**Computer-Aided Diagnostics of Prostate Cancer with Automated Image Analysis using Machine Learning**

**Benjamin Warrick**

**Abstract**

Prostate cancer is the most common cancer among men, other than skin cancer. Prostate cancer screening is done with a blood test that measures levels of prostate-specific antigens (PSA) in the blood and by palpating the prostate through the rectum. If indications are found in either test, the next step is a biopsy of the prostate, which involves using one or more thin needles to remove cores from multiple areas of the prostate. Three sections are taken from each core, and each section is prepared on a glass slide. A pathologist then renders a final diagnosis (i.e. benign vs cancer) after examining each of the slides though a microscope. Because the number of samples can be high per patient, and a given pathologist may see many patients’ biopsies in a day, prostate biopsies are one of the most time consuming case types pathologists must process. Here we show how an automatic system can assist the pathologist in prostate cancer diagnosis by sifting through numerous samples and flagging those suspicious for cancer. Experimental results have produced repeatable accuracies averaging as high as 85%.

**Introduction**

Other than skin cancer, prostate cancer is the most prevalent form of cancer in men, with a risk that rises with age. For this reason, men who seek preventative healthcare treatment will eventually be screened for prostate cancer. Because they are at a higher than average risk, screening should start between the ages of 40 and 45 for African American men and men with a father, brother, or son who has been diagnosed with prostate cancer. Men with an average risk should start screening at age 50. Screening involves the digital rectal exam, the PSA, or both [cancer.org. (2012)]

The digital rectal exam involves a physician palpating the prostate through the rectum, feeling for abnormalities. The PSA (Prostate Specific Antigen) test is a blood test that checks the levels of prostate-specific antigens present. If either test indicates the possibility of cancer, a prostate biopsy process is likely the next step. Some cancers found in the prostate are fast growing, deadly, and are most effectively treated in the early stages. Others are slow growing and might not require treatment, especially in elderly patients. For these reasons it a timely diagnosis is important. [cancer.org]

The biopsy process consists of extracting cores of tissue from the prostate with thin needles. The number of cores varies. The standard initial biopsy retrieves 12. However, the science of prostate biopsy is sophisticated and inexact. There are a number of methods for deciding how many cores to take and what pattern to use, all with the goal of finding and retrieving a sample of the suspected cancerous tissue, but this is not always successful. At times a patient with elevated PSA or abnormal prostate examination will have biopsy results negative for cancer. In such cases, a patient returns a year or so later for a more thorough repeat biopsy, sampling with a great number of cores taken [Bjurlin].

The cores are then processed (a method of removing water from the tissue and replacing it with paraffin wax), cut into three 4-micrometer sections, placed on glass slides, and stained with hematoxylin and eosin. Each requires examination through a microscope by a pathologist trained in the skill of recognizing cancer. With the standard prostate biopsy producing 36 slides (12 cores x 3 sections), and other prostate biopsies resulting in numerous more, possibly hundreds of slides, prostate biopsies are some of the most arduous and time consuming cases a pathologist must process [Bjurlin].

The meticulous and time-consuming nature of such responsibilities makes them expensive. Pathologists represent a small population of physicians and require over a decade to train. Instead of paying them to perform these repetitive tasks it would be better for their effort to be focused on more complex tasks, such as determining the type of cancer once it’s found, or better, doing research that will find ways to prevent or cure cancer.

The repetitive and detailed nature of this problem makes it ideal for a computer. The challenge is in the technique and complexity. Recognizing the difference between healthy and cancerous tissue is easy for an experienced pathologist, so is recognizing the difference between a smooth road and a ditch – but both have only recently been accomplished by a computer, and not yet as effectively as a human.

Machine learning has emerged as a way of solving these types of problems which normally require human judgment and intuition to solve. The dramatic improvement in voice recognition systems over the past few years were accomplished with machine learning [G, R.L. (2014).]. Current progress in the development of the self-driving car is being accomplished with machine learning [Udacity.com]. And research is being conducted in how cancer might be detected in radiological and MRI images using computer vision and machine learning [Lee, Howard, & Chen, Yi-Ping Phoebe. (2015).]. Likewise, computer vision and machine learning can be applied to the diagnosis of prostate cancer. A system can be created that will process numerous biopsy images and flag images and areas of images suspicious for cancer, reducing the pathologists work load, and allowing her to focus on higher-level tasks.

**Methods**

**Image Data**

The data for the experiments consisted of 44 images of prostate tissue, 22 of which contained cancerous tissue and the remaining 22 contained only non-cancerous tissue. The images, which were each 2448 x 1920 pixels and 13.4MB in size, were obtained with the use of special digital camera incorporated into the microscope. Two sets of images where taken, one at 100x magnification and the other at 200x, for a total image count of 88.

The tissue samples were previously available, already prepared to microscope slides, and came from cancerous prostectomies – cancerous prostates that had been removed from the patient. The images were each taken of only part of the sample on the slide, in a deliberate way as to find clear examples of cancerous and non-cancerous prostate tissue. Below is an example of a cancerous image sample, Figure 1.



**Figure 1.** Cancerous prostate tissue image. Reduced to 612 x 480

**wndchrm**

The initial approach to verifying the feasibility of using machine learning to automate the diagnosis of prostate cancer centered on the use of wndchrm. wndchrm is an open source utility for biological image analysis developed by Dr. Lior Shamir of LTU (Lawrence Technology University). wndchrm can run as a stand-alone utility or as a C library, and was originally developed to analyze biological cells during experiments to find new medications. [Shamir, Lior. (2008)] In these experiments, numerous chemical combinations are applied to sets of unhealthy cell samples. wndchrm is capable of learning to recognize which samples have reacted well to a chemical combination by recognizing healthy cells. However, wndchrm is not a specific set of algorithms for this purpose, but rather a set of algorithms designed for the general purpose of classifying images, which makes it ideal for our purposes.

wndchrm works by analyzing sets of images with a list of algorithms to discover ‘features’ in each image. The feature set is recorded in a file. wndchrm can then test the feature set by using only the features from some of the images to classify the rest, then calculate and output the accuracy achieved. Because wndchrm takes a general approach to discovering features, many of the features it finds become noise and reduce accuracy. For example, an image of a cancerous tissue sample likely also contains healthy tissue. Features of the healthy tissue will make the difference between a healthy sample and unhealthy sample less evident. wndchrm addresses this by allowing only the most helpful features to be used, specified with a percent value. This allows for repeated experiments to test different percentages of features to determine the most accurate collection.

**Whole-Image Analysis**

wndchrm analyzes images at an individual pixel level, and all pixels in the image are processed. Thus, images with more pixels require more processing time. Images of the size in our data set, 2448 x 1920 pixels, require substantial processing power, and for the sake of time and practicality, the initial whole-image experiments were processed on a server cluster at LTU comprised of 400 processor cores, of which only one was recorded. Running these same experiments on a 4-core desktop computer proved too time consuming.

To increase efficiency and accuracy, wndchrm offers an option to 'tile' the image. This technique splits the total image into several smaller images and processes them individually. In our whole-image experiment, we used 4x4 tiling. This means that wndchrm separated each image into 16 images before processing. Naturally, when tests for accuracy were performed on the feature set, 4x4 tiling was specified so the images would once again be split into 16 before they were evaluated.

**ROI**

Processing only select portions of an image, ROI (Region of Interest), is a well-established method of increasing the accuracy and efficiency of image analysis [Shamir, L]. For the experiments using ROI, the approach was to split the image into a grid, do a quick analysis of each grid-cell, and based on that analysis select a subset of the grid-cells to be turning into separate images and processed by wndchrm.

The use of ROI increased the efficiency of experiments such that they could be processed on a 4-core desktop pc. This allowed for many more experiments to be performed when using ROI, and accounts for the greater number of ROI experiments produced here, of which seven where recorded. So far, three of these experiments were performed by selecting various numbers of ROI with the lowest standard deviation and using the 100x magnification data set. Three more experiments were performed likewise on the 200x magnification data set. One experiment chose 16 ROI based on high entropy and used the 100x magnification data set.

These experiments were chosen and recorded based on their success and the success of previous similar experiments.

**GUI Interface -- C# .Net OpenCV**

A GUI interface was constructed to facilitate the design and creation of different ROI experiments. It allows the user to see the image, a grid overlay, values analysis values in each grid, and which ROIs are being selected per the analysis. The different algorithms for selecting the ROI are added to a method in the interface’s code. To speed development of this interface it was written in C# using the .Net framework. Emgu CV, a .Net wrapper for OpenCV, which is a well-developed open source image processing library, written in optimized C/C++, for the purpose of computer vision, was also used [opencv.org].

**Results**

**Whole-Image Analysis**

Whole image processing produced the highest accuracy for classifying cancerous and benign prostate tissue sample images in the experiments, reaching averages of 85% when 4x4 tiling and 2% of features were used. See Table 1 and Figure 2. Accuracy was calculated with wndchrm’s test utility which splits the images into two sets; from one set the features are used to classify the other set. The images in the sets are determined randomly, so each run of the test utility can produce different accuracy values. wndchrm is able to run multiple tests and average the calculated average. For these experiments wndchrm was set to run 10 times and the average of the 10 was recorded.

**Table 1.** Whole image processed, tiled 4 x 4, 10-run average accuracy.

|  |  |
| --- | --- |
| Features | Average Accuracy |
| 0.6 | 0.75 |
| 0.55 | 0.82 |
| 0.5 | 0.72 |
| 0.45 | 0.78 |
| 0.4 | 0.76 |
| 0.3 | 0.83 |
| 0.2 | 0.85 |
| 0.1 | 0.84 |
| 0.05 | 0.86 |
| 0.02 | 0.81 |

**ROI**

*Low standard deviation*  
The method of processing only ROIs produced slightly less accurate results in the experiment thus far performed. The most accurate results appeared from the 100x magnification image set when 12 ROI images with lowest standard deviation were taken from an 8x8 windows grid and 1% of features were used. See Table 2.

Low standard deviation produces ROIs that best resembled the overall image, in other words, ROIs from this method best represent an average glimpse of the image. This is likely why the results closely resemble whole image processing.

**Table 2.** ROI Experiments – Lowest Standard Deviation.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Magnification | Windows | Num Imgs | Features | Ave. Accuracy 10-split |
| 100 | 8 | 12 | 0.6 | 0.778788 |
| 100 | 8 | 12 | 0.5 | 0.776515 |
| 100 | 8 | 12 | 0.4 | 0.774242 |
| 100 | 8 | 12 | 0.3 | 0.756818 |
| 100 | 8 | 12 | 0.2 | 0.780303 |
| 100 | 8 | 12 | 0.15 | 0.804545 |
| 100 | 8 | 12 | 0.1 | 0.843182 |
| 100 | 8 | 12 | 0.05 | 0.819697 |
| 100 | 8 | 12 | 0.04 | 0.835606 |
| 100 | 8 | 12 | 0.03 | 0.834091 |
| 100 | 8 | 12 | 0.02 | 0.818939 |
|  |  |  |  |  |
| 200 | 8 | 12 | 0.6 | 0.766667 |
| 200 | 8 | 12 | 0.5 | 0.772727 |
| 200 | 8 | 12 | 0.4 | 0.836364 |
| 200 | 8 | 12 | 0.3 | 0.777273 |
| 200 | 8 | 12 | 0.2 | 0.800758 |
| 200 | 8 | 12 | 0.15 | 0.789394 |
| 200 | 8 | 12 | 0.1 | 0.805303 |
|  |  |  |  |  |
| 100 | 16 | 24 | 0.7 | 0.717424 |
| 100 | 16 | 24 | 0.5 | 0.717424 |
| 100 | 16 | 24 | 0.4 | 0.751136 |
| 100 | 16 | 24 | 0.3 | 0.748864 |
| 100 | 16 | 24 | 0.2 | 0.750758 |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| 200 | 16 | 24 | 0.8 | 0.747727 |
| 200 | 16 | 24 | 0.5 | 0.759091 |
| 200 | 16 | 24 | 0.4 | 0.76553 |
| 200 | 16 | 24 | 0.3 | 0.756439 |
| 200 | 16 | 24 | 0.2 | 0.773864 |
| 200 | 16 | 24 | 0.1 | 0.793182 |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| 100 | 16 | 48 | 0.5 | 0.745455 |
| 100 | 16 | 48 | 0.4 | 0.754167 |
| 100 | 16 | 48 | 0.3 | 0.75 |
| 100 | 16 | 48 | 0.2 | 0.757765 |
| 100 | 16 | 48 | 0.1 | 0.741098 |
|  |  |  |  |  |
| 200 | 16 | 48 | 0.6 | 0.746212 |
| 200 | 16 | 48 | 0.5 | 0.752273 |
| 200 | 16 | 48 | 0.4 | 0.763636 |
| 200 | 16 | 48 | 0.3 | 0.752841 |
| 200 | 16 | 48 | 0.2 | 0.761932 |
| 200 | 16 | 48 | 0.1 | 0.793939 |
| 200 | 16 | 48 | 0.05 | 0.813447 |
| 200 | 16 | 48 | 0.04 | 0.804545 |

*High Standard Deviation*ROIs based on high standard deviation were also tried but tended to find areas of white space and so did not provide helpful results, although, high standard deviation in combination with other features holds promise and warrants further experimentation, which is planned.

*High Entropy*  
Initial experiments with high entropy were performed with the idea that finding areas of relatively high disorder might include finding areas with cancer, which occurred for the most part in images containing cancer. However, these experiments have so far produced lower accuracies compared with lowest standard deviation ROIs. These accuracies are still promising though, ranging in the low- to mid-70% when 24 ROI images of highest entropy are taken from a 16x16 window grid. See Table 3.

This ROI method was used for both cancer and non-cancer images, which may explain the reduced accuracy, because high entropy ROIs from non-cancer images will likely provide non-typical examples of non-cancerous tissue.

The success of our high entropy ROIs landing on cancer could be coincidental. The images containing cancer in these experiments contained a majority of cancerous tissue, so random ROIs would also mostly land on cancer. Also, high entropy ROIs in these cancer images may find especially disordered examples of cancer, or perhaps areas without cancer. Areas without cancer in a mostly cancer images may have high relative entropy, and explain why some ROIs did not land on cancer. Further experimentation is necessary to answer these questions, which is planned.

**Table 3.** ROI Experiments – High Entropy.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Magnification | Windows | Num Imgs | Features | Ave. Accuracy 10-split |
| 100 | 16 | 24 | 0.7 | 0.742803 |
| 100 | 16 | 24 | 0.6 | 0.722727 |
| 100 | 16 | 24 | 0.5 | 0.757576 |
| 100 | 16 | 24 | 0.4 | 0.739015 |
| 100 | 16 | 24 | 0.3 | 0.726136 |
| 100 | 16 | 24 | 0.2 | 0.724621 |
| 100 | 16 | 24 | 0.1 | 0.731439 |

**Conclusion**

The level of accuracy produces by all of the conducted experiments shows great promise for the feasibility of a practical cancer screening system, especially since the training data was limited to only 44 images per experiment. Although the data was intentionally chosen to present distinct examples of both classes, and the probabilities in excess of 80% may be overly optimistic, the fact that a majority of the experiments produced probabilities greater than 75% indicates that the system is likely recognizing the difference between cancer and non-cancer images. The addition of more challenging images may reduce this accuracy in experiments, but the planned increase in data by one or more orders of magnitude should prove to result in high overall accuracy.

**Boosting**

Furthermore, even if accuracy falls below 75% with additional data, the possibility of a highly accurate system is still feasible through boosting. If five separate techniques are discovered that each have accuracy in excess of 70%, all five could be used to classify actual biopsies with accuracy in excess of 95%. If 10 methods are used with the same accuracy, taking the majority would yield an accuracy approaching 99.9%. Additionally, a practical cancer screen would be tilted toward sensitivity and away from specificity, so the likelihood of cancer not being detected would be minuscule, which is the goal of a practical cancer screen.

**Planned Experiments**

Further experimentation should reveal new methods for capturing better ROI. To explore this, several experiments are planned for the next stage of this research project including the following:

* Combined standard deviation and entropy.
* Simplified feature detection. Essentially, making a first pass over the image looking for a simple set of features.
* Applying different ROI methods to each class of data. One ROI method might work well in directing the system to cancerous tissue in a cancer image, while a second better at selecting typical examples of non-cancer tissue from a benign image.

Also planned is further experimentation and development of the GUI interface used to generate ROIs from the test data.

**Hands-On Training Utility**

An alternative reliance on auto-generated ROI is human indicated ROIs. The system planned for the next stage of this research project will be similar to the following description: A GUI utility will display an image to the pathologist and allow him to select the cancerous regions, if present. If no cancer is present, random or method generated ROIs will be selected of the non-cancer tissue. Images from those ROIs will be created, processed, and their features added to the training set. Once a modest training set is established, a second stage of training will begin where the utility tries to classify new images, indicating the location of cancer on the image in the display, allowing the pathologist to correct any mistakes. The display will likely have a grid overlay, allowing the pathologist to indicate which grid cells contain cancer that are not flagged, and un-flag flagged regions of benign tissue. Corrected grid cells will be processed into features.

A system with this type of flexibility is implementable in a variety of laboratory settings and wouldn’t be dependent on a specific laboratory equipment, calibrations, or techniques because the system would be trained only on data from that source. Such a system might also be flexible enough to detect other types of cancer other than prostate.

**Cloud Feature**

If auto image analysis does prove successful in actual pathologist laboratories, the next stage in development might include collecting data from each instance of the system and consolidating it to a central location. This consolidated data might include entire images, only features, or other types of data, and could be used for a variety of machine learning purposes, such as distilling out a highly accurate set of image features for specific cancers and stages of these cancers.

**Challenges**

Two primary challenges face a practical and reliable cancer screening system such as is described here. First, images are not normally created of prostate biopsies. Normal day-to-day pathology laboratory activities include preparing the prostate tissues to slide, diagnosing it from the slide, and storing the slide. For image analysis to be useful, it would be necessary to add a step to this process where a photo is taken, or replace the glass slide preparation step with a direct imaging step. Second, wndchrm, as well as other approaches to machine learning, require a lot of computer resources, far more than are normally available in a pathologist laboratory. A central server cluster or a local server cluster may be necessary to automate image analysis in this way.

Other challenges include resistance posed by the pathology community; few professional are eager to see a portion of their skills and responsibilities replaced by a computer. Legal implications also exist. Although auto image analysis would improve the speed and accuracy with which diagnoses are made, when a mistake occurs, the question of liability is less clear when a computer is responsible for part of the diagnosis, especially when machine learning is involved and the point at which a positive or negative result is produced may not be traceable to a specific line of code. We hope that as the usefulness and eventually necessity of this and similar systems become apparent, such challenges will naturally be overcome.

**Acknowledgments**

Dr. Joshua Warrick was instrumental in gaining access to all the data for this project and continues to build a much larger set of data for the following stages of this research.

**References**

Bjurlin, M. A., Carter, H. B., Schellhammer, P., Cookson, M. S., Gomella, L. G., Troyer, D., ... & Taneja, S. S. (2013). Optimization of initial prostate biopsy in clinical practice: sampling, labeling and specimen processing. The Journal of urology, 189(6), 2039-2046.

cancer.org. (2012). Prostate Cancer Screening FAQ. Retrieved from <http://www.cancer.org/cancer/news/prostate-cancer-screening-faq>

Collins, Francis. (2015). Prostate Cancer: Designing a Smarter High-Tech Biopsy. Retrieved from <http://directorsblog.nih.gov/2015/02/03/prostate-cancer-designing-a-smarter-high-tech-biopsy/>

emgu.com. (n.d.) Main Page. Retrieved from <http://www.emgu.com/wiki/index.php/Main_Page>

G, R. L. (2014). How speech-recognition software got so good. Retrieved from <http://www.economist.com/blogs/economist-explains/2014/04/economist-explains-11>

Kandemir, Melih, & Hamprecht, FredA. (2014). Computer-aided diagnosis from weak supervision: A benchmarking study. Journal of Computerized Medical Imaging and Graphics, 42 (2015), 44-50.

Lee, Howard, & Chen, Yi-Ping Phoebe. (2015). Image based computer aided diagnosis systems for cancer detection. Journal of Expert Systems with Applications, 42 (2015), 5356-5365.

Mayo Clinic Staff. (2013). Prostate cancer screening: Should you get a PSA test? Retrieved from: <http://www.mayoclinic.org/diseases-conditions/prostate-cancer/in-depth/prostate-cancer/art-20048087>

opencv.org. (n.d.) OPENCV (OPEN SOURCE CMPUTER VISION). Retrieved from opencv.org

Shamir, L., Delaney, J. D., Orlov, N., Eckley, D. M., & Goldberg, I. G. (2010). Pattern recognition software and techniques for biological image analysis.PLoS Comput Biol, 6(11), e1000974-e1000974.

Shamir, Lior. (2008). Wndchrm – an open source utility for biological image analysis. Retrieved from <http://www.scfbm.org/content/3/1/13>

Udacity.com. (n.d.). ML in The Google Self-Driving Car. Retrieved from [https://www.udacity.com/course/viewer#!/c-ud120/l-2254358555/m-2374468553](https://www.udacity.com/course/viewer)