UCLA
UCLA Electronic Theses and Dissertations

Title

Prediction of Prostate Cancer in Biopsy Targets Using Multiparametric Magnetic Resonance Imaging

Permalink

https://escholarship.org/uc/item/4bf012w0

Author

Raman, Alex Gautham

Publication Date

2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Prediction of Prostate Cancer in Biopsy Targets

Using Multiparametric Magnetic Resonance Imaging

A thesis submitted in partial satisfaction

of the requirements for the degree Master of Science

in Bioinformatics

by

Alex Gautham Raman

2019

ABSTRACT OF THE THESIS

Prediction of Prostate Cancer in Biopsy Targets

Using Multiparametric Magnetic Resonance Imaging

by

Alex Gautham Raman

Master of Science in Bioinformatics

University of California, Los Angeles, 2019

Professor Xinshu Xiao, Chair

Prostate cancer is one of the leading causes of death due to cancer for men in the United States. Current diagnostic procedures have been shown to lead to overdiagnosis and overtreatment of the disease. Thus, a need exists for diagnostic protocols with higher positive predictive value (PPV). Magnetic resonance imaging (MRI) guided biopsy has emerged as a new diagnostic technology that has increased the ability to detect cancerous tissue in the prostate. In this study, we utilize a dataset of 555 patients who have undergone MRI-guided biopsy to answer two questions: 1) how accurate are radiologist drawn regions of interest (ROIs) on prostate MRI, and 2) can we map the location of a biopsy to an MRI and use MRI voxel intensities at that location to predict whether or not the biopsy core contains clinically significant cancer (csCaP). In answering the first question, we found that 50.35% of csCaP-containing cores are found inside ROIs while 49.65% of csCaP-containing cores are found outside of ROIs. This indicates room

for improvement in ROI delineation. We then trained support vector machine (SVM) and logistic regression classifiers using features from the MRI voxels corresponding to each biopsy core's location to predict whether cores would be cancer-positive or cancer-negative. The SVM achieved the best performance, with a negative predictive value of 0.93, a PPV of 0.23, and a test area under the curve (AUC) of 0.72.

The thesis of Alex Gautham Raman is approved.

Eleazar Eskin

Corey Wells Arnold

Xinshu Xiao, Committee Chair

University of California, Los Angeles

2019

To my family -- thank you for always being my source

of unconditional love and support

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would first like to thank my mentors Dr. Corey Arnold and Karthik Sarma for their advice, patience, and kindness over the last three years. They have provided me with nonstop encouragement ever since I was an undergraduate sophomore, fresh out of CS 32, and in doing so, have afforded me the opportunity to discover a deep passion for computer science and medicine. I would also like to thank my committee members Drs. Xinshu Xiao and Eleazar Eskin for their time and help with reviewing this thesis.

I am extremely grateful, and lucky to have had guidance from Drs. Alan Priester, Shyam Natarajan, Leonard Marks, and Steven Raman. Without the code and excellent data processing done by Alan and Shyam as well as the dataset generation and clinical annotations performed by Dr. Marks and Dr. Raman, this work would not have been possible. I would also like to thank my lab mate Wen Shi for his direct help on this topic as well.

Lastly, I would like to thank my parents, sister, and friends for their endless love and support. I would not be who I am today without you all.

CHAPTER 1

INTRODUCTION

Prostate cancer (CaP) is the most commonly diagnosed cancer in American men. It is also the second leading cause of death due to cancer for men in the United States.[1] Though CaP can be a severe condition, most men who have the disease do not die from it but rather, with it. The growth of CaP is often slow, and this has important implications for treatment decisions and screening methodologies.[2] Current treatment for the disease includes procedures such as radical prostatectomy, radiation, and chemotherapy, all of which have serious possible side-effects such as incontinence and impotence.[3,4] In addition, screening methodologies such as the prostate-specific antigen (PSA) test have been shown to lead to overdiagnosis and overtreatment.[5] These two aspects of CaP - the large decline in quality of life due to treatment and the trend of overdiagnosis – make it a unique type of cancer to be tackled. The costs and benefits of treatment must be carefully weighed on a patient-by-patient basis, and the development of diagnostic procedures with high PPV is critical.

Current screening procedures for CaP proposed by the National Comprehensive Cancer Network involve monitoring PSA levels, performing a digital rectal exam (DRE), conducting imaging studies such as multiparametric MRI (mpMRI), and then following up with a transrectal-ultrasound (TRUS) guided biopsy for at-risk patients.[6] PSA is a protein made by the prostate and can be detected by a simple blood test. Typically, the larger the size of the gland, the higher concentration of PSA (measured in ng/ml) there will be in the blood. Gland size can increase due to cancer or diseases such as benign prostatic hyperplasia. PSA velocity, a measure of how quickly a patient's PSA level increases over time, is often used as an indication of

problems in the prostate. However, it is seen as somewhat unreliable since PSA is also known to increase after exercise and sex.

Next, the DRE involves a physician inserting a gloved finger into the rectum and feeling for hard, lumpy areas that could be indicative of CaP. If PSA levels or the DRE are abnormal, a physician may then schedule an mpMRI. The mpMRI is typically done using either a 1.5 or 3 Tesla MRI scanner and consists of T2-weighted imaging, Diffusion Weighted Imaging (DWI), and Dynamic Contrast Enhanced (DCE) or "perfusion" imaging. The use of these different MRI modalities allows for the increased detection of prostate cancer.[7] In T2 and DWI, cancer is often seen as a low-intensity signal, while in DCE cancer is seen as a high-intensity signal.[8] Using mpMRI, a radiologist can denote regions of interest (ROIs) in the image that they think correspond to potentially suspicious lesions. The severity of these lesions can then be scored on a 1-5 scale using the Prostate Imaging – Reporting and Data Services (PIRADS) scoring system, with 1 being negligible and 5 being highly suspicious of cancer.[9]

Finally, with all of this information at hand, a physician will perform a TRUS-guided biopsy of the prostate. Biopsy cores consisting of prostate tissue will then be sent to a pathologist, who will examine each core and assign it a Gleason score. The Gleason score consists of two numbers – the primary and secondary score. Each score is a categorical score from 1 to 5 based on the histopathological structure of the tissue. A score of 1 indicates well-differentiated tissue and has the most favorable prognosis, while a score of 5 indicates the least-differentiated tissue and has a poor prognosis.[10] The two scores combined (e.g. 4+3 = 7) yields the total Gleason Score. A combined Gleason score of 7 or higher is deemed to be clinically significant CaP.[11]

From the early 2000s to the recent present, biopsy of the prostate has been performed using a systematic technique.[12] In this method, 12 cores are taken from pre-defined regions of the prostate and sent to the pathologist for Gleason grading. However, with the advent of MRI, a new technique called MRI-Ultrasound (MRI-US) Fusion Targeted Biopsy has emerged, which allows for biopsy cores to be taken not only systematically, but also from previously denoted ROIs.[13] In this technique, an MRI of the prostate is taken and sent to a radiologist for examination. The radiologist will denote ROIs as well as a contour of the prostate on the MRI so that a biopsy can be taken under MR guidance. Using a biopsy device such as the Artemis, produced by Eigen, this MRI, including the ROI and prostate contours, will be fused in real time with the ultrasound that the urologist uses to take the biopsy. In this manner, the urologist can take cores directly from the region of the prostate corresponding to the ROIs denoted by the radiologist. This technology allows cores to be taken with a spatial error of just 1-2 mm.[14] MRI-US Targeted Biopsy has been shown to increase the detection of high-risk CaP by 30% and decrease the detection of low-risk cancers by 17% when compared to TRUS systematic biopsy.[15] Thus targeted biopsy has proven to be an enormously useful technique that has improved diagnosis of CaP.
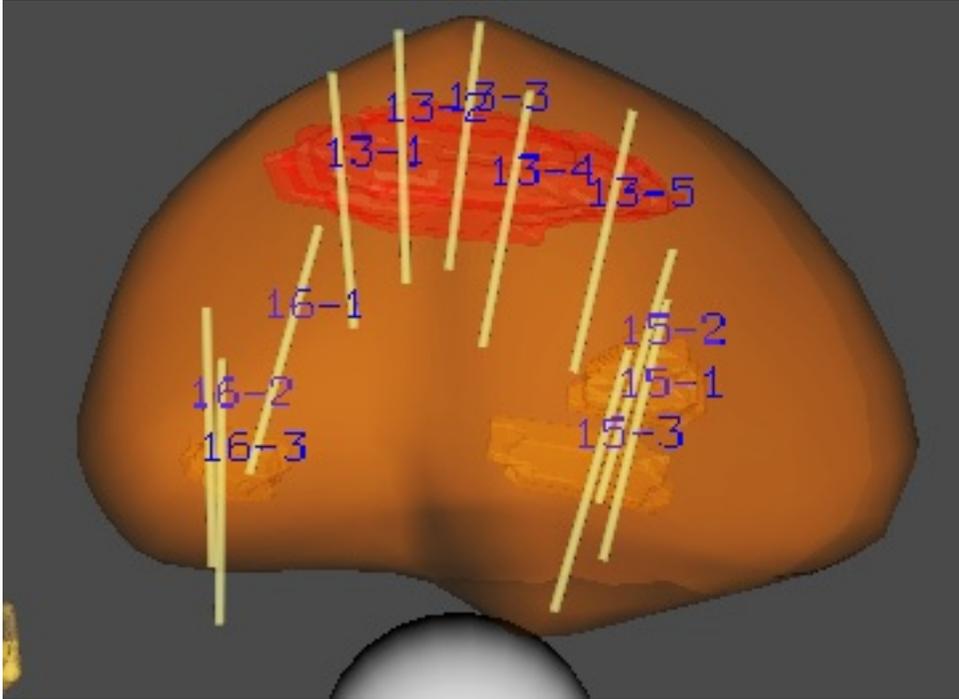
Figure 1: MRI-US Fusion Targeted Biopsy - The dark orange contour in this figure represents the prostate, while ROIs are denoted by the red and light orange regions within the prostate. Biopsy cores, seen in yellow, are taken from these ROIs. Figure from: Jessedle [CC BY-SA 3.0 (https://creativecommons.org/licenses/by-sa/3.0)].

This work utilizes a dataset of 555 MRI-US Fusion guided prostate biopsy procedures performed at UCLA between 2012 and 2016. This data consists of MR images with prostate and ROI contours, 3-Dimensional (3D) biopsy coordinates in both MR and Ultrasound (US) space, as well as Gleason scores for each biopsy core. Using the MRI-US Fusion technology developed by Eigen, the path of each biopsy core can be visualized in the 3D MRI. Voxels from the MRI corresponding to the spatial location of each biopsy can then be extracted. Using this data, we seek to answer two questions:

1. **How often do cancer-positive and cancer-negative cores occur outside of ROIs?**

2. **Can the result of a biopsy (cancer-positive or cancer-negative) be predicted from its corresponding MR voxel intensities alone?**

The purpose of answering the first question is to ascertain how effective radiologist-drawn ROIs are. An analysis will be conducted on how often cancer-positive cores occur inside vs. outside ROIs as well as how often cancer-negative cores occur inside vs. outside ROIs. We will also examine how far away the positive and negative cores that occur outside of ROIs are from the ROIs.

By answering the second question, we hope to enable a method for conducting "virtual" biopsies of the prostate. We aim to build a machine learning based classifier that can accurately predict the presence of prostate cancer in a biopsy core from nothing but MR voxel intensities. Biopsy Gleason score is still considered to be the best ground truth for presence of CaP apart from whole-mount histopathology (which requires resection of the prostate gland). So, if Gleason scores can be predicted from an MRI, we can improve the detection of *clinically significant* lesions and better inform physicians about potential biopsy targets.

The use of MR-derived parameters has already been shown to improve the identification of risk for csCaP upon biopsy.[16] Specifically, using MRI-derived prostate volume as well as the PI-RADS score of the ROI decreases the false positive rate of a model designed to predict the presence of csCaP during biopsy. Additionally, Haralick textures, which are features of an image based on spatial grayscale variation, have been shown to be able to differentiate between

cancerous and non-cancerous tissue in T2 MRIs of the prostate.[17] These features include quantities like homogeneity and contrast, and can be computed directly from the grayscale intensity values at a region of an image. Thus, it is clear that MR features, both seen through the eye and derived computationally, have the potential for detecting prostate cancer. We hypothesize that applying these types of features to a volume of pixels representing a biopsy core will allow us to differentiate between cores containing cancerous and cores containing non-cancerous tissue, thereby allowing us to more accurately localize cancer in the prostate.

CHAPTER 2

DATA PREPROCESSING

§ 2.1 Introduction to the Artemis Dataset

The dataset used for all the analyses described in this thesis consists of patients from Ronald Reagan Hospital who have undergone MRI-US Fusion guided prostate biopsy using the Artemis device produced by Eigen. This dataset consists of 1810 patients who have undergone MRI-US Fusion guided biopsy. However, not all these patients were used in the subsequent analyses. Various exclusion criteria were used to select patients with a complete set of biopsy coordinates, Gleason scores, and MR data. The exclusion criteria and resulting patient cohort are described below.
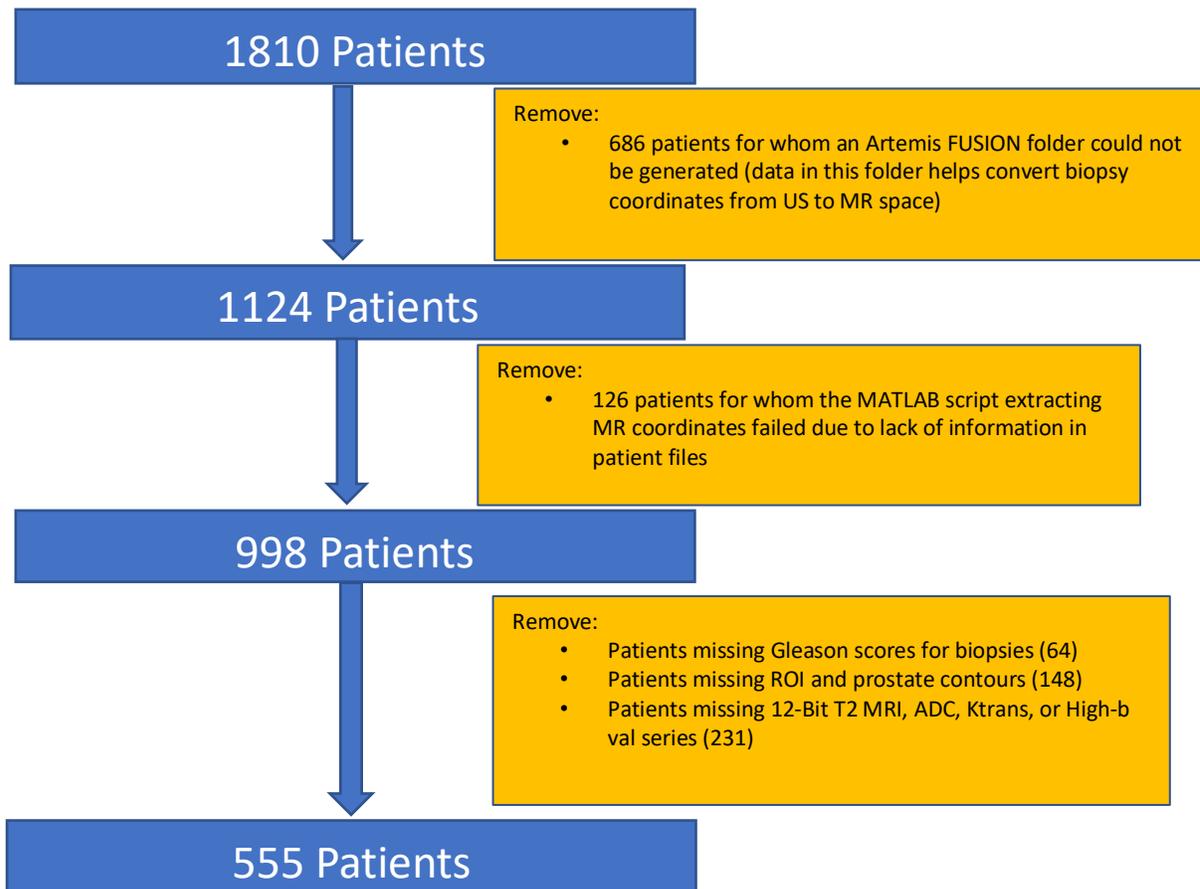
Figure 2: Patient Exclusion Criteria – Patients are excluded from the final analysis performed in this study largely due to missing information. This information includes certain MR series, segmentation files, Gleason scores, biopsy coordinates, and Artemis files.

The final dataset used in the following analyses consisted of 555 patients who met the above criteria. Their age, biopsy, and ROI information are provided below.
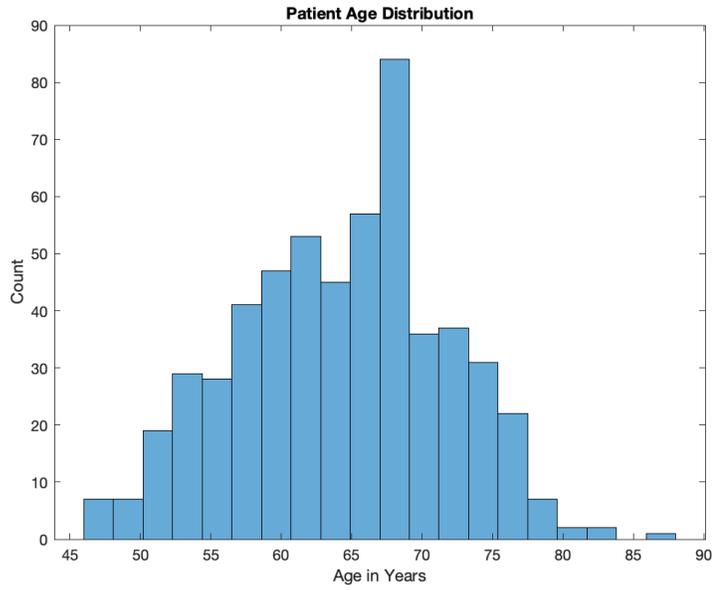
Figure 3: Patient Age Distribution – The age distribution of the patients in this study is roughly normally distributed with a mean of 64.27 years and a standard deviation of 7.36 years.
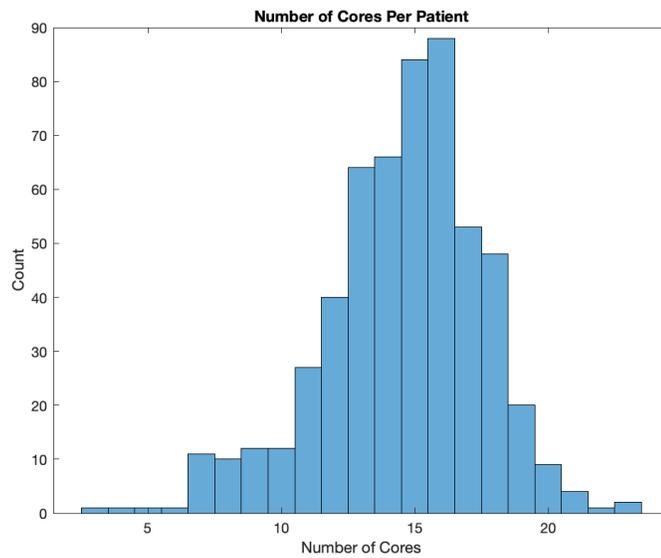
Figure 4: Distribution of Cores Per Patient - The distribution of the number of cores per patient is roughly normal with a mean of 14.56 and a standard deviation of 3.00.
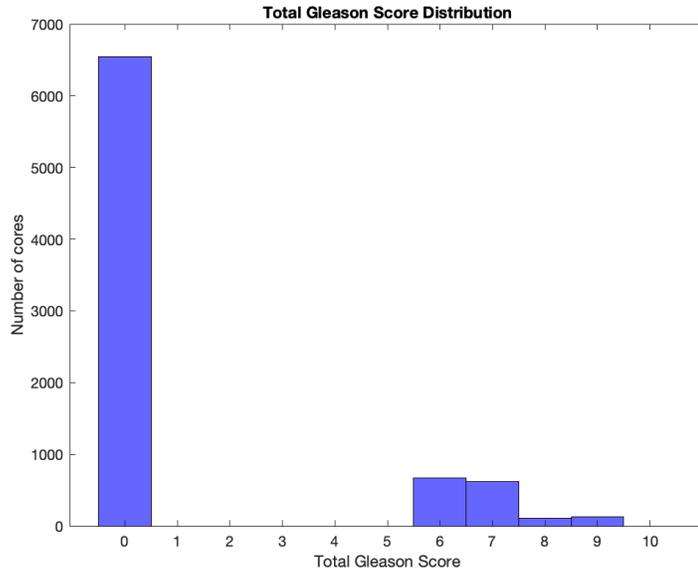
Figure 5: Gleason Score Distribution – The distribution of total Gleason score (primary + secondary score) for all the cores in the dataset is bimodally distributed. Every core was labeled as either 0+0, or 3+3 or higher.
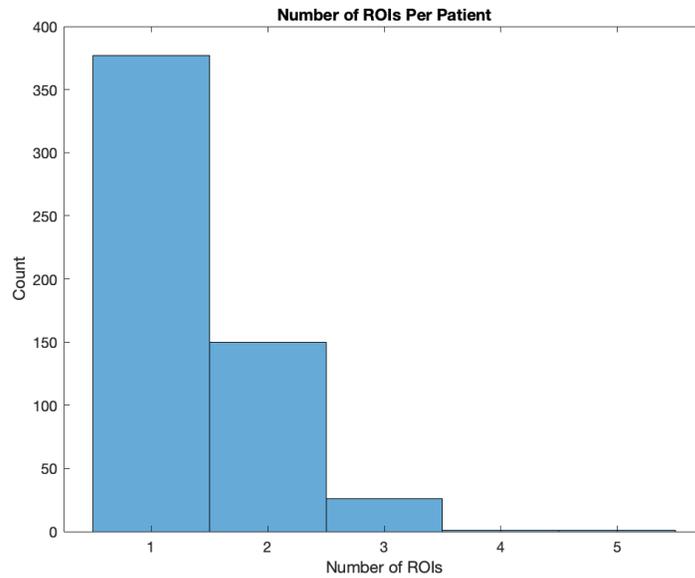
Figure 6: Distribution of the Number of ROIs per Patient – The distribution of the number of ROIs per patient is skewed right with a mean of 1.38 and a standard deviation of 0.60.

Number of Cores with Gleason ≥ 7 (csCaP)	858
Number of Cores with Gleason <7 (non-csCaP)	7223
Number of Cores with Gleason ≥ 6	1531
Number of Cores with Gleason < 6	6550
Total Number of Cores	8081

Table 1: Number of Cores in Different Gleason Score Categories – Since the distribution of Gleason scores is bimodal, with cores being either 0+0 or ≥ 3+3, but csCaP is defined as ≥ 3+4, we present the number of cores that are ≥ 3+3 as well as the number of cores that are denoted as csCaP.

§ 2.2 Generating 3-D Prostate Biopsy Visualizations

The generated dataset is a set of MR voxels corresponding to each biopsy core as well as an associated primary and secondary Gleason score for those voxels. We also obtain the radiologist drawn ROI and prostate contours registered to the same MRI coordinates. The biopsy is performed using an US probe along with an 18-gauge (0.838 mm inner diameter) needle.[18,19] The Artemis software is then able to transform the US coordinates of each biopsy core to MR coordinates. In the Artemis system, we start with patient files that include all of the biopsies' needle tip and needle base coordinates as well as all of the ROI and prostate contour coordinates in MR space. All of these coordinates are contained in an Extensible Markup Language (XML) file that is parsed using MATLAB and turned into a MATLAB-friendly data structure.
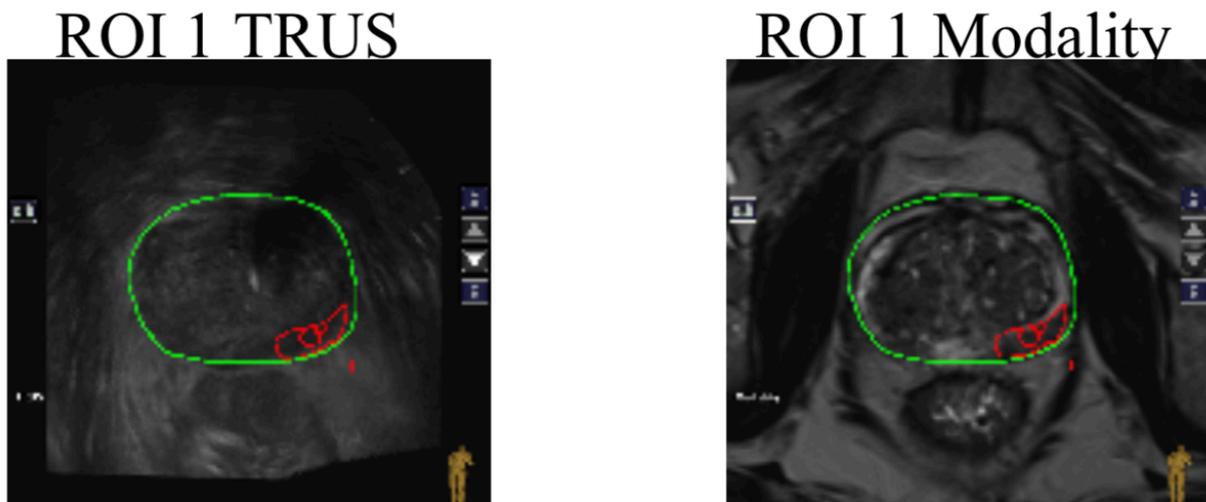
Figure 7: ROI Seen in US and MRI – The radiologist-drawn ROI (marked in red) in the T2 image (right) is superimposed using the Artemis software to the US image (left) seen by the urologist performing the biopsy. Biopsy cores can then be registered from US space to MR space using an inverse of this transformation.

12

After the MR coordinates of the biopsy cores as well as the prostate contour and ROIs for each patient are obtained, these coordinates, which are measured in millimeters along the x, y, and z planes, are then registered to each particular slice of the MR image. Each MRI in this dataset consists of 60 slices, each of size 256 by 256 pixels. After this final registration is performed, images like the figure below can be obtained for each patient.
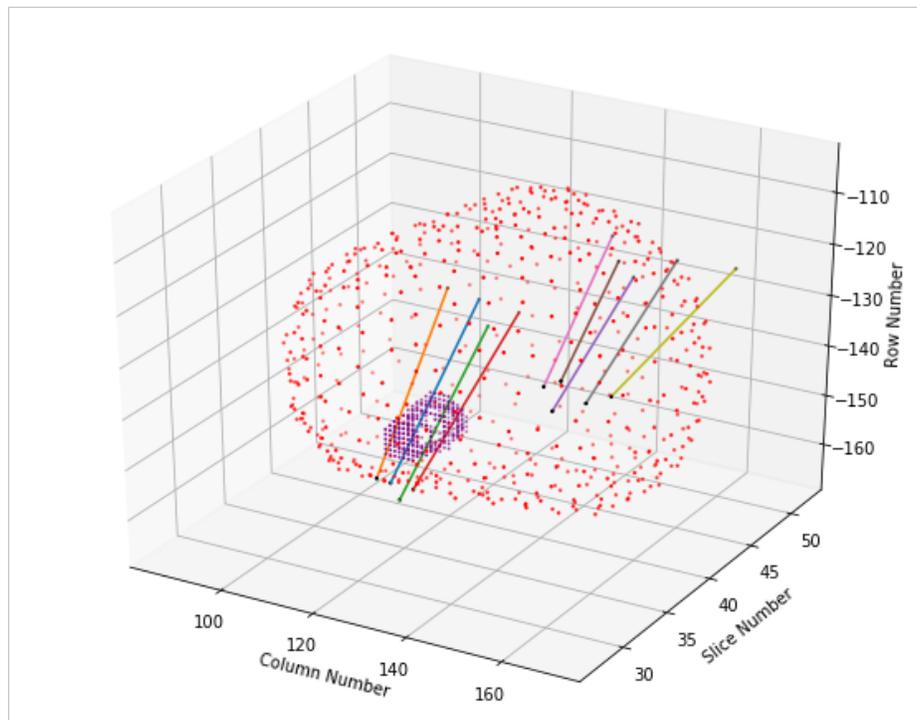
Figure 8: Prostate Contour, ROI, and Needle Coordinates for a Single Patient –The prostate contour (red dots), radiologist drawn ROI (purple dots), and biopsy cores (multicolored lines) are all plotted in the same coordinate space. The coordinates for each of these points directly correspond to the row, column, and slice number of the MRI for this patient.

Now that each biopsy core is registered to its corresponding rows, columns, and slice locations in the MRI, the pixels associated with that core can be obtained. The biopsy core consists of a needle tip and needle base coordinate as two separate points in 3D space. We connect these two points with 18 additional points between them that are linearly spaced in each of the 3 dimensions (using Numpy's linspace function) to generate 20 points for each biopsy core.[20] Though each individual biopsy coordinate is mapped to a single pixel, we extract a small region of pixels around that initial coordinate in order to account for needle registration error and to extract more meaningful intensity features from the MRI. Thus, the set of pixels used to represent a biopsy will be a rectangular prism of MR pixels.
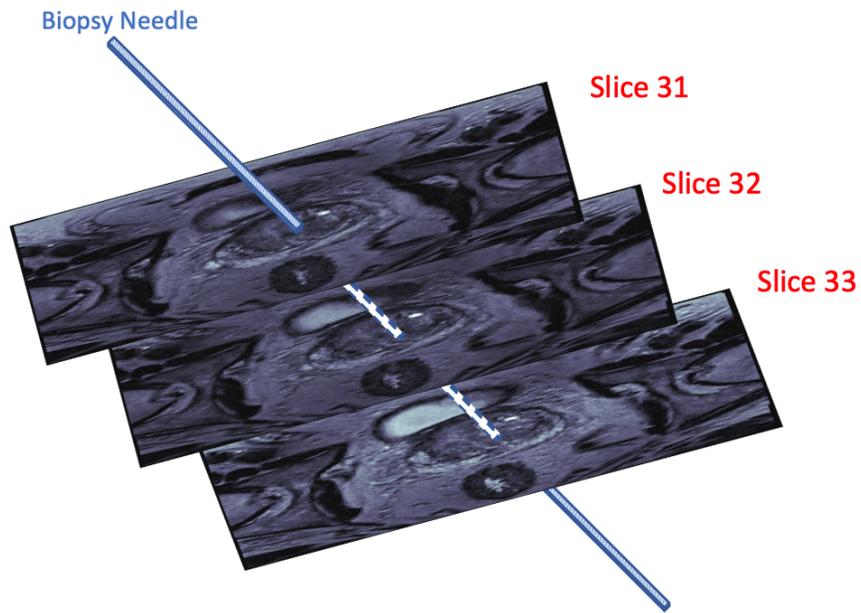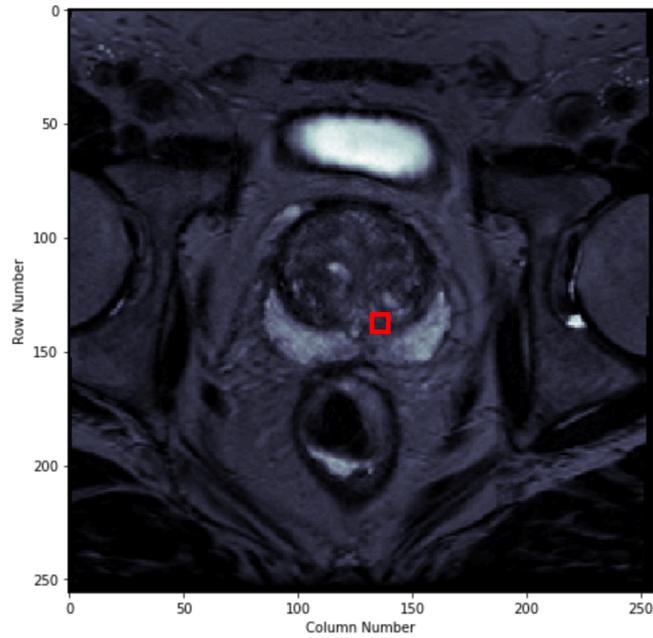
Figure 9: A Biopsy Core Visualized Going Through an MRI – The top image shows a single slice of an MRI and in red, the two-dimensional region of pixels extracted from this slice that is representative of part of a biopsy core. In the bottom image, we see a 3-D representation of the rectangular prism of pixels extracted from multiple slices of an MRI.

§ 2.3 Final HDF5 Dataset

The data for each of the patients is initially saved in an XML file that contains biopsy coordinates, as well as Digital Imaging and Communications in Medicine (DICOM) files that contain the T2 MRI, prostate contours, and ROI segmentations. In addition to the T2 images, additional DIOCM series for each patient study are pulled from the UCLA Picture Archiving Communication Systems (PACS). These series include Diffusion Weighted Imaging (DWI) in the form of the ADC (Apparent Diffusion Coefficient) and High b-val MR modalities, as well as Perfusion Weighted Imaging (PWI) in the form of the Ktrans modality.

Once these different series are pulled, all of the necessary biopsy information for each patient including biopsy coordinates, Gleason scores, prostate and ROI contours, and MR series are reformatted into Hierarchical Data Format (HDF5) files. These files allow for easy manipulation of the high dimensional data that we are dealing with. An explanation of the dimensionality of each of these patient-specific elements is provided below.

Element	Size
biopsy_grades	14 x 2
biopsy_masks	14 x 256 x 256 x 60
prostate_mask	256 x 256 x 60
roi_masks	1 x 256 x 256 x 60
T2	256 x 256 x 60
ADC	256 x 256 x 60
Ktrans	256 x 256 x 60
High b-val	256 x 256 x 60

Table 2: The Elements Associated with an Example Patient – This example patient had 14 biopsy cores, which can be seen by the (14 x 2) biopsy grades as well as the (14 x 256 x 256 x 60) biopsy_masks and one ROI which can be seen by the (1 x 256 x 256 x 60) roi_masks element. These masks are created to easily associate cores, ROIs, and contours with the different MR series.

Each of the masks described in the table above is an extremely sparse matrix, with one's wherever a biopsy, ROI, or prostate contour coordinate occurs (for the biopsy_masks, roi_masks, and prostate_mask respectively), and zeros everywhere else. Thus, all of these masks are kept in compressed HDF5 files to avoid using unnecessary memory. The biopsy_grades element corresponds to the primary and secondary Gleason score of each core, which accounts for its size of (number of cores x 2).

Lastly, the pixel intensities in each of the MR series are normalized using a z-score so that intensities across the series are all on the same relative scale. Every pixel intensity in each of the series is thus recalibrated according to the following equation:

$$p_{final} = \frac{p_{initial} - \mu}{\sigma}$$

where p_{final} = new pixel intensity, $p_{initial}$ = initial pixel intensity, μ = mean pixel intensity across the entire 256 x 256 x 60 series, and σ = standard deviation of pixels across the entire series. The mean and standard deviation of the pixel intensities for each series are stored as attributes of the HDF5 dataset in case they need to be denormalized later on.

CHAPTER 3

METHODS

§ 3.1 Core Distances from ROIs

In this section, we seek to answer two questions that will motivate the rest of the experiments in this work. The first question is how often are there cancer-containing cores found outside of ROIs (what we will call "MR-invisible" lesions). The second question is how far away from the ROIs are these cores found. Since the goal of this work is to improve detection of prostate cancer from MRI, it is worthwhile to investigate whether radiologists miss cancerous lesions in the MRI in the first place. Thus, it is important to rigorously define what it means to be "inside" vs "outside" an ROI as well as the "distance" of a core from a set of ROIs.

Each ROI is loaded into MATLAB as a 3D-triangulated surface with faces and vertices. The in_polyhedron function in MATLAB is used to detect whether any of the 20 points used to represent a biopsy core lie within the ROI surface.[21] This algorithm works using the following method:

For each point:

Shoot a random ray from that point in a random direction

    For each face:

$$Solve\ the\ equation \begin{bmatrix} -d \\ v_1 - v_0 \\ v_2 - v_0 \end{bmatrix} \begin{bmatrix} t \\ u \\ v \end{bmatrix} = o - v_0$$

    Where $u, v, w = 1 - u - v$, are barycentric coordinates, d is the ray direction, o is the ray origin, and t is the distance from the ray origin.

    The ray and triangle intersect if u, v, w, t are all positive

19

If the number of intersections is even, the query point is inside the ROI, if it's odd then the query point is outside the ROI

If the ray hits the edge of a face, repeat the process with a new ray

If any of the 20 points for the biopsy core lie within the ROI, then that core is said to lie within the ROI. If none of the 20 points lie within the ROI, then that core is said to lie outside the ROI. For cores outside of ROIs, we calculate the distance of a core to an ROI with the following equation:

$$\min \left[L2\_dist(R, C) \right]$$

where L2_dist is the L2 distance of every possible pair of points in two sets, R is the set of all points for all ROIs for that patient, and C is the set of points in that biopsy core

Using these two methods, we will present how many cancer-positive cores are found outside of ROIs as well as how far away these cores are from the ROIs. This will help determine the significance of the MR-invisible lesion problem.

§ 3.2 Generating MR-based Features of Biopsy Cores

      We analyze the rectangular prism of MR voxels representing each biopsy core to try to find features of these voxels that are different between cancer-positive and cancer-negative cores. The first step to this procedure is extracting this prism of voxels from the MRI along the angle of the biopsy. We find the biopsy angle using polar coordinates, and then sample a region of pixels around each coordinate. The chosen region size for these analyses is 5 pixels, creating an 11 x 11 pixel square for each biopsy coordinate.
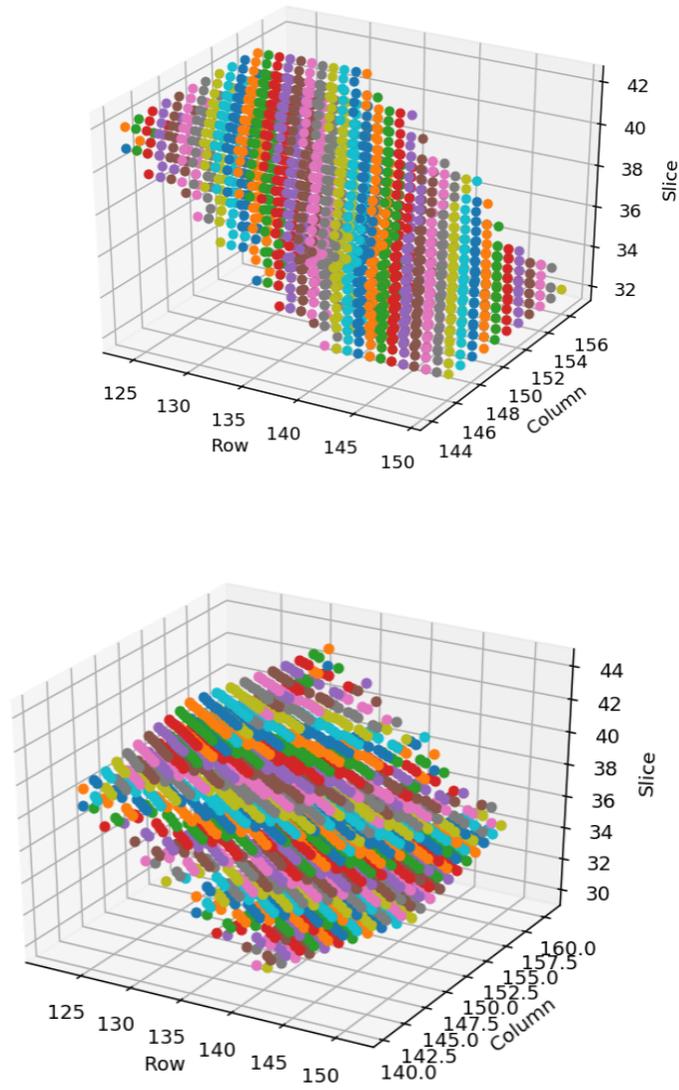
Figure 10: Selecting Voxels along the Angle of the Biopsy – The top image shows a voxel sampling using pixels adjacent to the initial coordinate but along the plane of the MR slices. The bottom image shows a sampling when voxels are chosen from the planes orthogonal to the biopsy angle.

By choosing voxels along the angle of the biopsy, we are sampling intensity levels that are more closely related to and thus, more representative of, that biopsy core. The biopsy angle is calculated in polar coordinates using the following method[22]:

$$\gamma = \sin^{-1}\left(\frac{\Delta d * s_1}{\sqrt{(\Delta d * s_1)^2 + (\Delta r * s_2)^2 + (\Delta c * s_2)^2}}\right) - \frac{\pi}{2}$$

$$\theta = \tan^{-1}\left(\frac{\Delta c * s_2}{\Delta r * s_2}\right) - \frac{\pi}{2}$$

where γ and θ are the two polar coordinate angles in radians, Δd, Δr, and Δc are the (final − initial) depth, row, and column coordinates of the biopsy, s_1 is the MRI slice thickness (1.5 mm), and s_2 is the MRI pixel spacing (0.664 mm)

Once the angle of the biopsy is calculated, each of the biopsy coordinates will be transformed into a square patch orthogonal to this angle. Thus, for each coordinate, calculating the intensity value of its $patch[i, j]$ will involve two steps:

1. Finding a new row, column, and depth value for that patch[i,j] to extract an MR intensity value from based on the biopsy angle and the radius of the region we are looking at

2. Performing a trilinear interpolation of that new row, column, and depth value in the case that these new values fall in between voxels of the MRI (i.e. they are decimals and not whole numbers)

These two steps are performed using the following method:

$$For\ i \in [0, 2 * radius]\ and\ For\ i \in [0, 2 * radius]$$

$$row_f = row_i + \frac{(j - radius)\cos(\theta) + (i - radius)\cos(\gamma)\cos(\theta)}{s_2}$$

$$col_f = col_i + \frac{(j - radius)\sin(\theta) + (i - radius)\cos(\gamma)\sin(\theta)}{s_2}$$

$$dep_f = dep_i + \frac{(i - radius)\sin(\gamma)}{s_2}$$

$$patch[i, j] = trilinear\_interpolation(row_f, col_f, dep_f)$$
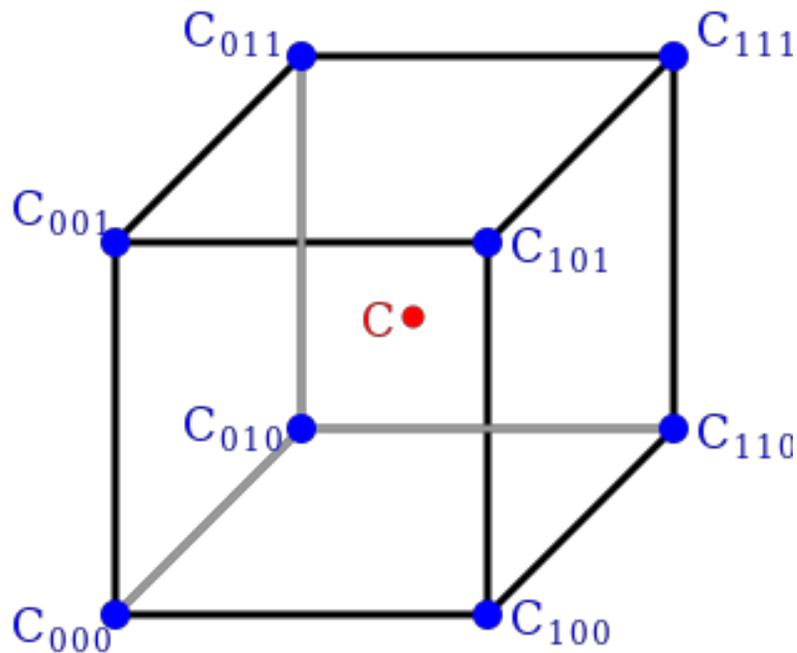
Figure 11: Performing Trilinear Interpolation For Coordinates Between Voxels – Using the

change in orientation formulas listed above, if a particular coordinate is mapped to a voxel that

lies between an integer row, column, or slice, (i.e. one of its x,y, or z coordinates is a decimal)

then the voxel intensity value for that coordinate is determined by the weighted average of the

intensities of the 8 surrounding voxels. The weights come from the distance of each of these

surrounding voxels to the coordinate to be interpolated. This diagram provides a visualization of

this procedure. The coordinate to be interpolated is the red C, and the 8 surrounding coordinates are represented by the blue C's.

For some cores, when coordinates are transformed from the millimeter-based MR coordinates to the actual row, column, and slice of the MR images, two coordinates can be mapped to the same voxel. Thus, some biopsies that may have started out with 20 separate coordinates, wind up with fewer than 20 coordinates after this transformation is applied. However, because we compute features from the prism that are independent of dimensionality, no mean padding is necessary before inputting into the model. Next, the pixel intensity values of each biopsy are denormalized based on the mean and standard deviation of the MR image from which they came and are renormalized onto a 0 to 256 grayscale.

Once the rectangular prism for each core is determined, we split the core into two halves of equal length and extract two types of features from each half – Haralick textures and statistical features. We perform this split because it is undetermined where in the core the cancer may be, so we compute features for each half and then average the features between the halves to get a resulting feature vector for the full core. This increases the relative importance of MR features from csCaP tissue in the overall feature vector even if the cancer is located in just a small portion of the core.

Haralick textures are imaging features derived from the gray level co-occurrence matrix of an image – a matrix that is computed by observing identical pixel values at a certain offset.[23] We calculate the Haralick textures for each core using the mahotas package in Python.[24] This package calculates the following 13 features in four directions. The value for each feature is taken as the mean of the values across the four directions. These textural features are:

1. Angular Second Moment

2. Contrast

3. Correlation

4. Sum of Squares: Variance

5. Inverse Difference Moment

6. Sum Average

7. Sum Variance

8. Sum Entropy

9. Entropy

10. Difference Variance

11. Difference Entropy

12. Information Measure of Correlation 1

13. Information Measure of Correlation 2

For each of the core halves, we also calculate statistical features, and again, take the average of the respective features between the two halves to represent the final feature for the core. The following statistical features are calculated:

1. The 5th, 20th, 50th, 80th, and 95th percentile of all the intensity values in the core (5 features)

2. The mean intensity of each bin when the intensities in the prism are sorted into 16 equally spaced bins (16 features)

3. Variance of the intensities

4. Mean of the intensities

It is important to note that when a core is mapped to its final rectangular prism of voxels, we extract the intensity levels from each of the four mpMRI modalities – T2, ADC, Ktrans, and High b-value - using the same calculated row, column, and slice coordinates. The biopsy coordinates are technically only registered to the T2 modality; thus, an assumption is made that corresponding row, column, and slice values between modalities are identical. Upon visual inspection of the four modalities for each patient, this largely seems to be true, with perhaps a very minor registration error.
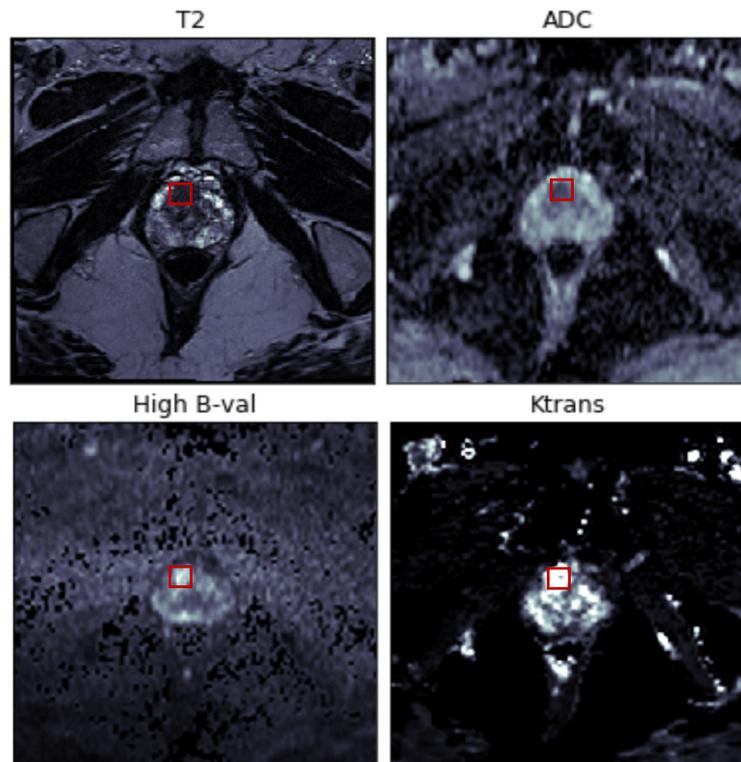
Figure 12: Visualization of the Four mpMRI Modalities – In this image we see the four mpMRI modalities from which biopsy voxels are extracted – T2, ADC, High b-value, and Ktrans. A lesion from which a csCaP core was extracted from is shown using the red box. Registration differences between the modalities is seen to be quite minor.

Each core is thus associated with 13 Haralick texture features and 23 statistical features for each of the four mpMRI modalities. This totals to 144 features per biopsy core. These features are subsequently used to train three machine learning classifiers in order to learn differences between cancer-positive and cancer-negative biopsies.

§ 3.3 Machine Learning Techniques and Evaluation Metrics

In this last section, we train and test two machine learning based classifiers – an SVM and a logistic regression model on the binary classification task of distinguishing between biopsy cores that are cancer-positive (Class 1) and biopsy cores that are cancer-negative (Class 0) based on their respective Haralick texture and statistical features. Two different definitions of cancer-positive are applied, and results are presented for each definition. The first definition involves any core having a Gleason score $\geq$ 3+3 being labeled as cancer-positive, and the second is to call a core cancer-positive only if the Gleason score is $\geq$ 3+4. These two definitions are commonly used in the literature for tasks that binarize cancer/no cancer.

The dataset of 555 patients is split 70% into a training set (388 patients) and 30% into an independent test set (167 patients). This totaled to 5642 cores in the training set and 2439 cores in the test set. For the Gleason $\geq$ 3+3 classifiers, this corresponded to 18.1% of cores in the training set being labeled as cancer-positive and 21.0% of cores in the test as being labeled positive. For the Gleason $\geq$ 3+4 classifiers, this corresponded with 9.9% of cores in the training set being labeled cancer-positive, and 12.4% of cores in the test set being labeled cancer-positive. The features of the training set were z-score normalized using their respective means and standard deviations and the same transformation was applied (using the same means and standard deviations from the training set) to the test set before evaluation. Each of the three classifiers were trained using 5-fold cross validation on the training set and evaluated on the test set. Each classifier was also trained with "balanced" class weights to correct for the imbalance in the size of the two classes. This means that mistakes in classifying the less frequent class, the cancer-positive class, will be penalized *P* times more than mistakes in classifying the more frequent class, the cancer-negative class. P is defined as:

$$P = \frac{(Number\ of\ Examples\ in\ More\ Frequent\ Class)}{(Number\ of\ Examples\ in\ Less\ Frequent\ Class)}$$

The classifiers will then be evaluated using the following metrics:

- Accuracy

- Precision

- Recall

- F1-score

- Area Under the Curve (AUC)

- Precision-Recall Curve

CHAPTER 4

RESULTS

§ 4.1 Detection of MR Invisible Cores

Using the spatial locations of the biopsy cores and ROIs, we first look at the number of cancer-positive cores outside ROIs (PO), the number of cancer-positive cores inside ROIs (PI), the number of cancer-negative cores outside of ROIs (NO), and the number of cancer-negative cores inside ROIs (NI). For each of these analyses, we use two different definitions of "cancer-positive." First, we define cancer-positive to be cores with $\geq$ 3+3 Gleason score, then we define cancer-positive to be cores with $\geq$ 3+4 Gleason score. For the 555 patients used in this study, using both of these criteria, the results are as follows.

Positive Outside	923
Positive Inside	608
Negative Outside	5,420
Negative Inside	1,130

Proportion of Cores Inside ROI that are Positive	34.98%
Proportion of Cores Outside ROI that are Positive	14.55%

Proportion of Positive Cores found Inside ROI	39.71%
Proportion of Positive Cores found Outside ROIs	60.28%

Number of Patients with no Positive Targeted Biopsies	275
Number of Patients with no Positive Targeted Biopsies, but with at least one Positive Systematic Biopsy	51
Total Number of Patients	555

Table 3: Cores Inside and Outside ROIs (≥ 3+3 Criteria) – When positive cores are defined as cores with Gleason ≥ 3+3, the above statistics are obtained for PO, PI, NO, and NI cores as well as the number of patients with no positive targeted biopsies but with at least one positive systematic biopsy.

Positive Outside	426
Positive Inside	432
Negative Outside	5,917
Negative Inside	1,306

Proportion of Cores Inside ROI that are Positive	24.86%
Proportion of Cores Outside ROI that are Positive	6.71%

Proportion of Positive cores found inside ROIs	50.35%
Proportion of Positive cores found Outside ROIs	49.65%

Number of Patients with no Positive Targeted Biopsies	349
Number of Patients with no Positive Targeted Biopsies, but with at least one Positive Systematic Biopsy	33
Total Number of Patients	555

Table 4: Cores Inside and Outside of ROIs (≥ 3+4 Criteria) – When positive cores are

defined as being cores with Gleason ≥ 3+4, the above statistics are obtained for PO, PI, NO, and

NI cores as well as the number of patients with no positive targeted biopsies but at least one
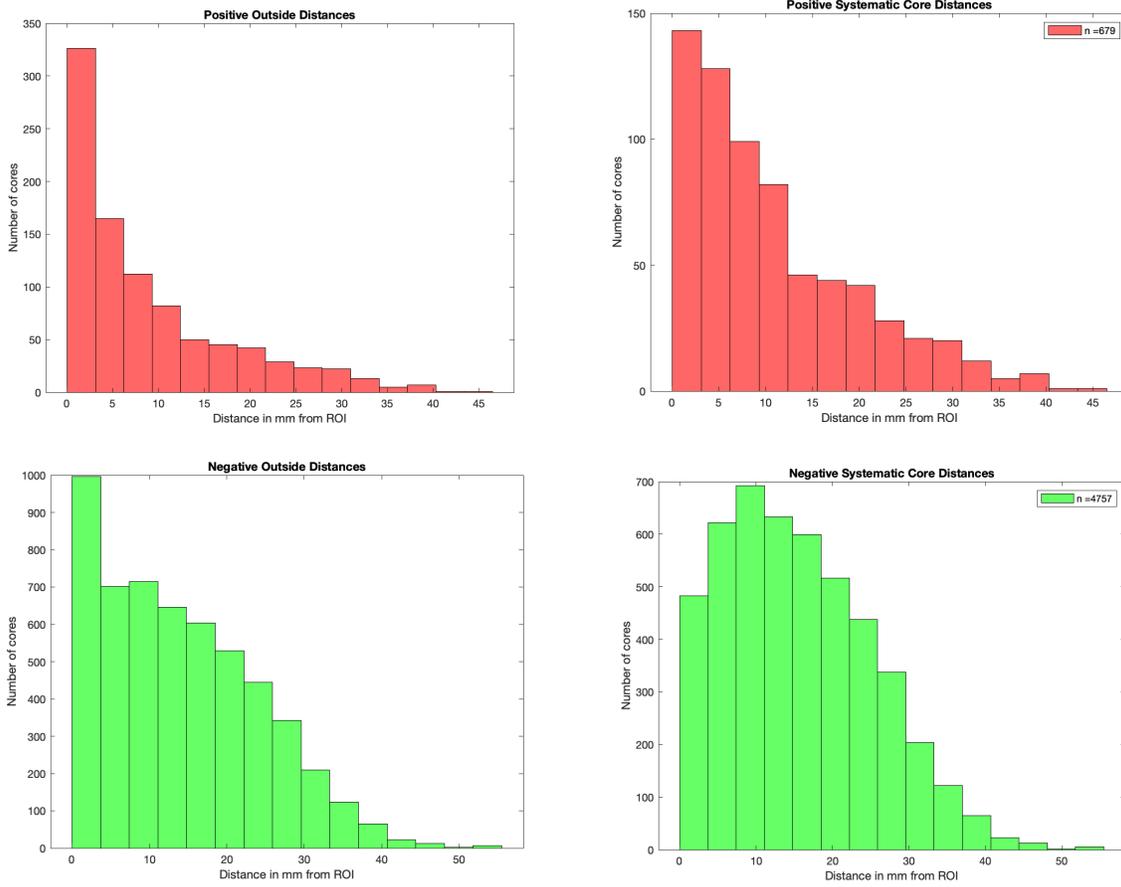
systematic biopsy.

Figure 13: PO and NO Core Distances from ROIs (≥ 3+3) Criteria – When Gleason ≥ 3+3 is used as the criteria for determining positive cores, the above plots are obtained for the number of cores observed at varying distances from the ROI.
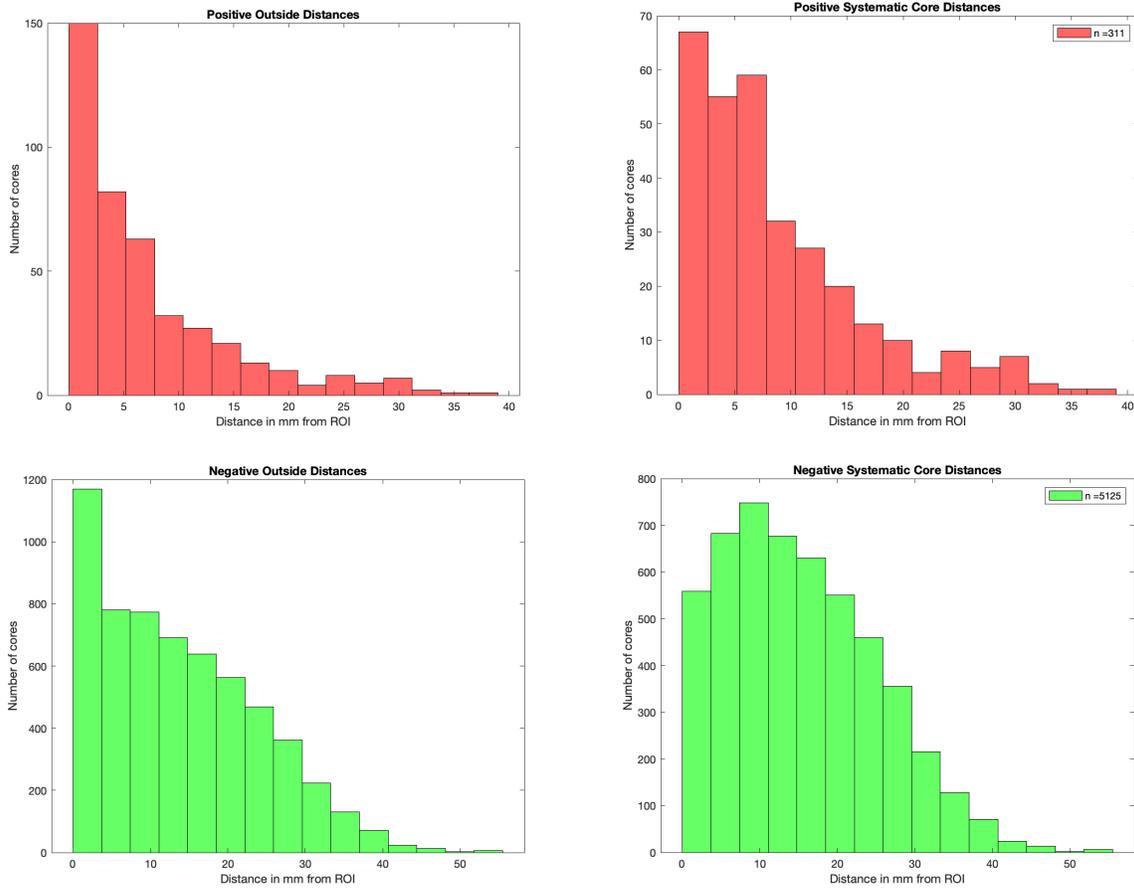
Figure 14: PO and NO Core Distances from ROIs (≥ 3+4) Criteria – When Gleason ≥ 3+4 is used as the criteria for determining positive cores, the above plots are obtained for the number of cores observed at varying distances from the ROI.

§ 4.2 Feature Analysis of Cancer-Positive and Cancer-Negative Biopsies

In this section, we analyze the features that are statistically significantly different between cores with csCaP and cores without csCaP. Thus, Gleason $\geq$ 3+4 is used as the criteria for positivity in this section.
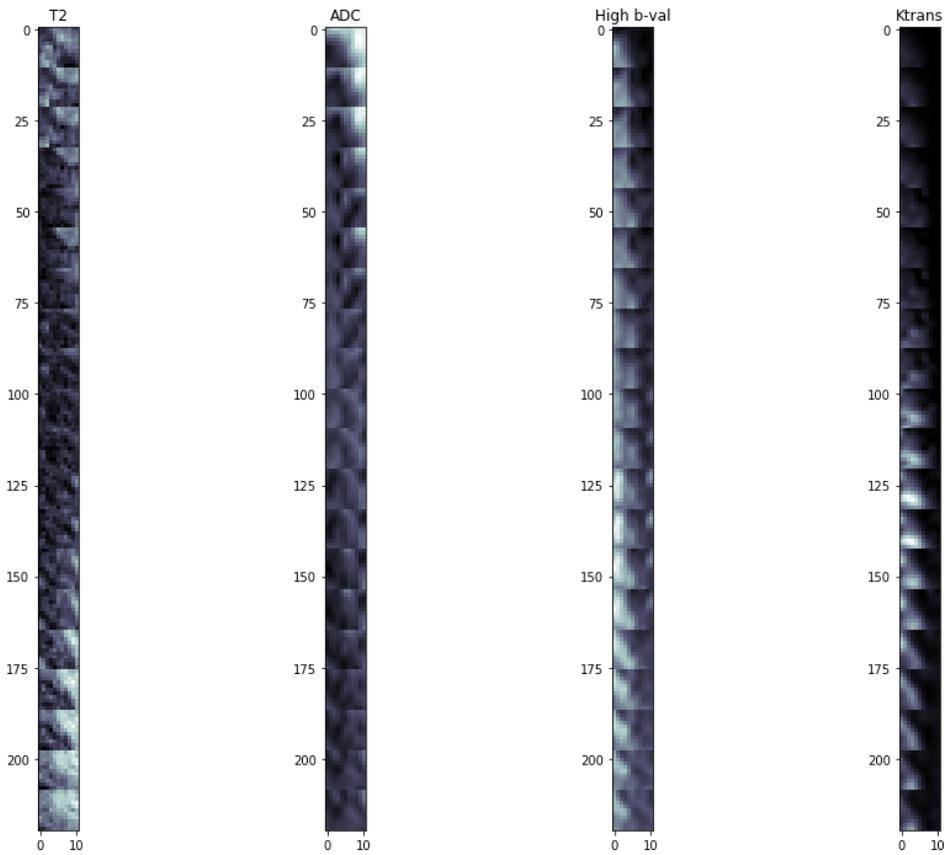
Positive Core
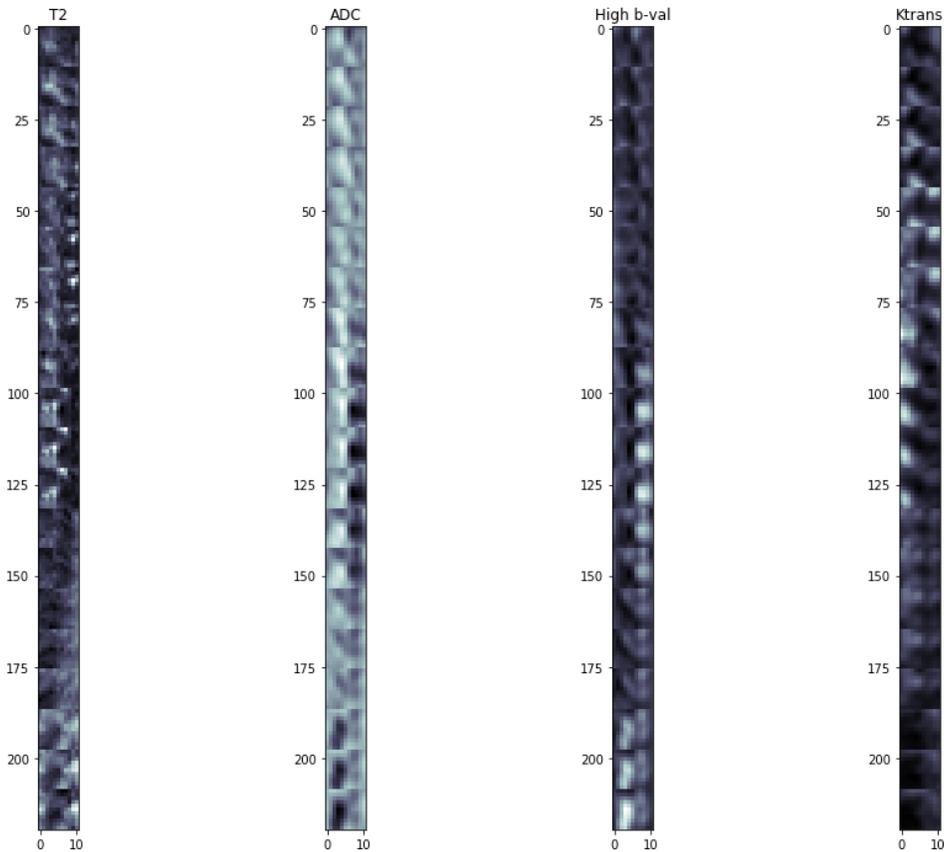
Figure 15: Unrolled Biopsy Vectors in Each Modality – A single positive and negative core seen in the 4 different MR modalities. Each core is normally an 11x11x20 matrix, but for the purpose of this plot, each of the 20 11x11 squares are stacked after one another so that the core can be viewed in 2D. Slightly different features can be seen between the corresponding modalities of the cancer-positive and cancer-negative cores including relative hypointensity and hyperintensity.

Feature Name	P-value	Positive Class Mean	Negative Class Mean	Positive Class Std	Negative Class Std
ADC 20th percentile	2.24E-27	70.077	78.498	16.618	15.795
ADC 6th Haralick (Sum Average)	1.68E-25	169.318	184.482	29.544	31.852
ADC Mean	2.43E-25	84.741	92.171	15.665	14.576
ADC 50th percentile	1.11E-24	84.665	92.476	16.82	15.326
ADC 5th percentile	1.24E-22	57.421	65.417	16.392	17.183
T2 5th Haralick (Inverse Difference Moment)	8.66E-21	0.1	0.0862	0.0318	0.0262
T2 10th Haralick (Difference Variance)	6.63E-20	0.000342	0.000253	0.000246	0.000171
T2 Histogram Bin 4	1.71E-19	32.402	37.954	10.576	12.654
T2 Histogram Bin 5	3.09E-19	40.069	46.708	12.815	15.089
T2 Histogram Bin 3	3.82E-19	24.835	29.296	8.589	10.422

Table 5: Features with Greatest Statistical Significance – The top 10 features with the lowest p-value between the cancer-positive and cancer-negative class. These p-values are generated by selecting a random sample of 858 non-csCaP cores (the same number as the number of csCaP cores) and performing a Wilcoxon signed-rank test.[25]

§ 4.3 Predictive Power of Machine Learning Models

      SVM and logistic regression classifiers are trained, using Haralick textures and statistical features as input, to predict whether a biopsy will be cancer-positive or cancer-negative. Results are presented for both definitions of cancer positivity. First, we show the results of the two models when Gleason $\geq$ 3+3 is used as the positivity criteria. Then we show the results of the two models when Gleason $\geq$ 3+4 is used as the positivity criteria. These models were trained, and plots were generated using Python's Scikit-learn library.[26]

Gleason ≥ 3+3 Results

SVM
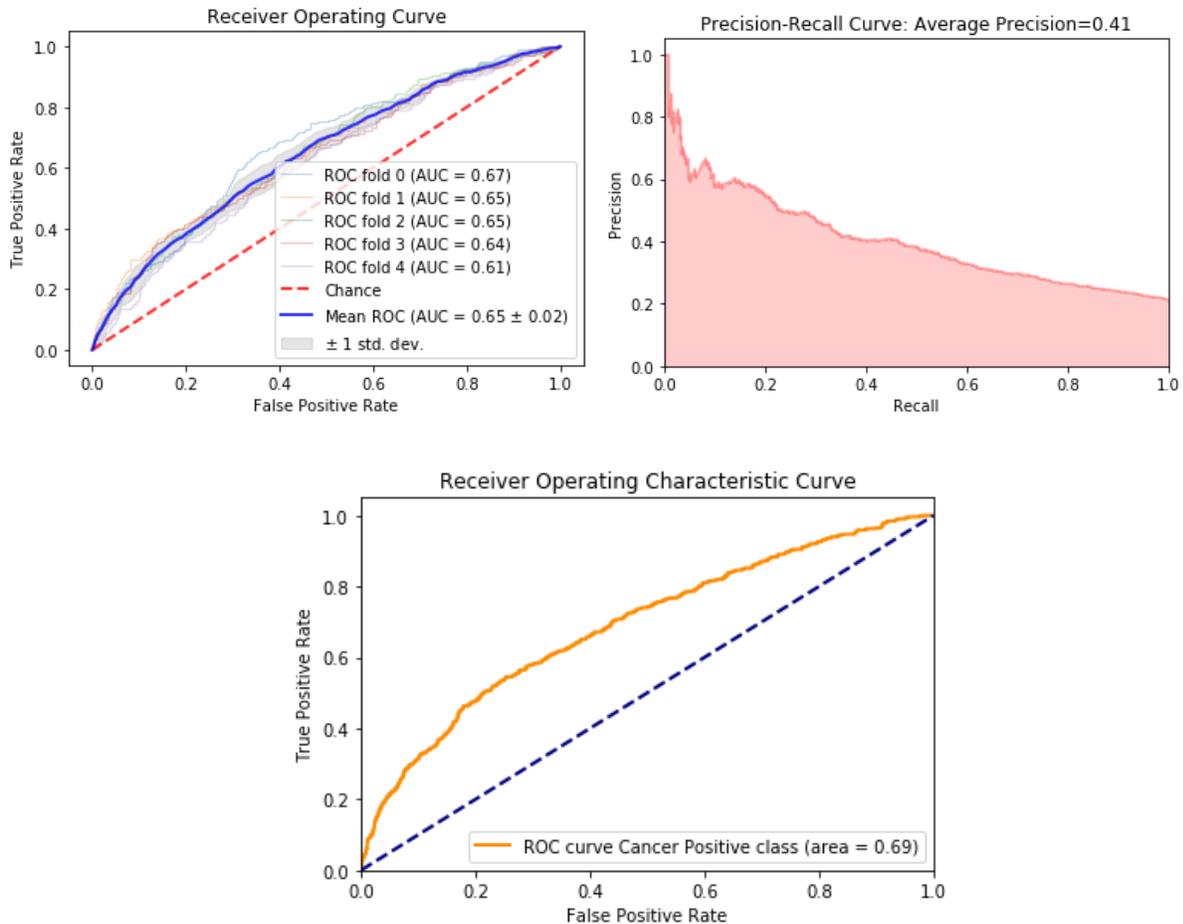
	precision	recall	f1-score	support
Cancer Negative	0.86	0.73	0.79	1927
Cancer Positive	0.35	0.55	0.43	512

Test Accuracy: 69.04%

Figure 16: SVM Results for Predicting Gleason ≥ 3+3 – The 5-fold cross validation ROC,

precision recall curve, test ROC, precision and recall statistics, and test accuracy for the SVM

trained with Gleason ≥ 3+3 as the positive class.

	precision	recall	f1-score	support
Cancer Negative	0.85	0.63	0.73	1927
Cancer Positive	0.30	0.59	0.40	512

Test Accuracy: 62.48%

Figure 17: Logistic Regression Results for Predicting Gleason ≥ 3+3 – The 5-fold cross

validation ROC, precision recall curve, test ROC, precision and recall statistics, and test

accuracy for the logistic regression model trained with Gleason ≥ 3+3 as the positive class.

Gleason ≥ 3+4 Results

SVM

	precision	recall	f1-score	support
Cancer Negative	0.93	0.79	0.85	2137
Cancer Positive	0.27	0.55	0.36	302

Test Accuracy: 75.72%

Figure 18: SVM Results for Predicting Gleason ≥ 3+4 – The 5-fold cross validation ROC,

precision recall curve, test ROC, precision and recall statistics, and test accuracy for the SVM

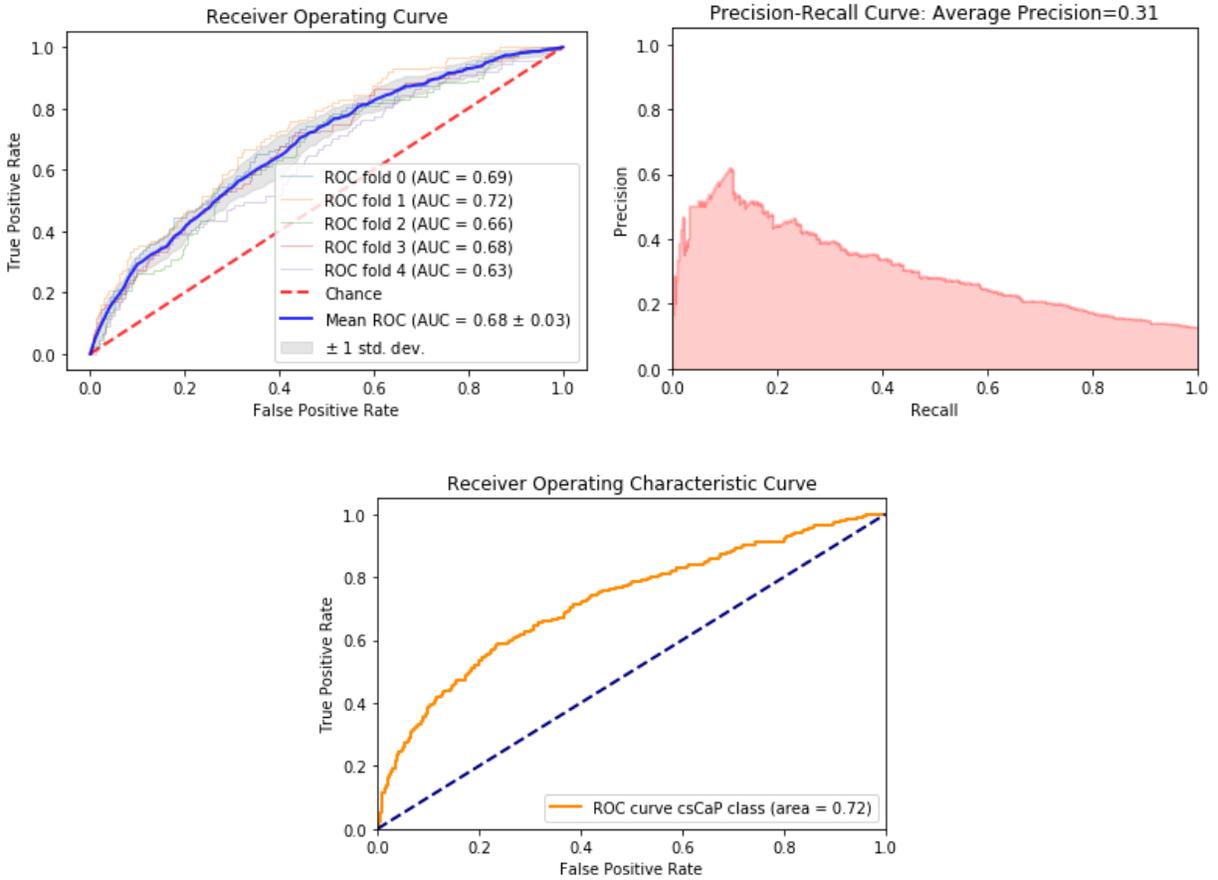trained with Gleason ≥ 3+4 as the positive class.

Logistic Regression
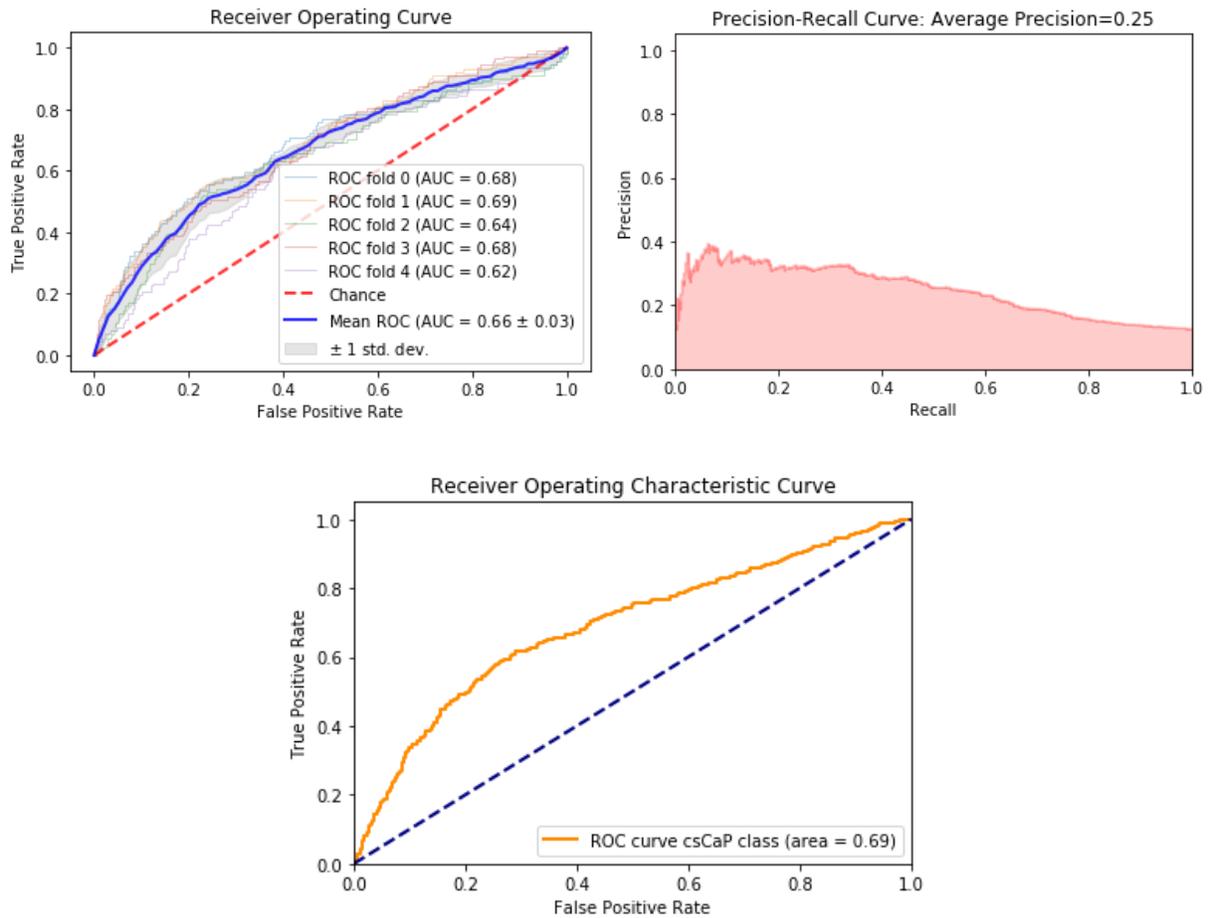
	precision	recall	f1-score	support
Cancer Negative	0.93	0.69	0.79	2137
Cancer Positive	0.22	0.62	0.33	302

Test Accuracy: 68.47%

Figure 19: Logistic Regression Results for Predicting Gleason ≥ 3+4 – The 5-fold cross

validation ROC, precision recall curve, test ROC, precision and recall statistics, and test

accuracy for the SVM trained with Gleason ≥ 3+4 as the positive class.

CHAPTER 5

DISCUSSION

In the first section of our results, we find that regardless of the criteria for "cancer-positivity", whether it be Gleason $\geq$ 3+3 or Gleason $\geq$ 3+4, the proportion of cores outside ROIs that are positive is much lower than the proportion of cores inside ROIs that are positive. This supports the claim that radiologist-drawn ROIs are indicative of underlying cancerous pathology and certainly gives confidence to ROI delineation. However, when the proportion of positive cores inside ROIs is compared to the proportion of positive cores outside of ROIs, we find that those two numbers are about the same. In the Gleason $\geq$ 3+3 criteria scheme, about 60% of positive cores were found outside of ROIs, and in the Gleason $\geq$ 3+4 criteria scheme, about 50% of positive cores were found outside of ROIs. This statistic implies that, though there is a higher chance of finding positive cores within ROIs, there is a significant proportion of positive cores that fall outside of ROIs, and this could be a potential area of improvement for the delineation of these ROIs.

Since so many positive cores were found outside of ROIs, we next looked at how far away from the ROIs these positive cores were. In both criteria schemes, the number of positive cores seems to fall off dramatically as distance from the ROI is increased. This lends credence to the idea that even if cores do fall outside of ROIs, most of them fall fairly close to these ROIs. When the same type of histogram is observed for negative cores that fall outside of ROIs, we see this same pattern of the number of cores dropping off as the distance from the ROI is increased. If it is true that the ROI is a ground truth hub of cancer from which the lesion originates as the PO plot suggests, we would expect to see the number of negative cores increase with distance from the ROI instead of decrease. However, because there is sampling bias in favor of regions

44

close to the biopsy (since the dataset involves targeted biopsy) this result is not surprising. To combat this sampling bias, we then plot the distance of systematic cores that are cancer-negative from ROIs. Systematic cores are taken uniformly from regions of the prostate, so we would not expect to see the same sampling bias of more cores taken close to ROIs. From this plot, we see that the number of negative systematic cores does increase with distance from the ROI, then trails off as that distance gets much larger (presumably because fewer cores are taken from very distant areas of the prostate). However, the fact that 51 out of 555 patients had positive systematic biopsies and no positive targeted biopsies using the $\geq 3+3$ criteria is indicative of the fact that it is not uncommon to find cancer outside of targets, and that systematic biopsy is still necessary for a reason. Altogether, these statistics indicate areas for improvement in ROI delineation.

Next, we go on to examine the MR features of csCaP cores vs. non csCaP cores. To do this, we use the Wilcoxon rank test (since these features are not normally distributed within the classes) to obtain p-values for each feature between the different classes. We present the top 10 most statistically significant features, and all of them have p-values far below an alpha of 0.05. These p-values are even far below an alpha of 0.05/144, applying the Bonferroni correction for multiple comparisons.[27] Interestingly, we see that the top 5 features all come from the ADC modality, while the next 5 features come from the T2 modality. High b-val and Ktrans features did not make it into the top ten most statistically significant features. The low p-values seen between these features established confidence in the hypothesis that machine learning based classifiers could be trained to differentiate between the two classes.

When the SVM and logistic regression models were trained and tested using the two different criteria for cancer-positivity, results seemed fairly similar across the board. The models

all had relatively steady AUC's during the 5-fold cross validation training and relatively similar test AUCs as well. Notably, both the SVM and the logistic regression models had higher precision on the cancer-positive class when the criteria for cancer-positive was lower, i.e. Gleasons 3+3 were considered as cancer-positive. In terms of overall test accuracy, and test AUC, the SVM trained to distinguish between csCaP (Gleasons $\geq$ 3+4) and non csCaP (Gleasons < 3+4) performed the best out of all four classifiers. It had a high negative predictive value of 0.93 and a fairly strong test AUC of 0.72. However, the PPV of all the classifiers (the same as the precision on the cancer-positive class) was still exceedingly low, none reaching more than 0.35. This metric, in the context of the problems associated with prostate cancer, is the most valuable. Since prostate cancer tends to get overdiagnosed, we aim to build models that have high PPV. In other words, of all the cores we predict to be cancer-positive with the model, a large majority of them should ideally be cancer-positive in the ground truth. If this metric is not high for a given model, then the model fails to address the clinical problem of overdiagnosis. Thus, these models provide a good baseline for building classifiers that can achieve results with higher PPV.

CHAPTER 6

CONCLUSIONS AND FUTURE DIRECTIONS

In this work, we first determine that there is room for improvement in the delineation of ROIs on prostate MRI. Cancerous biopsy cores are often found outside of ROIs, and thus, the delineation of ROIs can benefit from using the pathology results of biopsies as ground truth. To enable this improvement, we propose a unique methodology of mapping biopsy cores to their location in the MRI, extracting features of voxels in these biopsy regions, and trying to correlate these features with the ground truth pathology of these biopsies. The classifiers we built had high negative predictive values and moderately high AUCs, but rather low positive predictive values. Since one of the biggest problems in the arena of prostate cancer is overdiagnosis, attempting to raise the low PPV of our classifiers will motivate future work in this topic.

Some of this future work could entail trying deep learning methods such as Convolutional Neural Networks on the raw voxel intensity levels extracted from each MR modality to have a classifier learn its own features that are representative of cancer-positive and cancer-negative cores.[28] In addition to this, the method by which we deal with the fact that we do not know where prostate cancer actually is in the core could be examined more closely. It may be useful to split the core up into smaller parts rather than just halves and use features such as minimums and maximums at these regions to boost the classifier results. In this vein, a useful analysis that could be done would be trying to correlate pathology features of biopsy cores with their MR features, to directly determine the type of MR features that denote prostate cancer. Lastly, additional types of data including time series perfusion data could be incorporated as additional features to this model.

47

REFERENCES

1. "Key Statistics for Prostate Cancer | Prostate Cancer Facts." n.d. Accessed Jan 7, 2019. https://www.cancer.org/cancer/prostate-cancer/about/key-statistics.html.

2. Schmid, H. P., J. E. McNeal, and T. A. Stamey. 1993. "Observations on the Doubling Time of Prostate Cancer. The Use of Serial Prostate-Specific Antigen in Patients with Untreated Disease as a Measure of Increasing Cancer Volume." *Cancer* 71 (6): 2031–40.

3. Stanford, J. L., Z. Feng, A. S. Hamilton, F. D. Gilliland, R. A. Stephenson, J. W. Eley, P. C. Albertsen, L. C. Harlan, and A. L. Potosky. 2000. "Urinary and Sexual Function after Radical Prostatectomy for Clinically Localized Prostate Cancer: The Prostate Cancer Outcomes Study." *JAMA: The Journal of the American Medical Association* 283 (3): 354–60.

4. Roth, Andrew J., Mark I. Weinberger, and Christian J. Nelson. 2008. "Prostate Cancer: Psychosocial Implications and Management." *Future Oncology* 4 (4): 561–68.

5. Heijnsdijk, E. A. M., A. der Kinderen, E. M. Wever, G. Draisma, M. J. Roobol, and H. J. de Koning. 2009. "Overdetection, Overtreatment and Costs in Prostate-Specific Antigen Screening for Prostate Cancer." *British Journal of Cancer* 101 (11): 1833–38.

6. Carroll, Peter R., J. Kellogg Parsons, Gerald Andriole, Robert R. Bahnson, Erik P. Castle, William J. Catalona, Douglas M. Dahl, et al. 2016. "NCCN Guidelines Insights: Prostate Cancer Early Detection, Version 2.2016." *Journal of the National Comprehensive Cancer Network: JNCCN* 14 (5): 509–19.

7. Ghai, Sangeet, and Masoom A. Haider. 2015. "Multiparametric-MRI in Diagnosis of Prostate Cancer." *Indian Journal of Urology: IJU: Journal of the Urological Society of India* 31 (3): 194–201.

8. Preston, David. 2006. "MRI Basics." n.d. Accessed Jan 10, 2019. http://casemed.case.edu/clerkships/neurology/web%20neurorad/mri%20basics.htm.

9. Weinreb, Jeffrey C., Jelle O. Barentsz, Peter L. Choyke, Francois Cornud, Masoom A. Haider, Katarzyna J. Macura, Daniel Margolis, et al. 2016. "PI-RADS Prostate Imaging-- Reporting and Data System: 2015, Version 2." *European Urology* 69 (1): 16–40.

10. Chen, Ni, and Qiao Zhou. 2016. "The Evolving Gleason Grading System." *Chinese Journal of Cancer Research = Chung-Kuo Yen Cheng Yen Chiu* 28 (1): 58–64.

11. Urology, Hopkins. n.d. "PROSTATE CANCER UPDATE -The James Buchanan Brady Urological Institute - Urology at Hopkins." Accessed March 7, 2019. http://urology.jhu.edu/newsletter/prostate_cancer69.php.

12. Scattoni, Vincenzo. 2006. "Systematic Prostate Biopsies Are More and More Often Becoming Saturation Biopsies." *European Urology* 50 (2): 202–4.

13. Marks, Leonard, Shelena Young, and Shyam Natarajan. 2013. "MRI-Ultrasound Fusion for Guidance of Targeted Prostate Biopsy." *Current Opinion in Urology* 23 (1): 43–50.

14. Susil, Robert C., Cynthia Ménard, Axel Krieger, Jonathan A. Coleman, Kevin Camphausen, Peter Choyke, Gabor Fichtinger, Louis L. Whitcomb, C. Norman Coleman, and Ergin Atalar. 2006. "Transrectal Prostate Biopsy and Fiducial Marker Placement in a Standard 1.5T Magnetic Resonance Imaging Scanner." *The Journal of Urology* 175 (1): 113–20.

15. Siddiqui, M. Minhaj, Soroush Rais-Bahrami, Baris Turkbey, Arvin K. George, Jason Rothwax, Nabeel Shakir, Chinonyerem Okoro, et al. 2015. "Comparison of MR/ultrasound Fusion-Guided Biopsy with Ultrasound-Guided Biopsy for the Diagnosis

of Prostate Cancer." *JAMA: The Journal of the American Medical Association* 313 (4): 390–97.

16. Mehralivand, Sherif, Joanna H. Shih, Soroush Rais-Bahrami, Aytekin Oto, Sandra Bednarova, Jeffrey W. Nix, John V. Thomas, et al. 2018. "A Magnetic Resonance Imaging-Based Prediction Model for Prostate Biopsy Risk Stratification." *JAMA Oncology* 4 (5): 678–85.

17. Wibmer, Andreas, Hedvig Hricak, Tatsuo Gondo, Kazuhiro Matsumoto, Harini Veeraraghavan, Duc Fehr, Junting Zheng, et al. 2015. "Haralick Texture Analysis of Prostate MRI: Utility for Differentiating Non-Cancerous Prostate from Prostate Cancer and Differentiating Prostate Cancers with Different Gleason Scores." *European Radiology* 25 (10): 2840–50.

18. "Hypodermic Needle Gauge Chart: Medical Tube Technology, Inc." n.d. Accessed Jan 10, 2019. http://www.medtube.com/hypo_chrt.htm.

19. Hu, Jim C., Edward Chang, Shyam Natarajan, Daniel J. Margolis, Malu Macairan, Patricia Lieu, Jiaoti Huang, Geoffrey Sonn, Frederick J. Dorey, and Leonard S. Marks. 2014. "Targeted Prostate Biopsy in Select Men for Active Surveillance: Do the Epstein Criteria Still Apply?" *The Journal of Urology* 192 (2): 385–90.

20. "Numpy.linspace — NumPy v1.16 Manual." n.d. Accessed March 7, 2019. https://docs.scipy.org/doc/numpy/reference/generated/numpy.linspace.html.

21. "In_polyhedron - File Exchange - MATLAB Central." 2018. August 20, 2018. https://www.mathworks.com/matlabcentral/fileexchange/48041-in_polyhedron?focused=7020435&tab=example.

22. Weisstein, Eric W. 2005. "Spherical Coordinates."

    http://mathworld.wolfram.com/SphericalCoordinates.html.

23. Haralick, R. M., K. Shanmugam, and I. Dinstein. 1973. "Textural Features for Image

    Classification." *IEEE Transactions on Systems, Man, and Cybernetics* SMC-3 (6): 610–

    21.

24. "Mahotas: Computer Vision in Python — Mahotas 1.4.3+git Documentation." n.d.

    Accessed March 7, 2019. https://mahotas.readthedocs.io/en/latest/.

25. Wilcoxon, Frank. 1945. "Individual Comparisons by Ranking Methods." *Biometrics

    Bulletin* 1 (6): 80–83.

26. Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand

    Thirion, Olivier Grisel, Mathieu Blondel, et al. 2011. "Scikit-Learn: Machine Learning in

    Python." *Journal of Machine Learning Research: JMLR* 12 (November): 2825–30.

27. Bland, J. M., and D. G. Altman. 1995. "Multiple Significance Tests: The Bonferroni

    Method." *BMJ* 310 (6973): 170.

28. Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. 2012. "ImageNet

    Classification with Deep Convolutional Neural Networks." In *Advances in Neural

    Information Processing Systems 25*, edited by F. Pereira, C. J. C. Burges, L. Bottou, and

    K. Q. Weinberger, 1097–1105. Curran Associates, Inc.

.