Bag of Bags: Nested Multi-Instance Classification for Prostate Cancer Detection

Conference Paper · December 2016

4 authors, including:

Farzad Khalvati
University of Toronto
64 PUBLICATIONS **177** CITATIONS
SEE PROFILE

Junjie Zhang
University of Toronto
22 PUBLICATIONS **62** CITATIONS
SEE PROFILE

Alexander Wong
University of Waterloo
263 PUBLICATIONS **1,333** CITATIONS
SEE PROFILE

Bag of Bags: Nested Multi Instance Classification for Prostate Cancer Detection

Farzad Khalvati, Junjie Zhang, Alexander Wong, Masoom A. Haider

***Abstract*— Computer-aided detection (CAD) algorithms have been proposed for auto-detection of different types of cancer. CAD algorithms rely on machine learning methods to classify regions of interest in images into cancerous and healthy regions. In cancer screening, the foremost problem to solve is whether a patient has cancer, regardless of the location of cancerous regions in the organ. This allows early detection of the disease leading to a right course of action in terms of treatment to be taken. In machine learning, this problem has been formulated as multi-instance learning (MIL) where bags of instances are classified rather than the individual instances. In this paper, we propose a bag of bags (BoB) nested MIL algorithm where high-level bags (or parent bags), each contains multiple smaller bags of instances. We applied the proposed BoB MIL algorithm to prostate cancer detection problem using magnetic resonance imaging data to first detect which patients have cancer and consequently, to detect which slices in the 3D volume imaging data of the detected patients contain cancerous regions. Experimental results obtained from the imaging data of 30 patients with ground-truth data based on biopsy results show that the proposed algorithm is not only capable of detecting prostate cancer at patient level, it is also able to detect the cancerous regions at slice level of imaging data with high accuracy.**

I. INTRODUCTION

Computer-aided detection (CAD) algorithms have been proposed for auto-detection of different types of cancer. The goal of CAD algorithms is to increase the accuracy and consistency of cancer detection. In cancer management, early and accurate detection is key for a right course of action to be taken. Prostate cancer is the most common form of cancer and third leading cause of cancer death diagnosed in Canadian men, with more than 24,000 new cases and 4,100 deaths in 2015 [1]. Nevertheless, early and accurate detection usually leads to high probability of survival [2]. Thus, it is crucial to have a reliable and accurate screening method for prostate cancer to improve the likelihood of survival.

Multi-parametric magnetic resonance imaging (mpMRI), which usually combines T2-weighted MRI (T2w) and diffusion-weighted imaging (DWI), is increasingly gaining traction in the clinical workflow of prostate cancer diagnosis. This is due to high accuracy in detection and non-invasive nature, as opposed to biopsy which is both painful and harmful [3], and the fact that imaging is capable of creating 3D images enabling the study of heterogeneity in cancerous regions, a well described phenomenon in cancer analysis with varying cell phenotypes. The conventional interpretations of mpMRI data are usually qualitative in nature which lead to high variability among readers and some clinically significant cancers are still missed [4].

To improve the accuracy and consistency of diagnosis, CAD tools have been proposed to aid the clinicians in interpreting the massive imaging data available in the current clinical settings for prostate cancer diagnosis [5], [6]. These CAD algorithms usually rely on pixel-level classification: first, a set of features (e.g., texture features) is computed for local neighbourhoods of pixels (e.g., 3×3). Due to high number of features, feature selection algorithms are usually applied to reduce the feature space. The selected features are used to train a classifier which is then used to classify new (unseen) windows of pixels into cancerous or healthy regions [5], [6], [7], [8], [9], [10], [11].

The proposed approaches in almost all of these CAD algorithms for prostate cancer are usually bottom-up; the classification is performed at pixel level (e.g., 3×3 neighbourhoods). Once the pixels in each slice of the 3D imaging data of a patient are classified, they are combined together (aggregated) to classify each slice to cancerous/non-cancerous, and then to decide whether the patient has cancer.

The drawback for bottom-up approaches is the fact that mpMRI data can be noisy, and pixel-level classification may result in many false positives leading to over-detection of prostate cancer (low specificity). In addition, the training data is usually heavily imbalanced at pixel-level; only a small fraction of pixels (e.g., less than 1%) are cancerous. Thus, in a 3D volume of imaging data for one patient, the vast majority of pixels are non-cancerous, making the classification a challenging task.

On the other hand, for cancer screening, the utmost important diagnosis action is to first determine whether the patient has cancer regardless of the location in the organ. Once correctly diagnosed, the consequent course of action may include the precise localization of the tumour. Inspired by this, in this paper, we propose a top-down approach for detection of prostate cancer.

Multi-instance learning or MIL aims for classifying bags of instances as the main classification goal [12]. In this method, each bag contains multiple feature vectors, each representing an instance. In training, the labelling is done at instance level. Nevertheless, the task at hand in MIL is to classify bags of instances.

The motivation behind MIL is scenarios where the task at hand is to classify each bag of instances rather than each

Farzad Khalvati, Junjie Zhang, and Masoom A. Haider are with the Department of Medical Imaging, University of Toronto and Sunnybrook Research Institute, Toronto, Ontario, Canada, M4N 3M5. farzad.khalvati@sri.utoronto.ca

Alexander Wong is with the Department of Systems Design Engineering, University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1.

individual instance. The standard assumption in MIL is that for a bag to be considered positive, it should contain at least one positive instance [13]. This is the case in cancer screening as well where a small cancerous region in prostate (e.g., 3×3 pixels) is enough to diagnose the patient as having cancer.

Vocabulary-based techniques are a subset of MIL in which the instances are classified first in an unsupervised way, which are then used to embed instance level information into the bag level [12], [14]. This enables the classification of unseen data at bag level, in a sense, with no direct correspondence to the classification of the instances of the bag.

In this paper, we extend upon vocabulary-based MIL to solve the problem of screening for prostate cancer. We introduce the concept of bag of bags in MIL where instances are grouped into small bags (sub-bags) which are then grouped to create high-level (parent) bags. This way, once the parent bags are classified, the sub-bags are classified with a higher accuracy. In our proposed method, the bags in training data are also sorted based on their labels, in order to increase the performance of clustering algorithm in terms of convergence.

The proposed bag of bags (BoB) nested MIL method in this paper is a top-down method in the context of prostate cancer detection. First, the detection is performed at patient level. Next, the imaging data of those patients with cancer are studied and analyzed further to determine which 2D slices in the 3D imaging data contain cancerous regions. We applied the proposed method to mpMRI data of 30 patients (15 with cancer and 15 healthy) and the proposed algorithm was able to detect cancer at both patient and slice level with high accuracy.

This paper is organized as follows: in Section II, the proposed BoB MIL method is presented. Section III presents the testing methodology which includes the description of imaging data, feature model, and the evaluation metrics used in this research. Sections IV and V present the experimental results and conclusions, respectively.

II. METHODOLOGY

The proposed algorithm in this paper is an extension on multi-instance learning (MIL) method where a classifier is learned via a set of bags as training data and each bag comprises of multiple feature vectors or instances. As discussed before, in vocabulary-based MIL techniques, the instances are first clustered in an unsupervised way, which are then used to embed instance level information into the bag level [12].

Let X be a set of all bags of instances $X = \{X_1, ..., X_n\}$ and each bag X_i has multiple instances $X_i = \{\overrightarrow{x_1}, ..., \overrightarrow{x_m}\}$. In order to compute the vocabulary, a clustering algorithm such as K-means (K_m) with hard assignment is applied to X to cluster the original instances into K clusters:

$$V = K_m(X) \tag{1}$$

which yields $V = \{\overrightarrow{c_1}, ..., \overrightarrow{c_K}\}$ where $\overrightarrow{c_j}$ is the centroid of j^{th} cluster of original instances. Once the vocabulary V was built, in the next step, a mapping function M is used to map each bag X_i to vocabulary V:

$$H = M(X, V) \tag{2}$$

where $H = \{\overrightarrow{H_1}, ..., \overrightarrow{H_n}\}$; and $\overrightarrow{H_i} = (h_1, .., h_K)$ is the histogram indicating the relationship between instances in X_i and vocabulary V. To calculate $\overrightarrow{H_i}$, first, it is determined which centroids $\overrightarrow{c_j}$ in V are the closest to each instance $\overrightarrow{x_z}$ in X_i, defined by function f [12]:

$$f_k(\overrightarrow{x_z}) = \begin{cases} 1, & \text{if } k = argmin_{k=1,..,K}||\overrightarrow{x_z} - \overrightarrow{c_k}|| \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

h_k is defined as:

$$h_k = \sum_{\overrightarrow{x_z} \in X_i} f_k(\overrightarrow{x_z}), \tag{4}$$

The calculated set of histograms H, acts as the training set where the labels are the corresponding bags labels. An unseen bag, X_{test}, is first mapped to the vocabulary V to yield $H_{test} = M(X_{test}, V)$, and H_{test} is then classified using a classifier G.

$$cl = G(H, H^{test}) \tag{5}$$

where cl is the classification results (label).

Our proposed BoB MIL algorithm in this paper has two features. First, the quality of the training set H in terms of separating different classes of bags is highly dependent on the performance of the clustering function K_m in Equation 1. In order to improve the convergence of K_m to meaningful centroids, we propose to sort the bags in X based on their labels before feeding to the clustering function. This yields highly converged centroids by K_m leading to more meaningful patterns in H. As shown in results section, this step is key for the proposed method for accurate classification of bags of instances.

$$V_s = K_m(S(X)) \tag{6}$$

where S is a sorting function based on bags labels. Thus, the training set H_s is generated by mapping X to vocabulary V_s:

$$\begin{aligned} H_s &= M(X, V_s) \\ cl &= G(H_s, H_s^{test}) \end{aligned} \tag{7}$$

where $H_s^{test} = M(X^{test}, V_s)$

The second and main feature of our proposed BoB MIL algorithm in this paper deals with the fact that data may have bags within bags. Let X be the original set of instances which belong to n bags of X_i. Each bag of instances X_i can contain sub-bags of instances $X_i = \{X_{i1}, ..., X_{ip}\}$. Thus, X can be rewritten in terms of these sub-bags (Y):

$$Y = \{X_{11}, ..., X_{1p_1}, X_{12}, ..., X_{2p_2}, ..., X_{n1}, ..., X_{np_n}\} \quad (8)$$

$\bar{Y}$ is a subset of Y that only contains those sub-bags whose parent bags have positive label.

Using mapping function M, $\bar{Y}$ can be mapped to vocabulary V_s:

$$HY_s = M(\bar{Y}, V_s) \quad (9)$$

HY_s is a new training set based on set of sub-bags $\bar{Y}$. In order to classify an unseen sub-bag X_{ij}^{test}, first, the parent bag X_i^{test} is mapped to $H_{i\_s}^{test}$ and then classified using training set H_s:

$$\begin{aligned} H_{i\_s}^{test} &= M(X_i^{test}, V_s) \\ cl_i &= G(H_s, H_{i\_s}^{test}) \end{aligned} \quad (10)$$

If the classification result cl_i is positive for the parent bag X_i^{test}, then sub-bag X_{ij}^{test} is classified using training set HY_s, otherwise the classification result for sub-bag X_{ij}^{test} is assumed to be negative:

$$cl_{ij} = \begin{cases} 0, & \text{if } cl_i = 0 \\ G(HY_s, HY_{ij\_s}^{test}) & \text{otherwise} \end{cases} \quad (11)$$

where $HY_{ij\_s}^{test} = M(X_{ij}^{test}, V_s)$.

In other words, to classify a sub-bag, first the parent bag is classified. If the classification result is negative, it is concluded that the sub-bag will also be classified as negative. Otherwise, the sub-bag is mapped and classified using the mapping of the instances that belong to the sub-bag. It should be noted that the training data for classifying a sub-bag only contains instances from the parent bags which have positive labels. Figure 1 illustrates the block diagram of the proposed BoB MIL algorithm.

In the prostate cancer mpMRI data context, the parent bags are the 3D imaging data of each patient and sub-bags represent each 2D image slice in the 3D imaging data of the patient, and instances are the imaging feature vectors extracted from local neighbourhoods of pixels in the images (e.g., 3×3).

III. Experimental Setup

In this section, the details of image acquisition protocols, the imaging data, feature model, and the performance measures used in this study are described.

A. Image Data

To evaluate the performance of the proposed algorithm for detection of prostate cancer, the imaging data of 30 patients were used. The images were acquired using a Philips Achieva 3.0T machine at Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada. All data was obtained under the local institutional research ethics board. For each patient, the following MRI modalities were obtained: T2w, DWI, and CDI [15], [16]. Table I shows the summary of the imaging parameters for the imaging modalities used for the experiments (e.g., displayed field of view (DFOV), resolution, echo time (TE), and repetition time (TR)).

TABLE I: Summary of imaging parameters

Modality	DFOV (cm^2)	Resolution (mm^3)	TE (ms)	TR (ms)
T2w	22×22	$0.49 \times 0.49 \times 3$	110	4,687
DWI	20×20	$1.56 \times 1.56 \times 3$	61	6,178
CDI	20×20	$1.56 \times 1.56 \times 3$	61	6,178

B. Feature Model

The instances in the experimental data are feature vectors extracted from local neighbourhoods of pixels in different images of mpMRI. For feature model, we used the one proposed in [5] where the following imaging modalities were used:

- T2-weighted MRI: the main MRI modality in detecting prostate cancer [17].
- Apparent diffusion coefficient (ADC) map: calculated from diffusion-weighted imaging (DWI) data and widely used for detecting and diagnosing prostate cancer [18].
- Computed high-b diffusion-weighted imaging (CHB-DWI): high-b DWI calculated computationally using low-b value DWI, rather than by actual DWI acquisition, which may not be possible due to MRI hardware limitation [19].
- Correlated diffusion imaging (CDI): a new diffusion MRI modality, which takes advantage of the joint correlation in signal attenuation across multiple gradient pulse strengths and timings to distinguish the healthy and cancerous tissues in prostate [15], [16].
- Individual b-value DWI images including b-values at 0, 100, 400, $1000 s/mm^2$.

The feature model used for our experiment is the one proposed in [5] where a comprehensive texture feature model was built; we used 4 well-known classes of texture features, which include first- and second-order statistical features, steerable Gabor filter features, Kirsch filter features [5]. The first-order statistical features include mean and standard deviation of grey-level intensity, skewness, and kurtosis. Second-order statistical features such as entropy and contrast are extracted from the gray-level co-occurrence matrix (GLCM). As a result, the mpMRI texture feature model consists of a total of 50 features for each imaging modality totaling 400 features for each instance. A feature selection algorithm [20] was used to reduce the feature space down to 50 features per instance.

C. Evaluation Metrics

The ground-truth data was obtained by radiology reports, confirmed by biopsy data. The performance evaluation was conducted using a leave-one-patient-out cross validation. The experiments were done at both slice (2D image) and patient (3D image) level, using conventional bottom-up feature model approach and the proposed top-down method (details in Section IV). The number of clusters in Equation 6 was set to 5. As classifier, Support Vector Machine (SVM) was used.

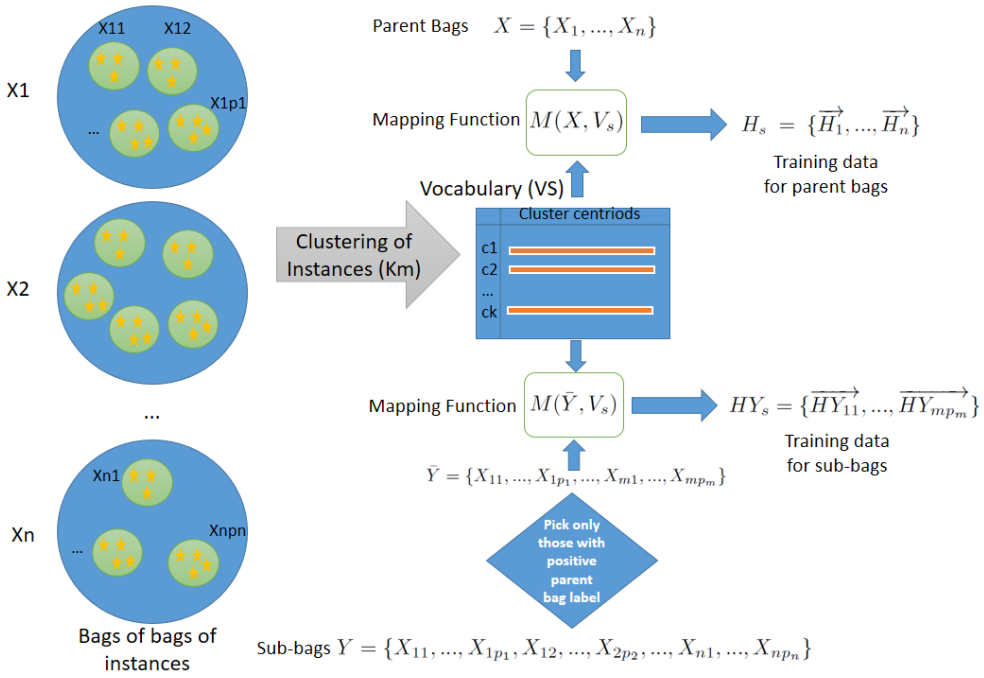

Fig. 1: Block diagram of bags of bags (BoB) nested MIL algorithm

For our experiments, mpMRI 3D data of 30 patients (15 healthy and 15 with cancer) were used. This amounted to 353 2D slices (308 healthy and 45 cancerous), and 2,132 instances (2,010 healthy and 122 cancerous).

IV. Experimental Results

To evaluate the performance of the proposed method in this paper, we applied it to mpMRI data of prostate for cancer detection. First, we used the conventional bottom-up feature model where instances are classified first and then aggregated to yield the classification results for 2D slices and 3D mpMRI images (i.e., patient level) [12]. Next, we applied vocabulary-based MIL algorithm (as discussed in Section II) to classify 2D images as well as 3D mpMRI data at patient level. Note that these two classifications were done separately where in each case, the bags of instances were 2D and 3D images, respectively. Finally, we applied our proposed BoB nested MIL method to classify 2D slices and 3D mpMRI data at patient level where 3D images and 2D images were considered as parent bags and sub-bags, respectively.

Table II shows the sensitivity, specificity, and accuracy of 3D (patient level) and 2D images using conventional bottom-up method.

TABLE II: Evaluation results for prostate cancer detection using conventional bottom-up method

Bags	Sensitivity	Specificity	Accuracy
3D mpMRI (Patient level)	0.69	0.39	0.54
2D Slices	0.86	0.54	0.58

Table III shows the sensitivity, specificity, and accuracy of 3D (patient level) and 2D images using vocabulary-based MIL method.

TABLE III: Evaluation results for prostate cancer detection using vocabulary-based MIL method

Bags	Sensitivity	Specificity	Accuracy
3D mpMRI (Patient level)	0.81	0.92	0.87
2D Slices	0.84	0.26	0.34

Table IV shows the sensitivity, specificity, and accuracy of 3D (patient level) and 2D images using our proposed BoB nested MIL method.

TABLE IV: Evaluation results for prostate cancer detection using proposed BoB nested method

Bags	Sensitivity	Specificity	Accuracy
3D mpMRI (Patient level)	0.97	0.95	0.96
2D Slices	0.89	0.77	0.79

As it can be seen, the vocabulary-based MIL method (Table III) improves the results for both 3D and 2D images compared to conventional bottom-up approach (Table II). Nevertheless, it still struggles with large amount of false positives (low specificity) at 2D slice level, as predicted. Furthermore, it misses several true positives at both patient and slice level (moderate sensitivity).

The proposed BoB nested MIL method yields the best results for both 3D images (patient level) and 2D slices for sensitivity, specificity, accuracy (Table IV). The improvement in results for parent bags for the proposed method is partially due to the fact that the bags are sorted based on their labels before being clustered (Equation 6). More importantly, the contributing factor to high accuracy of results for the proposed BoB MIL algorithm is the fact that the proposed

algorithm uses bags of bags of instances, where parent bags are classified first, and then the sub-bags are classified. Furthermore, the training data for sub-bags only contains the instances that belong to parent bags with positive labels. This leads to a cleaner and balanced training data for sub-bags (HY_s in Equation 11) and thus, highly accurate results are achieved.

Figures 2 and 3 show the sensitivity, specificity, and accuracy for bottom-up , vocabulary-based MIL (MIL), and proposed BoB nested MIL methods for 3D mpMRI data (patient level) and 2D mpMRI data (slice level), respectively.

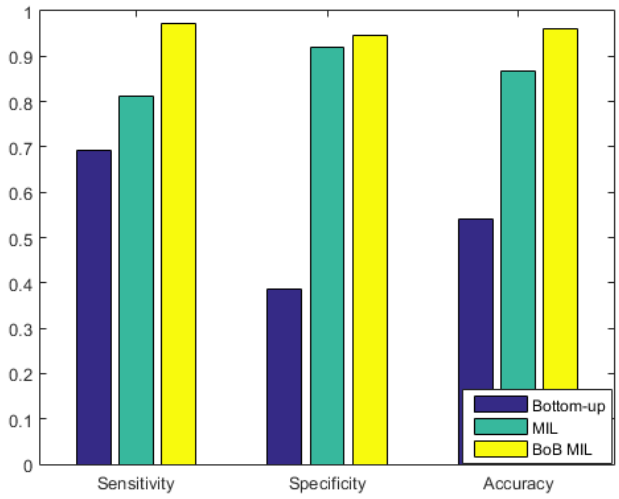

Fig. 2: Sensitivity, specificity, and accuracy for bottom-up, MIL, and proposed BoB nested MIL methods for 3D mpMRI data (patient level)

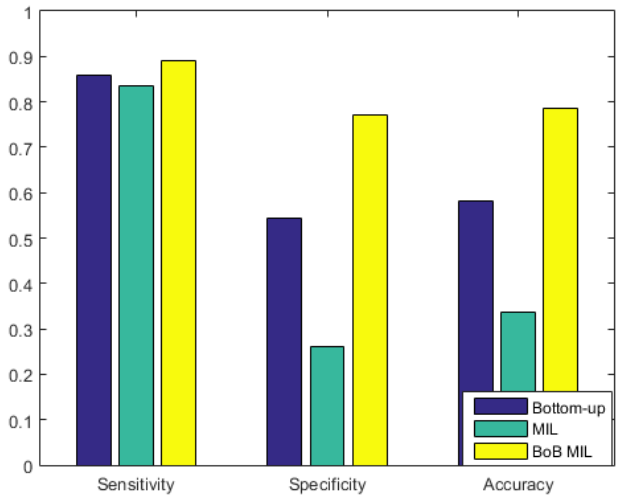

Fig. 3: Sensitivity, specificity, and accuracy for bottom-up, MIL, and proposed BoB nested MIL methods for 2D mpMRI data (slice level)

Figure 4 shows ADC maps from two patients mpMRI data where the one on the left has cancer and the one on the right has no cancer. As it can be seen, detecting cancerous region in prostate mpMRI images is complicated and not a trivial task.

V. CONCLUSIONS

In this paper, a novel bag of bags (BoB) nested multi-instance learning method was proposed to enable the classification of both bags of instances and sub-bags of instances for a given dataset. The motivation behind this approach was the fact that in cancer screening, imaging data is used with the foremost goal of detecting cancer at patient level first, before localizing tumours in the organ. Our proposed BoB nested MIL method allows the mpMRI data of patients to be classified at both 3D (patient) level and 2D slice level. This means that, first, patients are classified into cancerous and healthy in the first run and then, the cancerous regions are localized within 2D images of the data. Experimental results show that the proposed method is a promising approach for prostate cancer detection with high accuracy, which can be used for cancer screening of population with a significant impact on early diagnosis and treatment of prostate cancer.

REFERENCES

[1] Canadian Cancer Society, "Canadian Cancer Statistics Special topic : Predictions of the future burden of cancer in Canada," Tech. Rep., 2015.

[2] R. Siegel, D. Naishadham, and A. Jemal, "Cancer Statistics, 2013," *Ca Cancer J CLIN*, vol. 37, no. 2, pp. 408–14, 2013.

[3] S Loeb, A Vellekoop, HU Ahmed, J Catto, M Emberton, R. Nam, DJ Rosario, V Scattoni, and Y Lotan, "Systematic Review of Complications of prostate biopsy," *Expert review of anticancer therapy*, vol. 13, no. 7, pp. 829–837, 2013.

[4] N. Arumainayagam, H. U. Ahmed, C. M. Moore, A. Freeman, C. Allen, S. A. Sohaib, A. Kirkham, J. van der Meulen, and M. Emberton, "Multiparametric MR imaging for detection of clinically significant prostate cancer: a validation cohort study with transperineal template prostate mapping as the reference standard." *Radiology*, vol. 268, no. 3, pp. 761–9, 2013.

[5] F. Khalvati, A. Wong, and M. A. Haider, "Automated Prostate Cancer Detection via Comprehensive Multi-Parametric Magnetic Resonance Imaging Texture Feature Models," *BMC Medical Imaging*, pp. 15–27, 2015.

[6] G. Lemaitre, R. Marti, J. Freixenet, J. C. Vilanova, P. M. Walker, and F. Meriaudeau, "Computer-Aided Detection and diagnosis for prostate cancer based on mono and multi-parametric MRI: A review," *Computers in Biology and Medicine*, vol. 60, pp. 8–31, 2015.

[7] R. Leijenaar et al, "Radiomics: extracting more information from medical images using advanced feature analysis." *Eur J Cancer*, vol. 62, no. 4, pp. 441–6, 2015.

[8] H. W. L. Aerts et. al, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach." *Nat Commun*, vol. 45, no. 4, 2014.

[9] A. Cameron, F. Khalvati, M. Haider, and A. Wong, "MAPS: A Quantitative Radiomics Approach for Prostate Cancer Detection." *IEEE transactions on bio-medical engineering*, vol. 63, no. 6, pp. 1145–1156, 2016.

[10] A. G. Chung, F. Khalvati, M. J. Shafaiee, M. A. Haider, and A. Wong, "Prostate Cancer Detection via a Quantitative Radiomics-Driven Conditional Random Field Framework," *IEEE Access*, vol. 3, pp. 2531–2541, 2015.

[11] F. Khalvati, A. Modhafar, A. Cameron, A. Wong, and A. Haider, "A Multi-Parametric Diffusion Magnetic Resonance Imaging Texture Feature Model for Prostate Cancer Analysis," in *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), Computational Diffusion MRI*, 2014, pp. 79–88.

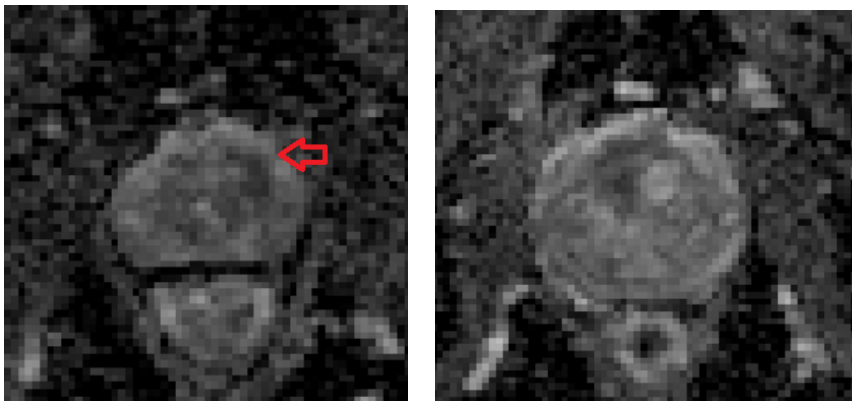

Fig. 4: Visual comparison of prostate tumour in ADC images from two patients: Left: arrow points to the cancerous region. Right: a healthy case with no cancer.

[12] J. Amores, "Multiple instance classification: Review, taxonomy and comparative study," *Artificial Intelligence*, vol. 201, no. August, pp. 81–105, 2013.

[13] J. Foulds and E. Frank, "A review of multi-instance learning assumptions," *The Knowledge Engineering Review*, pp. 1–25, 2010.

[14] Y. Chen, J. Bi, and J. Z. Wang, "Miles: Multiple-instance learning via embedded instance selection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 28(12), p. 19311947, 2006.

[15] A. Wong, F. Khalvati, and M. A. Haider, "Dual-Stage Correlated Diffusion Imaging," in *IEEE International Symposium on Biomedical Imaging*, 2015, pp. 75–78.

[16] F. Khalvati, J. Zhang, M. A. Haider, and A. Wong, "Enhanced Dual-Stage Correlated Diffusion Imaging," in *International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2016, pp. 5537–5540.

[17] M. A. Haider, T. H. Van Der Kwast, J. Tanguay, A. J. Evans, A. T. Hashmi, G. Lockwood, and J. Trachtenberg, "Combined T2-weighted and diffusion-weighted MRI for localization of prostate cancer," *American Journal of Roentgenology*, vol. 189, no. 2, pp. 323–328, 2007.

[18] D. M. Koh and A. Padhani, "Diffusion-weighted mri: a new functional clinical technique for tumour imaging," *Br J Radiol*, vol. 79(944), pp. 633–635, 2006.

[19] J. Glaister et al, "Quantitative investigative analysis of tumour separability in the prostate gland using ultra-high b-value computed diffusion imaging." vol. 2012, 2012, pp. 420–3.

[20] H. Peng et al, "Feature selection based on mutual information criteria of maxdependency, max-relevance, and min-redundancy," *IEEE Trans Pattern Anal Mach Intell*, vol. 27, pp. 1226–1238, 2005.