

Influence of Rician Noise on Cardiac MR Image Segmentation using Deep Learning

Chien-Cheng Wu*, Chao-Hsiung Hsu*, Paul C. Wang, Tsang-Wei Tu, and Yi-Yu Hsu†

Abstract. Precision in segmenting cardiac MR images is critical for accurately diagnosing cardiovascular diseases. Several deep learning models have been shown useful in segmenting the structure of the heart, such as atrium, ventricle and myocardium, in cardiac MR images. Given the diverse image quality in cardiac MRI scans from various clinical settings, it is currently uncertain how different levels of noise affect the precision of deep learning image segmentation. This uncertainty could potentially lead to bias in subsequent diagnoses. The goal of this study is to examine the effects of noise in cardiac MRI segmentation using deep learning. We employed the Automated Cardiac Diagnosis Challenge MRI dataset and augmented it with varying degrees of Rician noise during model training to test the model's capability in segmenting heart structures. Three models, including TransUnet, SwinUnet, and Unet, were compared by calculating the SNR-Dice relations to evaluate the models' noise resilience. Results show that the TransUnet model, which combines CNN and Transformer architectures, demonstrated superior noise resilience. Noise augmentation during model training improved the models' noise resilience for segmentation. The findings under-score the critical role of deep learning models in adequately handling diverse noise conditions for the segmentation of heart structures in cardiac images.

Keywords: MRI Image Segmentation, Noise Augmentation, Rician Noise.

1. Introduction

Magnetic Resonance Imaging (MRI) segmentation is pivotal in medical image analysis, which offers tissue localization that facilitates diagnosis and interpretation of the outcome. The accuracy of segmentation is dependent on an in-depth understanding of tissue structures and the associated MRI contrast. Compared to manual segmentation, computer-aided image segmentation greatly improves efficiency when analyzing large image dataset. Traditional unsupervised clustering algorithms, such as Fuzzy C-Means (FCM), have been used for segmenting brain images with noise [1, 2]. Despite some FCM-based algorithms incorporating spatial information, clustering methods struggle to separate the connected tissues that exhibit the same contrast in MRI images.

Recently, deep learning has demonstrated significant potential in medical image segmentation. A few deep learning models have been developed to automatically identify and segment tissue structures. While the image quality may be highly variable in cardiac MRI acquired across different clinical settings, the effect of different noise levels on the accuracy of deep learning image segmentation remains unclear.

The objective of this study is to investigate the impact of image noise on the segmentation accuracy of different deep learning models. We utilized the publicly available Automated Cardiac Diagnosis Challenge (ACDC) MRI dataset [3] and augmented with various Rician noise [4, 5] to test three segmentation models, including Convolutional Neural Networks (CNN) [6–8], Transformers [9, 10], and CNN-Transformer hybrid architectures [11, 12].

2. Related work

Segmentation of MRI images under noisy conditions has been explored using both traditional methods and deep learning techniques. Traditional methods like FCM have been applied for MRI brain imaging segmentation. Gharieb et al. enhanced the FCM algorithm by incorporating local data and membership details, leading to the creation of LMREFCM and LDMREFCM algorithms [1], which significantly improved noise resilience and homogeneity in region clustering of noisy MRI images. Rahman et al. advanced segmentation techniques by integrating a power mean function and a fuzzy-membership function into a variational framework, thereby increasing accuracy in noisy, multi-object scenarios [13].

In recent years, deep learning has realized remarkable advances in medical image analysis. The seminal U-Net [6] architecture substantially improved segmentation accuracy over earlier methods. Variants including ResU-Net [8] and U-Net++ [7] were later introduced. Beyond CNNs, Transformers leverage attention to model long-range dependencies and expand contextual learning [14]. In particular, Swin Transformers capture hierarchical feature representations through shifted windows [10, 15]. Hybrid CNN-Transformer networks synergize the complementary strengths of both architectures [11, 12].

Image quality critically impacts segmentation outcomes. Noise pervades clinical MRI acquisition [16]. Studies have shown deep learning performance deteriorates with increased noise levels [17]. Noise augmentation during training provides partial mitigation [5]. These developments indicate deep learning is a promising direction for addressing challenges posed by noise in medical imaging segmentation.

3. Methodology

3.1 MRI Dataset

The ACDC MRI dataset contains 3D images of 100 patients acquired from two temporal frames in a cardiac cycle, end diastole (ED) and end systole (ES) [3]. Images were acquired using Steady-State Free Precession (SSFP) pulse sequence by 1.5T or 3T clinical scanners, with slice thickness ranging from 5 to 10 mm (some with a 5 mm gap). The in-plane image resolution ranged from 1.34 to 1.68 mm². Three cardiac regions were labeled in the dataset as the ground truth to delineate right ventricle (RV), myocardium (MY), and left ventricle (LV). This study employed 2D images for model training, since not all dataset contains complete 3D images. The dataset was divided into training, validation, and test sets with a ratio of 70-10-20, respectively.

3.2 Rician Noise in MRI Image

MRI utilizes magnetic field gradient to generate spatially encoded frequencies and phase encoding of nuclear spin precessions within the sample. The yielding data includes the real and imaginary parts of the free induction decay. Through Fourier transformation, signals from local nuclear spins are obtained and converted into a magnitude image for display. Hence, it can be assumed that the image contains noise from both real and imaginary parts [18, 19]:

$$s = (A_R + n_R) + i(A_I + n_I) \quad (1)$$

In Eq. 1, $A_{R,I}$ represents the real and imaginary parts of the signal without noise, while $n_{R,I}$ represents the noise in those parts. It is assumed that the noise follows a Gaussian distribution:

$$P(n_{R,I}) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{n_{R,I}^2}{2\sigma^2}\right) \quad (2)$$

Here, σ represents the standard deviation of the Gaussian noise distribution. After converted to magnitude image, MRI signals with noise follow a Rician distribution:

$$m = \frac{\sqrt{(A_R + n_R)^2 + (A_I + n_I)^2}}{\sqrt{(A \cos \varphi + n_R)^2 + (A \sin \varphi + n_I)^2}} \quad (3)$$

The probability density function (PDF) for the Rician distribution is given by:

$$P(m) = \frac{m}{\sigma^2} \exp\left(-\frac{A^2 + m^2}{2\sigma^2}\right) I_0\left(\frac{Am}{\sigma^2}\right) \quad (4)$$

Where m represents the pixel signal in the magnitude image, A is the image amplitude without noise, and σ is the Gaussian noise distribution standard deviation for both the real and imaginary parts. I_0 is the Bessel function of the first kind, defined as:

$$I_0(z) = \frac{1}{2\pi} \int_0^{2\pi} e^{z \cos \alpha} d\alpha \quad (5)$$

Since the ACDC dataset provides only image amplitude, it is assumed that the amplitude image (*Image*) contributes solely from the real part. Rician noise can be added to the magnitude image using Eq. 6:

$$M = \sqrt{(Image + n_R)^2 + n_I^2} \quad (6)$$

Signal-to-Noise Ratios (SNR)[20] was calculated for each image by:

$$SNR = 0.66 S / \sigma_{noise} \quad (7)$$

Here, S represents the averaged signal intensity of an image within the Region of Interest (ROI) in the LV, RV and MY, and σ_{noise} represents the standard deviation of noise in the background.

3.3 AI Segmentation Models

Deep learning image segmentation models typically utilize an encoder-decoder (U-Net) structure to extract spatial features of objects in images. Both encoder and decoder can be composed of different variations of the backbone. In this study, we compared three deep learning models:

Unet (U): A CNN with a ResNet backbone for both the encoder and decoder [8].

SwinUnet (S): A pure Transformer architecture for image segmentation [10].

TransUnet (T): A CNN-Transformer hybrid architecture for image segmentation [11].

These models were evaluated in the context of their effectiveness in segmenting heart structures (RV, MY, LV) within the images by Dice coefficient.

Models were trained with images without noise augmentation (wo), or with noise augmented in either ~12% (w12), ~30% (w30), or ~48% (w48) of the images. Then, a random choice of noise level in a range of 0.05 to 0.3 of the noise's standard deviation was added.

3.4 Assessing the Noise Resilience of the Model

To evaluate a model's resilience to noise, SNR-Dice characteristic curves were derived. The curves were divided into high-precision and low-precision prediction range for comparison. A parameterized logistic function was used to fit the curves:

$$Dice = \frac{D_{max}}{1 + \exp(-k(SNR - S_{50}))} \quad (8)$$

, where D_{max} is the maximum Dice coefficient under high SNR conditions, S_{50} represents the SNR at which Dice decreases to 50% of D_{max} , and k is the curve's logistic growth rate or steepness. As the simulated noise values do not encompass scenario where SNR reaches zero, a minimum boundary condition was established for S_{50} during the fitting process. This minimum boundary condition was based on the lowest simulated SNR to prevent errors arising from the absence of low-precision data. The model performance was evaluated by obtaining the SNR threshold, where the Dice coefficient started to decrease rapidly. SNR_{80} was defined when the Dice coefficient dropped to 80% of D_{max} :

$$SNR_{80} = S_{50} - \frac{\log(\frac{1}{0.8} - 1)}{k} \quad (9)$$

4. Result

Distribution of the added Rician noise in this study can be observed by subtracting the original image from the noisy image (Figure 1).

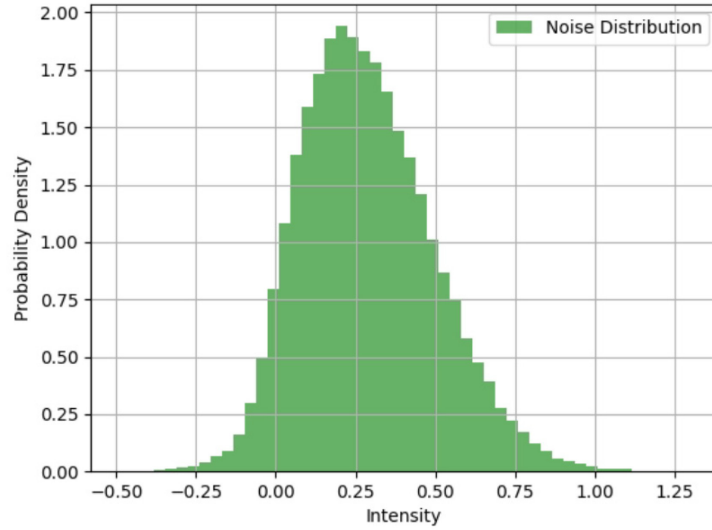


Fig. 1. Simulated Rician Noise distribution in MRI magnitude images with standard deviation of 0.3.

Figure 2 illustrates the TransUnet segmentation results trained without noise augmentation (Two) and with 30% noise augmentation (Tw30), tested under varying noise levels. It is noticeable that, in the Two models, RV was segmented with significant distortion at SNR 2.38, where it was completely indistinguishable at SNR 1.94. In contrast, the Tw30 model retained the ability to predict the RV region at SNR 1.94. The experiment demonstrates that noise augmentation during model training can improve segmentation results in high-noise images.

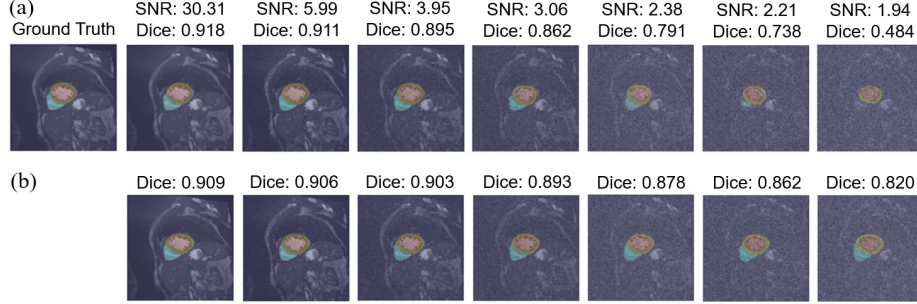


Fig. 2. Segmentation results from TransUnet trained (a) without noise augmentation and (b) with 30% noise augmentation. The three ROIs' average SNR and Dice index are displayed above each image.

Figure 3 compares the segmentation results from the three models trained with 30% noise augmentation for low-SNR images. Among the tested models, TransUnet demonstrated superior performance in segmenting cardiac structures in noisy images. Unet exhibited significant shortcomings in segmenting the RV and MY regions, while SwinUnet showed partial deficiencies in defining the RV.

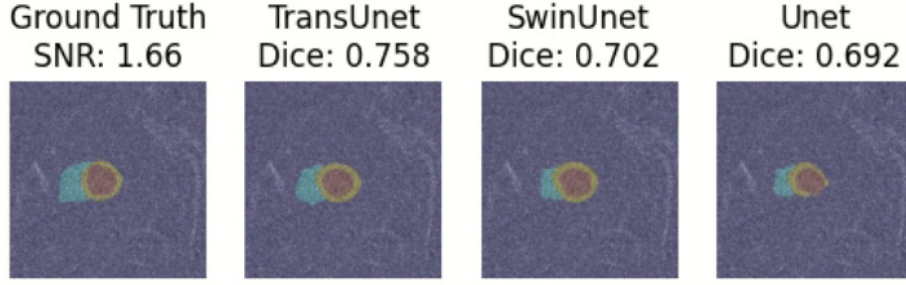
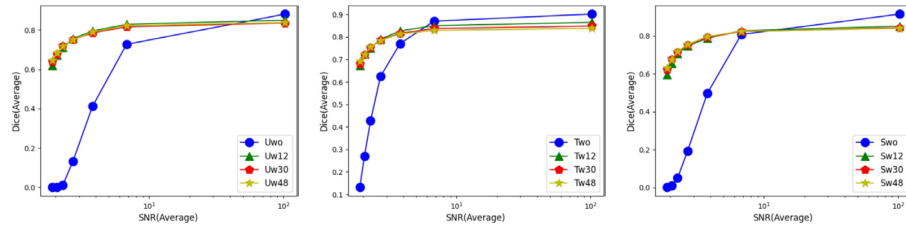


Fig. 3. TransUnet (Tw30), SwinUnet (Sw30), Unet (Uw30) on high-noise data of 30% noise augmentation.

The comparison of noise's influence on models subjected to different training conditions can be observed in Figure 4. The averaged Dice coefficient from three regions that were segmented by the models was calculated and plotted against the average SNR of the corresponding ground truth. Since the image quality in the ACDC dataset varied widely, the SNR distribution was broad (103 ± 95). Dice coefficients of the models that were trained without noise augmentation showed a rapid decrease as SNR decreased. On the other hand, models that were trained with noise augmentation demonstrated noise resilience. Nevertheless, training models with a higher proportion of noisy images resulted in reduced segmentation accuracy on raw image data.

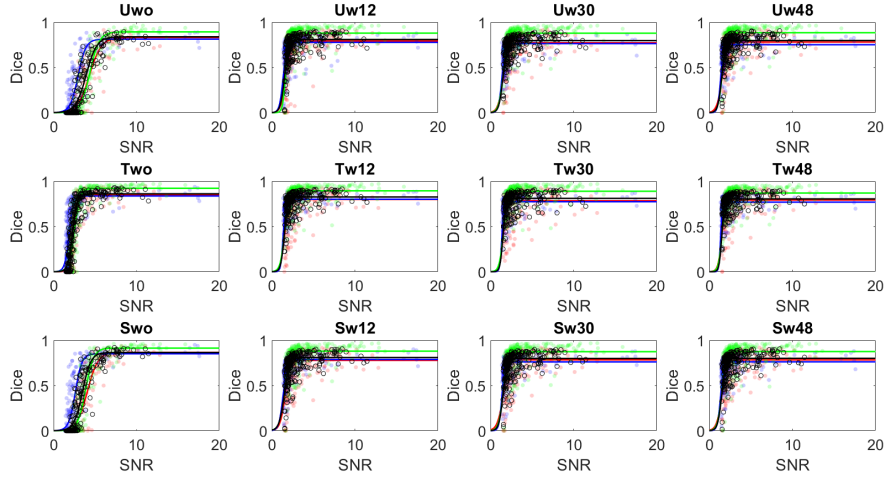


(a) (b) (c)

Fig. 4. A comparison of the segmentation models trained by different proportions of noise augmentation (a) Unet, (b) TransUnet, and (c) SwinUnet.

In Figure 5, the models' Dice coefficient at high SNR, increased in the order of LV (green line) > RV (red line) > MY (blue line). At low SNR, among the models trained without noise augmentation, the MY regions were mostly observable. Compared to other models, TransUnet demonstrated better noise resistance at low SNR conditions. For the models trained with noise augmentation, a significant improvement was observed when segmenting the cardiac structures in the images with higher noise levels.

Table 1 presents the results of SNR-Dice curve fitting (black line in Figure 5) between models. Among them, the SwinUnet model without noise augmentation achieved the highest D_{max} of 0.864. On the other hand, the TransUnet model with 48% noise obtained the lowest SNR_{80} at 1.709, implying its excellent performance in high-noise images. Considering all the performance metrics, TransUnet demonstrated the best overall performance, validating the advantages of a model architecture that combined CNN and Transformer for handling noisy images.

**Fig. 5. Results of SNR-Dice curve fitting using Equation (10) for Unet, TransUnet, and SwinUnet, in respect to noise augmentation strategies.****Table 1. Comparison of Average SNR-Dice Curve Fitting Metrics for Model Noise Robustness**

Model	D_{max}	S_{50}	k	SNR_{80}
Uwo	0.839	3.941	1.569	4.825
Uw12	0.809	1.537	4.049	1.879
Uw30	0.797	1.500	4.199	1.831
Uw48	0.798	1.500	4.412	1.815
Two	0.861	2.283	3.586	2.669
Tw12	0.826	1.500	4.918	1.782

Tw30	0.811	1.500	5.941	1.734
Tw48	0.805	1.500	6.639	1.709
Swo	0.864	3.484	1.981	4.184
Sw12	0.809	1.500	3.224	1.931
Sw30	0.799	1.500	4.156	1.834
Sw48	0.800	1.500	4.200	1.831

5. Discussion

Segmentation of MRI images relies significantly on the tissue contrast to distinguish various anatomical regions. Factors, such as the proton density, T1 and T2 relaxation, presence of diffusion or flow, and imaging acquisition parameters, may influence MR image contrast. MRI signals exhibit rapid decay as image resolution increases. When acquiring cardiac images, it is essential to maintain sufficient temporal resolution to capture cardiac phase cycles, making it challenging to increase SNR by taking more averaging. Low resolution images may not reveal subtle pathologies, but high image resolution can lead to low tissue contrast due to decreased SNR, affecting image segmentation accuracy.

In this study, Rician noise was introduced to simulate low-SNR MRI images. The performance of three deep learning segmentation models (Unet, TransUnet, and SwinUnet) was evaluated in the presence of noise. The ROI analysis revealed similar signal intensities in the LV and RV, but the LV exhibited a higher Dice coefficient (Fig. 5). This phenomenon may be due to a better-defined MY area surrounding LV, which provided better contrast and made it easier for pattern recognition. In low-SNR conditions, the model demonstrated better noise resilience in segmenting MY, potentially due to the learning from the training data with MY feature in lower SNR. Consequently, even when the signal in the MY region was exceedingly noisy, the model could still use the remaining signal from the adjacent areas (e.g., LV) to make predictions for that region.

Among the three tested models, the exceptional performance of TransUnet model may be attributed to the integration of both Transformer and CNN architectures. Compared to the pure CNN model Unet, TransUnet introduces a Transformer encoder, which expands the receptive field and enhances the utilization of global semantic information. This architecture allows TransUnet to mitigate high-frequency noise to some extent. In contrast, a pure CNN encoder specializes in local features, making it more susceptible to high-frequency noise and potentially capturing erroneous features. They could lead to receiving incorrect features in the decoder. Compared to SwinUnet, TransUnet, with its combination of Transformer and CNN encoder, reduces the interference from high-frequency noise, which allows accurate extraction of the structural boundary. By using a CNN decoder, the models can recover local details more effectively.

With respect to noise augmentation strategies, increasing the proportion of noisy training data improved the model's performance on segmenting low-SNR images. However, it also reduced the segmentation accuracy on original clean data. These results might be due to the inclusion of a large proportion of noisy data during the model training, resulting in insufficient fit to the data without noise augmentation. Additionally, the incomplete MRI signal in the ACDC dataset, lacking real and imaginary parts, may limit our simulation's accuracy in replicating real-world noise distributions.

6. Conclusion

In this study, we investigated the impact of MRI Rician noise on the performance of medical image segmentation using deep learning. The deep learning models that were trained without noise augmentation showed a significant decrease in performance in segmenting a low-SNR image. In contrast, models trained with noise augmentation exhibited superior noise resilience. These results indicate that noise augmentation strategies were effective in improving a model's resistance to noise. We introduced the SNR-Dice curve as an indicator to assess model performance and noise resilience. Through curve fitting analysis, we could quantitatively evaluate the performance of different models under various noise levels. The approach offered a useful technique for analyzing model performance. The TransUnet model performed exceptionally well in high-noise environments, possibly due to its integration of the advantages of CNN and Transformer architectures, allowing it to extract local features and model long-range dependencies simultaneously. In summary, our study emphasizes the importance of noise handling in medical image segmentation tasks and provides an effective noise augmentation strategy to enhance the noise resilience of deep learning models.

Acknowledgments

This study was supported by NIH grants of NINDS R01NS112294, R01NS123442, NICHD P50HD105328, NIMHD U54MD007597, NSF grants of EES 2200489, CNS 2200585, and NSTC grants of NSTC 112-2425-H-006-001 -.

References

1. Gharieb RR, Gendy G, Abdelfattah A (2017) C-means clustering fuzzified by two membership relative entropy functions approach incorporating local data information for noisy image segmentation. *Signal Image Video Process* 11:541–548. <https://doi.org/10.1007/s11760-016-0992-4>
2. Dubey YK, Mushrif MM (2016) FCM Clustering Algorithms for Segmentation of Brain MR Images. *Advances in Fuzzy Systems* 2016
3. Bernard O, Lalande A, Zotti C, et al (2018) Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved? *IEEE Trans Med Imaging* 37:2514–2525. <https://doi.org/10.1109/TMI.2018.2837502>
4. Momeny M, Neshat AA, Hussain MA, et al (2021) Learning-to-augment strategy using noisy and denoised data: Improving generalizability of deep CNN for the detection of COVID-19 in X-ray images. *Comput Biol Med* 136:. <https://doi.org/10.1016/j.compbiomed.2021.104704>
5. Akbiyik ME (2023) Data Augmentation in Training CNNs: Injecting Noise to Images
6. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer Verlag, pp 234–241
7. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J (2018) UNet++: A Nested U-Net Architecture for Medical Image Segmentation
8. Diakogiannis FI, Waldner F, Caccetta P, Wu C (2019) ResUNet-a: a deep learning framework for semantic segmentation of remotely sensed data. <https://doi.org/10.1016/j.isprsjprs.2020.01.013>
9. Strudel R, Garcia R, Laptev I, Schmid C (2021) Segmenter: Transformer for Semantic Segmentation
10. Cao H, Wang Y, Chen J, et al (2021) Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation

11. Chen J, Lu Y, Yu Q, et al TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation
12. Zhang Y, Liu H, Hu Q (2021) TransFuse: Fusing Transformers and CNNs for Medical Image Segmentation
13. Rahman A, Ali H, Badshah N, et al (2022) Power mean based image segmentation in the presence of noise. Sci Rep 12:.. <https://doi.org/10.1038/s41598-022-25250-x>
14. Vaswani A, Shazeer N, Parmar N, et al (2017) Attention Is All You Need
15. Liu Z, Lin Y, Cao Y, et al Swin Transformer: Hierarchical Vision Transformer using Shifted Windows
16. Dietrich O, Raya JG, Reeder SB, et al (2007) Measurement of signal-to-noise ratios in MR images: Influence of multichannel coils, parallel imaging, and reconstruction filters. Journal of Magnetic Resonance Imaging 26:375–385. <https://doi.org/10.1002/jmri.20969>
17. Szegedy C, Zaremba W, Sutskever I, et al (2013) Intriguing properties of neural networks
18. Cárdenas-Blanco A, Tejos C, Irarrazaval P, Cameron I (2008) Noise in magnitude magnetic resonance images. Concepts Magn Reson Part A Bridg Educ Res 32:409–416
19. Gudbjartsson H, Patz S The Rician Distribution of Noisy MRI Data
20. Firbank MJ, Coulthard A, Harrison RM, Williams ED (1999) A comparison of two methods for measuring the signal to noise ratio on MR images