0.25
SUSAN	4	3.5	3	2.5	3.25±0.65
BELTRAMI	**4.5**	**4.5**	**4**	**3.5**	**4.13±0.48**
Wiener	4.25	3.75	2.75	2.75	3.38±0.75

BF, AD, Kuwahara, WT and LMMSE have low values of MOS at this level. Only TV and Beltrami filters have appreciable MOS at noise levels $\sigma = 0.005$ and 0.01. Scores of NLM, BF, AD, Kuwahara, WT, LMMSE and SUSAN filters are not appreciable at this level. TV and Beltrami filters have shown high MOS over all the four noise levels. NLM and Wiener filters have shown good performance at low noise levels $\sigma = 0.001$ and 0.0025. BF, Kuwahara and SUSAN filters have good scores only at low noise level $\sigma = 0.001$. AD, WT and LMMSE filters do not exhibit appreciable performance on any of the four noise levels. Among the filters, only TV and Beltrami have shown consistent performance on all noise levels.

The PSNR, EPI, SSIM and NQM values calculated between ground-truth and Phantom images restored with various denoising schemes, at four different noise levels ($\sigma = 0.001$, 0.0025, 0.005 and 0.01) are given in Table 2. At low noise levels, $\sigma = 0.001$ & 0.0025, BF, Kuwahara and SUSAN have shown high PSNR values. NLM and Wiener filters also have appreciable PSNR values at this noise level. At noise level $\sigma = 0.005$ and 0.01, BF shows highest PSNR value. Kuwahara, SUSAN and NLM filters also have appreciable PSNR values at these noise levels. PSNR value of BF is consistently high overall noise levels. Kuwahara and SUSAN filters also have shown appreciable high values over the four noise levels. NLM shows good values at low noise levels only. TV, AD, WT, LMMSE and Beltrami filters do not show appreciable PSNR values in any of the noise levels.

Considering EPI values at noise level $\sigma = 0.001$, BF, Kuwahara, SUSAN and Beltrami filters have shown high values. NLM and Wiener filters also have good EPI values compared to other filters. At noise level $\sigma = 0.0025$, BF has shown highest EPI. Wiener, Kuwahara, SUSAN, Beltrami and NLM filters have shown good EPI values. At $\sigma = 0.005$, highest EPI values are exhibited by BF and SUSAN filter. NLM, Kuwahara and Beltrami filters also exhibited appreciable scores. BF, SUSAN and Beltrami filters have shown high EPI values at noise level $\sigma = 0.01$. BF and SUSAN filter have shown high EPI values consistently over all four noise levels. NLM, Kuwahara and Beltrami filters also have shown appreciable scores. Wiener filter has good scores at low noise levels. TV, AD, WT and LMMSE filters do not show appreciable scores at any of the four noise levels.

SSIM values of all filters except WT are appreciable at noise level $\sigma = 0.001$. At $\sigma = 0.0025$, Kuwahara and Beltrami filters have shown high scores. NLM, BF, SUSAN and Wiener filters also have shown appreciable scores. At $\sigma = 0.005$, Kuwahara, SUSAN and Beltrami filters have high scores. Wiener, BF and NLM filters also have good scores. The highest SSIM value at noise level $\sigma = 0.01$ is shown by Beltrami filter. NLM, BF, Kuwahara, SUSAN and Wiener filters have also shown appreciable scores. Kuwahara and Beltrami have high SSIM values at all noise levels. SUSAN, NLM, BF, and Wiener filters also have shown appreciable SSIM values. AD filter has shown good SSIM value at low noise level only. TV, LMMSE

Table 2
PSNR, EPI, SSIM and NQM values between ground-truth and Phantom images restored with various filters at four different noise levels.

FILTERS	PSNR				EPI				SSIM				NQM			
	$\sigma = 0.001$	$\sigma = 0.0025$	$\sigma = 0.005$	$\sigma = 0.01$	$\sigma = 0.001$	$\sigma = 0.0025$	$\sigma = 0.005$	$\sigma = 0.01$	$\sigma = 0.001$	$\sigma = 0.0025$	$\sigma = 0.005$	$\sigma = 0.01$	$\sigma = 0.001$	$\sigma = 0.0025$	$\sigma = 0.005$	$\sigma = 0.01$
NLM	37.83	34.61	31.97	29.08	0.9848	0.9786	0.9697	0.9504	0.9973	0.9940	0.9890	0.9788	30.33	26.82	23.92	20.88
BF	**40.85**	**36.88**	**34.37**	**31.28**	**0.9968**	**0.9922**	**0.9867**	**0.9655**	0.9978	0.9947	0.9897	0.9806	**31.50**	27.95	**25.34**	**22.91**
TV	31.17	29.42	28.88	27.42	0.9575	0.9507	0.9234	0.9079	0.9840	0.9758	0.9721	0.9663	27.90	26.81	**25.60**	**22.46**
AD	35.91	31.27	27.92	24.84	0.9659	0.8973	0.8001	0.6723	0.9961	0.9892	0.9773	0.9544	30.29	26.48	23.59	20.6637
KUWAHARA	**40.21**	**36.19**	33.16	30.21	0.9939	0.9834	0.9668	0.9404	**0.9985**	**0.9962**	**0.9924**	0.9853	**31.49**	**28.07**	25.26	22.05
WT	32.46	29.39	26.73	23.72	0.9211	0.9074	0.9050	0.7282	0.6290	0.6248	0.6233	0.6041	28.26	23.50	20.66	18.45
LMMSE	32.35	28.57	23.29	22.80	0.9382	0.8337	0.7264	0.6082	0.9904	0.9792	0.9608	0.9235	26.47	23.83	13.91	17.99
SUSAN	**40.85**	**36.39**	33.56	30.42	**0.9970**	0.9887	**0.9818**	0.9614	0.9980	0.9949	0.9901	0.9807	**31.51**	27.82	25.09	22.10
BELTRAMI	26.82	25.75	26.49	24.84	0.9936	0.9885	0.9798	**0.9661**	**0.9989**	**0.9976**	**0.9957**	**0.9919**	21.10	20.70	20.02	18.93
Wiener	37.11	33.87	31.4534	28.8847	0.9879	0.9701	0.9439	0.8921	0.9969	0.9937	0.9888	0.9799	29.77	26.46	24.07	21.3476

and WT filters have not shown good SSIM values for any of the four noise levels.

Considering NQM values at $\sigma = 0.001$, BF, Kuwahara and SUSAN filters have high values. NLM, Wiener and AD filters also have shown good values. At $\sigma = 0.0025$, Kuwahara filter has highest NQM value. BF and SUSAN filter also have shown good scores. BF, TV, Kuwahara and SUSAN filters have good scores at noise levels $\sigma = 0.005$ and 0.01. Wiener filter's NQM values are also appreciable at these noise levels. BF, Kuwahara and SUSAN filters have consistently good NQM values overall four noise levels. Wiener filter also has shown appreciable score at all noise levels. NLM and AD filters have good NQM value at the low noise level ($\sigma = 0.001$) only. The TV filter has shown appreciable value at high noise levels ($\sigma = 0.005$ and 0.01). WT, LMMSE and Beltrami filters never exhibit appreciable NQM scores in any of the noise levels.

As per MOS values, the filters performing well on Phantom images are Beltrami, TV and NLM. However, the filters which have shown high scores equally for PSNR, EPI, SSIM and NQM, are BF, Kuwahara and SUSAN. TV, AD, WT and LMMSE filters are failed to give good scores for any of these evaluation metrics. It is very clear that the objective evaluation of filters on Phantom images (Table 2) is not on par with visual quality evaluation through MOS (Table 1). Another observation is that, some filters show different performance for different noise levels. For example, NLM and Wiener filter have shown good PSNR values for low noise levels only. SUSAN filter has good EPI values only for two noise levels ($\sigma = 0.001$ and 0.005). Beltrami filter have shown high EPI value at noise level $\sigma = 0.01$ only. Kuwahara filter has high NQM values only at noise levels $\sigma = 0.001$ and $\sigma = 0.0025$.

PSNR is considered as a common image quality measure metric. However, very low PSNR values are observed for Beltrami and TV filters which have shown high visual quality and thus high MOS values. For example, Fig. 1(e), 1(f), 2(e) and 2(f) shown high visual clarity have low PSNR values among other filters. It is also observed that some images showing equal PSNR values have extremely different visual quality. For example Fig. 3(k) and 3(h) of Beltrami and WT filters with PSNR values 26.49 and 26.73 respectively have shown extremely different visual quality. Noise is visible in Fig. 3(h) but Fig. 3(k) is noise-free. Another typical example is between Fig. 2(e) and 2(h) of filters TV and WT. Both figures have almost equal PSNR values 29.42 and 29.39 respectively. However, on visual analysis, it is very evident that Fig. 2(e) is absolutely noise free and Fig. 2(h) contains a severe amount of noise. The same problem can be observed with Fig. 3(e) and 3(f). Fig. 3(e) is completely noise free and Fig. 3(f) is highly noise, but both have almost equal PSNR values.

The values of MetricQ, AQI and BRISQUE on Phantom images denoised by different filters and their computational time in seconds are given in Table 3. At noise level $\sigma = 0.001$, NLM, TV and WT filters have shown high MetricQ values. BF, Kuwahara and SUSAN filters also have shown appreciable scores. NLM and TV filters have shown high MetricQ values at $\sigma = 0.0025$ and 0.005. BF, Kuwahara and SUSAN filters also have good scores at these noise levels. At noise level $\sigma = 0.01$, NLM filter has the highest score. BF and SUSAN filters also have comparably good scores at this noise level. NLM filter has shown highest MetricQ value at all four noise levels. TV and Kuwahara filters have shown high scores at $\sigma = 0.001$, 0.0025 and 0.005. At high noise level $\sigma = 0.01$, the scores of those filters are not appreciable. WT has a good score only at low noise level $\sigma = 0.001$. BF and SUSAN filter exhibit appreciable scores at all the noise levels consistently. AD, LMMSE and Beltrami filters do not show appreciable MetricQ values in any of the four noise levels.

AQI values of all filters at four different noise levels are observed to be the same. Therefore AQI is improper for the quantitative evaluation of Phantom images.

Table 3
MetricQ, AQI, BRISQUE values and computational time in seconds of Phantom images restored with different filters, at different noise levels.

FILTERS	Metric Q				AQI				BRISQUE				COMP. TIME (s)
	$\sigma = 0.001$	$\sigma = 0.0025$	$\sigma = 0.005$	$\sigma = 0.01$	$\sigma = 0.001$	$\sigma = 0.0025$	$\sigma = 0.005$	$\sigma = 0.01$	$\sigma = 0.001$	$\sigma = 0.0025$	$\sigma = 0.005$	$\sigma = 0.01$	
NLM	**0.0817**	**0.0766**	**0.0742**	**0.0722**	0.0016	0.0016	0.0016	0.0016	38.9	36.46	39.82	39.85	53.10±0.12
BF	0.0782	0.0751	0.0736	0.0696	0.0016	0.0016	0.0016	0.0016	43.1	42.29	34.35	25.26	2.66±0.23
TV	**0.0814**	**0.0772**	**0.0749**	0.0688	0.0016	0.0016	0.0016	0.0016	34.88	**21.32**	31.05	**17.53**	3.82±0.10
AD	0.0747	0.0732	0.0703	0.0644	0.0016	0.0016	0.0016	0.0016	49.69	46.19	44.52	44.4	1.06±0.12
KUWAHARA	0.0786	0.0748	0.0721	0.0685	0.0016	0.0016	0.0016	0.0016	**22.46**	31.17	34.36	33.13	45.50±0.08
WT	**0.08**	0.0694	0.0664	0.0610	0.0016	0.0016	0.0016	0.0016	30.12	22.75	**18.89**	32.48	**0.13±0.001**
LMMSE	0.0755	0.0704	0.0536	0.0619	0.0016	0.0016	0.0016	0.0016	38.7	42.75	40.37	42.99	**0.04±0.01**
SUSAN	0.0785	0.0749	0.0731	0.0698	0.0016	0.0016	0.0016	0.0016	44.24	38.54	43.54	41.56	2.63±0.19
BELTRAMI	0.0698	0.0670	0.0647	0.0634	0.0016	0.0016	0.0016	0.0016	49.55	49.1	48.87	48.78	12.79±10.84
Wiener	0.0762	0.0687	0.0643	0.0574	0.0016	0.0016	0.0016	0.0016	37.45	35.41	38.53	38.41	**0.015±0.001**

At noise level $\sigma = 0.001$, Kuwahara filter has appreciable low BRISQUE value. WT also has shown good BRISQUE score at this noise level. TV and WT filters express good scores at noise level $\sigma = 0.0025$ and $\sigma = 0.005$. At $\sigma = 0.01$, TV and BF filters have shown good scores. BF have appreciable score only at $\sigma = 0.01$. At the same time, Kuwahara filter has shown good scores only at low noise level $\sigma = 0.001$. The TV filter has appreciable scores at noise levels $\sigma = 0.0025$, 0.005 and 0.01. WT have appreciable scores at $\sigma = 0.001$, 0.0025 and 0.005. NLM, AD, LMMSE, SUSAN, Wiener and Beltrami filters do not exhibit good scores at any of the four noise levels.

As per MetricQ values, NLM, TV, BF, Kuwahara and SUSAN filters have shown appreciable scores. But these observations in Table 3 are not on par with those MOS values given in Table 1. AQI may not suit for quality evaluation of Phantom images as it offers constant value for all filters. Even BRISQUE values are not on concordance with visual quality evaluation. None of the quality metric evaluation values on Phantom images is on par with the subjective evaluation.

While analyzing the computational time of all filters used in this work, Wiener filter is observed to be the fastest one. Another filter showing less computational time is LMMSE, but its denoising performance is not appreciable. The filter which has good denoising performance, NLM, is computationally very complex. Kuwahara filter is also has high computational time. TV and Beltrami filters which have good MOS values have moderate computational time.

3.2. Simulated MR images

Simulated MR images at three different noise levels ($\sigma = 1\%$, 3% and 5%) restored with NLM, Bilateral, TV, AD, Kuwahara, WT, LMMSE, SUSAN and Beltrami filters are shown in Figs. 5–7. On the visual evaluation of denoised images at noise level $\sigma = 1\%$ (Fig. 5(b)), NLM and SUSAN filters (Fig. 5(c) and 5(j)) have shown noise-free images without blur and structural loss. Fig. 5(d), 5(e) and 5(f) of BF, TV and AD filters have not shown residual noise. But images are blurred and they have no ground-truth similarity. Cartoon-like effect is visible in Kuwahara filter's image (Fig. 5(g)). As though residual noise is not there in the image, the similarity with ground-truth is very less. Noise is not visible on Fig. 5(h), 5(i), 5(k) and 5(l) corresponds to WT, LMMSE, Beltrami and Wiener filters respectively. Images are not blurred but it does not maintain ground-truth similarity as such.

At noise level $\sigma = 3\%$ (Fig. 5(b)), NLM and SUSAN filters (Fig. 6(c) and 6(j)) have produced noiseless, blur-free images without structural loss. Restored image of BF in Fig. 6(d) have shown residual noise and blur. Fig. 6(e) and 6(f) of TV and AD filters are over-blurred but noise is absent. Kuwahara filtered image (Fig. 6(g)) is suffered from cartoon-effect and thus it loses similarity with ground-truth. Residual noise is also present in this image. Slight noise and blur are visible in image denoised with WT (Fig. 6(h)). Fig. 6(i) of LMMSE filter has shown a moderate amount of noise but blur-free. Fig. 6(k) of the Beltrami filter is over-smoothened but noise-free. Fig. 6(l) of the Wiener filter is slightly noisy. It is free from blur and preserves similarity with ground-truth.

At noise level $\sigma = 5\%$ (Fig. 7(b)), NLM filter has produced noise-free image (Fig. 7(c)) without blur. But it has no ground-truth similarity. Image denoised by BF filter is (Fig. 7(d)) blur-free but it retains noise. TV and AD filter have produced blurred images (Fig. 7(e) and 7(f)) without residual noise. Fig. 7(g) of Kuwahara contains noise and has cartoon artefact. Slight amount of noise and blur is visible in WT's image (Fig. 7(h)). Output of the LMMSE filter contains a slight amount of noise in Fig. 7(i). SUSAN filter has produced a noise-free image (Fig. 7(j)) with no blur and structural loss. Fig. 7(k) of Beltrami filter is over-smoothened but noise is absent. The Wiener filter (Fig. 7(l)) has produced a slightly nosily image with negligible structural loss. The image is free from blur.

Table 4
Mean Opinion Score of simulated images denoised by different filters.

FILTERS	Noise Level			Summary
	$\sigma = 1\%$	$\sigma = 3\%$	$\sigma = 5\%$	
NLM	**4**	**4**	3.75	**3.92±0.14**
BF	2.5	3	3	2.83±0.29
TV	2	2	2	2 ± 0
AD	2.5	2.5	3.5	2.83±0.58
KUWAHARA	2	2	2	2 ± 0
WT	3.75	3.75	**4**	3.83±0.14
LMMSE	3	3	3	3 ± 0
SUSAN	**4**	**4**	**4**	**4 ± 0**
BELTRAMI	3.5	3.5	3.5	3.5 ± 0
Wiener	3.75	3.5	3.5	3.58±0.1443

MOS of each denoised images in Figs. 5–7 are given in Table 4. At low noise level ($\sigma = 1\%$), denoised images of NLM and SUSAN filters have shown high MOS score. Beltrami, Wiener and WT filters also have appreciable MOS values. MOS values expressed by other filters at low noise level are not appreciable. At $\sigma = 3\%$, NLM and SUSAN filters have shown high scores. WT, Wiener and Beltrami filters are also expressed appreciable MOS values. At $\sigma = 5\%$, WT and SUSAN filters have shown high MOS values. NLM, Wiener, Beltrami and AD filters also have good scores. Considering MOS of all filters at three different noise levels, SUSAN has the highest MOS score. The MOS value of NLM filter is also good. NLM has high scores at $\sigma = 1\%$ and 3% level, but its score reduced slightly on $\sigma = 5\%$. WT and Beltrami filters have shown good scores at three noise levels. BF, TV and Kuwahara filters do not show appreciable MOS at any of the noise levels. AD filter has shown good score only at $\sigma = 5\%$. At all other noise levels, its MOS is not good.

The PSNR, EPI, SSIM and NQM metrics values of simulated images denoised with different filters are given in Table 5. On analyzing PSNR values at noise level $\sigma = 1\%$, LMMSE and SUSAN filters have exhibit high values. WT and Beltrami filters also have shown appreciable PSNR values at this noise level. NLM and SUSAN filters have shown good performance at noise level $\sigma = 3\%$. AD, WT, LMMSE, Wiener and Beltrami filters are also exhibited good PSNR scores. At noise level $\sigma = 5\%$, NLM and AD have high PSNR values. WT, LMMSE, SUSAN and Wiener filters also have appreciable PSNR values at this level. BF, TV and Kuwahara filters have not shown appreciable PSNR values at any of the three noise levels. WT, LMMSE and SUSAN filters have shown appreciable scores at all the three noise levels. NLM filter has good scores at noise levels $\sigma = 3\%$ and 5%. Beltrami filter has good scores at $\sigma = 1\%$ and $\sigma = 3\%$.

Considering EPI scores at noise level $\sigma = 1\%$, LMMSE and Beltrami filters have high scores. Other filters showing appreciable performance at this level are WT and SUSAN. NLM has the highest EPI value at noise level $\sigma = 3\%$. WT, LMMSE and SUSAN filters also have appreciable scores. At noise level $\sigma = 5\%$ NLM has highest EPI value. Other filters shown good scores at this noise level are WT, BF and AD. WT filter has shown good EPI values over three noise levels. NLM filter has shown good scores at $\sigma = 3\%$ and $\sigma = 5\%$. BF and AD have shown good values only at high noise level $\sigma = 5\%$. LMMSE, SUSAN and Beltrami filters have appreciable EPI scores over noise levels $\sigma = 1\%$ and 3%.

LMMSE filter has highest SSIM value at noise level $\sigma = 1\%$. Other filters have good scores at this noise level are NLM, WT, SUSAN and Beltrami. NLM also has an appreciable score at this noise level. WT, Wiener and LMMSE filters have high values at noise level $\sigma = 3\%$. Beltrami, NLM, BF and SUSAN filters also have appreciable good scores at this level. WT and Wiener filters have high

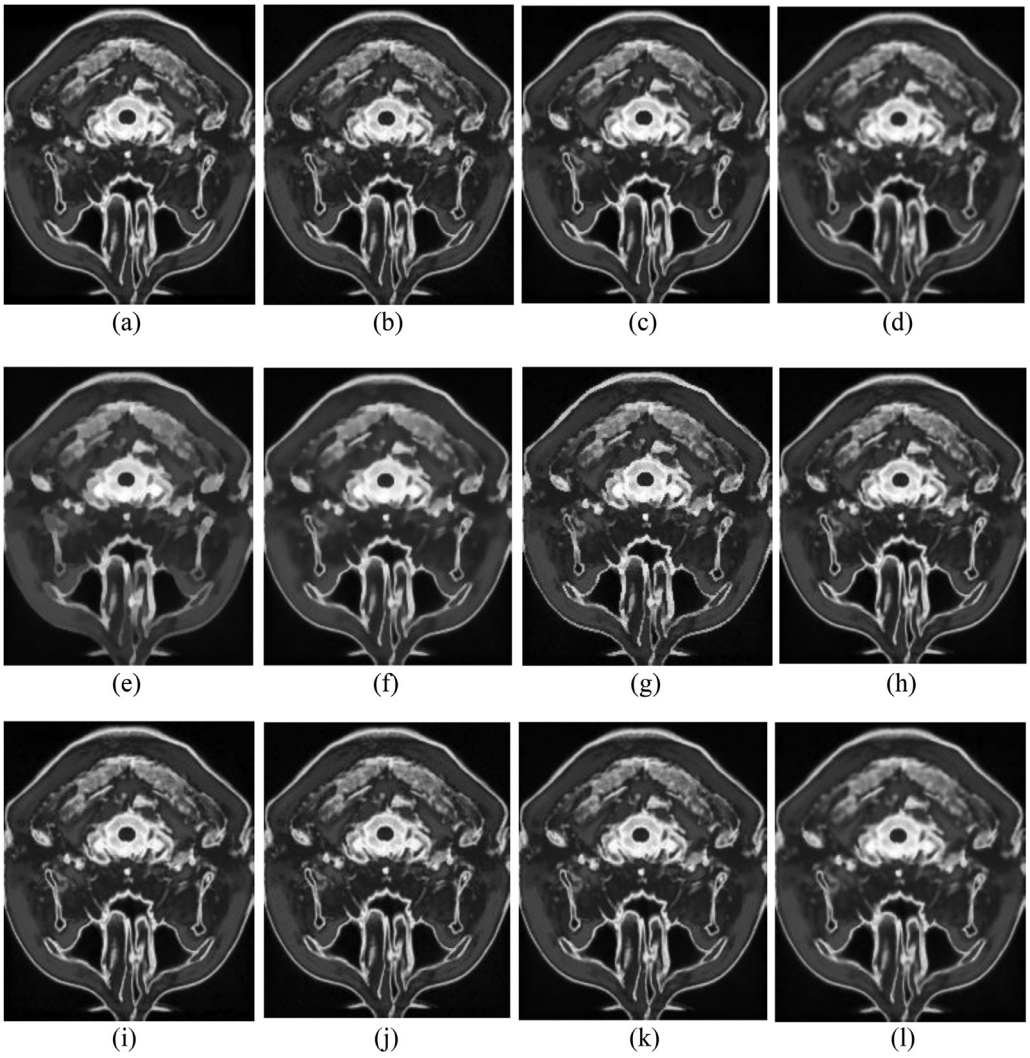


Fig. 5. Simulated noisy MR images denoised with different filters (a) Ground-truth (b) Noisy image ($\sigma = 1\%$) (c) NLM (d) Bilateral filter (e) TV (f) AD (g) Kuwahara (h) Wavelet thresholding (i) LMMSE (j) SUSAN (k) Beltrami (l) Wiener.

Table 5
PSNR, EPI, SSIM, NQM values between ground-truth and simulated MR images restored with different filters, at three different noise levels.

FILTERS	PSNR			EPI			SSIM			NQM		
	$\sigma = 1\%$	$\sigma = 3\%$	$\sigma = 5\%$	$\sigma = 1\%$	$\sigma = 3\%$	$\sigma = 5\%$	$\sigma = 1\%$	$\sigma = 3\%$	$\sigma = 5\%$	$\sigma = 1\%$	$\sigma = 3\%$	$\sigma = 5\%$
NLM	34.79	**30.58**	**27.09**	0.9578	**0.9380**	**0.9071**	0.9967	0.9819	0.9502	27.24	**22.15**	**18.65**
BF	28.44	27.17	25.87	0.8921	0.8810	0.8536	0.9869	0.9817	0.9744	19.77	17.60	15.85
TV	24.73	23.91	22.98	0.7297	0.6871	0.6301	0.9687	0.9624	0.9533	18.46	17.47	15.82
AD	30.25	29.07	**27.02**	0.9085	0.8910	0.8510	0.9896	0.9751	0.9467	24.13	20.98	17.93
KUWAHARA	27.08	25.78	23.82	0.7891	0.7486	0.6927	0.9830	0.9691	0.9405	21.15	18.30	15.39
WT	37.33	29.57	26.27	0.9793	0.9299	0.8756	0.9984	**0.9896**	**0.9771**	**29.37**	19.70	16.97
LMMSE	**38.81**	29.87	26.28	**0.9836**	0.9120	0.8218	**0.9986**	**0.9851**	0.9576	**29.12**	20.47	16.64
SUSAN	**38.37**	**30.54**	26.97	0.9792	0.9122	0.8371	0.9983	0.9817	0.9471	**29.71**	21.72	17.88
BELTRAMI	37.56	29.15	25.60	**0.9801**	0.9021	0.8119	0.9982	0.9820	0.9439	28.87	20.00	15.90
Wiener	32.03	29.14	26.70	0.9435	0.9067	0.8425	0.9944	**0.9886**	**0.9793**	24.29	19.99	17.04

SSIM values at noise level $\sigma = 5\%$. BF's score is also appreciable at this noise level. SSIM value of WT is observed to be appreciable at all noise levels. BF and Wiener filter have appreciable scores at $\sigma = 3\%$ and 5%. SUSAN, NLM and Beltrami filters have good scores at $\sigma = 1\%$ and 3%.

While analysing NQM scores at noise level $\sigma = 1\%$, it is observed that the filters WT, LMMSE and SUSAN have high values. Beltrami filer also has the appreciable score at this noise level. NLM filter has the highest NQM values at $\sigma = 3\%$ and 5%. SUSAN and AD filters also have good scores at these noise levels. SUSAN and NLM filters have shown good NQM scores over all the three noise levels. AD filter has shown good scores at noise levels $\sigma = 3\%$ and $\sigma = 5\%$. LMMSE, WT and Beltrami filters have good scores only at $\sigma = 1\%$. BF, TV, Wiener and Kuwahara filters have not shown good NQM scores over any of the three noise levels.

MetricQ, AQI and BRISQUE values of simulated MR images denoised by different filters and their computational time are given in Table 6. NLM filter has shown the highest MetricQ value at noise level $\sigma = 1\%$, 3% and 5%. LMMSE, SUSAN and WT filters also have shown good scores at $\sigma = 1\%$. At noise level, $\sigma = 3\%$ SUSAN filter

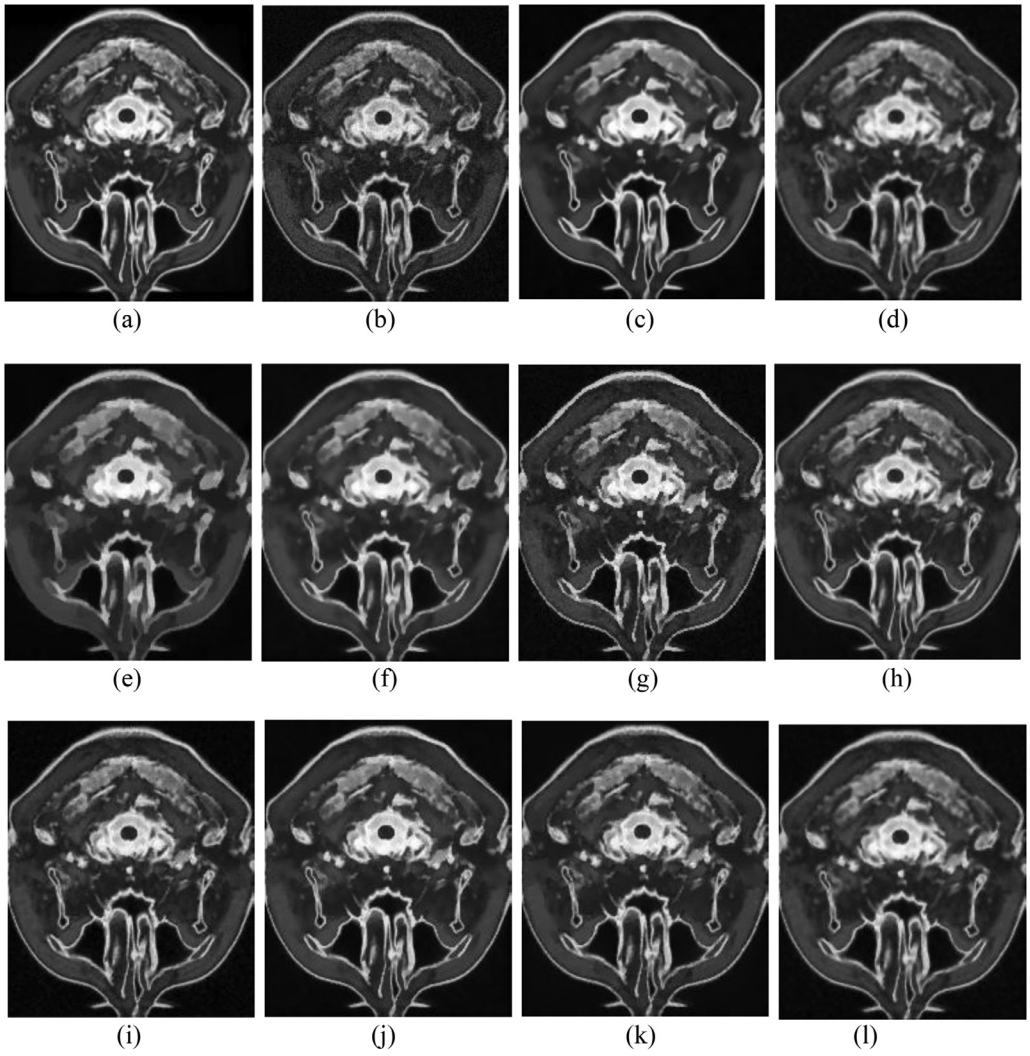


Fig. 6. Simulated noisy MR images denoised with different filters (a) Ground-truth (b) Noisy image ($\sigma = 3\%$) (c) NLM (d) Bilateral filter (e) TV (f) AD (g) Kuwahara (h) Wavelet thresholding (i) LMMSE (j) SUSAN (k) Beltrami (l) Wiener.

Table 6
MetricQ and AQI values of simulated images restored with various filters, at different noise levels and the computational time in seconds of all filters.

FILTERS	Metric Q			AQI			BRISQUE			COMP. TIME (s)
	$\sigma = 1\%$	$\sigma = 3\%$	$\sigma = 5\%$	$\sigma = 1\%$	$\sigma = 3\%$	$\sigma = 5\%$	$\sigma = 1\%$	$\sigma = 3\%$	$\sigma = 5\%$	
NLM	**0.3030**	**0.2909**	**0.2385**	0.0059	0.0057	0.0054	**25.87**	31.58	39.69	7.9203±0.0579
BF	0.2356	0.2108	0.1742	0.0053	0.0052	0.0051	43.41	43.46	43.46	0.2470±0.0011
TV	0.1878	0.1657	0.1303	0.0057	0.0056	0.0053	43.46	43.46	43.46	0.1697±0.0110
AD	0.2570	0.2356	0.1926	0.0060	0.0059	0.0056	37.03	42.92	42.63	0.1898±0.0058
KUWAHARA	0.2459	0.2098	0.1635	**0.0064**	**0.0064**	**0.0060**	43.46	43.46	43.46	6.7821±0.1149
WT	0.2904	0.2457	0.1957	0.0055	0.0056	0.0052	28.44	**23.15**	**14.39**	**0.0316±0.0004**
LMMSE	0.2917	0.2441	0.1864	0.0056	0.0054	0.0051	33.76	42.50	42.70	**0.0501±0.0037**
SUSAN	0.2948	0.2571	0.1938	0.0058	0.0059	0.0054	32.70	38.31	39.68	0.4456±0.0444
BELTRAMI	0.2880	0.2374	0.1821	0.0055	0.0055	0.0054	28.08	26.6	29.44	0.7239±0.7988
Wiener	0.2604	0.2283	0.1826	0.0056	0.0056	0.0053	30.31	28.95	31.13	**0.0050±0.0051**

also exhibit good values. At $\sigma = 5\%$ AD, WT and SUSAN filters also have shown appreciable MetricQ values. AD filter has good score only at noise level 5%. Beltrami filter has shown good score only at noise level $\sigma = 1\%$. BF, Kuwahara, Beltrami, Wiener and TV filters do not show good scores at any of the three noise levels.

While analyzing AQI values of different filters, it is observed that Kuwahara filter has the highest value at noise levels $\sigma = 1\%$, 3% and 5%. AD and NLM filters also have appreciable scores at noise levels $\sigma = 1\%$. Moreover, AD and SUSAN filters also have shown good scores at noise level $\sigma = 3\%$. AD filter has exhibited appreciable scores at $\sigma = 5\%$. BF, TV, WT, Beltrami, Wiener and LMMSE filters have not shown good scores at any of the noise levels. Beltrami filter has good score only at $\sigma = 5\%$. NLM filter has good score only at noise level $\sigma = 1\%$.

NLM has appreciable BRISQUE value (low) at noise level $\sigma = 1\%$. WT and Beltrami filters also have shown good scores at this noise level. But at noise level, $\sigma = 3\%$ and 5% WT have shown the appreciable scores. Beltrami filter also has shown appreciable scores good at noise level 3%. Beltrami, SUSAN and NLM have shown good scores at this level $\sigma = 3\%$ and 5%. The filters which

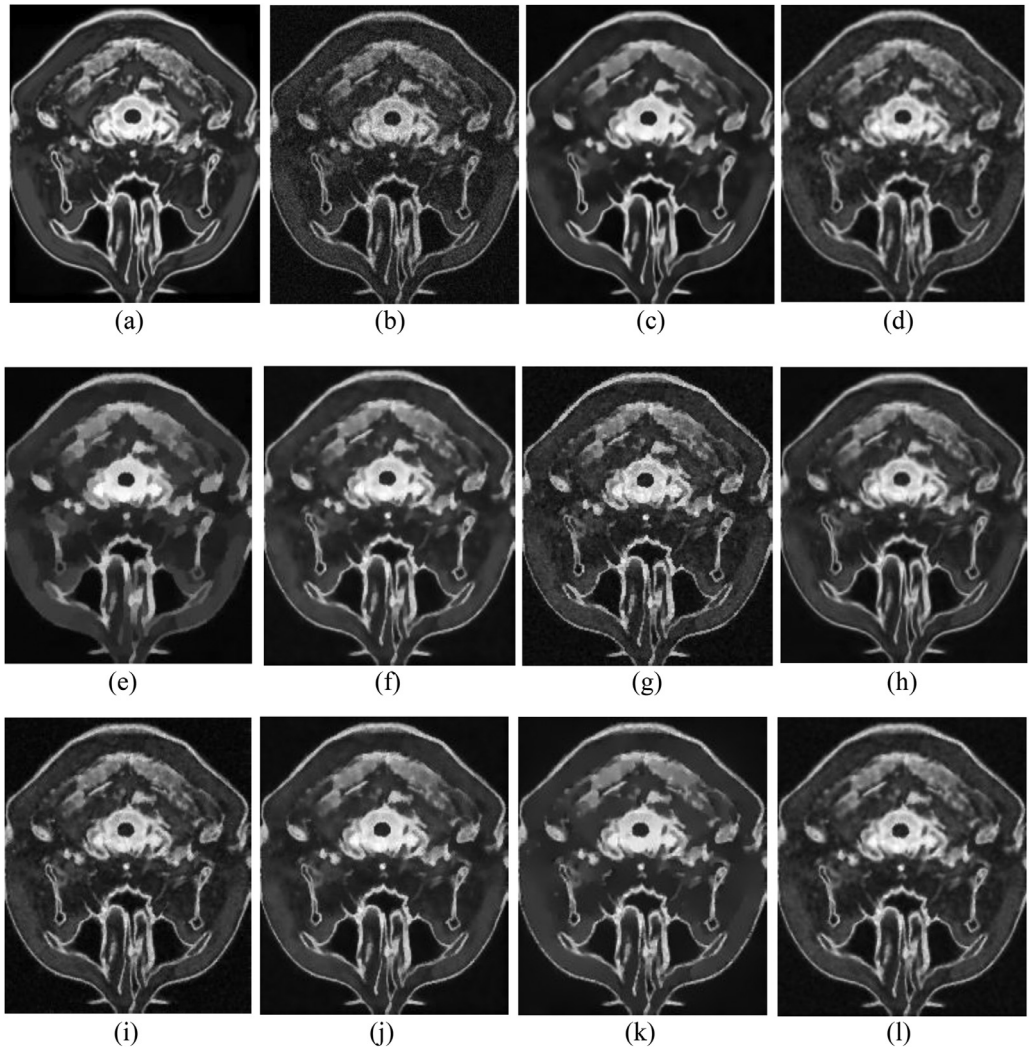


Fig. 7. Simulated noisy images denoised with different filters (a) Ground-truth (b) Noisy image ($\sigma = 5\%$) (c) NLM (d) Bilateral filter (e) TV (f) AD (g) Kuwahara (h) Wavelet thresholding (i) LMMSE (j) SUSAN (k) Beltrami (l) Wiener.

have not shown good BRISQUE at any of the noise levels are BF, TV, AD, Kuwahara, SUSAN, Wiener and LMMSE. NLM filter has good score only at noise level $\sigma = 1\%$. Beltrami filter has shown good scores at all three noise levels. WT has good scores at noise level $\sigma = 3\%$ and 5%.

On simulated images also, Wiener filter is the computationally fastest one. LMMSE filter also have shown less computational time but its denoising performance is not appreciable. NLM and Kuwahara filters are computationally very complex filters. WT and SUSAN filters shown good restoration result on simulated images have appreciable processing time also.

From the above analysis, it is very clear that filters performance on simulated images is different for different noise levels. The PSNR values of different filters produced in Table 5 are not on par with observations of MOS in Table 4. The TV and Beltrami filters have shown high MOS values on all noise levels and have good visual clarity. However, image quality metrics have failed to produce values on par with quality evaluation results. For example, while observing the PSNR values on Table 5; it is seen that LMMSE and AD filters have high scores. However, the denoised images correspond to these filters (Figs. 5(f), 5(i), 6(f) and 6(i) do not offer high visual clarity and edge-preservation as per the PSNR values. EPI, SSIM, NQM, MetricQ values given in Table 5 are on par with MOS values in Table 4. AQI and BRISQUE metrics have shown completely mismatching scores with MOS values. Example AQI and BRISQUE scores show KUWAHARA and WT as best filters and this is not on par with MOS values.

3.3. Real-time MR images

Real MR images denoised with different edge-preserving filters NLM, Bilateral, TV, AD, Kuwahara, wavelet thresholding, LMMSE, SUSAN and Beltrami are shown in Figs. 8–10. Among denoised images in Fig. 8, the output image of NLM filter (Fig. 8(b)) contains no residual noise, no blur and negligible structural loss. Fig. 8(c) of BF is blurred and it contains noise and structural loss. Slight structural loss is there in image denoised with TV filter (Fig. 8(d)), but it is free from noise and blur. Mean brightness of the image is slightly changed by TV filter. Fig. 8(e) of AD filter has exhibited no blur, no residual noise and negligible structural loss. Fig. 8(f), image denoised by Kuwahara, is suffered from cartoon-artefact and structural loss. But it does not contain residual noise. WT over smoothens the image (Fig. 8(g)), but noise is not visible in it. LMMSE filter's denoised image (Fig. 8(h)) contains noise but without blur. Image 1 denoised by SUSAN filter (Fig. 8(i)) contains slight noise. But it has no edge-blur or structural loss. Beltrami filter's denoised image (Fig. 8(j)) is completely free from residual noise, edge-blur and structural loss. Slight blur is visible

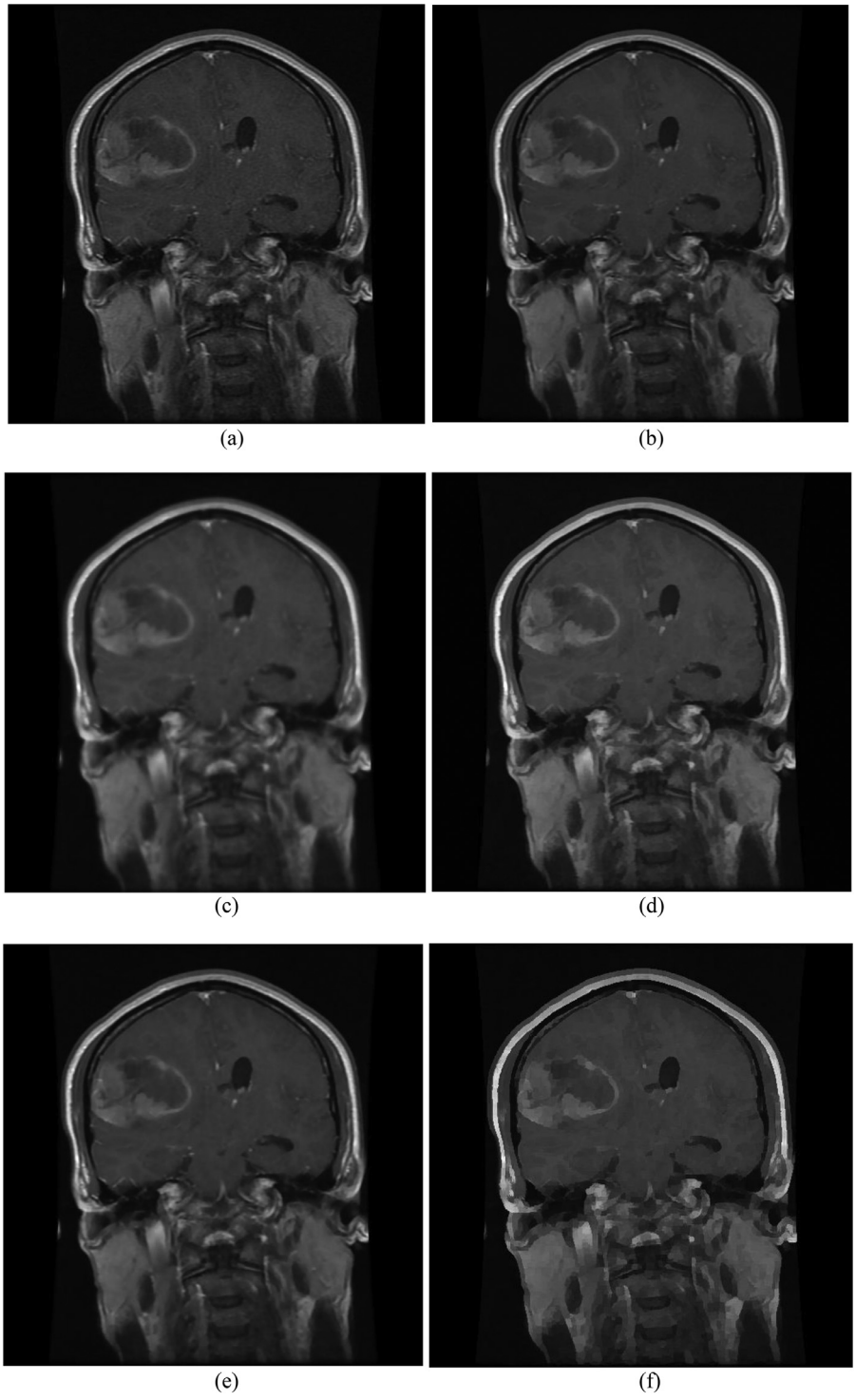
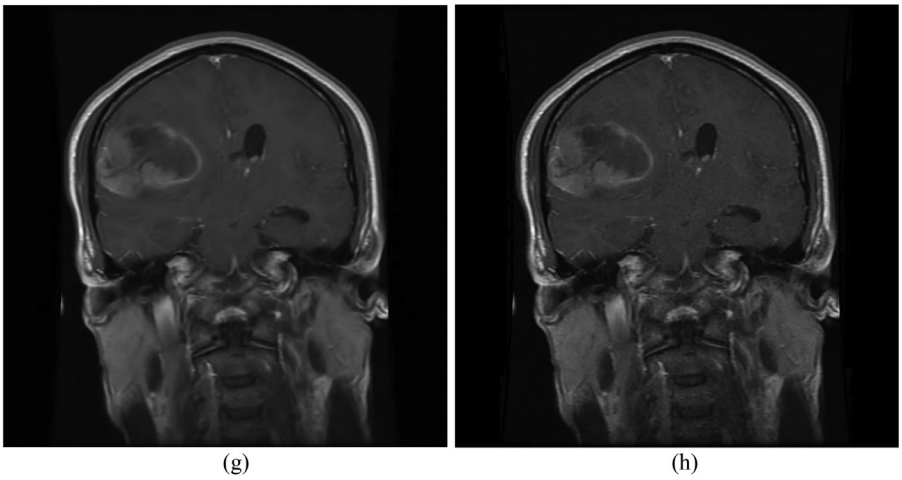


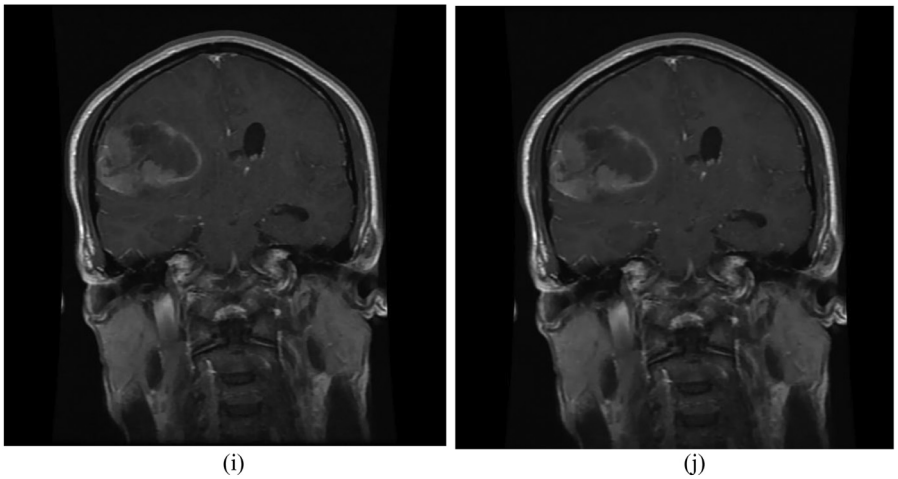
Fig. 8. Real MR images denoised with different filters (a) MR Image 1 (b) NLM (c) Bilateral filter (d) TV (e) AD (f) Kuwahara (g) Wavelet thresholding (h) LMMSE (i) SUSAN (j) Beltrami (k) Wiener.

in Wiener filter's denoised image. Image is not absolutely noise free.

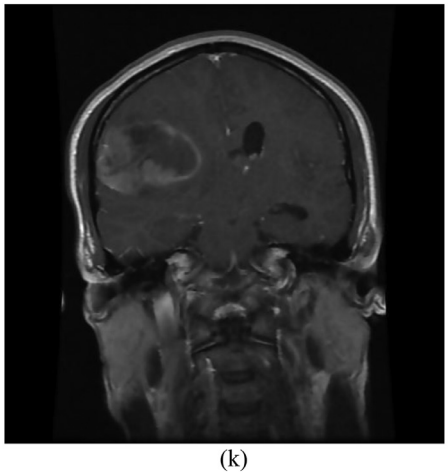
In Fig. 9, NLM filter (Fig. 9(b)) has generated a noise-free image with structural loss. Output image of BF (Fig. 9(c)) has exhibited slight noise, edge-blur and structural loss. Fig. 9(d), the output image of TV filter is free from noise and blur. But minute structural loss is there in it. Output image of AD filter (Fig. 9(e)) is noise-free but edge blur and structural loss are visible. In Kuwahara filter's denoised image (Fig. 9(f)) cartoon artefact is severe. It has edge-blur and structural loss but the image is free from noise. Fig. 9(g)), the output image of WT filter is noise free but shows over-smoothening with structural loss. Slightly noisy image (Fig. 9(h)) is generated by the LMMSE filter. Edges are not preserved well in that image but free from blur. The output image of SUSAN filter (Fig. 9(i)) has shown a negligible structural loss, contains no noise and no edge-blur. Beltrami filter generates a



(g) (h)



(i) (j)



(k)

Fig. 8. Continued

complete noise-free image (Fig. 9(j)) without edge-blur and structural loss. Wiener filter has generated a noise-free image (Fig. 9(k)). But it has edge-blur.

Image denoised by NLM (Fig. 10(b)) is observed to be free from noise, blur and structural loss. Edge-blur and structural loss are visible in the denoised image of BF (Fig.10(c)), but it is free from noise. It is noticeable that TV filter changes the mean brightness of the image 2 while denoising (Fig. 10(d)). Even though mean brightness changed, denoised image is free from noise, blur and structural loss. AD filtered image (Fig. 10(e)) shows structural loss and slight edge-blur, but noise free. Image denoised by Kuwahara (Fig. 10(f)) is noise-free but it has cartoon artefact. Mean brightness of the image is slightly changed and have structural loss. Fig. 10(g) of WT has slight blur and structural loss but it is free from noise. LMMSE filter's denoised image (Fig. 10(h)) retains noise and shows structural loss but blur-free. Fig. 10(i), denoised image of SUSAN

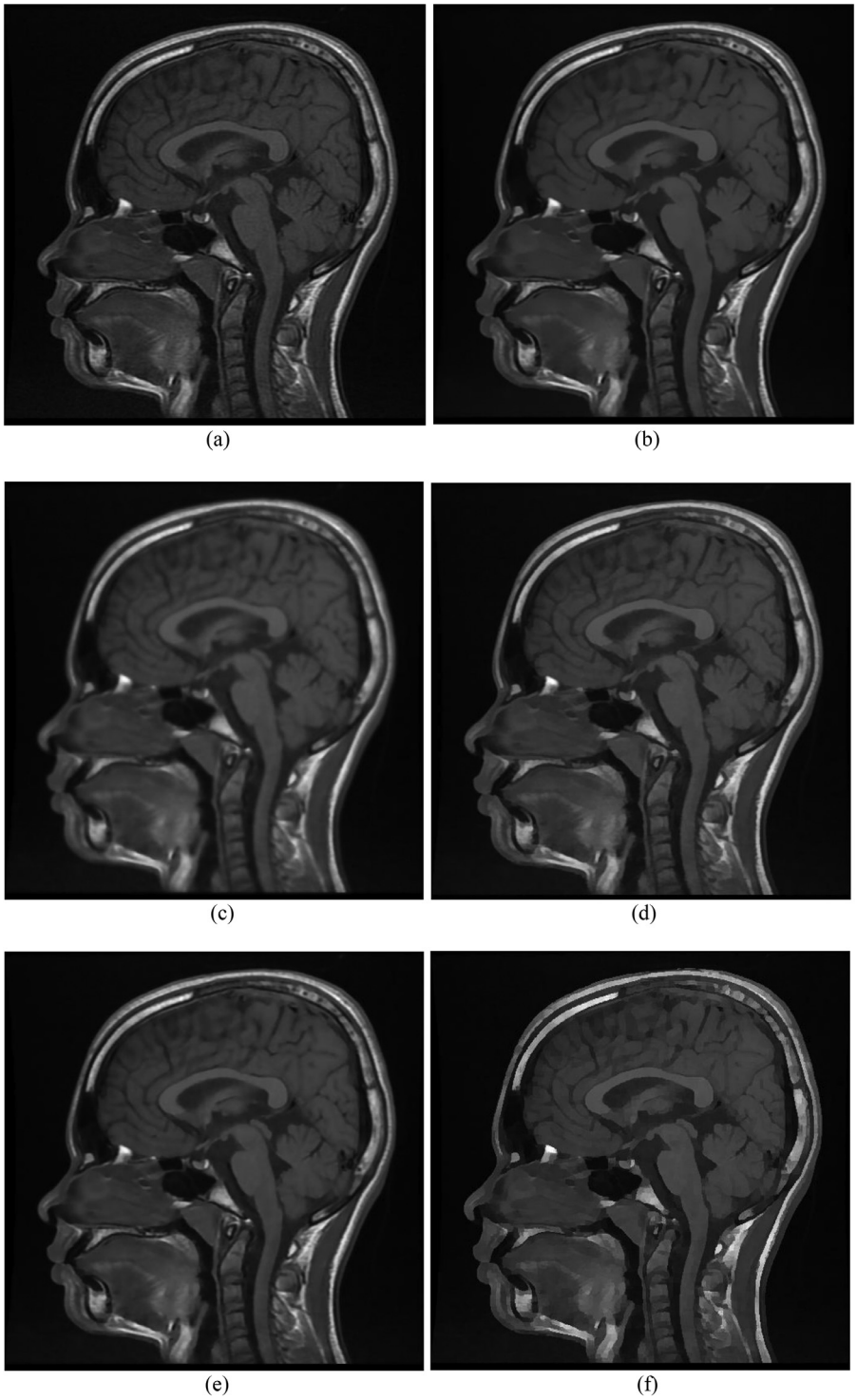


Fig. 9. Real MR images denoised with different filters (a) MR Image 3 (b) NLM (c) Bilateral filter (d) TV (e) AD (f) Kuwahara (g) Wavelet thresholding (h) LMMSE (i) SUSAN (j) Beltrami (k) Wiener.

filter, is observed to be free from noise and blur. But it has a slight structural loss. Output image of Beltrami filter (Fig. 10(j)) has noise-free image without blur but a slight structural loss is there. Wiener filter has generated noise-free, blur-free images with structural loss.

Mean Opinion Score (MOS) of the denoised MR images is given in Table 7. Considering image 1, Beltrami filter has high MOS. NLM and AD also have good MOS values. On image 2, both SUSAN and Beltrami filters have shown high scores. NLM and TV also have equally good MOS values. On image 3, NLM shows better performance. SUSAN and Beltrami filters also have an equally good performance on image 3. NLM and Beltrami filters have consistently appreciable performance on three images. The performance of BF, Kuwahara, WT and LMMSE filters are not appreciable on three images. TV filter's MOS is appreciable only in image 2. AD filter has good performance on image 1 only. SUSAN filter's MOS is good in image 2 and image 3, but not appreciable in image 1.

The quality metric (MetricQ, AQI and BRISQUE) values on real-time MRI's denoised by different filters are given in Table 8. On image 1, highest MetricQ value is shown by NLM and TV filters. WT

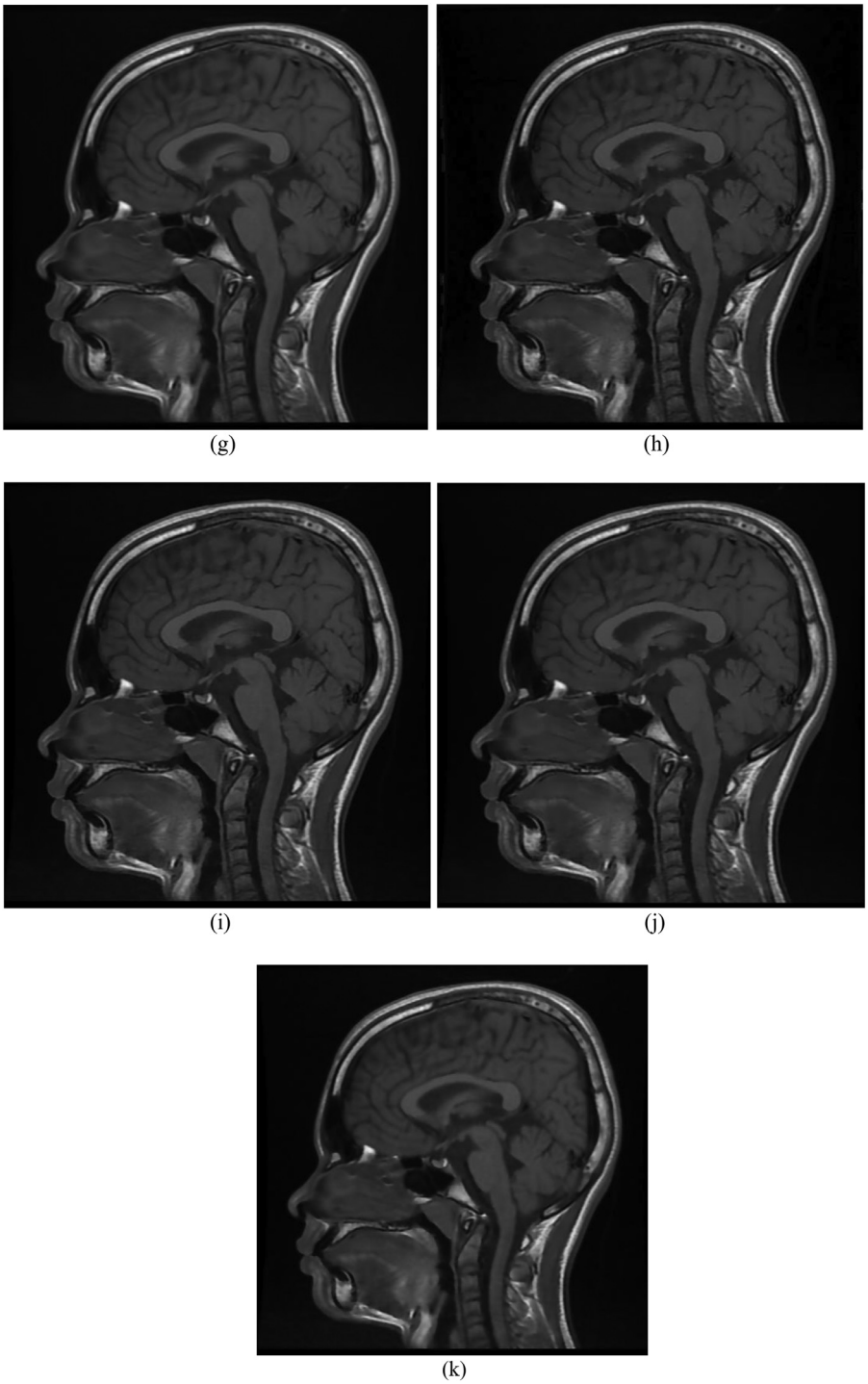


Fig. 9. Continued

and SUSAN filters have also shown good MetricQ values. On image 2 also NLM filter has highest MetricQ value. AD, WT and SUSAN filters are also exhibited good scores. On image 3, TV, Kuwahara, and NLM filters have shown first, second and third highest MetricQ values respectively. AD, Beltrami and LMMSE filters scores are not good for any of the MR images.

Considering AQI values on image 1, TV filter has the highest value. Kuwahara has shown good score. On image 2, TV and Kuwahara have good AQI values than other filters. BF and NLM filter also have shown good scores. On image 3, TV filter has the highest value. AD and SUSAN filters also have good scores. All other filters have equal AQI values.

On three MR images (Image 1, Image 2 and Image 3) Kuwahara has shown appreciable (low) value. TV and NLM filters also have shown appreciable scores on image 1. On image 2, TV and SUSAN filters also have good scores. In image 3, TV filter has good BRISQUE value other than Kuwahara.

In terms of computational time, LMMSE filter is the fastest among others. But LMMSE's restoration performance is not appreciable on real-time MR images. Wiener filter also has shown less

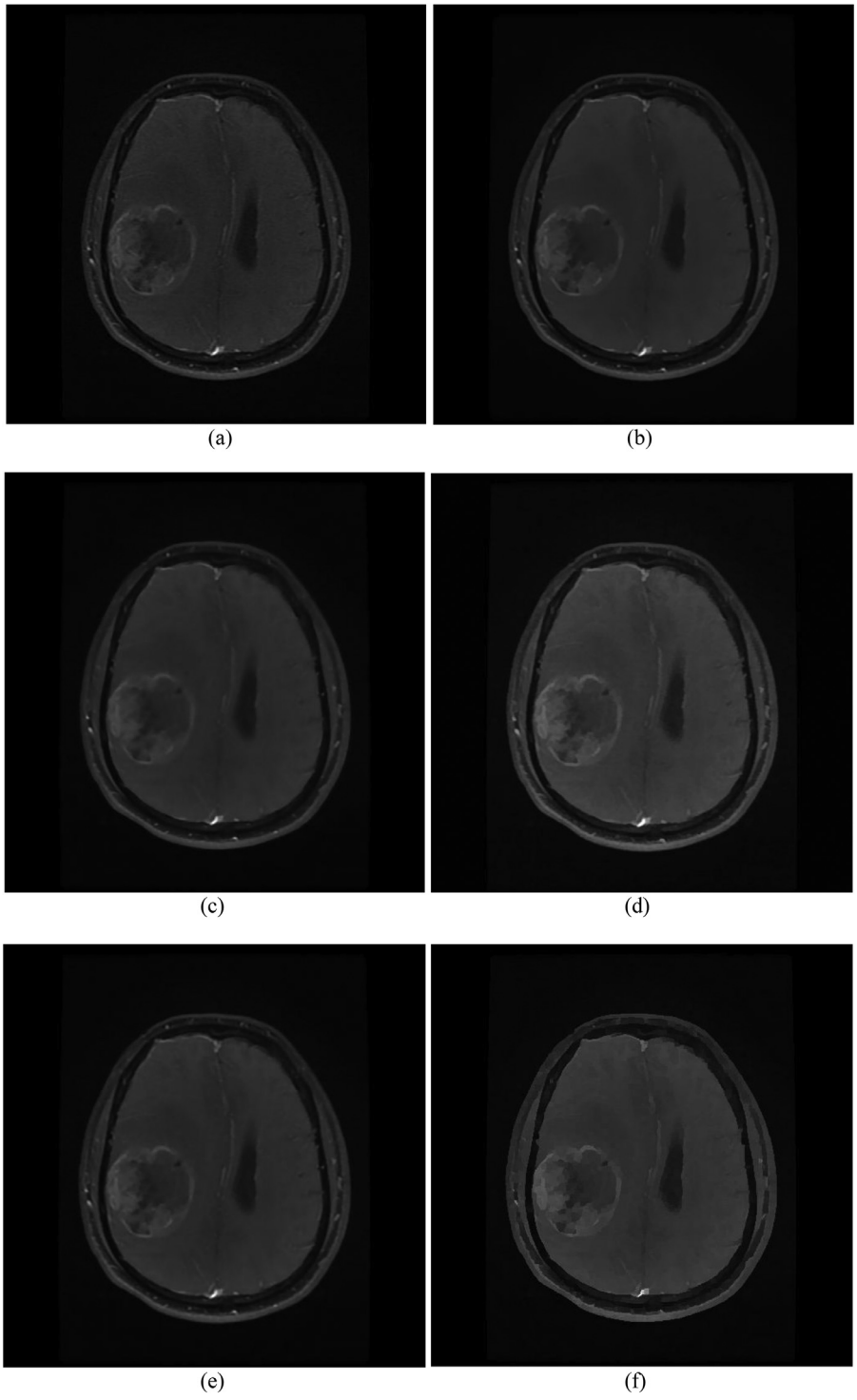


Fig. 10. Real MR images denoised with different filters (a) MR Image 3 (b) NLM (c) Bilateral filter (d) TV (e) AD (f) Kuwahara (g) Wavelet thresholding (h) LMMSE (i) SUSAN (j) Beltrami (k) Wiener.

computational time. NLM filter whose performance is good in real-time MR images has shown the highest computational time. Another filter performing well on MR images, Beltrami, have shown appreciable computational time.

On real-time MR images, NLM and Beltrami have shown better performance on denoising as well as edge-preservation. NLM is computationally complex but Beltrami filter's computational time is appreciable. Susan and TV filer also have shown good performance. It can be observed that TV, AD, and SUSAN filter's performance on real-time MR images are image specific also. MetricQ, AQI and BRISQUE values of real-time MR images are not on par with MOS values.

4. Discussions

Filters which have shown good performance on subjective assessment have not shown good performance on objective evaluation. The quantitative evaluation indices computed from

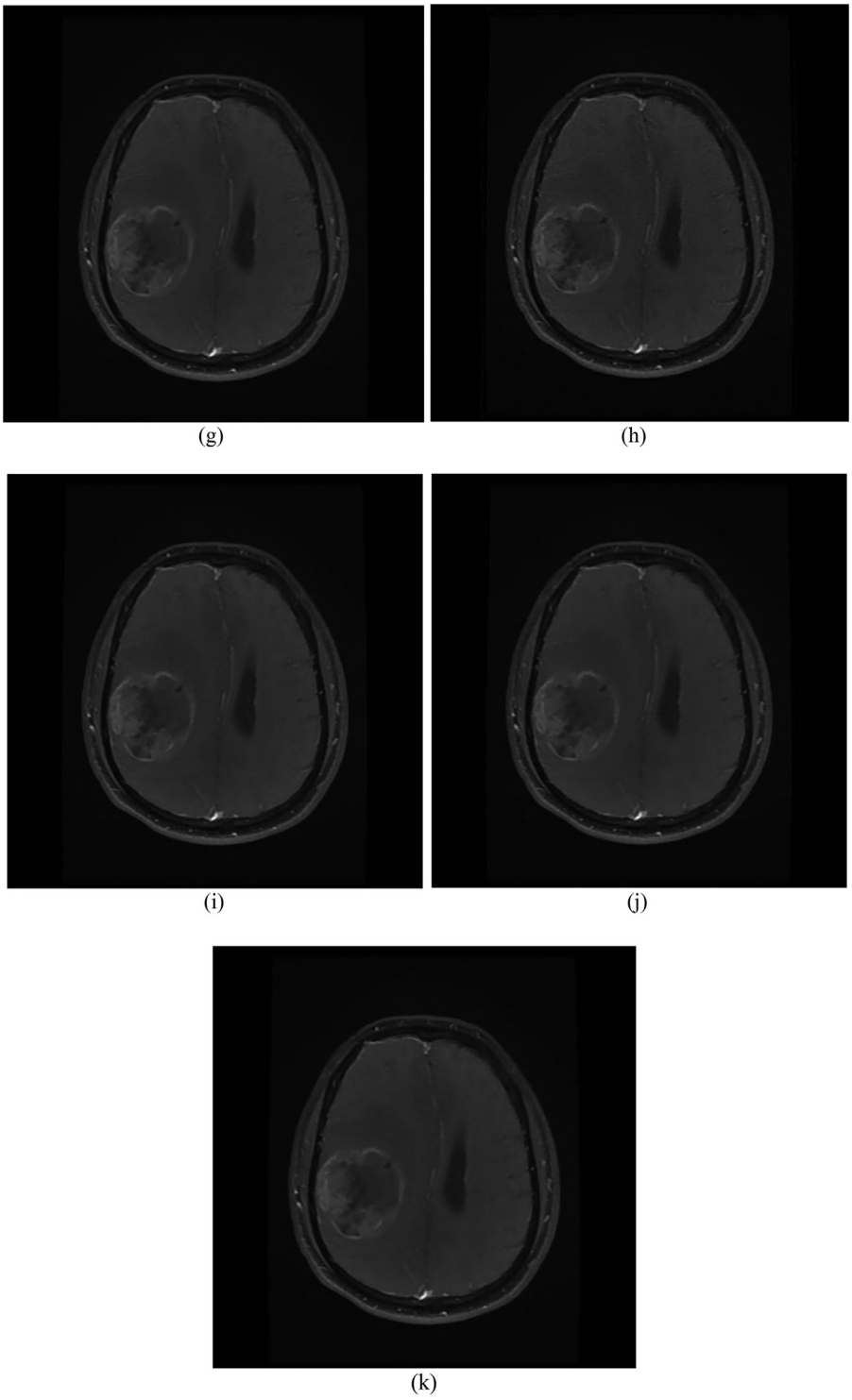


Fig. 10. Continued

denoised Phantom, simulated and MR images are not on par with the qualitative evaluations done with MOS. Similarly, no-reference image quality indices often used in denoising studies were not in correspondence with full-reference image quality metrics. There is an urgent need for a no-reference image quality metric to measure the quality of restored image for MR image denoising applications.

Even though PSNR between ground-truth and denoised images is a full-reference metric commonly used objective measure for denoised image quality, sometimes PSNR values show significant difference between two visually similar or indistinguishable images. Similarly, denoised images with equal values of PSNR show significant difference on visual inspection. None of the quantitative measures discussed in this work are not particularly well at

Table 7
Mean Opinion Score of MR images denoised by different filters.

FILTERS	MOS on different images			Summary
	Image 1	Image 2	Image 3	
NLM	**4**	**4**	**4.5**	**4.17±0.29**
BF	2	3	2.5	2.50±0.50
TV	3.5	**4**	3.5	3.67±0.29
AD	**4**	3	3.5	3.50±0.50
KUWAHARA	3	3	3	3 ± 0
WT	2.75	3.5	3	3.08±0.38
LMMSE	2.5	3.5	2.75	3.08±0.80
SUSAN	3	**4.5**	**4**	3.83±0.76
BELTRAMI	**4.5**	**4.5**	**4**	**4.33±0.29**
Wiener	2.75	3.25	3.5	3.17±0.38

predicting human visual response to quality of denoised MR images. Filters which have shown high PSNR values between denoised and ground-truth images do not show good visual quality.

There is a usual practice of doing design and validation of the filters on Phantom or simulated MR images based on the assumption that they are the best alternatives of real-time MR images. It has been observed that the performance of each filter is completely different on Shepp–Logan, simulated MR and real-time MR images. The filter performance is critically sensitive to the strength of noise also. No filter which can offer good performance equally on Phantom, simulated MR image and real-time MR images, is available in literature. Filter designs optimized on Phantom or simulated MR using maximum PSNR between denoised and ground-truth images as objective function (minimum error sense in general) do not perform well of real-time MRI. As the performance of filters on MR images and simulated or Phantom images are different, the assumption that Phantom or simulated MR images are best alternatives of real-time MR images is not valid. Even on real-time MR images also, there is no consistency in the performance of filters.

Even though the performance of filters are sensitive to the type of the image (Phantom, simulated MR or real-time MR), their merits and demerits can be generalized up to a certain extent as follows. Among the edge preserving spatial filters, Kuwahara filter is the one which is very simple to implement. Because of high smoothing characteristics, Kuwahara filter causes shading or cartoon artefacts. Dislocation of edges and gradient reversal are other two major issues associated with Kuwahara filter. Among the edge preserving spatial filters, NLM filter offers the best trade-off between edge reservation and noise removal. It selectively smoothens homogenous regions in the input image without compromising sharpness of edges. The search procedure for finding similar pixels within the search window makes the NLM computationally complex. However, the performance of the NLM filter is not appreciable at higher noise levels. Bilateral and SUSAN filters have very good edge preserving characteristics. Staircase artefacts and gradient reversal are two limitations of bilateral and SUSAN filters. Both of them introduce false edges. Total variation and Beltrami filters are faster than NLM, bilateral, SUSN and AD filters. TV and Beltrami filters have the ability to restore comparatively better noise-free estimates, at higher noise levels. However, the sharpness of edges is compromised both in TV and Beltrami filters. This makes TV and Beltrami filters inferior to bilateral, NLM and AD filters at low SNR. Beltrami filter significantly alters mean brightness characteristics of the input image during the restoration. However, Beltrami regularisation is free from staircase artefacts usually seen on TV. AD is computationally slow as it is an iterative denoising algorithm. Edge preservation capability of AD filter is less than NLM, SUSAN and bilateral filters.

Apart from the class of edge-preserving spatial filters comprising Kuwahara, AD, bilateral, SUSAN, NLM, Beltrami and TV filters, wavelet thresholding is a transform domain denoising technique which is believed to have edge preservation characteristics. In fact, WT is not up to the mark of AD, bilateral, NLM filters in terms of edge preservation at low SNR. Similarly, the quality of output images produced by WT is comparatively less than those produced by TV or Beltrami filters at a high noise level. Moreover, the selection of mother wavelet, level of decomposition and the wavelet sub-band to be thresholded determine the reliability of WT. Wiener filter is a simple and computationally fast inverse filtering algorithm outside the category of edge-preserving filters. The inverse filter algorithm in wiener needs the prior of local and global noise estimates. Hence, the performance of Wiener filter critically depends on the accuracy of the noise estimation model used.

5. Conclusion

Performance of different edge reserving filters was evaluated both objectively and subjectively on Phantom, simulated and real-time MR images. It was observed that the performance of denoising filters heavily depends on type of the image and the strength of noise in it. No filter which can offer good performance equally on the Phantom, simulated MR and real-time MRI, at all noise levels was found. Objective indices generally used to quantify the quality of the denoised images were not in concordance with subjective quality ratings. Similarly, no-reference image quality indices often used in denoising studies were not in correspondence with full-reference image quality metrics. There is always a probability that the performance of the filters designed on Phantom or simulated MR images can be sub-optimal on real-time MR images. Above inferences are helpful as reliable guidelines for the selection of edge-preserving filters, formulation of new denoising algorithms and optimization of operational parameters of denoising filters, for MR images.

Table 8
MetricQ, AQI and BRISQUE values on MR images denoised with different filters, and the computational time in seconds for all filters.

FILTERS	MetricQ			AQI			BRISQUE			COMP. TIME (s)
	Image 1	Image 2	Image 3	Image 1	Image 2	Image 3	Image 1	Image 2	Image 3	
NLM	**0.0673**	**0.0998**	0.0267	0.0012	0.0020	0.00002	37	41.69	47.55	231.72±1.21
BF	0.0568	0.0797	0.0232	0.0014	0.0021	0.00002	51.28	56.23	52.73	2.07±0.21
TV	**0.0673**	0.0893	**0.0314**	**0.0017**	**0.0022**	**0.00004**	33.48	37.12	34.62	3.73±0.03
AD	0.0637	0.0922	0.0256	0.0013	0.0020	0.00003	41.81	42.53	50.94	0.95±0.04
KUWAHARA	0.0595	0.0812	0.0282	0.0015	**0.0022**	0.00002	**30.28**	**26.73**	**33.17**	43.52±0.32
WT	0.0659	0.0952	0.0262	0.0013	0.0019	0.00002	46.87	45.48	49.75	**0.18±0.002**
LMMSE	0.0574	0.0773	0.0231	0.0014	0.0018	0.00002	44.41	42.03	46.71	**0.10±0.006**
SUSAN	0.0657	0.0931	0.0256	0.0013	0.0017	0.00003	37.94	37.35	52.38	2.87±0.06
BELTRAMI	0.0633	0.0907	0.0245	0.0012	0.0017	0.00002	40.14	48.04	48.91	0.42±0.10
Wiener	0.0560	0.0832	0.0233	0.0012	0.0017	0.00002	40.12	39.91	42.12	**0.018±0.003**

Conflict of interest

Corresponding author of the manuscript entitled, "Analysis of Controversies in the Formulation and Evaluation of Restoration Algorithms for MR Images", submitted to Expert Systems with Applications, Ms. Simi V.R. on behalf of my co-author as well, state that this endeavour is not funded or financially supported by any agency and the authors have no conflicts of interest in this concern.

Credit authorship contribution statement

Simi V.R.: Conceptualization, Software, Validation, Formal analysis, Investigation, Writing - original draft. **Damodar Reddy Edla:** Resources, Data curation, Supervision, Project administration. **Justin Joseph:** Methodology, Validation, Formal analysis. **Venkatanareshbabu Kuppili:** Validation, Software.

References

Akar, S. A. (2016). Determination of optimal parameters for bilateral filter in brain MR image denoising. *Applied Soft Computing, 43*, 87–96.

Bartyzel, K. (2016). Adaptive Kuwahara filter. *Signal, Image and Video Processing, 10*(4), 663–670.

Benou, A., Veksler, R., Friedman, A., & Raviv, T. R. (2017). Ensemble of expert deep neural networks for spatio-temporal denoising of contrast-enhanced MRI sequences. *Medical Image Analysis, 42*, 145–159.

Chow, L. S., & Rajagopal, H. (2017). Modified-BRISQUE as no reference image quality assessment for structural MR images. *Magnetic Resonance Imaging, 43*, 74–87.

Cocosco, C. A., Kollokian, V., Kwan, R. K. S., Pike, G. B., & Evans, A. C. (1997). Brainweb, Online interface to a 3D MRI simulated brain database. *NeuroImage, 5*(4), 425.

Damera-Venkata, N., Kite, T. D., Geisler, W. S., Evans, B. L., & Bovik, A. C. (2000). Image quality assessment based on a degradation model. *IEEE Transactions on Image Processing, 9*(4), 636–650.

Djurović, I. (2017). Combination of the adaptive Kuwahara and BM3D filters for filtering mixed Gaussian and impulsive noise. *Signal, Image and Video Processing, 11*(4), 753–760.

Fernández, J. J., & Martínez, J. A. (2010). Three-dimensional feature-preserving noise reduction for real-time electron tomography. *Digital Signal Processing, 20*(4), 1162–1172.

Gabarda, S., & Cristóbal, G. (2007). Blind image quality assessment through anisotropy. *JOSA A, 24*(12), B42–B51.

Gabarda, S., Cristóbal, G., & Goel, N. (2018). Anisotropic blind image quality assessment: Survey and analysis with current methods. *Journal of Visual Communication and Image Representation, 52*, 101–105.

Golshan, H. M., & Hasanzadeh, R. P. (2015). An optimized LMMSE based method for 3D MRI denoising. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 12*(4), 861–870.

Hermessi, H., Mourali, O., & Zagrouba, E. (2019). Deep feature learning for soft tissue sarcoma classification in MR images via transfer learning. *Expert Systems with Applications, 120*, 116–127.

Jain, A. K. (1989). *Fundamentals of digital image processing*. Englewood Cliffs, NJ: Prentice Hall.

Joseph, J., Anoop, B. N., & Williams, J. (2019). A modified unsharp masking with adaptive threshold and objectively defined amount based on saturation constraints. *Multimedia Tools and Applications, 78*(8), 11073–11089.

Joseph, J., & Periyasamy, R. (2018a). A fully customized enhancement scheme for controlling brightness error and contrast in magnetic resonance images. *Biomedical Signal Processing and Control, 39*, 271–283.

Joseph, J., & Periyasamy, R. (2018b). An image driven bilateral filter with adaptive range and spatial parameters for denoising Magnetic Resonance Images. *Computers & Electrical Engineering, 69*, 782–795.

Kang, M., Jung, M., & Kang, M. (2018). Rician denoising and deblurring using sparse representation prior and nonconvex total variation. *Journal of Visual Communication and Image Representation, 54*, 80–99.

Khatami, A., Khosravi, A., Nguyen, T., Lim, C. P., & Nahavandi, S. (2017). Medical image analysis using wavelet transform and deep belief networks. *Expert Systems with Applications, 86*, 190–198.

Kuppusamy, P. G., Joseph, J., & Jayaraman, S. (2019). A customized nonlocal restoration scheme with adaptive strength of smoothening for magnetic resonance images. *Biomedical Signal Processing and Control, 49*, 160–172.

Kuppusamy, P. G., Joseph, J., & Sivaraman, J. (2017). A full reference Morphological Edge Similarity Index to account processing induced edge artefacts in magnetic resonance images. *Biocybernetics and Biomedical Engineering, 37*(1), 159–166.

Lahmiri, S. (2015). Image denoising in bidimensional empirical mode decomposition domain: The role of Student's probability distribution function. *Healthcare Technology Letters, 3*(1), 67–71.

Lahmiri, S. (2017a). Denoising techniques in adaptive multi-resolution domains with applications to biomedical images. *Healthcare Technology Letters, 4*(1), 25–29.

Lahmiri, S. (2017b). An iterative denoising system based on Wiener filtering with application to biomedical images. *Optics & Laser Technology, 90*, 128–132.

Lahmiri, S., & Boukadoum, M. (2014). Biomedical image denoising using variational mode decomposition. In *Proceedings of the IEEE biomedical circuits and systems conference (BioCAS)* (pp. 340–343). IEEE.

Lahmiri, S., & Boukadoum, M. (2015a). A weighted bio-signal denoising approach using empirical mode decomposition. *Biomedical Engineering Letters, 5*(2), 131–139.

Lahmiri, S., & Boukadoum, M. (2015b). Physiological signal denoising with variational mode decomposition and weighted reconstruction after DWT thresholding. In *Proceedings of the IEEE international symposium on circuits and systems (ISCAS)* (pp. 806–809). IEEE.

Lahmiri, S., & Boukadoum, M. (2016). Combined partial differential equation filtering and particle swarm optimization for noisy biomedical image segmentation. In *Proceedings of the IEEE 7th Latin American symposium on circuits & systems (LASCAS)* (pp. 363–366). IEEE.

Rundo, L., Tangherloni, A., Nobile, M. S., Militello, C., Besozzi, D., & Mauri, G. (2019). MedGA: A novel evolutionary method for image enhancement in medical imaging systems. *Expert Systems with Applications, 119*, 387–399.

Simi, V. R., Edla, D. R., & Joseph, J. (2018). A Fuzzy Sharpness Metric for Magnetic Resonance Images. *Journal of Computational Science, 29*, 1–8.

Singh, C., & Bala, A. (2019). A local Zernike moment-based unbiased nonlocal means fuzzy C-Means algorithm for segmentation of brain magnetic resonance images. *Expert Systems with Applications, 118*, 625–639.

Smith, S. M., & Brady, J. M. (1997). SUSAN—a new approach to low level image processing. *International Journal of Computer Vision, 23*(1), 45–78.

Tong, C., Sun, Y., Payet, N., & Ong, S. H. (2012). A general strategy for anisotropic diffusion in MR image denoising and enhancement. *Magnetic Resonance Imaging, 30*(10), 1381–1393.

Yang, J., Fan, J., Ai, D., Zhou, S., Tang, S., & Wang, Y. (2015). Brain MR image denoising for Rician noise using pre-smooth non-local means filter. *Biomedical Engineering Online, 14*(1), 2.

Zhang, X., Feng, X., Wang, W., & Xue, W. (2013). Edge strength similarity for image quality assessment. *IEEE Signal Processing Letters, 20*(4), 319–322.

Zhu, X., & Milanfar, P. (2010). Automatic parameter selection for denoising algorithms using a no-reference measure of image content. *IEEE Transactions On Image Processing, 19*(12), 3116–3132.