Distinguishing prostate cancer from benign confounders via a cascaded classifier on multi-parametric MRI

G.J.S. Litjens\textsuperscript{a}, R. Elliott\textsuperscript{b}, N. Shih\textsuperscript{c}, M. Feldman\textsuperscript{c}, J.O. Barentsz\textsuperscript{a}, C.A. Hulsbergen - van de Kaa\textsuperscript{a}, I. Kovacs\textsuperscript{a}, H.J. Huisman\textsuperscript{a} and A. Madabhushi\textsuperscript{b}

\textsuperscript{a}Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; \textsuperscript{b}Case Western Reserve University, Cleveland, USA; \textsuperscript{c}University of Pennsylvania, Philadelphia, USA;

ABSTRACT

Learning how to separate benign confounders from prostate cancer is important because the imaging characteristics of these confounders are poorly understood. Furthermore, the typical representations of the MRI parameters might not be enough to allow discrimination. The diagnostic uncertainty this causes leads to a lower diagnostic accuracy. In this paper a new cascaded classifier is introduced to separate prostate cancer and benign confounders on MRI in conjunction with specific computer-extracted features to distinguish each of the benign classes (benign prostatic hyperplasia (BPH), inflammation, atrophy or prostatic intra-epithelial neoplasia (PIN)). In this study we tried to (1) calculate different mathematical representations of the MRI parameters which more clearly express subtle differences between different classes, (2) learn which of the MRI image features will allow to distinguish specific benign confounders from prostate cancer, and (2) find the combination of computer-extracted MRI features to best discriminate cancer from the confounding classes using a cascaded classifier. One of the most important requirements for identifying MRI signatures for adenocarcinoma, BPH, atrophy, inflammation, and PIN is accurate mapping of the location and spatial extent of the confounder and cancer categories from ex vivo histopathology to MRI. Towards this end we employed an annotated prostatectomy data set of 31 patients, all of whom underwent a multi-parametric 3 Tesla MRI prior to radical prostatectomy. The prostatectomy slides were carefully co-registered to the corresponding MRI slices using an elastic registration technique. We extracted texture features from the T2-weighted imaging, pharmacokinetic features from the dynamic contrast enhanced imaging and diffusion features from the diffusion-weighted imaging for each of the confounder classes and prostate cancer. These features were selected because they form the mainstay of clinical diagnosis. Relevant features for each of the classes were selected using maximum relevance minimum redundancy feature selection, allowing us to perform classifier independent feature selection. The selected features were then incorporated in a cascading classifier, which can focus on easier sub-tasks at each stage, leaving the more difficult classification tasks for later stages. Results show that distinct features are relevant for each of the benign classes, for example the fraction of extra-vascular, extra-cellular space in a voxel is a clear discriminator for inflammation. Furthermore, the cascaded classifier outperforms both multi-class and one-shot classifiers in overall accuracy for discriminating confounders from cancer: 0.76 versus 0.71 and 0.62.

Keywords: CAD, feature selection, cascading classification, prostate cancer, MRI, confounders

1. INTRODUCTION

One of the most difficult tasks in diagnosing prostate cancer on MRI is separation between prostate cancer and benign confounders like atrophy, inflammation, benign prostatic hyperplasia (BPH) and prostatic intra-epithelial neoplasia (PIN).\textsuperscript{1,2} Atrophy occurs generally in the aging prostate and results in the loss of the normal lobular characteristics of the prostate glands; resulting in different structures in pathology. Inflammation of the prostate occurs for various reasons, for example urinary retention due a blocked urethra or biopsy. Inflammation tends to cause vaso-dilation and apoptotic or necrotic processes. Benign prostatic hyperplasia is the benign growth of the prostate transition zone (i.e. proliferation of epithelial and stromal cells), which occurs in many older men, although the exact cause is unknown. This results in a nodular, high density area of either stromal or glandular cells. Finally, PIN is essentially a carcinoma in situ, an abnormal proliferation of epithelial cells which yet remain contained inside the gland. Some PIN progresses to cancer, some doesn’t. One reason that it is hard to separate
Table 1: Description of overlap in imaging attributes on multi-parametric MRI for adenocarcinoma and confounding tissue classes.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>T2W</td>
<td>High-resolution allows texture characterization, cancer exhibits so-called ‘erased charcoal’ sign. Different confounding classes might also exhibit different textures.</td>
</tr>
<tr>
<td>DWI</td>
<td>Characterizes tissue properties like cell density. Due to fast proliferation of cells prostate cancer tends to exhibit low diffusivity. Benign prostatic hyperplasia is also known to have low diffusivity.</td>
</tr>
<tr>
<td>DCE</td>
<td>Enhancement over time is related to micro-vessel density and permeability. Prostate cancer tends to have a high vascularization with leaky vessels. Confounders like inflammation are also known to exhibit a vascular response.</td>
</tr>
</tbody>
</table>

Figure 1 illustrates the challenges in distinguishing between cancer and all the various confounding categories in "one shot". In Figure 1(a) we show what happens when we group all confounders as a single class. A linear discriminant classifier attempting to identify every image location as belonging to either cancer or benign confounding classes, ends up assigning every sample to the non-cancer class. In Figure 1(b), we show what happens when you use the same features, and attempt a pair-wise classification (in this case atrophy versus cancer). We see that in this 2-class separation problem, the LDA classifier is able to create an accurate decision boundary. Figure 1(c)-(e) further illustrate the utility of the pair-wise classification approach. However, it is important to note that the choice of features for distinguishing the classes of interest is as important as the specific choice of hierarchy within the cascade for separating adenocarcinoma from benign confounding classes.

The types of features we use to characterize the benign confounders are based on the diagnostic guidelines presented in. A short summary is also presented in Table 1. The T2-weighted imaging is mostly used for its high resolution and contrast, allowing detailed visualization of the tissue anatomy. From clinical guidelines we know that the T2-weighted images are especially useful to assess the texture of prostate lesions. Prostate cancer exhibits a so-called 'erased charcoal sign', a smudge-like dark texture on T2-weighted images. Different benign confounders might also express different types of textures.

Diffusion weighted imaging is specifically useful for characterizing tissue at a microscopic level, enabling us to assess traits like cellular density at a macro level. Prostate cancer has a high cellular density compared to the normal glandular structure of the prostate. This results in a reduced diffusivity in cancerous tissue and thus a high signal in high b-value images and subsequently a low apparent diffusion coefficient values.

Finally, dynamic contrast-enhanced MRI result in signal over time curves and shows the uptake of contrast agent in tissue. This allows us to measure attributes of the tissue vasculature, like the relative fraction of extra-vascular, extra-cellular space in each voxel and micro-vessel permeability. Prostate cancer lesions tend to have leaky micro-vasculature, which results in fast initial enhancement and wash-out. It is known that for example inflammation also causes a vascular response, however, it is unknown in which way this differs from prostate cancer.
Figure 1: Figure (a) shows the results for training a linear discriminant classifier to separate all the benign confounding classes (0) and cancer (1) at once given two features. As it is unable to discriminate the classes, it assigns every sample to class 0. The other figures show that in a cascaded setting, separating out one benign confounder per step, gives a reasonable decision boundary given discriminative features. In figures (b)-(e) class 0 is atrophy, class 1 is cancer, class 2 is BPH, class 3 is inflammation and class 4 is HGPIN. An explanation of the features used in these plots can be found in section 3.3.

However, as previously suggested the scanner derived MRI parameters by themselves might not possess sufficient discriminability to distinguish adenocarcinoma from the confounding tissue classes, motivating the need for computer extracted image features. The first goal of this study was (1) to obtain computer-extracted features which are designed to better capture specific subtle differences between the classes. For the second goal we need to (2) establish which combination of regular and computer-extracted features is the most suitable for each benign confounding class.
To be able to assess whether these identified features are useful in discriminating the different confounders and cancer, we can integrate them into different classification strategies. One option is to use a single shot classification approach, considering all the benign confounders as a single class and trying to differentiate this class from prostate cancer. The second option is to use a multi-class classification approach in which each benign confounder is a separate class. In this paper we will also investigate a third strategy: the use of a cascaded classifier to separate the confounding classes and prostate cancer step-by-step. Similar studies have shown cascaded approaches to outperform regular one-shot classification or multi-class classification approaches because it sub-divides the difficult task of separating confounders and cancer into easier sub-problems, thus circumventing issues with similar classes. In the paper by Doyle et al. they showed that cascaded approaches resulted in better discrimination of histopathologic tissue classes compared to multi-class classification. In Figure 1 we presented an illustrative example of such an approach. Furthermore, it allows us to perform feature selection for each problem separately. As such the third goal of this paper is (3) to establish if cascaded classifiers outperform regular single-shot or multi-class classifiers.

The rest of the structure of this paper is as follows: first we will discuss previous work and our novel contributions. Second, we will discuss the methodology and experimental design. This will cover histology/MRI co-registration, feature extraction and selection and subsequently classification. The next section will present and discuss the results after which we will end with concluding remarks and future work.

2. PREVIOUS WORK AND NOVEL CONTRIBUTIONS

Only little work has been done on accurately characterizing the appearance of different benign confounders in prostate MRI, most likely due to difficulty of obtaining accurate annotations. Several groups have investigated specific confounding classes using a single modality, for example benign prostatic hyperplasia on diffusion-weighted imaging (Liu et al., Oto et al.). Another example is the differentiation of prostatitis using diffusion-weighted imaging (Nagel et al.). None of these groups have looked at all the benign confounders or all the modalities.

There has been some previous work on computer-extracted features for the detection of prostate cancer. investigated the use of magnetic resonance spectroscopy in combination with T2-weighted imaging to identify the voxels that are affected by prostate cancer. They also introduced the use of wavelet embedding to map MRS and T2-W texture features into a common space. This work was further expanded and evaluated in. Niaf et al. presented the use of computer-aided diagnosis in the peripheral zone of the prostate using DCE features (similar to Vos et al.). They confirmed the results in discriminating prostate cancer from normal regions (area under the ROC curve (AUC)=0.89) and discriminating prostate cancer from suspicious benign regions (AUC of 0.82). Singanamalli et al. investigated the use of DCE-MRI/histopathology fusion to assess feature correlations between prostate cancer of different grades. Lastly, Viswanath et al. investigated the use of texture features to discriminate prostate cancer from normal and benign regions. They also found different texture features were important depending on the originating prostatic zone of the cancer.

In this paper we have several novel contributions:

1. Use of pathology annotated benign confounder classes to identify corresponding regions on the MRI
2. Identifying important imaging features per confounding class over all prostate MRI parameters
3. Establishing computer-extracted features which better capture subtle differences between the confounding classes
4. Improving computerized prostate cancer diagnosis using a cascaded classifier which incorporates per-stage feature selection
3. METHODOLOGY

3.1 Brief overview
These methods are evaluated in a unique MRI/histology data set with annotations of cancer and the benign confounding classes on the histopathologic slides from 31 patients with biopsy-confirmed prostate cancer. The prostatectomy slides were then carefully co-registered to the MRI to obtain the MR regions corresponding to the different classes. The first goal was to (1) extract texture, pharmacokinetic and intensity features for each of the classes to establish the imaging signatures, after which we (2) used maximum relevance minimum redundancy (mRMR) feature selection\(^{22}\) to identify the best feature combination for each class. Using this information a cascading classifier which will remove each confounding class in a step-by-step approach was created. Subsequently, (3) the cascaded classifier was evaluated with respect to accuracy of classifying cancer and compared to single-shot two-class and multi-class classification.

3.2 Annotation and co-registration of histopathology
A pathologist contoured distinct examples of each benign confounding class and prostate cancer on the histopathology slides using the Aperio ImageScope software, if present. The pathology annotations were transferred to the MRI by registering the whole-mount slide using a thin plate spline registration technique.\(^{21}\) The process, in a step-by-step fashion, goes as follows:

1. The slice in the MRI which corresponds to the prostatectomy slide is established by an image analysis researcher under the supervision of a radiologist by comparing landmarks on the pathology and the MRI.\(^{23}\)

2. Corresponding points are indicated on the prostate boundary for both the prostatectomy slide and the MRI slice.

3. A b-spline transformation is calculated to move from the prostatectomy coordinate space to MRI coordinate space.\(^{24}\)

4. The histopathology image is transformed to the MRI space using this b-spline transformation. A visual assessment of whether the registration is accurate is made.

5. The annotations of the pathologist are transferred to the MRI using this b-spline transformation.

3.3 Feature extraction and feature selection

Correction of intensity drift
Intensity drift and bias field are distinct issues that are well known in MRI.\(^{21}\) Intensity drift means that intensities differ from scanner to scanner and even from protocol to protocol on the same scanner. The bias field caused by the coil causes signal intensity fluctuation of the same tissue across the field of view. To circumvent both issues in T2-weighted images we can calculate a (pseudo)T2-map using the transverse T2W-image and the proton density-weighted image as described in.\(^{13}\) This approach uses MR signal equations and a muscle reference region of interest to reduce intensity drift between the different studies.

T2-weighted imaging
For the T2-weighted texture features we calculated several often used filter types: 13 Haralick texture features using 3 kernel sizes (3, 5 and 7 voxels), Gabor texture features using 4 different angles and 3 different wavelengths between 1mm and 6mm and Gaussian derivatives up to second order using 4 different scales between 1mm and 6mm.\(^{21}\) The texture features were all calculated on the (pseudo)T2-map.

Diffusion-weighted imaging
From the diffusion-weighted imaging we directly incorporate the apparent diffusion coefficient and b800 image. Furthermore, as prostate cancer lesions tend to have a focal appearance on diffusion weighted imaging, we implemented the multi-scale blobness filter proposed by Li et al.\(^{25}\) and calculated the filter using 4 scales between 1 and 6 mm on the b800 and ADC images.
<table>
<thead>
<tr>
<th>Category</th>
<th>Feature name</th>
<th>MR sequence</th>
<th>Feature settings</th>
<th>Captures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intensity</td>
<td>(Pseudo)T2-map$^{27}$</td>
<td>T2W</td>
<td>Up to 2nd order, $\sigma=2.0, 2.67, 4.1$ and 6.0 mm</td>
<td>Molecular structure, Cellular density</td>
</tr>
<tr>
<td></td>
<td>ADC</td>
<td>DWI</td>
<td>Kernel sizes 3, 5, 7 voxels</td>
<td>Cellular density</td>
</tr>
<tr>
<td></td>
<td>b-800</td>
<td>DWI</td>
<td>Four angles: $0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}$, $\lambda=1.5, 2$ and 4 voxels</td>
<td>Homogeneity, quantitative texture measures</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\sigma=2.0, 2.67, 4.1$ and 6</td>
<td>Edge strengths, stability of edges</td>
</tr>
<tr>
<td></td>
<td>Gaussian Derivatives$^{27}$</td>
<td>T2Map</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Haralick$^{21}$</td>
<td>T2Map</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gabor$^{21}$</td>
<td>T2Map</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Multi-scale blobness$^{25}$</td>
<td>All</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pharmacokinetic</td>
<td>Curve fitting parameters$^5$</td>
<td>DCE</td>
<td>Time-to-peak, maximum enhancement, wash-out rate</td>
<td>Semi-quantitative Micro-vasculature permeability, extra-cellular space</td>
</tr>
<tr>
<td></td>
<td>Std. Tofts PK model$^5$</td>
<td>DCE</td>
<td>Ktrans, Kep, Ve</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Table of feature and feature settings calculated on the different MR sequences and used in the different classifiers. Features are chosen such that they represent physiological quantities which we hypothesize are different in the different confounding classes. For example, inflammation is known to cause vaso-dilation and as such result in a different response compared to the other confounders.

**Dynamic contrast-enhanced imaging**

Dynamic contrast-enhanced MRI also tends to suffer from scanner and protocol dependency.$^{12,19,26}$ To remove this dependency and extract the most useful information from these curves we implemented curve fitting and pharmacokinetic modeling routines as presented in $^{5,12,26}$ The curve fitting routine uses a bi-exponential curve model and was implemented to enable faster and more robust pharmacokinetic modeling. The pharmacokinetic model that was implemented was the standard Tofts model,$^{26}$ which ignores the vascular component in each voxel. The temporal resolution of the DCE time series was 4 seconds. To capture characteristics on the microvasculature we included a total of 3 curve features (time-to-peak, washout rate and maximum enhancement)$^{12}$ and 3 pharmacokinetic features (Ktrans, Kep, Ve)$^5$. Furthermore, as cancer also tends to have a focal appearance on DCE MRI$^3$ we also calculate the Li blobness filter on the Ktrans, Kep, Ve, maximum enhancement, time-to-peak and wash-out rate images.

**Feature selection**

The maximum relevance, minimum redundancy (mRMR) feature selection technique$^{22}$ was used to determine features for distinguishing each of the following two-class classification tasks (BPH vs. cancer, inflammation vs. cancer, HGPIN vs. cancer and atrophy vs. cancer). This mRMR feature selection tries to maximize the mutual information between the individual features and the class labels and minimize the mutual information between features to select the most relevant features without redundancy. After mRMR feature selection, unique features discriminating each benign class from cancer were identified. mRMR feature selection was selected because it tends to give good results at reasonable computation time and does not need parameter optimization or nested cross-validation loops like for example sparse coding.$^{28}$
Cascading Classifier

The feature selection step can be used to improve the classification of unknown samples in prostate cancer and benign disease by using a cascaded classification scheme. At each step of the cascade the classifier attempts to separate out one type of benign disease. This is schematically shown in Figure 2. The order of the cascade was empirically determined.

At each step of the cascade, mRMR feature selection is used. Training data at that step consists of the benign confounder to be detected at that stage (in step 1 inflammation), which is given class label 0 and all classes that are upstream in the cascade (given class label 1). The training folds were balanced before feature selection. A linear discriminant classifier is then trained with the selected features. In each step either a sample is identified as either belonging to the target benign confounding class and removed from further classification, or it moves on to the next stage. At the final stage of the classifier cascade, the remaining samples are either classified as atrophy or as prostate cancer.

4. EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Patient data, annotation, and co-registration

For this study we included 31 patients with an MRI including T2-weighted, dynamic contrast-enhanced, diffusion-weighted and proton density-weighted image sequences. Inclusion criteria were persistently high PSA, initial negative TRUS biopsy, multi-parametric MRI positive for prostate cancer and subsequent prostatectomy.

MRI acquisition was performed using a 3 Tesla MRI scanner (either a TrioTim or a Skyra; Siemens, Erlangen, Germany). Cases were acquired both with and without an endorectal coil. A pelvic phased array coil was always used. The multi-parametric protocol consisted of three orthogonal T2-weighted volumes, diffusion-weighted imaging (three b-values averaged over three orthogonal directions, 50, 400-500 and 800) and dynamic contrast-enhanced imaging (15 mL of Dotarem; Guerbet, France).

After prostatectomy, in total 44 histological H&E stained slides where digitized using a digital slide scanner (VS120-S5, Olympus, Japan) at 10x or 20x, corresponding to a resolution of 0.6 um and 0.3 um respectively, with at least 1 slide per patient; sometimes two slides per patient were digitized. The slides selected contained the largest tumor volume, based on the interpretation by an experienced pathologist. Areas of prostate cancer,
Figure 3: Representative images of a T2-weighted MRI ((a),(d),(g)) and a H&E stained prostatectomy slide ((b),(e),(h)) and the result of the subsequent MRI/histology registration with ((c),(f),(i)) annotations. In this example prostate cancer is indicated in yellow, a BPH nodule in red, inflammation in blue and atrophy in green.
Table 3: The top 2 selected features are shown in the first two columns for each of the benign class versus cancer classification tasks. Each row is a separate patient. The last two columns show likelihood maps for the single-shot classification (column 3) and the cascaded classification (column 4). The overlays indicate likelihood value: transparent, green and red are a zero, low and high value respectively between 0 and 1. The confounding class is indicated with a dark blue contour and prostate cancer with a purple contour. The corresponding over-expressing features (first, second column) are described in Table 4.

BPH, atrophy, inflammation and HGPIN were subsequently annotated on each slide by one of two pathologists using the Aperio ImageScope software. The annotations made on the histopathologic slides were transferred to the MRI using the methodology explained in subsection 3.2. Slice correspondences were manually established by an image analysis researcher based on landmarks and the relative position within the prostatectomy data set (e.g. if the middle slide in the prostatectomy was selected it most likely does not correspond to the first prostate slice in the MRI). This resulted in a total of 68 atrophy, 25 BPH, 58 PIN, 47 inflammation and 93 prostate cancer lesions on MRI. An example of co-registered annotations is shown in Figure 3.

4.2 Experiment 1: Feature Selection via mRmR

In Table 4, the top 5 selected features for each of the different benign vs. cancer classification tasks are presented. Each task has a number of distinct unique features (indicated with the gray background). We also show the top
Table 4: Selected features for each of the benign classes during the feature selection for each of the classification tasks (BPH vs. cancer, atrophy vs. cancer). Feature that are only selected in one classification task, and thus are specific for that specific confounder, are indicated with gray backgrounds. In the table \( o \) is the order of the Gaussian derivative, \( \sigma \) the scale, \( \theta \) the angle of the Gabor filter, \( \lambda \) the bandwidth and \( \omega \) the kernel size for the Haralick texture features.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Inflammation</th>
<th>BPH</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ve Gauss. Deriv (( o = yy, \sigma = 4.1 ))</td>
<td>Gauss. Deriv (( o = yy, \sigma = 4.1 ))</td>
</tr>
<tr>
<td>2</td>
<td>Haralick Sum Variance (( \omega = 7 ))</td>
<td>b800 intensity</td>
</tr>
<tr>
<td>3</td>
<td>Blobness ADC</td>
<td>Time-to-peak</td>
</tr>
<tr>
<td>4</td>
<td>Gauss. Deriv (( o = xx, \sigma = 4.1 ))</td>
<td>Blobness ADC</td>
</tr>
<tr>
<td>5</td>
<td>Max. Enhancement</td>
<td>Gabor (( \lambda = 1.5, \theta = 0.0 ))</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rank</th>
<th>Atrophy</th>
<th>HGPIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Gauss. Deriv (( o = -, \sigma = 2.8 ))</td>
<td>Gauss. Deriv (( o = xx, \sigma = 4.1 ))</td>
</tr>
<tr>
<td>2</td>
<td>Haralick Measure of Correlation1 (( \omega = 3 ))</td>
<td>Blobness b800</td>
</tr>
<tr>
<td>3</td>
<td>Blobness (Pseudo)T2Map</td>
<td>ADC intensity</td>
</tr>
<tr>
<td>4</td>
<td>Blobness Kep</td>
<td>Haralick Contrast (( \omega = 3 ))</td>
</tr>
<tr>
<td>5</td>
<td>Gabor (( \lambda = 2.0, \theta = 0.0 ))</td>
<td>Blobness time-to-peak</td>
</tr>
</tbody>
</table>

2 selected features in Table 3.

### 4.2.1 Distinguishing inflammation from cancer

Inflammation often results in marked vascular changes like increased blood flow and permeability (to allow leukocytes to move to the inflamed tissue) and increased extra-cellular, extra-vascular space due to dying (apoptotic or necrotic) cells. From Table 4 we can see that this has been picked up by the feature selection strategy (selected features 1 and 5, Ve and maximum enhancement) and that these features can help discriminate between inflammation and prostate cancer.

### 4.2.2 Distinguishing atrophy from cancer

Atrophy of the prostate results in a loss of the regular structure of the glands in the prostate (reduced infolding, cytoplasm volume and presence of corpora amylacea). Atrophy tends to be focal and looks texturally different from prostate cancer on pathology. This can also be recovered on the MRI if we look at Table 4, where 3 out of the top 5 features are related to textural attributes. Finally, an excellent discriminator of atrophy versus prostate cancer would be the blobness feature on the (pseudo)T2-map, as prostate cancer tends to have a diffuse appearance with irregular border, whereas the focal appearance of atrophy on pathology carries over to the T2-weighted MRI.

### 4.2.3 Distinguishing BPH from cancer

Benign prostatic hyperplasia is benign glandular or stromal proliferation in the transition zone of the prostate. Typically BPH is a difficult confounder to separate from prostate cancer, as features like ADC intensity tend to show similar changes. From our feature selection results we can appreciate that instead of using ADC intensity to discriminate prostate cancer and BPH, it is more useful to look at the textured appearance on T2-weighted imaging, in combination with a focal appearance on the ADC map. BPH tends to have a more nodular, compact shape compared to prostate cancer which is picked up in our experiments.

### 4.2.4 Distinguishing PIN from cancer

Prostatic intra-epithelial neoplasia is a cancer pre-cursor, sometimes it will progress into cancer, sometimes it will not. As such, it shares similar imaging characteristics with cancer. However, in our experiments we can still identify features which can specifically help discriminate this confounder. Directly using ADC intensity works well in discriminating PIN, as it tends not yet to exhibit the high cellular density which is apparent in aggressive prostate cancer. Furthermore, as it still localized and usually well-defined, it exhibits a more compact and focal appearance compared to prostate cancer on the time-to-peak images.
Figure 4: Two-class accuracies over the number of selected features for the simple two-class strategy (green), multi-class classification (blue) and the cascade classification (red) including 95% confidence intervals. The cascaded classifier has higher accuracy for almost the entire range of features.

### 4.3 Cascaded classifier based evaluation of discriminating features separating cancer from benign confounders

The evaluation is based on the overall accuracy in separating benign disease and prostate cancer on a lesion level and compared to direct two-class classification of benign disease and prostate cancer and multi-class classification with and without feature selection. In the multi-class and cascaded classification the resulting benign classes are grouped as one benign class after classification, after which accuracy is calculated. The number of features for the classifiers and in each stage of the cascade was varied between 2 and 20 in steps of 2 to allow a comparison across the number of selected features. The validation of the cascaded classifier was performed in a five-fold patient-based cross-validation (20 repeated runs).

The results are shown in Figure 4. The cascaded classifier outperforms the single-shot classifier and the multi-class classifier in the overall accuracy of correctly classifying voxels as either containing cancer or not containing cancer over almost the entire range of features. The overall maximum accuracy is 76.4% when using 4 features for the cascading classifier and 63.0% and 71.0% for the single-shot classification and the multi-class classifier at 10 and 8 features respectively. Overall accuracy is limited because some lesions annotated on pathology ended up being only a couple of voxels on the MRI and are indistinguishable from normal tissue.

### 5. CONCLUDING REMARKS

A major issue in rendering a cancer diagnosis on prostate MRI is the presence of confounding benign classes which have similar image characteristics as prostate cancer on multi-parametric MRI. Since many of these confounder categories (e.g. BPH, PIN, atrophy, inflammation) have overlapping appearance to prostate cancer, they can mask the appearance of true disease on MRI. Hence it can be difficult for radiologists to assess whether a suspicious region is benign or malignant. Computerized feature extraction might allow better assessment of subtle differences between classes by changing the representation of the multi-parametric MRI in such a way that accentuates these subtle differences.

By leveraging ex vivo histopathology from patients undergoing radical prostatectomy with a pre-operative MRI we were able to map annotations of benign confounders from ex vivo pathology onto MRI. This allowed us to identify features which might help in discriminating between specific confounders and prostate cancer. Furthermore, we combined mRMR feature selection with a cascaded classifier, which can step-by-step learn features for each of the confounders, thus simplifying the classification task for the following steps, as opposed to single-shot or multi-class classifiers, which are restricted to only learning features common to all benign classes.
or having to learn all features at once. We applied these techniques to a data set of 31 patients for which we obtained whole mount prostatectomy and full multi-parametric MRI.

The main results include a unique set of image attributes for each of the benign classes. These selected features can be related to the underlying pathology to achieve a better understanding of which pathological processes cause certain image characteristics on MRI. An example here is the fraction of extra-cellular, extravascular space, \( V_e \), which is discriminative for inflammation. Our results suggest that only very specific computer extracted image features are able to accurately discriminate between the different benign confounder and prostate cancer. Our results also indicate that the computer extracted features, as opposed to the original scanner derived parameters (e.g., T2w intensities) can better express the subtle differences between classes than the original MR images is valid. The fact that computer extracted features from all 3 imaging protocols are differentially expressed in discrimination of one or more confounding tissue classes and cancer suggests that multi-parametric MRI is needed for improved prostate cancer diagnosis.

This study also has some limitations. We did not discriminate on lesion size when converting the pathology annotations to the MRI. This resulted in some very small lesions (< 5 voxels) which might get misclassified due to issues like partial volume effect or small co-registration errors. We also did not take into account combined classes, i.e. sometimes areas of atrophy or cancer also show active inflammation. Finally, the splits in the hierarchical classifier were determined empirically. A future avenue of work is to automatically learn the optimal cascade hierarchy to distinguish the confounder and cancer classes.

In the future we would like to validate our approach on a larger cohort and experiment with more complex classifiers for each of the steps in the cascade. Furthermore, we would like to look into a soft per-class classification to take into account class overlap (e.g., inflammation occurring together with atrophy).

5.1 Acknowledgements

Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under award numbers R01CA136535-01, R01CA140772-01, and R21CA167811-01; the National Institute of Diabetes and Digestive and Kidney Diseases under award number R01DK098503-02, the DOD Prostate Cancer Synergistic Idea Development Award (PC120857); the QED award from the University City Science Center and Rutgers University, the Ohio Third Frontier Technology development Grant. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES


