33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Interplay between Bias Field Correction, Intensity Standardization, and Noise Filtering for T2-weighted MRI

Daniel Palumbo, Brian Yee, Patrick O'Dea, Shane Leedy, Satish Viswanath, Anant Madabhushi
Rutgers University, The State University of New Jersey
Department of Biomedical Engineering
599 Taylor Road, Piscataway, NJ, 08854

Abstract— **Magnetic Resonance Imaging (MRI) is known to be significantly affected by a number of acquisition artifacts, such as intensity non-standardness, bias field, and Gaussian noise. These artifacts degrade MR image quality significantly, obfuscating anatomical and physiological detail and hence need to be corrected for to facilitate application of computerized analysis techniques such as segmentation, registration, and classification. Specifically, algorithms are required to correct for bias field (intensity inhomogeneity), intensity non-standardness (drift in tissue intensities across patient acquisitions), and Gaussian noise, an artifact that significantly affects and blurs tissue boundaries (resulting in poor gradients). While clearly one needs to correct for all these artifacts, the exact sequence in which all three operations need to be applied in order to maximize MR image quality has not been explored. In this paper, we empirically evaluate the interplay between distinct algorithms for bias field correction (BFC), intensity standardization (IS), and noise filtering (NF) to study the effect of these operations on image quality in the context of 3 Tesla T2-weighted (T2w) prostate MRI. 7 different sequences comprising combinations of BFC, IS, and NF were quantitatively evaluated in terms of the percent coefficient of variation (%CV), a statistic which attempts to quantify the intensity inhomogeneity within a region of interest (prostate). The different combinations were also independently evaluated in the context of a classifier scheme for detection of prostate cancer on high resolution**

***in vivo* T2w prostate MRI. A secondary contribution of this work is a novel evaluation measure for quantifying the level of intensity non-standardness, called difference of modes (DoM). Experimental evaluation of the different sequences of operations across 22 patient datasets revealed that the sequence of BFC, followed by NF, and IS provided the best image quality in terms of %CV as well as classifier accuracy. The DoM measure was able to accurately capture the level of intensity non-standardness present in the images resulting from the different sequences of operations.**

I. Introduction

Acquisition artifacts associated with Magnetic Resonance Imaging (MRI) may be categorized into the following main categories: (1) bias field related intensity non-uniformity [1], (2) inter-acquisition non-standardness [2], and (3) Gaussian noise [3]. Both intensity non-uniformity (nonlinear signal intensity variation across an MR image, caused by the multiplicative effect of low pass signal intensity [1]) and intensity non-standardness (drift in image intensities across different acquisitions) are known to be caused mainly due to the effects of radio frequency field nonuniformity, patient anatomy, and scanner differences [4]. The aim of bias field correction (BFC) is to reduce MR image intensity non-uniformity either by a filtering based or bias field model based approach [1]. Intensity standardization (IS) approaches have been previously proposed to correct for the misalignment between intensity distributions across different MRI acquisitions [2]. Gaussian noise is typically addressed via noise filtering (NF), where the goal is to smooth image intensities within a tissue region while preserving (or accentuating) tissue boundaries [3].

Correcting for these three classes of image intensity artifacts has been shown to be highly relevant to improving image quality in order to facilitate application of computerized image analysis algorithms. For example, it has been previously shown that both image registration [5] as well as tissue classification [6] are significantly affected by these three artifacts. Similarly, segmentation performance has also been shown to be affected by the presence of bias field inhomogeneity and intensity non-standardness [7]. Thus, while it is clear that all three techniques need to be applied to MR imaging data to maximize image quality, the exact sequence of application for these operations is not immediately obvious.

Some prior studies have focused on identifying the correct sequence of a subset of operations (e.g BFC and IS [4] and NF and BFC [8]), but to our knowledge, no studies have attempted to identify the optimal sequence of all three operations simultaneously, for maximizing image quality. Madabhushi and Udupa [4] showed that BFC should be followed by IS from the perspective of improving MR image quality; succeeding IS by BFC resulted in the introduction of image intensity non-standardness. Similarly, Montillo et al [8] concluded that BFC should precede NF, since bias correction algorithms had the unfortunate side-effect of introducing noise into the image. By leveraging the conclusions drawn in [4], [8], we were able to eliminate several sequences of post-processing operations from consideration for this study.

Figure 1 shows the effect of applying BFC, IS, and NF (specific algorithms employed are described in Section II.B) on a cohort of endorectal, T2-weighted (T2w) prostate MRI studies. It may be observed that the BFC, IS, and NF

This work was made possible by grants by the Walter H. Coulter Foundation, National Cancer Institute (Grant Nos. R01CA136535, R01CA140772, and R03CA143991), Department of Defense Prostate Cancer Research Program, The Cancer Institute of New Jersey and the Society for Imaging Informatics in Medicine.

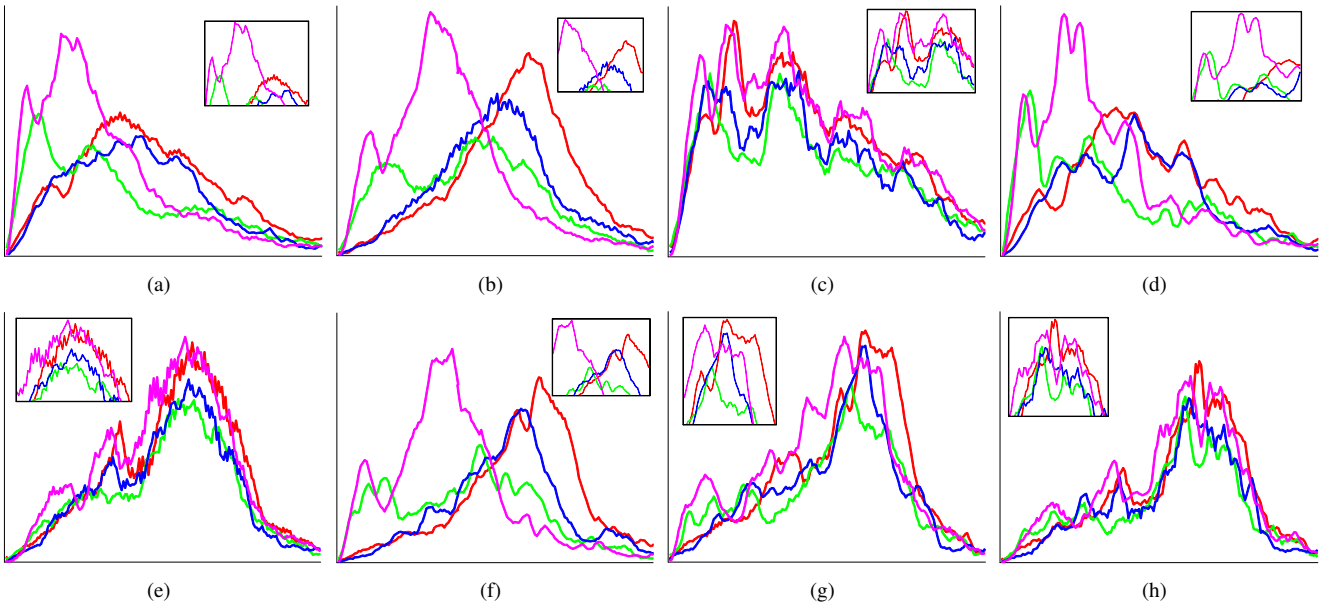

Fig. 1. Intensity distributions resulting from the application of post processing techniques on 4 different endorectal, prostate T2w MRI patient studies (each color corresponds to a different patient). Inlets depict regions of intensity distribution alignment, or lack thereof. Given the original uncorrected images (a) $\mathcal{C}$, results of post-processing are shown for (b) $\mathcal{C}^{\kappa}$ (BFC), (c) $\mathcal{C}^{\tau}$ (IS), and (d) $\mathcal{C}^{\eta}$ (NF). Different combinations of these operations such as (e) $\mathcal{C}^{\kappa\tau}$ (f) $\mathcal{C}^{\kappa\eta}$ (g) $\mathcal{C}^{\kappa\tau\eta}$ (h) $\mathcal{C}^{\kappa\eta\tau}$ are further evaluated from the perspective of optimal image quality. Note that different operations are seen to affect the intensity distributions in different ways. Scenarios where misalignment between distributions is observable ((a), (b), (d), (f), (g)) are reflective of the non-standardness artifact.

operations affect the subsequent image intensity distributions in very different ways (Figures 1(b), (c), (d)). The original, un-processed image (Figure 1(a)) reveals the effects of bias field (multiple modes in distribution), noise (non-smooth distributions), and non-standardness (mis-alignment between intensity distributions; modes do not line up). BFC (Figures 1(b), (e), (f), (g), (h)) and NF (Figures 1(d), (f), (g), (h)) result in smoother distributions with the modes corresponding to the bias field eliminated. Finally all sequences involving IS (Figures 1(c), (e), (g), (h)) result in removal of the intensity drift by aligning the intensity distributions. Note however, that in Figure 1(g) where NF succeeds IS, some mis-alignment in distributions (non-standardness) appears to have been introduced.

In [4], [9], [10], [11] performance measures such as percent coefficient of variation (%CV) and peak signal to noise ratio were employed to specifically evaluate the effects of BFC and NF, respectively. Relatively few performance measures have been proposed to evaluate the effect of IS. While Madabhushi and Udupa [4] utilized normalized mean intensity (NMI) to evaluate IS, this statistic is calculated for each acquisition individually, and hence may be sub-optimal when evaluating inter-acquisition non-standardness (i.e. amount of intensity drift between different studies). In this work, we introduce a novel measure to evaluate IS, termed difference of modes (DoM). In addition to evaluating the effects of NF, BFC, and IS individually, we also evaluate the effect of the sequences in the context of a domain specific task: detecting prostate cancer (CaP) on *in vivo* T2w MRI.

The contributions of this study are hence:

- A novel performance measure, difference of modes, for evaluating the effect of IS.
- Quantitatively determining the optimal sequence of BFC, IS, NF in order to maximize image quality and classifier accuracy for CaP detection on T2w, endorectal prostate MRI.

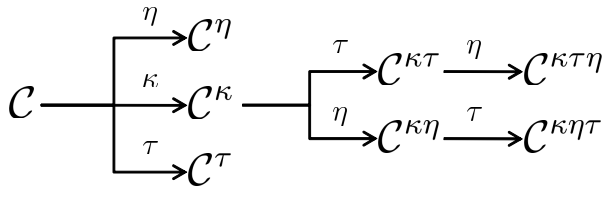

Fig. 2. A flowchart representing the 7 post-processing sequences we have experimentally compared, in addition to the uncorrected image. $\mathcal{C}$ represents an uncorrected image, κ represents BFC, τ represents IS, and η represents NF.

II. Experimental Design

A. Data Description and Notation

Twenty two *in vivo* 3 Tesla pre-operative endorectal T2w MRI patient datasets were acquired. For each patient, annotations of CaP presence and extent had been mapped onto T2w MRI (as described in [12]) via image registration with corresponding *ex vivo* histology, for a total of 94 distinct T2w MR images. These annotations formed the ground truth for disease extent when training and evaluating our classifier for CaP detection. We denote an uncorrected T2w prostate MR image as $\mathcal{C} = (C, f)$, where $f(c)$ is the measured signal intensity associated with every voxel c in a 3D grid C.

Type	Name	Evaluation of	Interpretation
Generalized post-processing evaluation	% Coefficient of Variation (ϕ^{CV})	BFC (BFC)	Lower ϕ^{CV} implies more effective BFC [10]
	Difference of Modes (ϕ^{DoM})	IS (IS)	Lower ϕ^{DoM} indicates improved IS
Application-based evaluation	Area Under the Curve (ϕ^{AUC})	classification performance	Higher ϕ^{AUC} implies superior classification performance [9]

TABLE I

PERFORMANCE MEASURES OF CORRECTION, STANDARDIZATION, AND FILTERING SCHEMES.

B. Review of post-processing operations

1) Bias Field Correction (BFC): The N3 method [1] was chosen for BFC based on its balance of calculation speed as well as robustness [9]. N3 seeks to find the multiplicative field that maximizes the high frequency components of the MR image intensity distribution. This is achieved by assuming the field is Gaussian, de-convolving the field into narrower Gaussian distributions and generating estimates of the true underlying intensity distribution. This process is performed iteratively until changes to the estimated intensity distribution are minimal [1]. The result of BFC on the original acquired MR image $\mathcal{C}$ (Figure 1(a)) is denoted by $\mathcal{C}^{\kappa}$ (Figure 1(b)).

2) Intensity Standardization (IS): Landmark-based intensity standardization [2] was used to align the image intensity distributions between patient studies. Intensity landmarks are first chosen at specific percentiles on a template distribution: the 1st percentile, 99th percentile, and every tenth percentile from 10 to 90 [2]. Test and template intensity distributions are then non-linearly aligned to one another by mapping intensity ranges between corresponding landmark intensities on the two distributions. The result of standardizing the scene $\mathcal{C}$ is denoted by $\mathcal{C}^{\tau}$ (Figure 1(c)).

3) Noise Filtering (NF): Anisotropic diffusion based filtering is an intra-region smoothing process for the removal of noise and the enhancement of edges. Regions are selected based on edge estimation using the gradient of intensity values within the MR image[3]. The regions are smoothed internally, effectively removing Gaussian noise. After application of NF to $\mathcal{C}$, the result is denoted by $\mathcal{C}^{\eta}$ (Figure 1(d)).

C. Outline of post-processing sequences

Fifteen possible post-processing sequences could be considered, utilizing different combinations of BFC, IS, and NF: $\mathcal{C}^{\kappa}$, $\mathcal{C}^{\tau}$, $\mathcal{C}^{\eta}$, $\mathcal{C}^{\kappa\tau}$, $\mathcal{C}^{\kappa\eta}$, $\mathcal{C}^{\tau\kappa}$, $\mathcal{C}^{\tau\eta}$, $\mathcal{C}^{\eta\kappa}$, $\mathcal{C}^{\eta\tau}$, $\mathcal{C}^{\kappa\tau\eta}$, $\mathcal{C}^{\tau\kappa\eta}$, $\mathcal{C}^{\tau\eta\kappa}$, $\mathcal{C}^{\eta\kappa\tau}$, and $\mathcal{C}^{\eta\tau\kappa}$. Based on previous work by Montillo et al [8] and Madabhushi and Udupa [4], we eliminated 8 of these combinations by ensuring that NF and IS were always preceded by BFC. The 7 remaining operations were each applied to the data and results were individually evaluated. We additionally evaluated the unprocessed image $\mathcal{C}$ to observe the exact improvements offered via post-processing MRI data.

1) Percentage co-efficient of variation (ϕ^{CV}): Employed to evaluate the effects of BFC [4], [9], [8], [10], and defined as

$$\phi^{CV} = \frac{\sigma}{\mu},$$

where σ is standard deviation, and μ is mean of a region of interest, which was the prostate for the purposes of this study. A decrease in ϕ^{CV} implies a reduction in the bias field inhomogeneity [10].

2) Difference of Modes ϕ^{DoM} for evaluating non-standardness: This measure is expressed as,

$$\phi^{DoM} = \frac{\mid \omega^{test} - \omega^{temp} \mid}{\omega^{temp}},$$

where ω^{test} is the principal mode of the intensity distribution of an MR image undergoing IS, and ω^{temp} is the principal mode of the template intensity distribution that is being standardized against. We posit that the average ϕ^{DoM} over all MR images that have been standardized will be higher if their intensity distributions remain misaligned. Note that the variance of ϕ^{DoM} must also be calculated to throughly test the extent to which intensity non-standardness has been eliminated following IS. A high ϕ^{DoM} is reflective of a high degree of intrinsic non-standardness.

3) Classifier-based evaluation: Automated classification of CaP on MRI data was considered for independent evaluation of post-processing operations. The classifier scheme [13], [9] comprises the following steps:

1) Extracting 110 texture features to model appearance of CaP and benign regions on prostate MR imagery [13].
2) Feature selection[14], to select the 10 most discriminatory texture features for CaP presence.
3) Performing automatic classification to identify presence or absence of CaP on a per-pixel basis via a Bayesian classifier upon the output of Step 2 [13].
4) Evaluating the classifier results from Step 3 against the ground truth extent for CaP on MRI on a per pixel basis [13].

Classification was performed in this manner for each of 94 slices (using a 3-fold cross validation approach to avoid training bias), and repeated for each of the 7 post-processing sequences (along with the uncorrected data) compared in this work. Cross validation entailed dividing the data set into 3 subsets, and each subset was used as a test set once. After all subsets were tested once, the subsets were randomized. 25 iterations of randomized cross validation were performed. Classifier accuracy was evaluated utilizing the classical Area Under the ROC Curve (AUC) parameter.

III. RESULTS AND DISCUSSION

Each sequence shown in Figure 2 was evaluated on a per-slice basis in terms of ϕ^{CV}, ϕ^{DoM}, as well as ϕ^{AUC}(derived from CaP classifier). One-way ANOVA tests were then conducted for each measure to determine statistical significance

Technique	ϕ^{CV}	ϕ^{DoM}	ϕ^{AUC}
$\mathcal{C}$	58.047±10.251	0.3968±0.220	0.6299±0.00470
$\mathcal{C}^{\kappa}$	37.719 ±8.253	0.2208±0.0483	0.6316±0.00446
$\mathcal{C}^{\tau}$	46.059±1.330	0.2202±0.0372	0.6182±0.01441
$\mathcal{C}^{\eta}$	57.001±10.75	0.6418±0.0221	0.6078±0.01707
$\mathcal{C}^{\kappa\tau}$	25.329±1.044	**0.0597±0.0033**	0.6084±0.00972
$\mathcal{C}^{\kappa\eta}$	35.883±8.922	0.2351±0.0410	0.6306±0.01410
$\mathcal{C}^{\kappa\tau\eta}$	23.838±1.218	0.0796±0.0028	0.6240±0.02580
$\mathcal{C}^{\kappa\eta\tau}$	**21.9324±1.002**	0.0996±0.0067	**0.6412±0.01212**

TABLE II

AVERAGES AND STANDARD DEVIATIONS OF ϕ^{CV}, ϕ^{AUC}, AVERAGE AND VARIANCE OF ϕ^{DoM}. VALUES AVERAGED OVER 94 DISTINCT MR IMAGES. BEST RESULTS FOR EACH COMPARISON HAVE BEEN BOLDED.

of the results with the null hypothesis being that there was no difference in performance between the 8 sequences considered over all 94 slices in each comparison. The Analysis of Variance (ANOVA) test compares the means of multiple samples to determine whether differences are the result of error, or assignable cause (statistical significance). A single ANOVA test fulfills the same purpose as multiple t-tests, without the accompanying complications.

A. Experiment 1: To determine which post-processing operation (κ, τ, η) maximally improves image quality

$\mathcal{C}^{\kappa}$ was found to significantly ($p < 0.05$) outperform both $\mathcal{C}^{\eta}$ and $\mathcal{C}^{\tau}$ in terms of ϕ^{CV} and ϕ^{AUC}, which resonates with previous results obtained by Madabhushi et al [4] and Montillo et al [8]. When inspecting the contribution of η and τ individually to the result of $\mathcal{C}^{\kappa}$, $\mathcal{C}^{\kappa\eta}$ and $\mathcal{C}^{\kappa\tau}$ were found to demonstrate mixed performance (optimal in terms of ϕ^{CV} and ϕ^{DoM} but not in terms of ϕ^{AUC}).

B. Experiment 2: To determine the sequence of post-processing operations that maximally improves image quality

$\mathcal{C}^{\kappa\eta\tau}$ demonstrated statistically significantly ($p < 0.05$) superior performance compared to the remaining 7 sequences in terms of ϕ^{AUC} and ϕ^{CV}. While $\mathcal{C}^{\kappa\eta\tau}$ performed marginally worse compared to $\mathcal{C}^{\kappa\tau\eta}$ in terms of ϕ^{DoM}, this may be on account of the fact that the decrease in ϕ^{DoM} for $\mathcal{C}^{\kappa\eta\tau}$ with respect to $\mathcal{C}^{\kappa\eta}$ is significantly larger as compared to the corresponding change from $\mathcal{C}^{\kappa\tau}$ to $\mathcal{C}^{\kappa\tau\eta}$ (where ϕ^{DoM} can be seen to increase). Note that when τ precedes η, lower quality (in terms of both ϕ^{CV} and ϕ^{AUC}) is observed in comparison to the unprocessed image $\mathcal{C}$, implying that NF is ideally applied prior to IS. Interestingly, our results concur with the combined conclusions of Madabhushi et al. [4] and Montillo et al. [8] who respectively found that (a) bias field correction should precede intensity standardization and (b) noise filtering should follow bias field correction.

C. Experiment 3: To determine the effectiveness of ϕ^{DoM} as an evaluation measure for image intensity standardization

ϕ^{DoM} was found to accurately quantify the effect of IS in all our experiments. In every case, both mean ϕ^{DoM} and variance of ϕ^{DoM} were found to significantly decrease after IS was applied to the data ($p < 0.05$). We also found that NF increases ϕ^{DoM} as did BFC.

Comparison	ϕ^{CV}	ϕ^{Dom}	ϕ^{AUC}
$\mathcal{C}$ $\mathcal{C}^{\kappa}$ $\mathcal{C}^{\tau}$, $\mathcal{C}^{\eta}$	$1.35\exp^{-54}$	$2.97\exp^{-16}$	$7.09\exp^{-11}$
$\mathcal{C}$ $\mathcal{C}^{\kappa\tau}$, $\mathcal{C}^{\kappa\eta}$	$3.17\exp^{-85}$	$5.88\exp^{-30}$	$1.30\exp^{-11}$
$\mathcal{C}$ $\mathcal{C}^{\kappa\tau\eta}$, $\mathcal{C}^{\kappa\eta\tau}$	$6.60\exp^{-174}$	$1.12\exp^{-45}$	.00235
All Sequences	$3.30\exp^{-132}$	$2.29\exp^{-56}$	$1.19\exp^{-16}$

TABLE III

P-VALUES FOR ANOVA TESTS, NULL HYPOTHESIS BEING THAT THERE IS NO DIFFERENCE BETWEEN SEQUENCES CONSIDERED IN EACH COMPARISON. THE COMPARISON IS STATISTICALLY SIGNIFIGANT

IV. CONCLUDING REMARKS

In this study, we have attempted to determine the optimal post-processing sequence for bias field correction, intensity standardization and noise filtering on 3 Tesla prostate T2-weighted MRI data. Our main conclusions were:

- The most optimal sequence was determined to be bias field correction, followed by noise filtering, and intensity standardization.
- Noise filtering, like bias field correction, introduces non-standardness.
- We have presented and demonstrated the effectiveness of a novel evaluation measure which quantifies the effect of intensity standardization, called difference of modes.

REFERENCES

[1] J.G. Sled et al, "A Non-parametric Method for Automatic Correction of Intensity Non-uniformity in MRI Data," IEEE Trans Med Imaging, vol. 17, pp. 87–97, 1998.

[2] L. Nzul et al, "New Variants of a Method of MRI Scale Standardization," IEEE Trans Med Imaging, vol. 19, pp. 143–150, 2000.

[3] P.Perona et al, "Scale-Space and Edge Detection Using Anisotropic Diffusion," IEEE Trans Pat Mach Intel, vol. 12, pp. 629–639, 1990.

[4] A. Madabhushi et al, "Interplay between Intensity Standardization and Inhomogeneity Correction in MR Image Processing," IEEE Trans Med Imaging, vol. 24, pp. 561–576, 2005.

[5] Ulas Bagci et al, "The role of intensity standardization in medical image registration," Pattern Recogn. Lett., vol. 31, 2010.

[6] B.M Dawant et al, "Correction of intensity variations in MR images for computer-aided tissue classification," IEEE Trans Med Imag, vol. 12, 1993.

[7] J.D Gispert et al, "Method for bias field correction of brain T1-weighted magnetic resonance images minimizing segmentation error," Hum Brain Mapp, vol. 22, 2004.

[8] A. Montillo et al, "Interaction between Noise Suppression and Inhomogeneity Correction in MRI," SPIE Med Imag, vol. 5032, pp. 1025–1036, 2003.

[9] S. Viswanath, A. Madabhushi et al, "Empirical Evaluation of Bias Field Correction Algorithms for Computer-Aided Detection of Prostate Cancer on T2w MRI," Proc. of SPIE, vol. 7963, 2011.

[10] B. Belaroussi et al, "Intensity non-uniformity correction in MRI: Existing Methods and their validation," Med Image Anal, vol. 10, pp. 234–246, 2006.

[11] C.P Loizou et al, "Comparative Evaluation of Despeckle Filtering in Ultrasound Imaging of the Carotid Artery," IEEE T Ultrason Ferr, vol. 52, pp. 1653–1669, 2005.

[12] J. Chappelow, A. Madabhushi et al, "Elastic Registration of Multimodal Prostate MRI and Histology via Multi-Attribute, Combined Mutual Information," Medical Physics, vol. 38, 2011.

[13] A. Madabhushi et al, "Automated Detection of Prostatic Adenocarcinoma from High Resolution Ex Vivo MRI," IEEE Trans Med Imaging, vol. 24, no. 12, pp. 1611–1625, 2005.

[14] H. Peng et al, "Feature selection based on mutual information: criteria of max-dependcy, max relevance, and min-redundancy," IEEE Trans Pat Mach Intel, vol. 27, 2005.