Objective Assessment of T2-based Liver Lesion Classifiers

C. Graff[1], E. W. Clarkson[2], and M. I. Altbach[2]
[1]Division of Imaging and Applied Math/OSEL/CDRH, U. S. Food and Drug Administration, Silver Spring, MD, United States, [2]Department of Radiology, University of Arizona, Tucson, AZ, United States

Introduction: Classification of lesions as benign or malignant is an important medical imaging task. It has been shown that in the liver the transverse relaxation time (T2) can be used to make such a classification, with malignant lesions typically having a shorter T2 than benign cysts and hemangiomas [1]. Obtaining a lesion T2 estimate can be a lengthy process, as data from multiple echo times (TEs) must be acquired. This is particularly problematic in regions of the body such as the liver which suffer from patient motion. Recently a radial fast spin-echo (Rad-FSE) technique has been developed to obtain T2 estimates within a single breath-hold during which an under-sampled set of radial k-space lines are acquired during a train of spin-echo periods [2,3]. Acquiring data within a breathhold reduces the effects of respiratory motion and misregistration; however there is time to acquire only 12-16 radial lines per spin-echo period. The high degree of under-sampling in the Rad-FSE technique has motivated the development of a variety of post-processing techniques that attempt to enforce prior information to compensate for the lack of data [3-5]. The goal of this work is to evaluate these proposed algorithms based directly on their lesion classification performance.

Theory and Methods: The receiver operating characteristic (ROC) curve has become a standard technique for the evaluation of 2-class classifiers in the medical imaging literature [6]. Here the 2 classes are benign and malignant and each proposed algorithm calculates a lesion T2 estimate, compares it to a threshold and decides if the lesion is benign or malignant based on whether the T2 estimate is greater or less than the threshold. By classifying many lesions, the true-positive fraction (TPF), or *sensitivity*, and false-positive fraction (FPF), equivalent to 1-*specificity*, for the given threshold may be estimated for each algorithm. Considering a range of T2 thresholds, the entire ROC curve for each algorithm is estimated as shown below in Fig. 2. A higher ROC curve indicates better performance, thus the area under the curve (AUC) is a common metric for ranking classifiers. A perfect classifier would have AUC=1, while a classifier that guesses randomly between benign and malignant would have AUC=0.5. To calculate the TPF and FPF, the true class of each lesion must be known. Thus a simulated imaging chain including a detailed computer-generated abdominal phantom has been developed which allows for the inclusion of random liver lesions (Fig. 1) [7]. A prior probability distribution for each lesion class, pr(T2|*benign*) and pr(T2|*malignant*) is assumed and by sampling from each distribution and randomly placing lesions within simulated abdominal slices, k-space data sets are created. The algorithms evaluated here are standard filtered back-projection (FBP), an alternating projection (POCS)-like technique that enforces prior information including object support, signal decay behavior and phase constraints [4] and an echo sharing algorithm (ES) which mixes data at different TEs to interpolate additional k-space data [3]. Two sizes of data sets were considered, 12 and 32 radial k-space lines per TE, with 256 samples per line and ETL=16 in each case. The echo spacing was 9 ms, TR=1000 ms and slice thickness = 8 mm. Noise levels consistent with in vivo acquisitions were employed in all simulations. Both 8 and 12 mm diameter spherical lesions were considered as evaluating performance for small lesions is of particular interest. For each lesion size approximately 1600 benign and 1600 malignant lesions were utilized to determine the ROC curves.

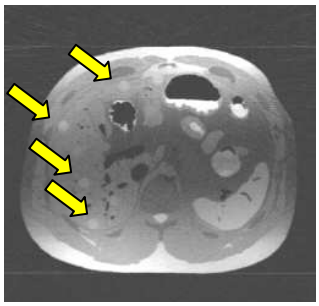

Figure 1: Reconstruction of a phantom slice with liver lesions (indicated by arrows).

Results and Discussion: The ROC curves obtained from the simulated data are given in Fig. 2. All algorithms have difficulty in classifying 8 mm lesions with only 12 lines per TE (Fig. 2a), where it is shown that each algorithm performs similarly, with AUC_{FBP} = 0.78, AUC_{POCS} = 0.76 and AUC_{ES} = 0.77. This poor performance is understandable considering that the lesion diameter is equal to the slice thickness, resulting in a large amount of partial volume in addition to the degree of under-sampling. For 12 mm lesions with the same amount of data, all algorithms improve their performance (Fig. 2b). In this case AUC_{FBP} = 0.82, AUC_{POCS} = 0.78 and AUC_{ES} = 0.94, indicating that the echo sharing algorithm performs best in this situation. If the amount of data is increased to 32 lines per TE, the performance on 8 mm lesions is significantly improved (Fig. 2c), with AUC_{FBP} = 0.90 and AUC_{ES} = 0.89. This increase in data from 12 to 32 lines can be achieved in vivo with a slightly increased breathhold and a bent trajectory [8]. In this case the POCS method was not considered due to its lengthy reconstruction time and relatively poor performance in other situations.

Conclusions: These results show that the choice of reconstruction algorithm can have a significant impact on performance and that choosing the best algorithm depends on lesion size and the amount of data available. The use of simulated data gives the knowledge of ground truth and allows for the consideration of small lesions and partial volume effects that are difficult to evaluate with physical phantoms or in vivo data. This type of task-based evaluation has clear advantages over traditional metrics of image quality such as SNR, CNR and MSE which do not directly measure the ability to classify lesions, the ultimate purpose of these imaging techniques. **Acknowledgements:** Work supported by NIH grants CA099074, HL085385, the Alliance Beverage Award and an appointment to the Research Participation Program at the Center for Devices and Radiological Health administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the U.S. Food and Drug Administration.

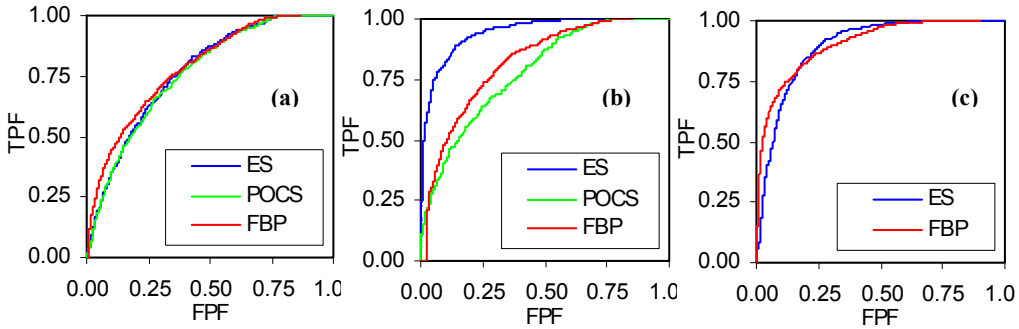

Figure 2: ROC curves for (a) 8mm lesions/12 views, (b) 12mm lesions/12 views and (c) 8 mm lesions/ 32 views.

References: [1] Goldberg, MA, AJR **157** 727-730 (1991). [2] Altbach, MI, JMRI **16** 179-189 (2002). [3] Altbach, MI, MRM **54** 549-559 (2005). [4] Graff, C, Proc. ISMRM 925 (2006). [5] Doneva, M., Proc. ISMRM 2812 (2009). [6] Barrett, HH, Foundations of Image Sci., Wiley (2003). [7] Graff, C, Proc. ISMRM 492 (2008). [8] Bilgin, A, Proc. ISMRM 2972 (2006).