PROSTATE CANCER LOCALIZATION WITH MULTISPECTRAL MRI BASED ON RELEVANCE VECTOR MACHINES

S. Ozer[1], M. A. Haider[2,3], D. L. Langer[2,3], T. H. van der Kwast[4], A. J. Evans[4], M. N. Wernick[1], J. Trachtenberg[2], I. S. Yetik[1]

[1]Medical Imaging Research Center (MIRC), Electrical & Computer Engineering Dept., Illinois Institute of Technology, Chicago, IL, USA

[2]Joint Department of Medical Imaging, Princess Margaret Hospital, University Health Network and Mount Sinai Hospital, Toronto, Ontario, Canada

[3]Institute of Medical Science, University of Toronto, King's College, Toronto, Ontario, Canada

[4]Department of Pathology and Laboratory Medicine, Toronto General Hospital, Toronto, Ontario, Canada

ABSTRACT

Prostate cancer is one of the leading causes of cancer death for men. However, early detection before cancer spreads beyond the prostate can reduce the mortality. Therefore, *invivo* imaging techniques play an important role to localize the prostate cancer for treatment. Although Magnetic Resonance Imaging (MRI) has been proposed to localize prostate cancer, the studies on automated localization with multispectral MRI have been limited. In this study we propose combining the pharmacokinetic parameters derived from DCE MRI with T2 MRI and DWI. We also propose to use Relevance Vector Machines (RVM) for automatic prostate cancer localization, compare its performance to Support Vector Machines (SVM) and show that RVM can produce more accurate and more efficient segmentation results than SVM for automated prostate cancer localization with multispectral MRI.

Index Terms— Relevance Vector Machine, Support Vector Machine, Prostate Cancer, localization, Multispectral MRI.

1. INTRODUCTION

PROSTATE cancer is one of the most prevalent cancer types and one of the leading causes of cancer death for elder men in the United States [1]. The detection of prostate cancer at an earlier stage is crucial for increased rate of successful treatment. Traditional prostate cancer detection uses digital rectal examinations and serum prostate specific antigen (PSA) levels [2]. The final clinical diagnosis of the prostate cancer is based on histological tissue analysis. This is currently performed using exhaustive needle biopsy by taking multiple tissue samples from the prostate. Therefore, the surveillance of the indolent prostate cancer regions on patients is an active problem, where imaging techniques can help to determine the size and location of the most significant cancer for targeted biopsy and thus reducing the number of tissue samples taken by the needle. Accurate localization of prostate cancer also opens the way to guided radiotherapy and surgery.

In order to reduce the number of tissue samples taken by the needle from the patients, the use of TRUS to find the presence of the prostate cancer has been proposed in e.g. [3]. However the accuracy of TRUS is limited; therefore, MRI has been proposed as an emerging alternative to TRUS [4], [5]. It is shown that using MR images can provide images to find the smaller volumes of the prostate cancer with a higher accuracy than TRUS [4], [6] and is a promising method for prostate cancer localization.

Although single type MR images, such as T2, can be used to localize the prostate cancer, the accuracy of such methods is not sufficient for clinical applications. There have been studies where T1 weighted, T2 weighted, diffusion weighted and dynamic contrast enhanced (DCE) MRI were individually discussed [4], [6]. A recent study compared T2 and apparent diffusion coefficient (ADC) derived from diffusion-weighted MRI (DWI) [7]. It is reported that combining T2 weighted imaging with DWI provides better results than that of T2 alone [8].

Therefore, in this study we combine the information coming from all 3 different types of MRI, T2 weighted MRI, ADC values derived from DWI, and the pharmacokinetic parameter k_{ep} derived from DCE MRI in order to localize the prostate cancer .

Although several image segmentation algorithms have been used to localize cancer regions on several types of human organs [10], [11], the studies for automated human prostate cancer segmentation with MRI has been very limited [12]. The study in [12] uses texture features derived from T2weighted MRI and ADC values and uses the generic support vector machines and Fisher linear discriminate to compare the results. However, instead of deriving the feature vectors including the tumor location, the mean values of the tumor and normal regions can be used as training feature vectors, as [13] reports that the mean values are different for benign and malignant tissues on peripheral zone for ADC values. In this paper, we also use parametric images derived from DCE MRI, which was not included in [12] for automated prostate cancer localization, Finally, we propose to use a Bayesian framework with RVM in contrast to previously used SVM for prostate cancer localization. Because RVM can handle the noise in this Bayesian framework, it has emerged as a more useful tool compared to SVM in several areas.

The rest of this paper is organized as follows: Section 2 and 3 briefly describe SVM and RVM algorithms respectively. In Section 4, the multispectral MRI data that we use is described. While Section 5 briefly explains the preprocessing steps needed to be applied on the raw data, Section 6 presents the experimental results. Finally, Section 7 concludes and discusses the results obtained in this study.

2. SUPPORT VECTOR MACHINES

Support Vector Machines (SVM) [14] belongs to the classifier group called large margin classifiers. As being a binary classifier, SVM uses the following formula to classify given test data [14]:

$$f(\mathbf{x}) = y = \mathrm{sgn}\left(\sum_{i=1}^{k} \alpha_i y_i K(\mathbf{x}, \mathbf{x}_i) + b\right), \quad (1)$$

where α_i is nonzero Lagrange multiplier for each support vector $\mathbf{x}_i$, k the support vector number, $y_i \in \{-1,+1\}$ the class label, b the bias value for the hyperplane and $K(\mathbf{x},\mathbf{x}_i)$ the kernel function. Polynomial kernel can be given as an example kernel function:

$$K(\mathbf{x}, \mathbf{x}_i) = (\langle \mathbf{x}, \mathbf{x}_i \rangle + 1)^d, \quad (2)$$

where d is the polynomial degree, which we select as one resulting in a linear SVM in this study based on our preliminary studies with other types of kernels. The dual problem for SVM can be solved by finding the α_i values by maximizing the cost function $Q(\alpha)$:

$$Q(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(\mathbf{x_i}, \mathbf{x_j}), \quad (3)$$

subject to $\sum_{i=1}^{n} \alpha_i y_i = 0$ and $C \geq \alpha_i \geq 0$.

The $\mathbf{x}_i$ input (training) vectors with the nonzero Lagrange multiplier α_i are called support vectors (SV).

3. RELEVANCE VECTOR MACHINES

RVM assumes that the targets are noisy observations of the actual values [15]:

$$t = y + \varepsilon, \quad (4)$$

where t is the noisy observed output, ε the noise value and y the actual target output. RVM can learn the noise variance along the training procedure as well. Because of its several advantages, RVM is getting more attention in today's machine learning applications [16], [17]. While SVM assumes -1 or +1 values for class labels, RVM classification assumes that the label values for t to be 0 or 1 for each class.

In this sparse Bayesian probabilistic model, for a given test vector $\mathbf{x}$, the class label y can be found as:

$$f(\mathbf{x}, \mathbf{w}) = y = \sum_{i=1}^{n} w_i K(\mathbf{x}, \mathbf{x}_i) + w_0, \quad (5)$$

where the w_i is the i^{th} weight for the i^{th} training input, $\mathbf{w}=[w_1,w_2,\ldots,w_n]^T$, and n the training sample number.

In this Bayesian approach, each w is assumed to have a Gaussian distribution parameterized by a hyperparameter α, whose most probable value is calculated iteratively through the RVM algorithm with the expectation maximization algorithm [15]. The hyperparameter α for RVM differs from the one defined for SVM.

$$p(\mathbf{w} \mid \alpha_1, \ldots, \alpha_n) = (2\pi)^{-n/2} \prod_{i=1}^{n} \alpha_i^{1/2} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} \alpha_i w_i^2\right), \quad (6)$$

where α_i is the hyperparameter corresponding to w_i, and is related to the inverse variance. The training vectors that have finite α value are called Relevance Vectors (RV). Therefore, in general, the RVs do not lie on the margin as opposed to SVs for SVM, and generally number of RVs is much smaller than the number of SVs providing a more efficient classifier.

4. MULTISPECTRAL MRI DATA

The multispectral MRI dataset is obtained from twenty patients with biopsy-confirmed prostate cancers. Each patient underwent endorectal MRI on a 1.5T GE Excite HD platform prior to prostatectomy, and the MR images are acquired by three different MRI techniques: T2 weighted MRI (T2w), DWI, and DCE MRI. Axial-oblique fast spin-echo (FSE) T2-weighted, echo planar DWI, multi-echo FSE, and DCE-MRI were performed before surgery on a 1.5-T MRI system (Echospeed or Excite HD; GE Healthcare, Milwaukee, WI) using a 4-channel phased-array surface coil coupled to an endorectal coil (MEDRAD, Warrendale, PA). All studies were obtained with the image plane perpendicular to the rectal wall/prostate interface. The endorectal coil was filled with a 100% w/v barium sulfate suspension (Liquid Polibar; E-Z-EM, Westbury, NY) to reduce susceptibility artifact (29). Median time between imaging and surgery was 33 days (range, 1-139 days).

Acquisition parameters for T2-weighted MRI were: TR/TE = 6550/101.5 ms; 320x256 matrix; echo-train length (ETL) = 16; bandwidth (BW) = 20.83 kHz; number of excitations (NEX) = 3; field of view (FOV) = 14 cm; no phase wrap. DWI parameters were: TR/TE = 4000/77 ms; 128x256 matrix; ETL = 144; BW = 166.7 kHz; NEX = 10; FOV = 14 cm; b = 0, 600 s/mm2. Multi-echo FSE images were acquired at ten echo times (9.0-90.0 ms, in 9 ms increments) for T2 mapping (TR = 2000 ms; 256x128 matrix; ETL = 10; BW = 31.25 kHz; NEX = 1; FOV = 20 cm). Datasets for DCE analyses consisted of T1-mapping from multi-slice, multi-flip fast spoiled gradient echo images (FSPGR) (flip-angles: 2^o, 5^o, 10^o, 20^o; TR/TE = 8.5/4.2 ms; 256x128 matrix; ETL = 8; BW = 31.25 kHz; NEX = 1; FOV = 20 cm), followed by 50 phases of multi-slice FSPGR MRI (flip-angle = 20^o; TR/TE = 4.3/1.9 ms; 256x128 matrix; BW = 62.5 kHz; NEX = 0.5; FOV = 20 cm; temporal resolution = 10s). Two phases were acquired before injection of 20 ml contrast agent (gadopentate dimeglumine (Magnevist; Bayer Schering Pharma, Berlin, Germany)) at a rate of 4 ml/s, followed by a 20 ml saline flush using a power injector (MEDRAD Spectris MR injection system). Twenty-six slices were obtained per phase. Phase encoding was left-right for T2-weighted, multi-echo FSE and DWI, and anterior-posterior for multi-slice FSPGR scans. All MRI datasets were obtained at identical slice locations with a slice thickness of 3 mm and no intersection gap.

Cancer regions exhibit low signal values when T2 weighted MRI is used, however common benign processes such as prostatic inflammation, post biopsy hemorrhage and bibrosis also exhibit low T2 signal. Recent work combines T2 imaging with ADC that is derived from diffusion weighted MRI and shows that using both imaging techniques can improve the accuracy for prostate cancer detection [8]. T2w provides morphologic information, and has been used for prostate cancer staging for many years. The T2 maps are calculated from a series of echo time measurements and remove the variations in signal intensity as a function of proximity to the endorectal coil seen in T2w, as well as providing quantitative

values. ADC maps are calculated from the DWI acquisition, and can assist in differentiating between T2 shine through effects or artifacts and real ischemic lesions.

We also include DCE MRI data for automated prostate cancer localization as opposed to [12]. To obtain DCE MRI data, contrast agent was injected into the patient, and multiple MR images were acquired at the same spatial location over approximately ten minutes. One goal of DCE MRI is to characterize tissue regions with the expectation that certain processes such as blood flow, vascular characteristics, or tissue integrity are different in pathological tissue with respect to normal tissue.

After the prostatectomy, each patient's prostate was fixed in formalin for 24 hours and embedded in HistOmer gel prior to ex vivo MRI. T2-weight images were taken at 5^o intervals, and the angle corresponding to the plane of in vivo imaging determined. A high-resolution 3D dataset for the sample was then obtained. The gel-embedded sample was cut in 3mm sections using a rotary blade, along the angled plane determined during the ex vivo imaging session. All sections were then prepared as hematoxylin and eosin (H&E) stained whole mount histologic slides using standard pathology techniques. The whole mount sections were assessed by a pathologist, and regions of tumor were outlined as the ground truth used for training and evaluation.

5. PREPROCESSING OF THE MULTISPECTRAL MRI DATA

Prostate has several zones such as the peripheral and transition zones. However, most of the cancer occurs on the peripheral zone [18]; therefore, this study focuses on the cancers on the peripheral zone.

As the images come from different imaging sessions, a preprocessing step including image registration is necessary. The image registration for different modalities is performed manually as well as the peripheral zone extraction from the images. Then, we apply a 5 by 5 median filter on each image to reduce noise on images.

As the intensity values for each image varies considerably across different images, we also need to normalize the intensity values. Therefore, we transform the image values to their z scores to have zero mean and unit variance with the following formula:

$$Datanormalized = \frac{Image - \mu_{PZ}}{\sigma_{PZ}}, \quad (7)$$

where the *Image* is the image to be normalized, μ_{PZ} and σ_{PZ} are the mean value and the standard deviation of the data within the peripheral zone respectively.

6. EXPERIMENTAL RESULTS

After the preprocessing steps, we use SVM and RVM algorithms separately in order to localize the prostate cancer on each patient. We use one-against-all approach for training, i.e., for each test patient we used remaining 19 patients' data as training data. The polynomial kernel with the polynomial order 1 is being used for both SVM and RVM algorithms resulting in linear classifiers based on preliminary tests with other types of kernels.

We calculate specificity, sensitivity and Dice coefficient (DSC) as measures of quantitative performance with the following equations, respectively:

$$Spec = \frac{DTN}{TN}, \quad (8)$$

where *DTN* is the detected true negative number and *TN* is the true negative number;

$$Sens = \frac{DTP}{TP}, \quad (9)$$

where *DTP* is the detected true positive number and *TP* is the true positive number and;

$$DSC = \frac{2DTP}{DTP + TP}. \quad (10)$$

After finding the specificity, sensitivity and DSC values for each patient, we calculate the average of all 20 patients as shown on Table 1.

	RVM	SVM	RVM	SVM	RVM	SVM
Methods	Spec/Sens	Spec/Sens	RV #	SV #	DSC	DSC
T2	0.68/0.77	0.55/0.89	2	21.5	0.46	0.44
ADC	0.74/0.70	0.67/0.75	2	23.3	0.45	0.44
T2,ADC	0.71/0.78	0.65/0.83	1.05	20.2	0.48	0.46
T2,ADC,k_{ep}	0.78/0.74	0.74/0.79	1.25	18.25	0.51	0.52

Table1: Comparison of RVM and SVM by using different sets of MR images for the average of 20 patients.

As Table 1 shows, the average DSC value for RVM is slightly higher than average SVM value. Moreover RVM finds considerably lower number of RVs, when compared to the average number of SVs. RVM results have higher specificity values in the price of decreased sensitivity. Higher specificity values provided by RVM may be more desirable as it can help to decrease the number of taken tissue samples for clinical diagnosis.

Figures 1.1 through 1.8 compares segmentation results obtained by SVM and RVM when different MR image types are used for one of the patients. Figure 1.9 shows the ground truth for the patient. Both SVM and RVM can locate the cancer region with a certain false positive region. However, RVM provides consistently smaller regions of false positive for each image combination, as is also validated on Table 1. As shown on the Figure (1.7), visually most accurate segmentation result is obtained when ADC, T2 and k_{ep} are used altogether with the RVM algorithm. This result is also in line with human-reader studies where multispectral MRI increases the localization accuracy.

7. CONCLUSION AND DISCUSSION

In this study we show that incorporating the pharmacokinetic parameter k_{ep} derived from DCE MRI, in addition to T2 weighted MRI and DWI, increases the accuracy of automatic segmentation methods as can be seen on Table 1.

Our results also show that we can increase the accuracy of localization and segmentation results for prostate cancer by using Relevance Vector Machines. Table 1 shows that RVM yields smaller regions of false positive (higher specificity values) compared to SVM. This is a promising result as higher specificity values can help decrease the number of tissue samples needed during the clinical diagnosis.

Moreover, RVM finds significantly lower number of vectors as relevance vectors when compared to the number of support vectors found by SVM. Lower number of relevance vectors helps to

decrease the computation time. This is a result of the Bayesian framework that RVM uses.

We can also conclude that the use of Multispectral MR images increases the accuracy of localization as Table 1 shows with DSC values. By definition, as the DSC value increases, the detected true positive numbers approaches to the number of true positives.

Future work includes using a richer set of MR images for prostate cancer localization, such as including different pharmacokinetic parameters derived from DCE MRI and working on alternative methods of learning algorithms.

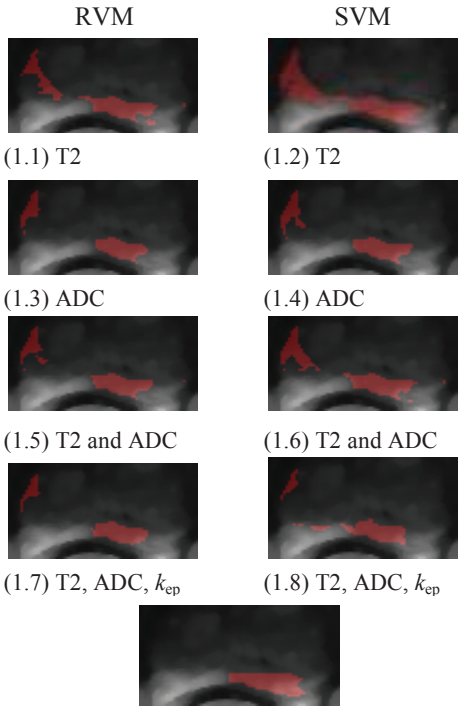

Figure 1: RVM and SVM segmentation results by using different modalities for the same patient.

8. ACKNOWLEDGEMENT

This work is partially supported by a grant from Prostate Research Foundation of Canada.

9. REFERENCES

[1] American Cancer Society. : Cancer Facts and Figures 2007. Atlanta, GA: American Cancer Society, 2007.

[2] W. Catalona et al., "Measurement of prostate-specific antigen in serum as a screening test for prostate cancer," New Engl. J. Med., vol. 324, no. 17, pp. 1156–1161, 1991.

[3] Lee F, Gray JM, McLeary RD, Meadows TR, Kumasaka GH, Borlaza GS, et al. "Transrectal ultrasound in the diagnosis of prostate cancer: location, echogenicity, histopathology, and staging. Prostate",. pp. 117-129, 1985.

[4] H. Hricak, P. L. Choyke, S. C. Eberhardt, S. A. Leibel, P. T. Scardino, "Imaging Prostate Cancer: A Multidisciplinary Perspective", *Radiology* Vol. 243 pp:28-53, 2007.

[5] MD Rifkin, EA Zerhouni, CA Gatsonis, LE Quint, DM Paushter, JI Epstein, U Hamper, PC Walsh, BJ McNeil, Comparison of magnetic resonance imaging and ultrasonography in staging early prostate cancer. Results of a multi-institutional cooperative trial,The New England Journal of Medicine, Vol. 323, pp. 621-626, 1990.

[6] SE Seltzer, DJ Getty, CM Tempany, RM Pickett, MD Schnall, BJ McNeil, JA Swets, "Staging prostate cancer with MR imaging: a combined radiologist- computer system", *Radiology*, Vol 202, pp: 219-226,1997.

[7] P. Gibbs, D. J. Tozer, G. P. Liney, L. W. Turnbull, "Comparison of quantitative T2 mapping and diffusion-weighted imaging in the normal and pathologic prostate",Magnetic Resonance in Medicine,Vol. 46, pp. 1054 - 1058, 2001.

[8] M. Haider, T.H. van der Kwast, J. Tanguay, et al., "Combined T2-weighted and diffusion weighted MRI for Localization of Prostate Cancer", American Journal of Roentgenology, vol 189, pp. 323-328, 2007.

[9] G. Brix, W. Semmler, R. Port, L. Schad, G. Layer, W. Lorenz, "Pharmacokinetic parameters in CNS Gd- DTPA enhanced MR Imaging", J Comput Assist Tomogr, vol. 15, pp. 621-628, 1991.

[10] L. O. Hall, A. M. Bensaid, L. P. Clarke, and etc., "A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain", *IEEE Trans. on Neural Networks*, vol. 3, pp. 672-682, 1992.

[11] S. Ruan, S. Lebonvallet, A. Merabet, J. M. Constans, "Tumor Segmentation from a Multispectral MRI Images by Using Support Vector Machine Classication", ISBI 2007, pp. 1236-1239, 2007.

[12] I Chan, W Wells, RV Mulkern, S Haker, J Zhang, KH Zou, SE Maier, CM Tempany,"Detection of prostate cancer by integration of line-scan diffusion, T2-mapping and T2-weighted magnetic resonance imaging; a multichannel statistical classifier",Med Phys, Vol. 30, No. 9, pp. 2390-2398, 2003.

[13] K. Hosseinzadeh, S. D. Schwarz, "Endorectal diffusion-weighted imaging in prostate cancer to differentiate malignant and benign peripheral zone tissue", Journal of Magnetic Resonance Imaging, Vol. 20, pp: 654 - 661, 2004

[14] Vapnik, V.: Statistical Learning Theory. Wiley-Interscience (1998)

[15] Tipping M. E, "Sparse Bayesian learning and the relevance vector machine", Journal of Machine Learning Research, pp. 211–244, 2001.

[16] Liyang W; Yongyi Y; Nishikawa, R.M.; Wernick, M.N.; Edwards, A.;"Relevance vector machine for automatic detection of clustered microcalcifications",IEEE Transactions on Medical Imaging,Vol. 24, Issue 10, pp. 1278 - 1285, Oct. 2005.

[17] Williams, O.; Blake, A.; Cipolla, R.;"Sparse Bayesian learning for efficient visual tracking", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 27, pp. 1292 - 1304, 2005.

[18] J. E. McNeal, E. A. Redwine, F. S. Freihas, T. A. Stamey, "Zonal Distribution of Prostatic Adenocarcinoma, Correlation with Histologic Pattern and Direction of Spread", American Journal of Surgical Pathology vol. 12, pp. 898-906, 1988.