An Automated Two-step Pipeline for Aggressive Prostate Lesion Detection from Multi-parametric MR Sequence

Josh Sanyal[1], Imon Banerjee, Ph.D[1,2,3,4], Lewis Hahn, MD[3], Daniel Rubin, MD[1,3]
[1]Department of Biomedical Data Science, Stanford University, Stanford, CA
[2]Department of Biomedical Informatics, Emory University, Atlanta, GA
[3]Department of Radiology, Stanford University, Stanford, CA
[4]Department of Radiology, Emory University, Atlanta, GA

Abstract

A substantial percentage of prostate cancer cases are overdiagnosed and overtreated due to the challenge in determining aggressiveness. Multi-parametric MR is a powerful imaging technique to capture distinct characteristics of prostate lesions that are informative for aggressiveness assessment. However, manual interpretation requires a high level of expertise, is time-consuming, and significant inter-observer variation exists for radiologists. We propose a completely automated approach to assessing pixel-level aggressiveness of prostate cancer in multi-parametric MRI. Our model efficiently combines traditional computer vision and deep learning algorithms, to remove reliance on manual features, prostate segmentation, and prior lesion detection and identified optimal combinations of MR pulse sequences for assessment. Using ADC and DWI, our proposed model achieves ROC-AUC of 0.86 and ROC-AUC of 0.88 for the diagnosis of aggressive and non-aggressive prostate lesions, respectively. In performing pixel-level classification, our model's classifications are easily interpretable and allow clinicians to infer localized analyses of the lesion.

Introduction

Prostate cancer is the second most common and second deadliest cancer among men in the United States, with an estimated 31,620 fatalities in 2019[1]. While approximately 90% of all prostate cancers are detected in local and regional stages, a substantial percentage (17-50%) of cases are overdiagnosed and overtreated[2]. Low-grade prostate cancers (Gleason Score $\leq$ 6; non-aggressive) are generally considered to be non-life threatening, and in many cases, the patient may not even require immediate treatment. Physicians often recommend patients with low-grade prostate cancer simply be monitored with active surveillance while the high-grade prostate cancers with a Gleason Score ranging between 7 and 10, however, are considered aggressive and may be life-threatening. Men diagnosed with aggressive prostate cancer can benefit from and should receive early treatment. The problem of overtreating men with nonaggressive cancer is often caused by the inability to distinguish aggressive cancer from non-aggressive cancer. Accurate assessment of prostate cancer is critical for determining suitable treatment in men with aggressive prostate cancer while sparing men with non-aggressive cancers the harms of aggressive treatment and improve quality of life.

Imaging of the prostate gland is traditionally performed by ultrasound that provides excellent images of the gland, but it does not provide reliable information about malignancy and has a relatively minor role in the characterization of prostate cancer[3]. Multi-parametric MR is a powerful imaging technique for which prostate lesions show distinct characteristics in different MR pulse sequences that are informative for assessing aggressiveness. However, interpreting MRI requires a high level of expertise, is time-consuming, and significant inter-observer variation exists for radiologists in interpreting lesions in prostate MRI[4]. The purpose of our study is to automate prostate tumor detection and assessment of lesion aggressiveness from multi-parametric MR sequence using a two-step deep learning approach - prostate segmentation followed by lesion detection and characterization.

Many machine learning approaches for prostate cancer classification with MR images use pre-defined image features to characterize lesions and trained a classifier[5-10]. A core limitation of such approach is the requirement for a segmented prostate lesion which is a time-consuming manual step. Banerjee et al.[5] classified the features into five types: intensity-based (mean, SD, entropy), texture features (Haralick features, Gabor features), shape features (compactness, eccentricity), histogram-based (LBP), and edge-based (sharpness, histogram). Using this wide range of quantitative imaging features, they developed an Elastic Net model to predict aggressiveness. Similarly, Fehr et al.[6] utilized first-order (mean, SD, skewness, and kurtosis) and second-order texture features (Haralick Features - energy, entropy,

correlation, homogeneity, and contrast) derived from ADC and T2 images to assess the Gleason scores. However, the selection of features is largely based on the previous clinical experiences or other unsupervised strategies, which may lead to uninformative, redundant, and incomplete information that does not capture the comprehensive characterization of the lesion within MR images.

To tackle these issues, we develop a deep learning model which jointly learns relevant features about the prostate lesion while training a classifier to assess its aggressiveness where the prostate lesion aggressiveness is determined by the Gleason score. The higher the Gleason Score (≥ 7), the more likely that the cancer will grow and spread quickly. However, there are a few challenges to applying deep learning models to our dataset. First, the small size of the prostate lesion relative to the entire image introduced significant noise within the training data. Second, the number of patients in the dataset is limited, which makes training difficult due to potential overfitting. To solve similar challenges, most recent work of Yuan et al.[11] employed transfer learning, data augmentation, and cropping smaller patches around the lesion, and achieved 0.896 ROC AUC in distinguishing between aggressive and non-aggressive lesions. However, this still requires knowledge of the lesion's location. Instead, we use a pixel-level approach that aims to classify each pixel inside of patches where the pixels will either be classified as inside of an aggressive lesion (Gleason score ≥ 7), inside of a non-aggressive lesion (Gleason score < 7), or outside of the lesion. The pixel-based approach vastly increases our training size while training with a limited dataset, eliminates unnecessary noise when examining pixels within the lesion, and provides a more exhaustive assessment of the lesion. The source code is available at https://github.com/joshsanyal/Prostate-Cancer-Lesion-Assessment.

The main contributions are as follows:

1. We developed an automated pipeline that takes input multi-parametric MR images[1], segment the prostate gland, and register the images into the same image-space based on the segmented prostate shape.

2. We propose a pixel-based deep learning model that detects pixels inside of the prostate lesion and classifies them as aggressive (Gleason score ≥ 7) or non-aggressive (Gleason score < 7)

3. To increase interpretability of the result, we propose a heatmap visualization of the aggressiveness score for the prostate tissue which has been computed by the model.

4. Using the proposed model, we experiment on our dataset to determine its performance with optimal combinations of MR pulse sequence and comment on the significance of the sequence.

Methodology

Figure 1 shows the five core components of proposed research workflow and their outputs: 1. Image Preprocessing, 2. Prostate Segmentation, 3. Image Registration, 4. Pixel-level classification, and 5. Evaluation. Each component is described in greater detail in the following subsections.

Dataset

We use a dataset of the multi-parametric MRI scans of 77 prostate cancer patients: 38 patients with aggressive cancer (Gleason score ≥ 7) and 39 patients with non-aggressive cancer (Gleason score < 7). The Gleason score, varying from 0 - 10, with 10 representing the most aggressive cancer and 0 representing benign lesion, was also recorded for each patient. For each patient, the multi-parametric MR image set includes T2-weighted (3.0 T, 3mm slice thickness) images, diffusion-weighted images (DWI b values 0 750s/mm^2) and apparent diffusion coefficient maps (ADC). All images were acquired in the axial plane. To create the ground truth for prostate segmentation and lesion detection, an expert radiologist identified and circumscribed the prostate gland and suspicious lesion in a single slice through the largest section of the lesion on each sequence.

---

[1]ADC: Apparent Diffusion Coefficient Maps, DWI: diffusion-weighted images, T2 MR: T2-weighted imaging
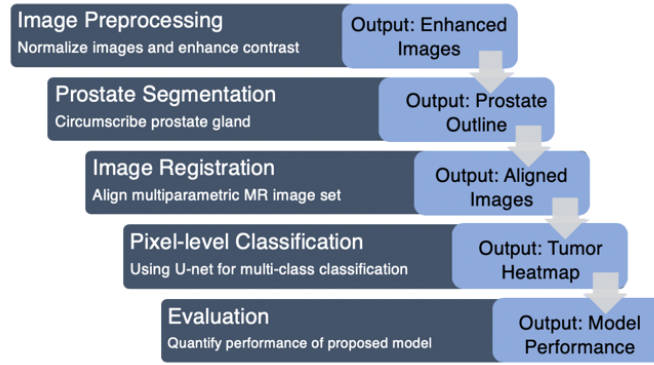
Figure 1: Proposed Research Workflow. The interaction between the sequential components is shown by the arrow. Output from the previous component is being passed to the next component for analysis.

Image Preprocessing

To reduce variance across the image sequences, they were normalized into a predefined range [0, 1]. As pre-processing steps, contrast-limited adaptive histogram equalization[12] was applied to enhance the image quality. The adaptive method computes several histograms, each corresponding to a distinct section of the same image, and uses them to redistribute the lightness values of the image. It is therefore suitable for improving the local contrast and enhancing the definitions of edges in each region of an image. The multi-parametric MR image set after preprocessing is shown in Figure 2.
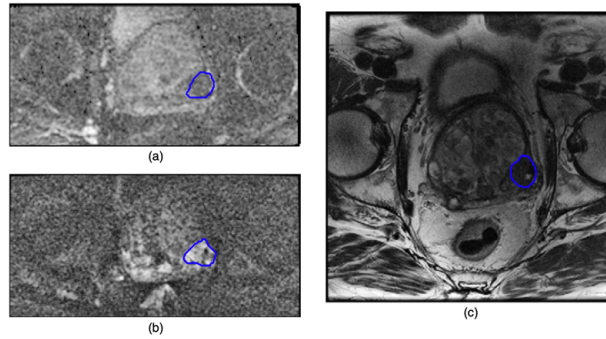

Figure 2: MR image sequences with the radiologist's lesion segmentation outlined (in blue) - (a) ADC, (b) DWI, and (c) T2

Prostate Segmentation

We utilize a U-Net architecture, introduced by Ronneberger et al.[13], to use the entire image as input and classify pixels as being inside or outside of the prostate. The architecture consists of 5 encoder subnetworks each with two sets of convolutional and ReLU layers and 5 decoder subnetworks each with a transposed convolutional layer for upsampling and two sets of convolutional and ReLU layers. The Adam optimizer was chosen with a cross entropy loss function and the learning rate was set to 10^{-3} and was decayed every 5 epochs, for 20 total epochs. Using this architecture and the ground truth prostate gland outlines, we trained three U-nets, one for each image sequence, to create a prostate gland heatmaps (probability map) as shown in Figure 3.

Due to classification errors, standard thresholding of the prostate gland heatmap resulted in segmentation noise with multiple distinct regions classified as prostate tissue. To combat this, we utilize the chan-vese active contour algorithm[14] to segment our heatmap. As an initial contour, the 10 x 10 px region with the highest total probability of

being inside the prostate was selected. The active contour algorithm then iteratively segmented the heatmap until the difference between the average probability of the pixels currently inside the segmentation and the average probability of the pixels begin added exceeded a pre-defined constant. The segmentation is then smoothened to more realistically represent the prostate's edge. An example of this iterative process is shown in Figure 4.
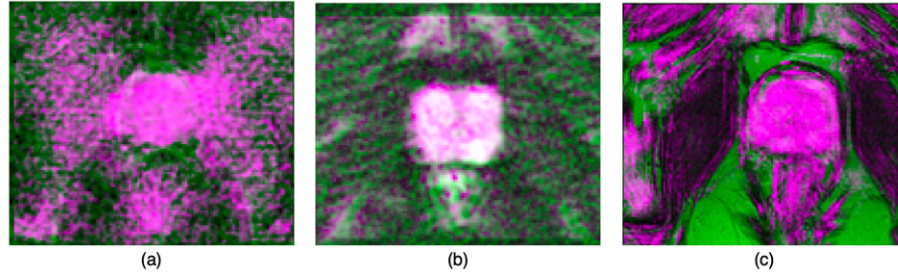

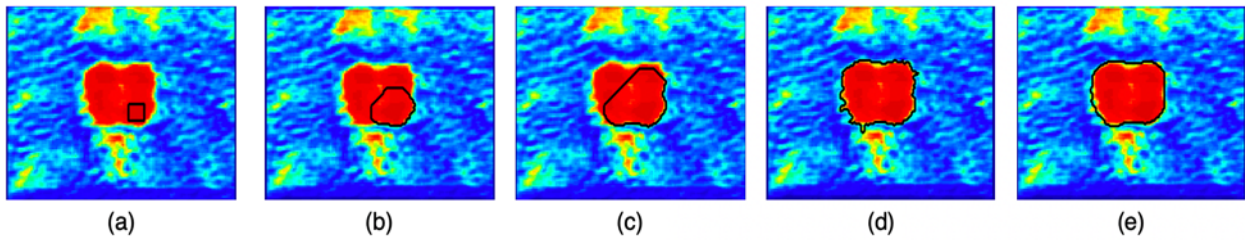**Figure 3:** Examples of prostate gland probability heatmaps generated by the network in (a) ADC, (b) DWI, (c) T2

Figure 4: Prostate segmentation at different iterations (a) 0, (b) 10, (c) 20, (d) final, (e) smooth

Image Registration

Image registration is needed to align the multi-parametric MRI sequence set such that the prostate gland and lesion captured in all three sequences are in the same space. This allows our model to leverage a combination of the different MRI sequences having unique characteristics for detection and aggressiveness classification. Due to varying image contrast and inconsistent intensity values among the MRI sequences, popular intensity-based and feature-point based approaches[14–18] to image registration were not judged as optimal during visual evaluation. Thus, we previously developed a shape-based registration algorithm[19] to align the three image sequences by using the segmented prostate gland. Our registration algorithm uses a rigid transformation, consisting of translation, rotation, and scaling, ensuring that the transformation is not affected by imperfect manual prostate segmentation which could lead to warped images. The algorithm first estimates a geometric transformation that minimizes mean square error to optimally align the prostate gland shapes captured in ADC and DWI images to the fixed T2 prostate gland. These two transformations are then applied to the original DWI and ADC images, register them with the fixed T2 image, as shown in Figure 5.

Pixel-level Classification using U-Net

To deal with a low number of training images, we utilize a U-Net architecture with three output channels to classify pixels as being outside of the tumor, inside an aggressive tumor (Gleason score ≥ 7), or inside a non-aggressive tumor (Gleason score < 7). The bounding box around the prostate with 10 pixels of padding is cropped to keep the peripheral tissue of the prostate and divided into smaller patches. These patches are then fed into the U-Net which calculates the probability of all pixels inside each patch being one of the three classes. A simplified representation of this architecture is shown in Figure 6 where the aggressive prostate tumor heatmap only displays the probability that the pixel is inside an aggressive lesion.

The U-Net architecture consists of 2 encoder subnetworks each with two sets of convolutional and ReLU layers and
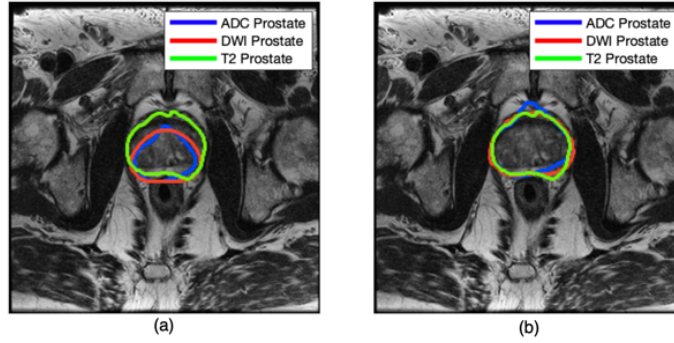
Figure 5: Prostate outline overlaid upon T2 sequence - (a) before, (b) after shape-based registration
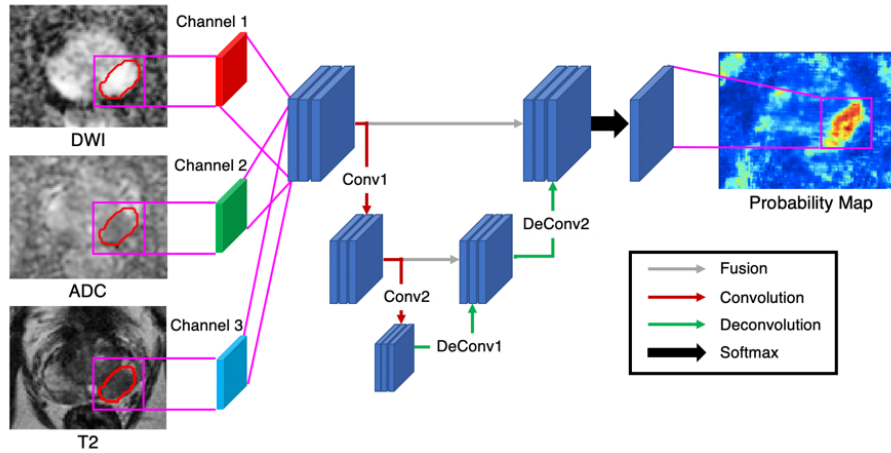

Figure 6: Simplified representation of the pixel-level aggressiveness assessment using 3-channel U-Net

2 decoder subnetworks each with a transposed convolutional layer for upsampling and two sets of convolutional and ReLU layers. The Adam optimizer was chosen with a cross entropy loss function and the learning rate was set to 10^{-3} and was decayed every 5 epochs, for 30 total epochs. Using this architecture, we experimented with models using patch sizes of 11 x 11 pixels, 21 x 21 pixels, and a combination of both where the model predicted probability were averaged. The patch size of 11 x 11 pixels was chosen to capture smaller regions of the lesion whereas the patch size of 21 x 21 pixels was chosen to capture the entirety of the lesion. We additionally tested models with different combinations of image sequences in which each image sequence is passed as a channel in the U-Net.

To train our U-Net, we obtained every patch of size n x n from our training images. The set of pixels within each image patch was then labeled either outside, non-aggressive, or aggressive based on its location and the aggressiveness of the lesion. To combat the high percentage of patches outside of the lesion and ensure class balance, we randomly sampled 3500 patches with the center pixel inside of an aggressive tumor, a non-aggressive tumor, and outside the tumor. This was done with an emphasis on patches with center pixels close to the edge of the lesion to ensure the model can detect the edge of the lesion by accurately distinguishing between pixels near the edge. When a new patch was sampled, patches of the same patient within its proximity were discarded to ensure diverse training examples. Once the initial patches were selected randomly, this randomness was fixed for all other experiments, namely different patch sizes, to ensure the same sampling of training examples. These classifications are then pieced back together and visualized using a prostate tumor heatmap.

Evaluation

To quantify the performance of our model, we split the dataset of 77 patients into 75% training set (57 images) and 25% test set (20 images). We first train our U-net for prostate segmentation using images from our training set and use the trained network to segment the prostate in our test set. Our shape-based registration algorithm then uses these segmentations to align our MR image set. Next, our U-net for pixel-level aggressiveness classification is trained using cropped patches from our training set as described in the previous section. Using the hold-out test set, we move cropped patches inside the prostate and feed them into our trained model to classify each pixel inside the prostate. This process is repeated with different patch sizes and combinations of image sequences using the same data split each time. In the results section, we describe our model's performance using the Dice similarity coefficient to quantify the accuracy of our prostate segmentations when compared to the radiologist's ground truth. The area under the receiver operating curve (ROC AUC) was chosen as a metric to quantify the classification performance based on its ability to accurately determine model performance despite class imbalance or a specific threshold. This allows clinicians to choose their preferred tradeoff between specificity and sensitivity. It is not legitimate to compare between previous lesion classifications methods and our pixel-level method since our method does not classify the whole lesion and does not require the lesion outline as input. We sought to directly compare our method with the recently proposed method of Yuan et al.[11] but were unable to obtain the code and replication of their method is not trivial.

Results

Prostate Segmentation

On the 20 images in our test set, our model was able to segment the prostate in T2 with a dice similarity coefficient (DSC) of 0.87 ± 0.04. For DWI images, our model achieved a 0.82 ± 0.06 DSC and for ADC images, our model achieved a 0.76 ± 0.05 DSC. This low variance in performance demonstrates the robustness of our segmentation method. In general, our model achieved the greatest performance in T2, indicating that this modality provided the most relevant information to delineate the prostate from its surrounding tissue. Additionally, when the modalities were registered based on the radiologist's ground truth segmentation beforehand, the U-net was able to use all three modalities as separate input channels, resulting in a 0.87 ± 0.04 DSC. The equivalence of this result to the U-net using T2 alone indicates that ADC and DWI were not informative modalities for distinguishing the prostate gland.

Pixel-level U-Net Classification

Significance of Patch Size

To assess the significance of patch size on the model's performance, we trained the model using patch sizes of 11 x 11 pixels, 21 x 21 pixels, and a combination of the two. This was done using each of the image sequences independently. As is shown in Figure 7, we assessed model performance by calculating the one-vs-all ROC AUC where we compared the aggressive class (pixels inside of an aggressive lesion) with all other classes as well as compared the non-aggressive class with all other classes. The comparison between the outside class to the other classes was excluded because of insignificant differences in ROC AUC because it was an easier task than detecting pixels inside of the lesion.

The proposed method was able to classify aggressive pixels with comparable ROC AUC across different patch sizes T2 (range of [0.67-0.69]), ADC (range of [0.76-0.81]), and DWI (range of [0.8-0.81]). Similar results occurred when classifying non-aggressive pixels in T2 (range of [0.69-0.71]), ADC (range of [0.8-0.82]), and DWI (range of [0.82-0.83]). This low variance in performance demonstrates the robustness of our method. In general, the model trained on the 21 x 21 patch size outperformed the 11 x 11 patch size, however, the average of the two predictions resulted in the best performance. This indicates the model was able to learn unique information from the different patch sizes, which when combined, were able to outperform the usage of a single patch size.

Significance of parametric maps

To determine the significance of parametric maps used in our model, we tested the model using all possible combinations of image sequences. We trained the model using a combination of the two patch sizes, due to its higher performance in the previous section. The results are shown in Figure 8, where both the aggressive and non-aggressive
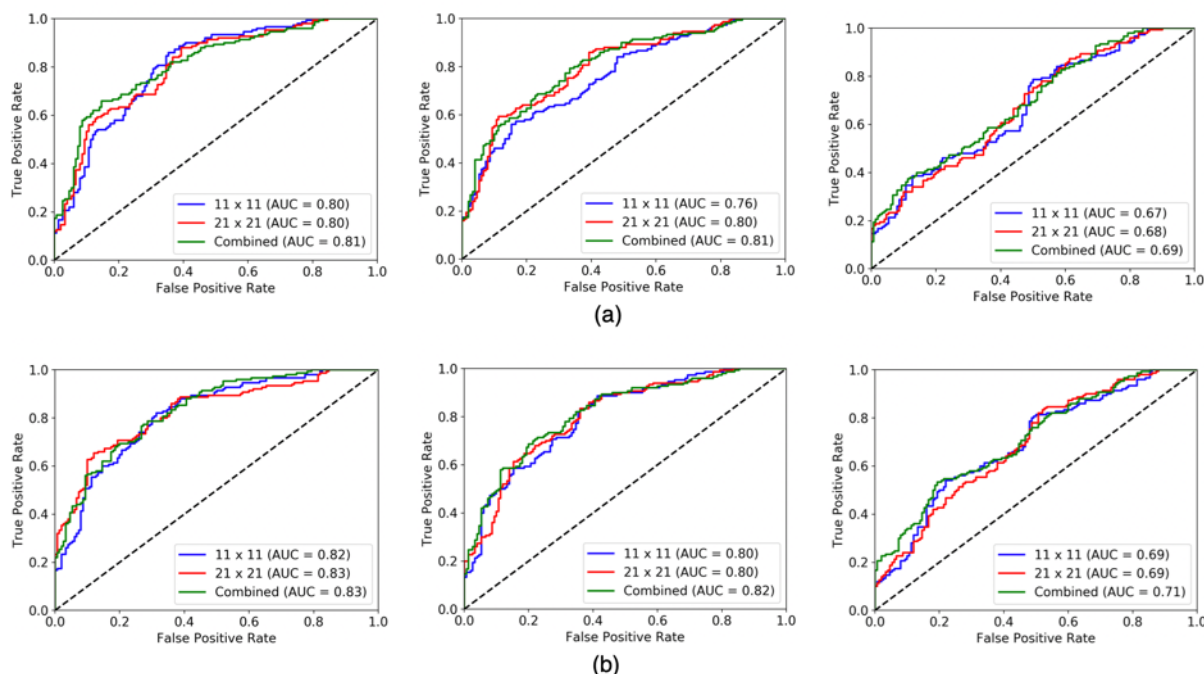
Figure 7: Model's ROC curves using different patch sizes in DWI (left), ADC (middle), T2 (right) - (a) Aggressive vs All, (b) Non-Aggressive vs All

ROC AUC values are displayed in 3 x 3 matrices where the modalities are indicated as the matrices' axes. The matrix elements along the diagonal from top left to bottom right, contain the performance of the model trained on a single modality, while all other elements contain the performance when two modalities are used.

The model was able to perform well using DWI and ADC individually when classifying aggressive pixels (both 0.81 ROC AUC) and non-aggressive pixels (0.83 and 0.82 ROC AUC, respectively). Using a combination of the two, led to the model's best performance with an aggressive ROC AUC of 0.86 and a non-aggressive ROC AUC of 0.88 (Figure 8.c). This indicates the model was able to learn unique and informative features from these modalities. However, the model was unable to perform well with the T2 image sequence, classifying aggressive pixels with 0.69 ROC AUC, classifying non-aggressive pixels with 0.71 ROC AUC, and decreasing the performance of any combinations of DWI and ADC it was incorporated in. This indicates that T2 was not an informative modality for distinguishing aggressive and non-aggressive lesion tissue. Adding T2 with DWI and ADC did not improve the performance (Figure 8.d).

Aggressiveness Visualization

The model which used the ADC and DWI images sequences and a combination of 11 x 11 and 21 x 21 px patch sizes, achieved the best performance with an aggressive ROC AUC of 0.86 and a non-aggressive ROC AUC of 0.88. This result suggests that the model which combines the ADC and DWI modalities performs well for distinguishing aggressive and non-aggressive lesion tissue. However, the model's classifications are likely to be doubted by a clinician, unless they are given a way to reason the model's outcomes. While this issue of traditional deep learning models functioning as a "black-box" without much transparency, our model's pixel-level approach allows direct visualization of computed probability as a heatmap. We present samples on two aggressive prostate cancer patients as well as two non-aggressive prostate cancer patients in Figure 9. Instead of providing the clinician just with a classification score of aggressive or non-aggressive for each patient, our pixel-level illustration allows the clinician infer localized analyses of the lesion by overlaying the prostate tumor heatmap on the patient's MR images. This can help in understanding the model's classifications outcome and thus, be utilized in a clinician's assessment and treatment planning.
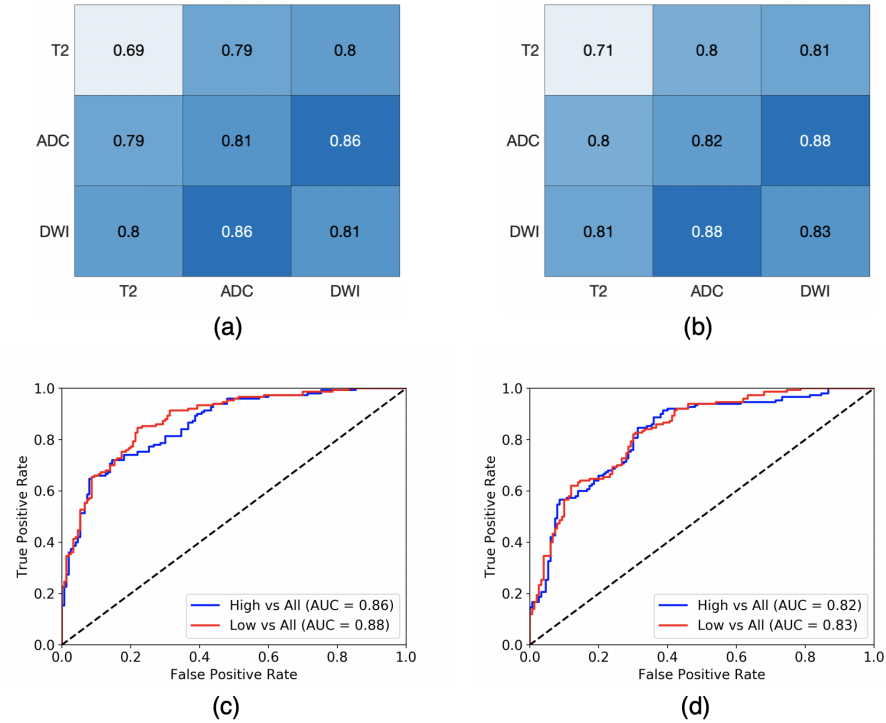
Figure 8: Model performance using different MR parametric maps - (a) Aggressive vs All, (b) Non aggressive vs All, (c) ROC Curve with Best Performance with DWI and ADC, (d) ROC curve using all modalities
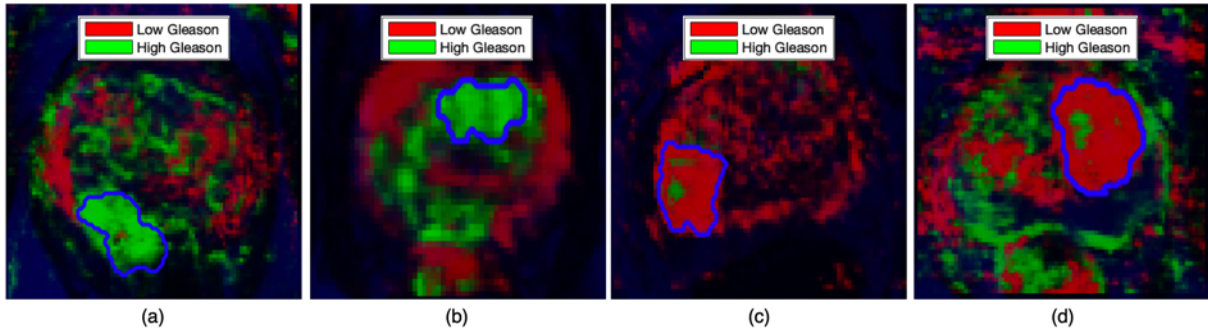

Figure 9: Prostate tumor heatmap visualized on T2 overlayed with manual outline in blue - (a, b) aggressive examples, (c, d) non-aggressive examples

Conclusion

We propose a completely automated approach to assessing pixel-level aggressiveness of prostate cancer in multi-parametric MRI. Our model efficiently combines traditional image registration, and deep learning algorithms - segmentation and detection, to remove reliance on manual feature engineering, prior lesion detection, and segmentation. Our experimental results show that using ADC and DWI, our proposed model achieved a ROC AUC of 0.86 for aggressive lesion and a ROC AUC of 0.88 for non-aggressive lesion. Further, our model is easily interpretable and allow clinicians to infer localized analyses of the prostate gland and lesion (Figure 9). The core limitation of the current study is that it only considered data from a single academic institution for training and validation. Studies in larger independent datasets would be helpful to confirm our results. Our study also was limited to the analysis of 2D images, and volumetric analysis could improve the results. For multi-lesion prostate, we considered a single lesion for each

study with the highest Gleason score for training due to the limitation of manual annotation. In future studies, we will train our model with multiple lesions circumscribed within the prostate gland. We will also compare our heatmap visualizations with histopathology to evaluate correspondence in spatial locations of aggressive tumor regions, utilize alternative MRI sequences which could boost model performance by providing unique information, and incorporate additional levels of aggressiveness for the model to detect.

References

1. American Cancer Society. *Cancer facts & figures 2019.* American Cancer Society, 2019.

2. Virginia A Moyer. Screening for prostate cancer: Us preventive services task force recommendation statement. *Annals of internal medicine*, 157(2):120–134, 2012.

3. John V Hegde, Robert V Mulkern, Lawrence P Panych, Fiona M Fennessy, Andriy Fedorov, Stephan E Maier, and Clare MC Tempany. Multiparametric mri of prostate cancer: an update on state-of-the-art techniques and their performance in detecting and localizing prostate cancer. *Journal of magnetic resonance imaging*, 37(5):1035–1054, 2013.

4. Shijun Wang, Karen Burtt, Baris Turkbey, Peter Choyke, and Ronald M Summers. Computer aided-diagnosis of prostate cancer on multiparametric mri: a technical review of current research. *BioMed research international*, 2014, 2014.

5. Imon Banerjee, Lewis Hahn, Geoffrey Sonn, Richard Fan, and Daniel L Rubin. Computerized multiparametric mr image analysis for prostate cancer aggressiveness-assessment. *arXiv preprint arXiv:1612.00408*, 2016.

6. Duc Fehr, Harini Veeraraghavan, Andreas Wibmer, Tatsuo Gondo, Kazuhiro Matsumoto, Herbert Alberto Vargas, Evis Sala, Hedvig Hricak, and Joseph O Deasy. Automatic classification of prostate cancer gleason scores from multiparametric magnetic resonance images. *Proceedings of the National Academy of Sciences*, 112(46):E6265–E6273, 2015.

7. Fusun Citak-Er, Metin Vural, Omer Acar, Tarik Esen, Aslihan Onay, and Esin Ozturk-Isik. Final gleason score prediction using discriminant analysis and support vector machine based on preoperative multiparametric mr imaging of prostate cancer at 3t. *BioMed research international*, 2014, 2014.

8. Emilie Niaf, Rémi Flamary, Olivier Rouviere, Carole Lartizien, and Stéphane Canu. Kernel-based learning from both qualitative and quantitative labels: application to prostate cancer diagnosis based on multiparametric mr imaging. *IEEE Transactions on Image Processing*, 23(3):979–991, 2013.

9. Geert Litjens, Oscar Debats, Jelle Barentsz, Nico Karssemeijer, and Henkjan Huisman. Computer-aided detection of prostate cancer in mri. *IEEE transactions on medical imaging*, 33(5):1083–1092, 2014.

10. A Vignati, S Mazzetti, V Giannini, F Russo, E Bollito, Francesco Porpiglia, M Stasi, and Daniele Regge. Texture features on t2-weighted magnetic resonance imaging: new potential biomarkers for prostate cancer aggressiveness. *Physics in Medicine & Biology*, 60(7):2685, 2015.

11. Yixuan Yuan, Wenjian Qin, Mark Buyyounouski, Bulat Ibragimov, Steve Hancock, Bin Han, and Lei Xing. Prostate cancer classification with multiparametric mri transfer learning model. *Medical physics*, 46(2):756–765, 2019.

12. Stephen M Pizer, E Philip Amburn, John D Austin, Robert Cromartie, Ari Geselowitz, Trey Greer, Bart ter Haar Romeny, John B Zimmerman, and Karel Zuiderveld. Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing*, 39(3):355–368, 1987.

13. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

14. Tony F Chan and Luminita A Vese. Active contours without edges. *IEEE Transactions on image processing*, 10(2):266–277, 2001.

15. Josien PW Pluim, JB Antoine Maintz, and Max A Viergever. Mutual-information-based registration of medical images: a survey. *IEEE transactions on medical imaging*, 22(8):986–1004, 2003.

16. Szymon Rusinkiewicz and Marc Levoy. Efficient variants of the icp algorithm. In *3dim*, volume 1, pages 145–152, 2001.

17. Mehdi Hedjazi Moghari and Purang Abolmaesumi. Point-based rigid-body registration using an unscented kalman filter. *IEEE Transactions on Medical Imaging*, 26(12):1708–1728, 2007.

18. Seth D Billings, Emad M Boctor, and Russell H Taylor. Iterative most-likely point registration (imlp): A robust algorithm for computing optimal shape alignment. *PloS one*, 10(3):e0117688, 2015.

19. Josh Sanyal, Imon Banerjee, and Daniel Rubin. Registration boost performance of aggressive prostate cancer diagnosis, 03 2019.