

PROCEEDINGS OF SPIE

SPIDigitalLibrary.org/conference-proceedings-of-spie

Prostate lesion detection and localization based on locality alignment discriminant analysis

Mingquan Lin
Weifu Chen
Mingbo Zhao
Eli Gibson
Matthew Bastian-Jordan
Derek W. Cool
Zahra Kassam
Tommy W. S. Chow
Aaron Ward
Bernard Chiu

Prostate lesion detection and localization based on locality alignment discriminant analysis

Mingquan Lin^a, Weifu Chen^{a,h}, Mingbo Zhao^g, Eli Gibson^{b,e,f}, Matthew Bastian-Jordan^c, Derek W. Cool^c, Zahra Kassam^{c,d}, Tommy WS Chow^a, Aaron Ward^d, and Bernard Chiu^a

^aDepartment of Electronic Engineering, City University of Hong Kong, China

^bBiomedical Engineering, University of Western Ontario, London, Ontario, Canada

^cDepartment of Medical Imaging, University of Western Ontario, London, Ontario, Canada

^dLawson Health Research Institute, London, Ontario, Canada

^eCentre for Medical Image Computing, University College London, London, UK

^fDepartment of Radiology, Radboud University Medical Centre, Nijmegen, Netherlands

^gSchool of Information Science and Technology, Donghua University, Shanghai, China

^hSchool of Mathematics, Sun Yat-sen University, Guangzhou, China

ABSTRACT

Prostatic adenocarcinoma is one of the most commonly occurring cancers among men in the world, and it also the most curable cancer when it is detected early. Multiparametric MRI (mpMRI) combines anatomic and functional prostate imaging techniques, which have been shown to produce high sensitivity and specificity in cancer localization, which is important in planning biopsies and focal therapies. However, in previous investigations, lesion localization was achieved mainly by manual segmentation, which is time-consuming and prone to observer variability. Here, we developed an algorithm based on locality alignment discriminant analysis (LADA) technique, which can be considered as a version of linear discriminant analysis (LDA) localized to patches in the feature space. Sensitivity, specificity and accuracy generated by the proposed algorithm in five prostates by LADA were 52.2%, 89.1% and 85.1% respectively, compared to 31.3%, 85.3% and 80.9% generated by LDA. The delineation accuracy attainable by this tool has a potential in increasing the cancer detection rate in biopsies and in minimizing collateral damage of surrounding tissues in focal therapies.

Keywords: Prostate cancer, Multiparametric MRI (mpMRI), Lesion localization, Locality alignment discriminant analysis (LADA)

1. INTRODUCTION

Prostate cancer is a leading cause of male cancer death worldwide,¹ but is the most curable cancer when it is detected early.² Clinically, transrectal ultrasound-guided (TRUS-guided) biopsy is the only definitive method to ascertain the presence of prostatic cancer. Due to the inability of TRUS in targeting lesions, 35% of detectable lesions were missed in the first biopsy, necessitating repeated biopsies, leading to increased emotion stress for patients. Multiparametric MRI (mpMRI) has allowed for localization of prostate cancer³ and has the potential to support personalized management in targeted biopsy procedures and focal cancer therapies. Currently, most investigations involve manual localization of prostate detection and localization,^{4,5} which is time-consuming and prone to observer variability. Therefore, parallel to the development of MRI imaging technology, an equally important requirement is to develop tools to automate precise and accurate prostate cancer localization. Here, we focuses on the development of an automated method to localize cancer with voxel precision. The delineation accuracy attainable by this tool will play a major role in increasing the cancer detection rate in biopsies and in minimizing collateral damage of surrounding tissues in focal therapies.

Correspondence should be addressed to Bernard Chiu, E-mail: bychiu@cityu.edu.hk

2. METHODS

2.1 Image acquisition and preprocessing

T2, diffusion weighted (DW) and dynamic contrast enhanced (DCE) images are acquired as previously described⁶ for six patients with prostate cancer histologically confirmed on previous biopsies. The images for each patient were registered by an expert observer and resampled to standardize the in-plane resolution and slice thickness. The in-plane voxel size is 0.5 mm x 0.5 mm and the slice thickness is 3 mm after resampling, with the peripheral zone of each gland represented by approximately 10,000 voxels. Lesions were manually delineated and scored by four radiologists based on the Prostate Imaging-Reporting and Data System (PI-RADS) detection guidelines.⁷ For each lesion, each radiologist reported a 5-point likelihood score for each of the T2W, DW and DCE MR images (i.e., 3 images x 4 observers = 12 scores per lesion), where 1 to 5 indicate that a clinically significant cancer is highly unlikely to highly likely. In this study, regions with PI-RADS ≥ 3 marked by any radiologist on any of the three images were considered cancerous, resulting in a binary classification available on a voxel-by-voxel basis. Our proposed voxel-based cancer localization technique were trained and validated using this voxel-based expert classification.

2.2 Voxel-based feature extraction

For each voxel of the peripheral zone, the following list of features were extracted: (i) Physical coordinates. The reason for including the coordinates as features was that there is a high spatial correlation in the status of voxels within a neighbourhood (i.e., if a voxel is within a cancerous lesion, there is a high probability that its neighbours is inside a lesion). (ii) Grayscale values in the three imaging sequences at the voxel and its 3×3 in-plane neighbourhood.^{8,9} (iv) Slope of DCE MR sequence, defined as $\kappa = \frac{I(t+1)-I(t)}{\Delta t}$, where $I(t)$ and $I(t+1)$ are two adjacent time points for DCE images acquisition. The voxel-based features described above were collectively represented by a high-dimensional data point.

2.3 Linear Discriminant Analysis (LDA)

Linear discriminant analysis (LDA) separates two or more classes of high-dimensional data points by finding a projection matrix to maximize a cost function quantifying the inter-class difference in relation to the intra-class variability. A commonly used function is the trace ratio cost function,¹⁰ and the development of the current algorithm is based on this cost function. Mathematically, the l D -dimensional data points in the training set are represented by a $D \times l$ matrix $\mathbf{X} = [x_1, x_2, \dots, x_l]$. Each data point belongs to a class c_i with $i \in \{1, 2, \dots, c\}$. Our application is a two-class problem (i.e., $c = 2$). We define l_i to be the total number of data points in c_i , μ_i to be the mean of all data points in c_i , and μ to be the mean of all data points. The within- and between-class scatter matrices, \mathbf{S}_w and \mathbf{S}_b respectively, are defined as follows:

$$\mathbf{S}_w = \sum_{i=1}^c \sum_{x \in c_i} (x - \mu)(x - \mu)^T, \quad (1)$$

$$\mathbf{S}_b = \sum_{i=1}^c l_i (\mu_i - \mu)(\mu_i - \mu)^T, \quad (2)$$

For later development of the Locality Alignment Discriminant Analysis (LADA) algorithm, \mathbf{S}_w and \mathbf{S}_b are expressed in terms of \mathbf{X} :

$$\mathbf{S}_w = \mathbf{X} \mathbf{L}_w \mathbf{X}^T, \quad (3)$$

$$\mathbf{S}_b = \mathbf{X} \mathbf{L}_b \mathbf{X}^T, \quad (4)$$

where L_w and L_b are $l \times l$ matrices related to class memberships of data points as described in Tang et al.¹¹

LDA finds a $D \times d$ matrix \mathbf{W} that projects each D -dimensional data point to a corresponding d -dimensional data point with $D \gg d$ that maximizes the trace-ratio cost function. Mathematically, $\mathbf{Y} = \mathbf{W}^T \mathbf{X}$, with

the columns of the $d \times l$ matrix \mathbf{Y} representing the projected d -dimensional data points. The optimal \mathbf{W}^* is represented by the following equation:

$$\mathbf{W}^* = \arg \max_{\mathbf{W}^T \mathbf{W} = \mathbf{I}} \frac{\text{Tr}(\mathbf{W}^T \mathbf{S}_b \mathbf{W})}{\text{Tr}(\mathbf{W}^T \mathbf{S}_w \mathbf{W})} = \arg \max_{\mathbf{W}^T \mathbf{W} = \mathbf{I}} \frac{\text{Tr}(\mathbf{W}^T \mathbf{X} \mathbf{L}_b \mathbf{X}^T \mathbf{W})}{\text{Tr}(\mathbf{W}^T \mathbf{X} \mathbf{L}_w \mathbf{X}^T \mathbf{W})} = \arg \max_{\mathbf{Y}} \frac{\text{Tr}(\mathbf{Y} \mathbf{L}_b \mathbf{Y}^T)}{\text{Tr}(\mathbf{Y} \mathbf{L}_w \mathbf{Y}^T)}, \quad (5)$$

in which we used Eqs. 3, 4 and $\mathbf{Y} = \mathbf{W}^T \mathbf{X}$. The last term in Eq. 5 was associated with a constraint of $\mathbf{W}^T \mathbf{W} = \mathbf{I}$ to ensure uniqueness of the solution. This equation serves as a building block of the LADA algorithm as will be shown in the next section.

2.4 Locality Alignment Discriminant analysis (LADA)

A disadvantage of LDA is that \mathbf{S}_w and \mathbf{S}_b were built globally based on the entire training set. The model would better fit the training data if the whole feature space is divided into *local patches* and the optimization carried out on a patch-by-patch basis. LADA has two major components, which are described in detail below. A generalization of this scheme to a number of dimensionality reduction algorithms were presented by Zhang et al.¹²

2.4.1 Patch-by-Patch Optimization

This step established a patch for each of the l data points and its $K - 1$ nearest neighbours, thereby forming l patches with K data points. For each data point x_i , we denote its $K - 1$ nearest neighbours by $x_{i_1}, x_{i_2}, \dots, x_{i_{K-1}}$. The K data points within the patch associated with x_i are represented as columns in the $D \times K$ matrix $\mathbf{X}_i = [x_i, x_{i_1}, x_{i_2}, \dots, x_{i_{K-1}}]$. Using the results established in Eq. 5, the cost function to be minimized in this patch-by-patch optimization is:

$$\frac{\text{Tr}(\mathbf{Y}_i \mathbf{L}_{b,i} \mathbf{Y}_i^T)}{\text{Tr}(\mathbf{Y}_i \mathbf{L}_{w,i} \mathbf{Y}_i^T)}, \quad (6)$$

where the definitions of $\mathbf{Y}_i, \mathbf{L}_{b,i}, \mathbf{L}_{w,i}$ are described in detail in Sec. 2.3, but now the application of the LDA algorithm is limited to the patch represented by \mathbf{X}_i .

2.4.2 Global Alignment

This step optimizes the sum of the cost function Eq. 6 associated with the l patches available. As described in Zhang et al.,¹² the solution can be expressed in the similar format as Eq. 5, except that now the matrices \mathbf{L}_b and \mathbf{L}_w depends on the groupings of the l patches. Expressing the sum of the l cost functions in the format of Eq. 5 would not only provide a better comparison between LDA and LADA, but would allow the same procedure used for maximizing the trace ratio in Eq. 5 to be directly applied to optimize the cost function associated with LADA. A major challenge is to find a way to express \mathbf{Y}_i associated with each patch in terms of the low-dimensional representation of the entire set of data point, denoted by \mathbf{Y} as described in Sec. 2.3. Zhang et al.¹² defined a $l \times K$ selection matrix for each patch i , denoted by \mathbf{S}_i , in order to pick out the K data points in \mathbf{Y}_i from \mathbf{Y} :

$$\mathbf{Y}_i = \mathbf{Y} \mathbf{S}_i, \quad (7)$$

with the pq entry of \mathbf{S}_i defined by:

$$(\mathbf{S}_i)_{pq} = \begin{cases} 1 & \text{if } p = F_i\{q\} \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

where $F_i = \{i, i_1, i_2, \dots, i_{K-1}\}$ are the indices of the data point within the patch \mathbf{X}_i .

With these definitions established, Eq. 6 can be written in terms of \mathbf{Y} :

$$\frac{\text{Tr}(\mathbf{Y} \mathbf{S}_i \mathbf{L}_{b,i} \mathbf{S}_i^T \mathbf{Y}^T)}{\text{Tr}(\mathbf{Y} \mathbf{S}_i \mathbf{L}_{w,i} \mathbf{S}_i^T \mathbf{Y}^T)}, \quad (9)$$

Summation of Eq. 9 for all patches results in the following cost function:

$$\frac{\sum_{i=1}^l \text{Tr}(\mathbf{Y}\mathbf{S}_i\mathbf{L}_{\mathbf{b},i}\mathbf{S}_i^T\mathbf{Y}^T)}{\sum_{i=1}^l \text{Tr}(\mathbf{Y}\mathbf{S}_i\mathbf{L}_{\mathbf{w},i}\mathbf{S}_i^T\mathbf{Y}^T)} = \frac{\text{Tr}\left(\mathbf{Y}\left(\sum_{i=1}^l \mathbf{S}_i\mathbf{L}_{\mathbf{b},i}\mathbf{S}_i^T\right)\mathbf{Y}^T\right)}{\text{Tr}\left(\mathbf{Y}\left(\sum_{i=1}^l \mathbf{S}_i\mathbf{L}_{\mathbf{w},i}\mathbf{S}_i^T\right)\mathbf{Y}^T\right)} = \frac{\text{Tr}\left(\mathbf{Y}\widetilde{\mathbf{L}}_{\mathbf{b}}\mathbf{Y}^T\right)}{\text{Tr}\left(\mathbf{Y}\widetilde{\mathbf{L}}_{\mathbf{w}}\mathbf{Y}^T\right)} \quad (10)$$

where $\widetilde{\mathbf{L}}_{\mathbf{b}} = \sum_{i=1}^l \mathbf{S}_i\mathbf{L}_{\mathbf{b},i}\mathbf{S}_i^T$ and $\widetilde{\mathbf{L}}_{\mathbf{w}} = \sum_{i=1}^l \mathbf{S}_i\mathbf{L}_{\mathbf{w},i}\mathbf{S}_i^T$. The optimal \mathbf{W}^* can be written in the same form as for LDA in Eq. 5:

$$\mathbf{W}^* = \arg \max_{\mathbf{W}^T\mathbf{W}=\mathbf{I}} \frac{\text{Tr}(\mathbf{W}^T\widetilde{\mathbf{S}}_{\mathbf{b}}\mathbf{W})}{\text{Tr}(\mathbf{W}^T\widetilde{\mathbf{S}}_{\mathbf{w}}\mathbf{W})}, \quad (11)$$

where $\widetilde{\mathbf{S}}_{\mathbf{b}} = \mathbf{X}\widetilde{\mathbf{L}}_{\mathbf{b}}\mathbf{X}^T$ and $\widetilde{\mathbf{S}}_{\mathbf{w}} = \mathbf{X}\widetilde{\mathbf{L}}_{\mathbf{w}}\mathbf{X}^T$. Compared to LDA, the between- and within-class matrices were computed by maximizing the trace-ratio cost function in local patches instead of globally and therefore were expected to result in a smaller fitting error.

2.5 Classification based on discriminant analyses

Classification was performed for each of the five prostates individually. Two axial images from each prostate were used for training, and the remaining axial images were reserved for validation. LDA and LADA were used to find projection matrices to optimally project data points onto a 2D feature space based on the training data. In the classification stage, data points were projected onto the 2D feature space using the projection matrices obtained in training. On the 2D feature space, a data point was classified as either cancerous or non-cancerous based on the labels of its nearest neighbour in the training set.

Since the result produced by the LADA algorithm depends on size of the patch K defined in Sec. 2.4, K was tuned using image slices of the sixth prostate not involved in the classification process described in the last paragraph. Training was performed as described above with K varying from 50 to 300 with a 50 increment. The K associated with the minimum sum of false native and false positive rates (FNR and FPR respectively) in this parameter tuning step was applied for the images associated with the other five prostates. Since the parameter K was not involved in LDA, parameter tuning was not required for LDA.

3. RESULTS

LADA were optimized when $K = 250$. Table 1 lists the sensitivity, specificity and accuracy achieved by LDA and LADA. Figs. 1-3 show cases where LDA failed to localize the lesion [Figs. 1(c), 2(c) and 3(c)]. The lesions were much better localized by the introduction of LADA as shown in Figs. 1(d), 2(d) and 3(d).

Table 1. Sensitivity, specificity and accuracy for LDA and LADA.

Method	Sensitivity	Specificity	Accuracy
LDA	0.31±0.07	0.85±0.04	0.81±0.07
LADA	0.52±0.10	0.89±0.04	0.85±0.04

4. DISCUSSION

We demonstrated that the flexibility introduced by LADA in the computation of the projection matrix has led to a large improvement in prostate lesion localization as compared to LDA. As demonstrated in the three examples shown in Figs. 1-3, LDA was unable to localize cancerous lesions, with identified voxels scattered throughout the entire peripheral zone as shown in Figs. 1(c), 2(c) and 3(c). LADA was more able to identify localize blobs as lesions as shown in Figs. 1(d), 2(d) and 3(d). Although we have shown that LADA was more accurate in identifying lesions than LDA, there is an obvious requirement to improve the sensitivity of the algorithm before it can be applied in clinical settings. As a margin of up to 9 mm was acceptable and even recommended in focal therapies,¹³ we will investigate how adding margins would improve the lesion detection sensitivity. In

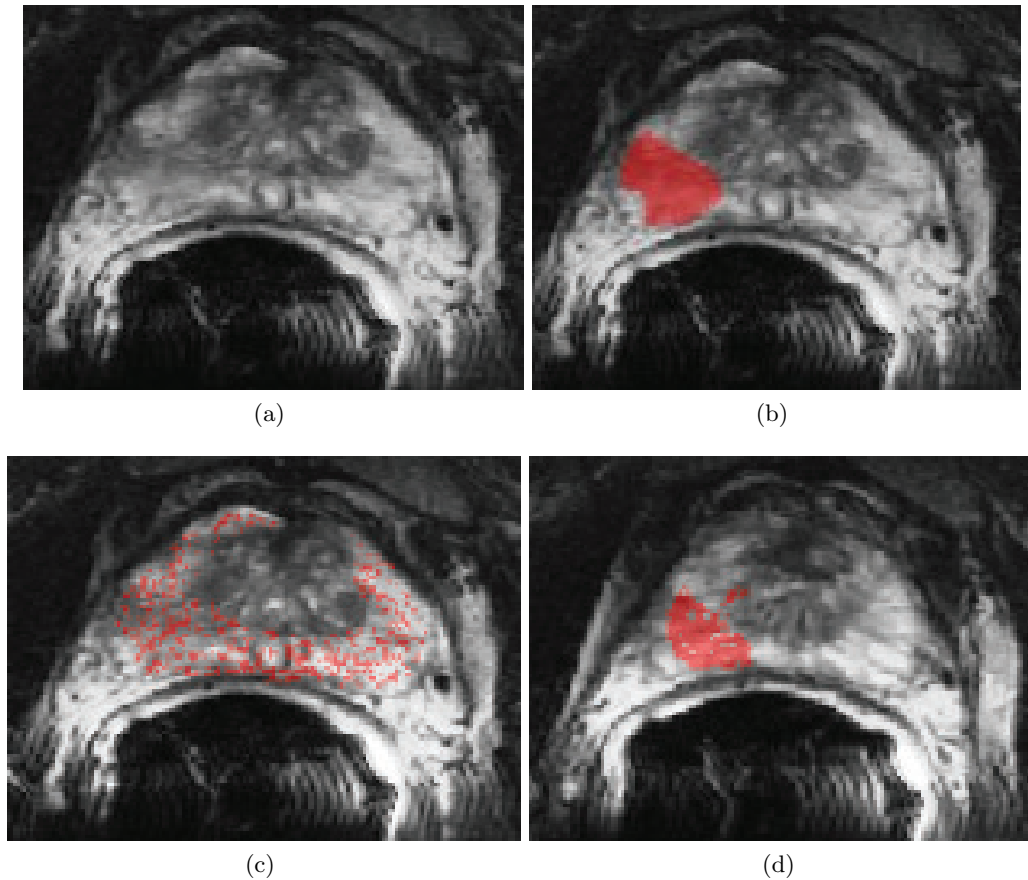


Figure 1. Lesion localization results in Example Axial Slice 1. (a) shows the T2-weighted image. (b), (c) and (d) show the cancerous regions identified by manual segmentation, LDA and LADA respectively.

In addition, we will incorporate prior knowledge of the lesions to further optimize the results provided by LADA. For example, a lesion should be at least 0.2 cm^3 in order to be considered clinically significant according to the Epstein's criteria.¹⁴ We will remove identified regions that are too small to reduce false positive detection. Secondly, holes inside a cancerous lesion is highly unlikely. Filling holes in the detected regions should be able to increase the accuracy of lesion identification. Investigations of these post-processing techniques will be the focus of a future investigation.

5. CONCLUSION

In this paper, LADA was introduced to localize the prostate lesion. Based on the preliminary results generated using mpMRI of five prostates, LADA detected lesions with greater sensitivity, specificity and accuracy as compared to LDA due to the more flexible patch-by-patch formulation. Further improvement in accuracy can be achieved by removing detected regions which are too small to be considered clinically significant (e.g., 0.2 cm^3 by Epstein's criteria¹⁴). A limitation of this study is the small sample size used to validate the algorithm, and as such, there is a requirement for a future evaluation study involving more subjects. Improvement on the sensitivity and the overall accuracy in lesion detection by LADA is required, and will be achieved through adding margin, remove small regions and filling holes inside lesions in a future investigation.

ACKNOWLEDGEMENT

Dr. Chiu is grateful for funding support from the Basic Research Free Exploration Program of the Science Technology and Innovation Committee of Shenzhen Municipality, China (Project No. JCYJ20160428155118212),

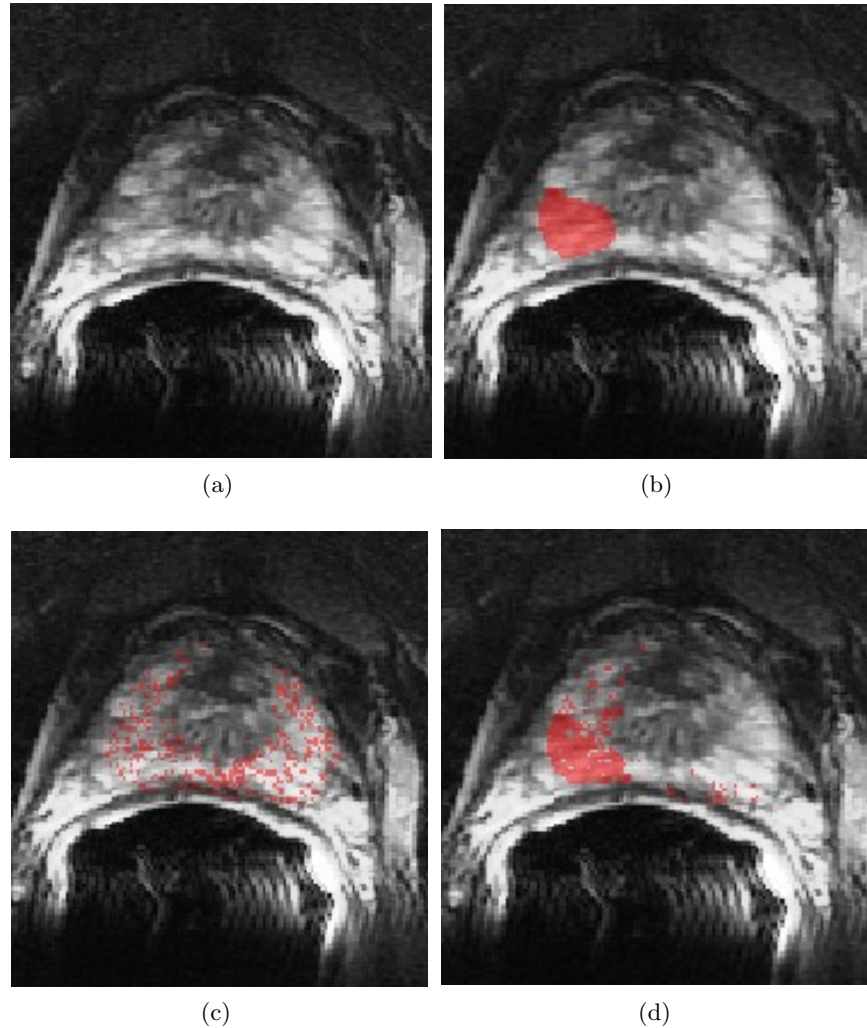


Figure 2. Lesion localization results in Example Axial Slice 2. (a) shows the T2-weighted image. (b), (c) and (d) show the cancerous regions identified by manual segmentation, LDA and LADA respectively respectively.

the Research Grant Council of the HKSAR, China (Project No. CityU 139713) and the City University of Hong Kong Strategic Research Grants (Nos. 7004229, 7004425 and 7004617).

REFERENCES

- [1] Haas, G. P., Delongchamps, N., Brawley, O. W., Wang, C. Y., and de la Roza, G., "The worldwide epidemiology of prostate cancer: perspectives from autopsy studies," *The Canadian journal of urology* **15**(1), 3866 (2008).
- [2] Turkbey, B., Albert, P. S., Kurdziel, K., and Choyke, P. L., "Imaging localized prostate cancer: current approaches and new developments," *AJR. American journal of roentgenology* **192**(6), 1471 (2009).
- [3] Hambrock, T., Somford, D. M., Huisman, H. J., van Oort, I. M., Witjes, J. A., Hulsbergen-van de Kaa, C. A., Scheenen, T., and Barentsz, J. O., "Relationship between apparent diffusion coefficients at 3.0-T MR imaging and gleason grade in peripheral zone prostate cancer," *Radiology* **259**(2), 453–461 (2011).
- [4] Tanimoto, A., Nakashima, J., Kohno, H., Shinmoto, H., and Kuribayashi, S., "Prostate cancer screening: The clinical value of diffusion-weighted imaging and dynamic mr imaging in combination with T2-weighted imaging," *Journal of Magnetic Resonance Imaging* **25**(1), 146–152 (2007).

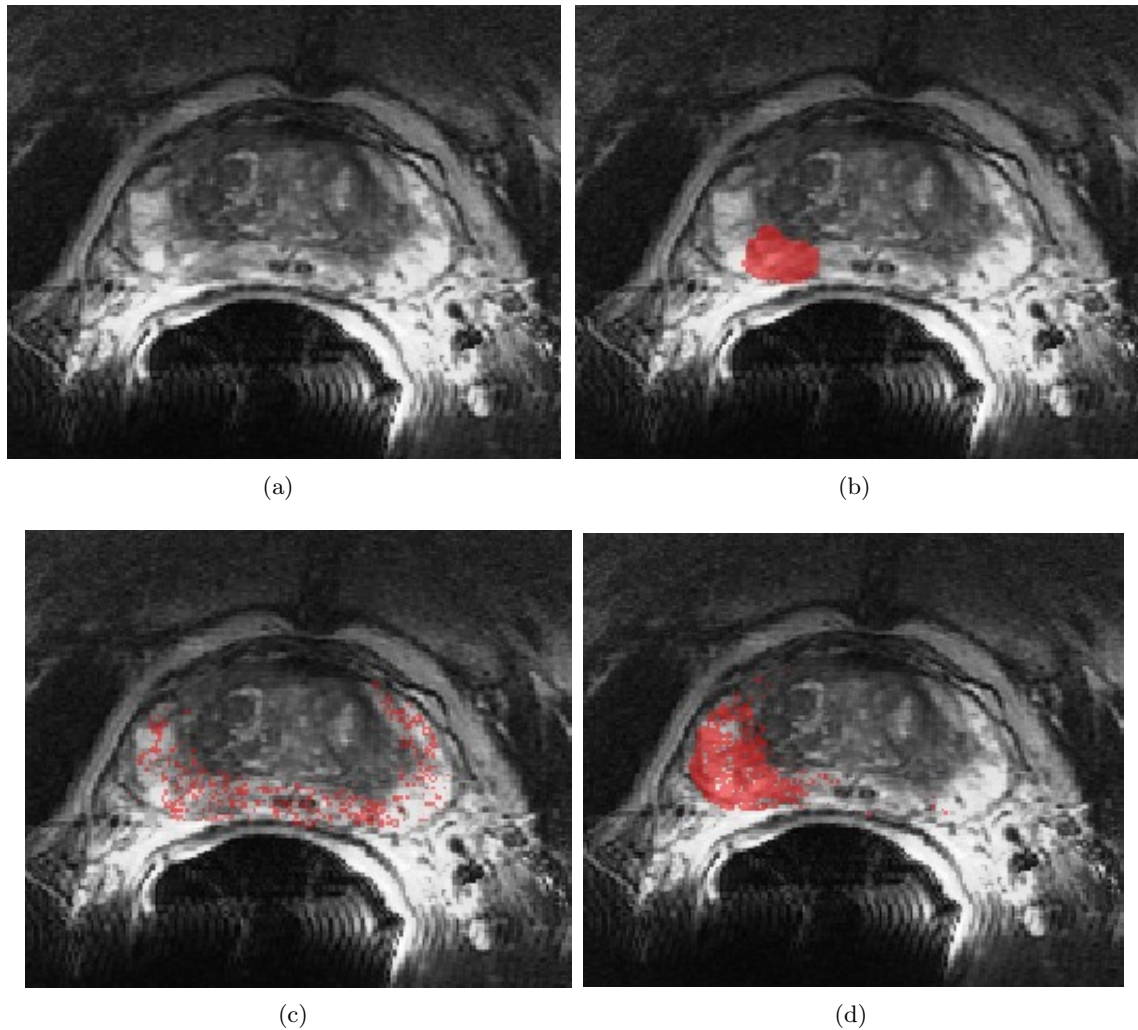


Figure 3. Lesion localization results in Example Axial Slice 3. (a) shows the T2-weighted image. (b), (c) and (d) show the cancerous regions identified by manual segmentation, LDA and LADA respectively respectively.

- [5] Turkbey, B., Pinto, P. A., Mani, H., Bernardo, M., Pang, Y., McKinney, Y. L., Khurana, K., Ravizzini, G. C., Albert, P. S., Merino, M. J., and Choyke, P. L., "Prostate cancer: value of multiparametric MR imaging at 3 T for detection–histopathologic correlation.," *Radiology* **255**, 89–99 (Apr 2010).
- [6] Gibson, E., Gaed, M., Hrinivich, T., Gómez, J. A., Moussa, M., Romagnoli, C., Mandel, J., Bastian-Jordan, M., Cool, D. W., Ghoul, S., et al., "Multiparametric MR imaging of prostate cancer foci: assessing the detectability and localizability of Gleason 7 peripheral zone cancers based on image contrasts," in [*SPIE Medical Imaging*], 90410N–90410N, International Society for Optics and Photonics (2014).
- [7] Barentsz, J. O., Richenberg, J., Clements, R., Choyke, P., Verma, S., Villeirs, G., Rouviere, O., Logager, V., and Fütterer, J. J., "ESUR prostate MR guidelines 2012," *Eur Radiol* **22**, 746–757 (Apr 2012).
- [8] Madabhushi, A., Feldman, M. D., Metaxas, D. N., Tomaszewski, J., and Chute, D., "Automated detection of prostatic adenocarcinoma from high-resolution ex vivo MRI," *IEEE transactions on medical imaging* **24**(12), 1611–1625 (2005).
- [9] Haralick, R. M., Shanmugam, K., and Dinstein, I., "Textural features for image classification," *IEEE Transactions on systems, man, and cybernetics* **6**, 610–621 (1973).
- [10] Wang, H., Yan, S., Xu, D., Tang, X., and Huang, T., "Trace ratio vs. ratio trace for dimensionality reduction," in [*2007 IEEE Conference on Computer Vision and Pattern Recognition*], 1–8 (2007).

- [11] Tang, P., Zhao, M., and Chow, T. W., “Locality alignment discriminant analysis for visualizing regional English,” *Neural Processing Letters* **43**(1), 295–307 (2016).
- [12] Zhang, T., Tao, D., Li, X., and Yang, J., “Patch alignment for dimensionality reduction,” *IEEE Transactions on Knowledge and Data Engineering* **21**(9), 1299–1313 (2009).
- [13] Le Nobin, J., Rosenkrantz, A. B., Villers, A., Orczyk, C., Deng, F.-M., Melamed, J., Mikheev, A., Rusinek, H., and Taneja, S. S., “Image guided focal therapy for magnetic resonance imaging visible prostate cancer: defining a 3-dimensional treatment margin based on magnetic resonance imaging histology co-registration analysis,” *The Journal of urology* **194**(2), 364–370 (2015).
- [14] Epstein, J. I., Walsh, P. C., Carmichael, M., and Brendler, C. B., “Pathologic and clinical findings to predict tumor extent of nonpalpable (stage T1c) prostate cancer.,” *JAMA* **271**, 368–374 (Feb 1994).