Automated detection of prostate cancer in digitized whole-slide images of H&E-stained biopsy specimens

G. Litjens, B. Ehteshami Bejnordi, N. Timofeeva, G. Swadi, I. Kovacs,
C. Hulsbergen - van de Kaa and J. van der Laak

Radboud University Medical Center, Nijmegen, The Netherlands;

ABSTRACT

Automated detection of prostate cancer in digitized H&E whole-slide images is an important first step for computer-driven grading. Most automated grading algorithms work on preselected image patches as they are too computationally expensive to calculate on the multi-gigapixel whole-slide images. An automated multi-resolution cancer detection system could reduce the computational workload for subsequent grading and quantification in two ways: by excluding areas of definitely normal tissue within a single specimen or by excluding entire specimens which do not contain any cancer. In this work we present a multi-resolution cancer detection algorithm geared towards the latter. The algorithm methodology is as follows: at a coarse resolution the system uses superpixels, color histograms and local binary patterns in combination with a random forest classifier to assess the likelihood of cancer. The five most suspicious superpixels are identified and at a higher resolution more computationally expensive graph and gland features are added to refine classification for these superpixels. Our methods were evaluated in a data set of 204 digitized whole-slide H&E stained images of MR-guided biopsy specimens from 163 patients. A pathologist exhaustively annotated the specimens for areas containing cancer. The performance of our system was evaluated using ten-fold cross-validation, stratified according to patient. Image-based receiver-operating characteristic (ROC) analysis was subsequently performed where a specimen containing cancer was considered positive and specimens without cancer negative. We obtained an area under the ROC curve of 0.96 and a 0.4 specificity at a 1.0 sensitivity.

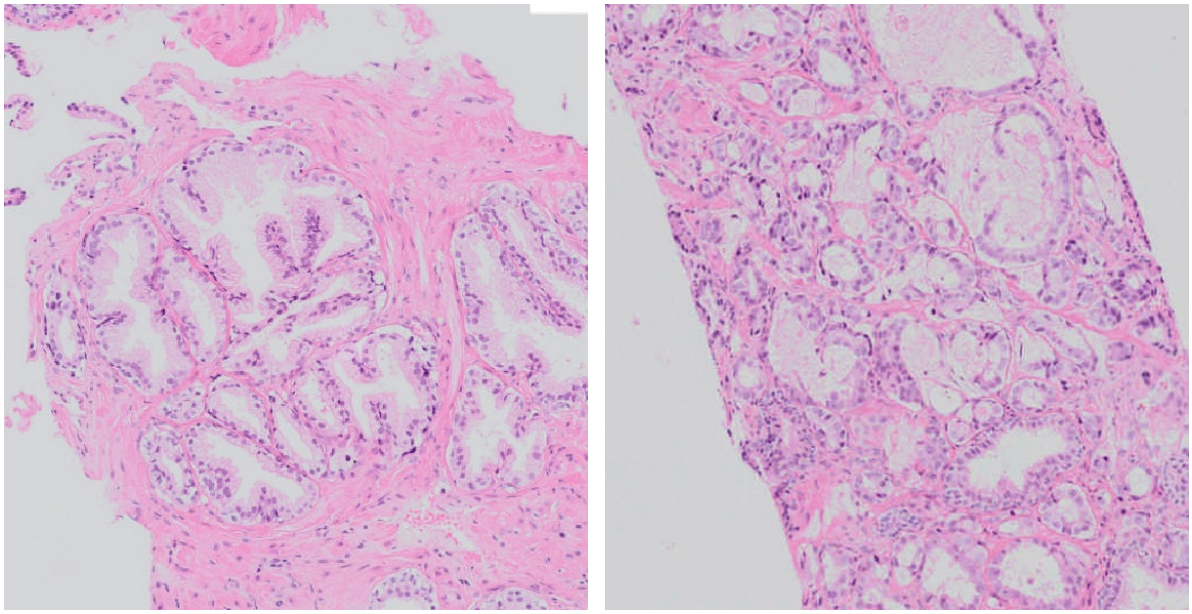
Keywords: histopathology, whole-slide imaging, computer-aided detection, prostate cancer

1. INTRODUCTION

In prostate histopathology, the Gleason score of cancer, determined on H&E-stained sections of biopsy specimens, is considered the most important marker for patient prognosis. However, visual assessment of tissue architecture underlying the Gleason grading system suffers from considerable intra- and inter-observer variation.[1] An example of normal prostate tissue and prostate cancer is shown in Figure 1. The recent advent of digital whole-slide scanning systems has opened the door for computerized analysis of histopathology slides. Computerized analysis is well suited to tackle the two most important problems in prostate histopathology: limited reproducibility and the possibility to overlook relevant tissue areas. Image analysis algorithms are inherently devoid of subjectivity. The exact same result is obtained with the same input time after time, enhancing reproducibility. Furthermore, computers never tire, allowing them to perform accurate analysis of the entire scanned tissue section.

Several groups have worked on automated grading and quantification of digitized prostate histopathology. Diamond et al.[2] and Khurd et al.[3] focused mostly on texture features (Haralick and multi-scale textons) to either discriminate cancer from normal tissue or to automatically grade cancer in predefined image patches. Gorelick et al.[4] and Doyle et al.[5] also used higher level features to discriminate different grades of cancer from normal tissue. Gorelick et al. implemented a superpixel approach and used tissue component histograms, Doyle et al. used a cascaded classifier in combination with texture and graph features to classify patches of tissue. However, most of these algorithms have only focused on predefined areas of interest or sub-sampled whole-slide images, mostly due to computational complexity. If we could reduce this complexity one could readily apply these grading algorithms to entire scanned slides without manual preselection. This can be done in two ways: by identifying cancerous areas automatically within a single specimen or by excluding entire specimens which do not contain any cancer. Doyle et al.[6] have previously implemented a method using a boosted Bayesian classifier which tries to address the identification of relevant regions within a single tissue section. They obtained per-pixel areas under the receiver-operating characteristic curve of up to 0.84.

Medical Imaging 2015: Digital Pathology, edited by Metin N. Gurcan, Anant Madabhushi
Proc. of SPIE Vol. 9420, 94200B · © 2015 SPIE · CCC code: 1605-7422/15/$18
doi: 10.1117/12.2081366



(a) Normal prostate glands (b) Prostate cancer (Gleason grade 4+3)

Figure 1: Example of normal prostate glands relative to prostate cancer

In this work we present a multi-resolution cancer detection algorithm which focuses on the latter problem: identifying entire specimens which do not contain cancer and as such do not require further analysis by automated grading or cancer quantification methods. The proposed system can not only be used as a pre-cursor step to further automated analysis, but could also reduce the workload for pathologists in clinical practice by pre-screening biopsy specimens at high sensitivity.

2. METHODOLOGY

2.1 Multi-resolution histopathology images

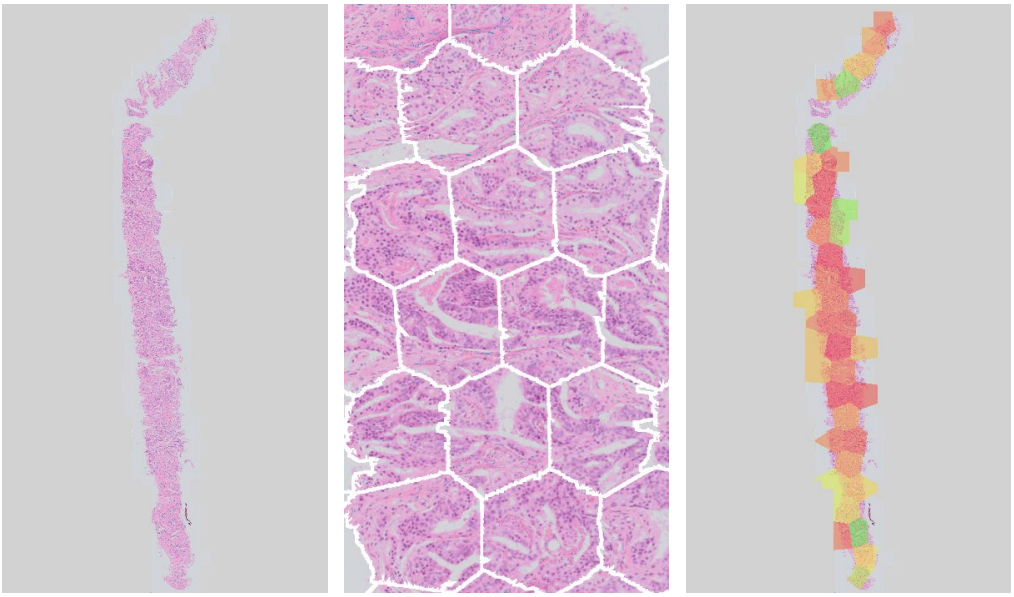
As whole-slide images are typically too big to fit in the memory of ordinary desktop computers, slide scanners also store a multi-resolution pyramid of the base image, which consists of consecutively downsampled levels. In our data, the base image is recorded at 40x objective magnification, corresponding to 0.16 microns per pixel. The pyramid is created with subsequent magnifications of 20x, 10x, 5x, 2.5x, 1.25x and 0.6125x. All images are stored in RGB format.

2.2 Coarse analysis

Our methodology is largely based on the workflow of pathologists, who also start off with a coarse analysis of the tissue at low magnification (usually 4x[1]). For the coarse analysis we use the 5x level in the multi-resolution pyramid. First, superpixels are calculated using the SLIC-algorithm[7] to obtain relatively homogenous patches of tissue. The SLIC algorithm uses both the Euclidean distance from a seed and the 'color distance' in the CIELab color space to define superpixels. The major parameter is the amount of seeds (superpixels) to calculate within an image. We set this parameter such that each superpixel contains around 8000 regular pixels. An example result is presented in Figure 2. After superpixel generation, a tissue mask was obtained by first calculating a density image.[8] The density image is constructed by calculating the optical density for each channel (R, G, B) via:

$$D = \frac{D_r + D_g + D_b}{3} \text{ where } D_c = -\ln\left(\frac{I_c}{I_c^0}\right) \quad (1)$$

where D_c is the optical density of a single channel, I_c the image intensity of a channel and I_c^0 the intensity of incident light of a channel (255 in our case). The density image has values between 0 and 2.4. Any superpixel



(a) Original image (b) Superpixels (zoomed) (c) Coarse classification

Figure 2: Results of the coarse analysis in the presented system

with an average density lower than 0.2 is considered a background pixel, all others are considered tissue pixels. Subsequently, color and local binary pattern histogram features are calculated for each remaining superpixel. For each color channel (R, G and B) the mean, standard deviation, sum, median, minimum, maximum, 25th-percentile and 75th-percentile are calculated. For the local binary patterns, a radius of 1 and neighborhood of 8 was used, resulting in 10 unique histogram bins per color channel.[9] After feature extraction a random forest classifier[10] (containing 100 regression trees) is used to calculate the likelihood that a superpixel contains cancer.

2.3 Detailed analysis

For each whole-slide image the five superpixels with the highest likelihood of being cancer were selected for subsequent analysis. The image data for each area covered by a superpixel was subsequently extracted at 40x, typically resulting in image patch sizes of around 1600 x 1600 pixels. For these patches more computationally intensive features are calculated, which can roughly be subdivided in three categories: RGB histogram, graph and gland features.

2.3.1 RGB histogram features

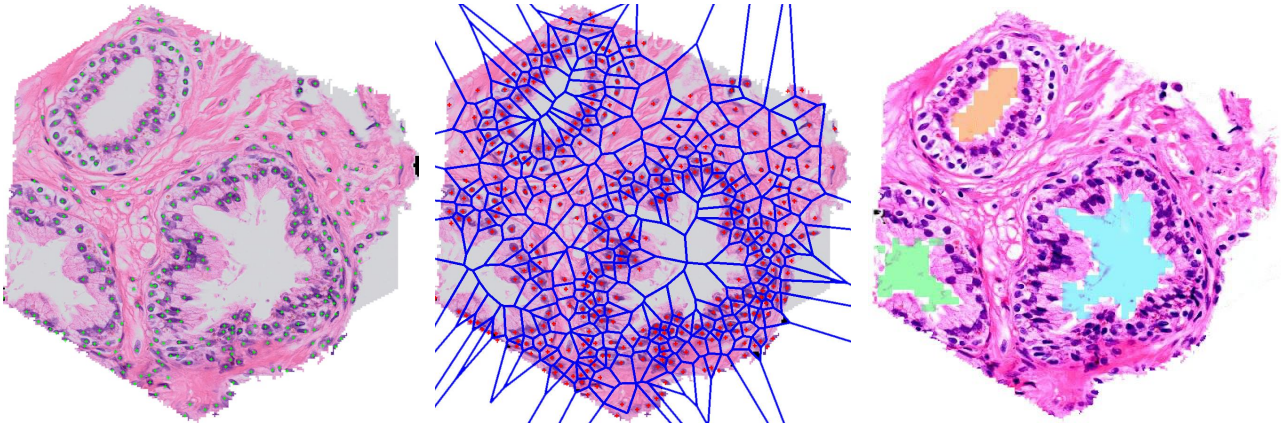
For the RGB histogram features we calculated the histogram of RGB values for the entire 40x image patch. Subsequently, for each color channel (R, G and B) the mean, standard deviation, sum, median, minimum, maximum, 25th-percentile and 75th-percentile were calculated and included as features.

2.3.2 Graph features

Graph features calculated on centroids of nuclei try to capture local tissue structure.[5] To calculate graph features, first we need to identify the centroids of the nuclei. We started with extracting the hematoxylin channel from the RGB image by applying color deconvolution.[11] The color stain vector were determined empirically by inspection of the deconvolved images.

Subsequently, we applied the method proposed by Veta et al.,[12] which uses the fast radial symmetry transform in combination with regional minima extraction to obtain nuclear centers. The fast radial symmetry transform enhanced areas were there is a consisted gradient pointing inwards in a circular area around a center (for a specified range of sizes). We use a range radii between 2 - 5 microns.

After the extraction of the nuclei centers, the Voronoi diagram,[5] Delaunay triangulation,[5] Ulam trees,[13] Gabriel graph, and the minimum spanning tree[5] were calculated. Although all these graphs are related and



(a) Original with detected nuclei indicated with green crosses

(b) Voronoi diagram calculated on the detected nuclei

(c) Gland segmentation

Figure 3: Results of the detailed analysis in the presented system

are sometimes even subsets or duals (e.g. Delaunay triangulation is a dual of the Voronoi diagram and Grabiel graphs a subgraph of the Delaunay triangulation), calculating certain features provides additional information. For example, calculating the mean edge lengths in a minimum spanning tree and in a Grabriel graph results in different features as they provide complimentary information.

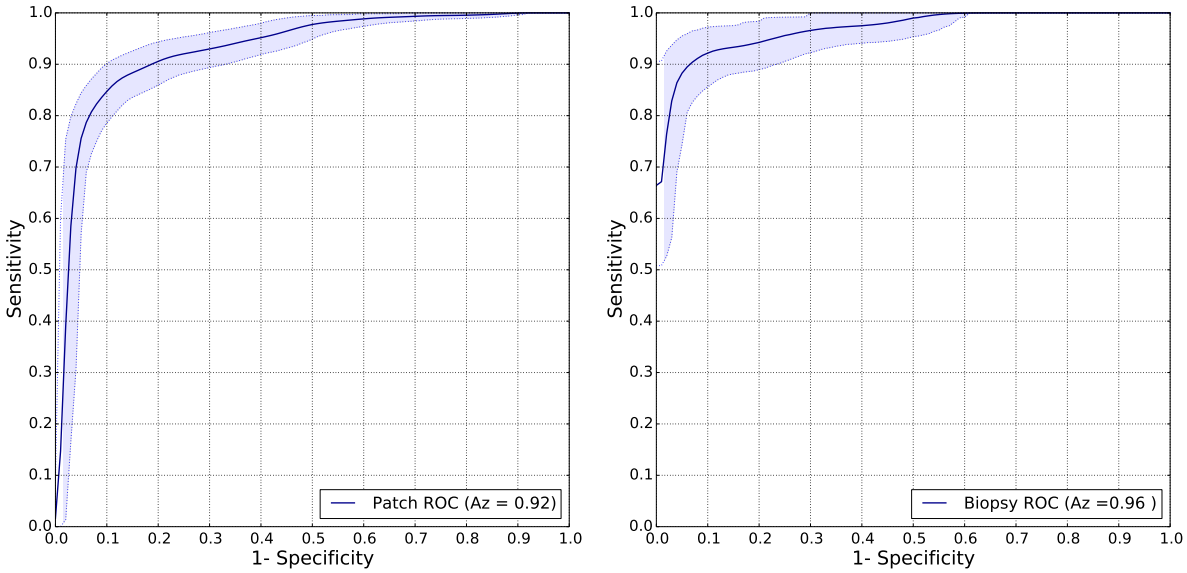
For the Voronoi diagram we calculated the number of sides, the perimeter and area per cell. For the Delaunay triangulation the area and maximum length were calculated per triangle. For the minimum spanning tree, the Gabriel graph and the Ulam trees the edge lengths were calculated. Both the mean and standard deviation for each feature were added.

2.3.3 Gland features

The last set of features was based on gland shape and distribution. As prostate cancer is inherently a glandular disease, extracting features from the glans could offer discriminative information. To capture these characteristics we implemented a gland segmentation technique which recognizes most glands. We start with our binary tissue mask, which contains holes at the gland locations. By flood filling the tissue mask from the image corners and subsequently inverting it, we obtain segmentations of the glands. Using a connected component analysis, we obtain the total number of glands and the sizes for the individual glands. In this paper we use the total number of glands, and the mean, median, standard deviation, 25th and 75th percentile of their sizes as features. An example result is presented in Figure 3. A random forest classifier was used to obtain a continuous likelihood (between 0 - 1) for each image patch that it contains cancer. The mean likelihood across the five superpixels per case was subsequently assigned as the per-case score.

2.4 Validation

In total 204 whole-slide images of H&E-stained biopsy specimens of 163 patients were used for validation of our algorithms. Out of 204 images, 115 contained cancer and 89 did not. Slides were digitized using an Olympus dotSlide system at a resolution of 0.16 um per pixel (40x). Subsequently, a pathologist carefully outlined all the regions on the slide containing cancer. All tissue that was not outlined was considered normal. For the training of the classifier in both levels superpixels and patches which were mostly within outlined cancer areas are considered positive, otherwise they were considered negative. The final performance of our system was evaluated using receiver-operating characteristic (ROC) analysis on the likelihoods after the detailed analysis stage. The ROC curves were generated via a ten-fold, patient-stratified cross-validation on a per-superpixel and per-biopsy level. Both the overall area under the ROC curve (AUC) and the performance at high sensitivity were inspected. Confidence interval were generated using patient-stratified bootstrapping with 10000 samples.



(a) Patch-based analysis (b) Biopsy-based analysis

Figure 4: ROC curves of the cancer detection system

3. RESULTS AND DISCUSSION

After coarse level classification, selecting the five superpixels with the highest likelihood results in a 1.0 per-biopsy sensitivity, meaning that if cancer is present in a section, at least one of the five selected superpixels contains it. The ROC curves after detailed analysis are presented in Figure 4. The per-patch AUC was 0.92 and the per-biopsy AUC was 0.96. Most importantly, at a sensitivity of 1.0 a specificity of 0.40 could be obtained at the per-biopsy level. This indicates that our system might have the potential to exclude up to 40% of all biopsy specimens from further analysis, be it by a pathologist or an automated grading system. The average run-time of the system per case was around 4 minutes on a single core of a standard desktop PC, where most time resides in the initial, coarse level classification. Although most likely not as fast as a pathologist, we feel this computation time is acceptable and could be further improved by implementing the algorithms in a multi-threaded or GPU-based fashion.

There are also some limitations. The system only performs a detection of cancer at the biopsy level, no accurate segmentation of the full cancer extent is performed. Expanding the system to classify all superpixel in the detailed fashion might allow us to provide this extra tool to pathologist when inspecting the remaining cases. Second, no statement about cancer grade is made. In general, detection and delineation of cancer in biopsy specimens is a task which pathologists can perform with high reproducibility and accuracy. However, grading of cancer has shown to be much more difficult to reproduce and as such expanding our system to be able to perform this step could potentially lead to improved inter-observer concordance in cancer grading. Last, the individual components like gland segmentation are still rather simple and could be further improved to reduce segmentation artifacts. An example of such an artifact is tears in the biopsy specimen being recognized as cancer. Summarizing, a fully automated system which can detect prostate cancer in digitized H&E biopsy specimens has been presented. Initial results show promising performance, especially in identifying biopsies for further evaluation.

REFERENCES

[1] Epstein, J. I., Allsbrook, W. C., Amin, M. B., Egevad, L. L., and ISUP Grading Committee, "The 2005 International Society of Urological Pathology (ISUP) consensus conference on Gleason grading of prostatic carcinoma," *Am J Surg Pathol* **29**, 1228–1242 (2005).

[2] Diamond, J., Anderson, N. H., Bartels, P. H., Montironi, R., and Hamilton, P. W., "The use of morphological characteristics and texture analysis in the identification of tissue composition in prostatic neoplasia," *Hum Pathol* **35**, 1121–1131 (2004).

[3] Khurd, P., Bahlmann, C., Maday, P., Kamen, A., Gibbs-Strauss, S., Genega, E. M., and Frangioni, J. V., "Computer-aided gleason grading of prostate cancer histopathological images using texton forests," *Proc IEEE Int Symp Biomed Imaging* **14-17 April 2010**, 636–639 (2010).

[4] Gorelick, L., Veksler, O., Gaed, M., Gomez, J. A., Moussa, M., Bauman, G., Fenster, A., and Ward, A. D., "Prostate histopathology: learning tissue component histograms for cancer detection and classification," *IEEE Trans Med Imaging* **32**, 1804–1818 (2013).

[5] Doyle, S., Feldman, M. D., Shih, N., Tomaszewski, J., and Madabhushi, A., "Cascaded discrimination of normal, abnormal, and confounder classes in histopathology: Gleason grading of prostate cancer," *BMC Bioinformatics* **13**, 282 (2012).

[6] Doyle, S., Feldman, M., Tomaszewski, J., and Madabhushi, A., "A boosted bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies," *IEEE Trans Biomed Eng* **59**, 1205–1218 (2012).

[7] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Susstrunk, S., "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Trans Pattern Anal Mach Intell* **34**, 2274–2282 (2012).

[8] van der Laak, J. A., Pahlplatz, M. M., Hanselaar, A. G., and de Wilde, P. C., "Hue-saturation-density (HSD) model for stain recognition in digital images from transmitted light microscopy," *Cytometry* **39**, 275–284 (2000).

[9] Ojala, T., Pietikainen, M., and Maenpaa, T., "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans Pattern Anal Mach Intell* **24**, 971–987 (2002).

[10] Breiman, L., "Random forests," *Machine Learning* **45**, 5–32 (2001).

[11] Ruifrok, A. C. and Johnston, D. A., "Quantification of histochemical staining by color deconvolution," *Anal Quant Cytol Histol* **23**, 291–299 (2001).

[12] Veta, M., van Diest, P. J., Kornegoor, R., Huisman, A., Viergever, M. A., and Pluim, J. P. W., "Automatic nuclei segmentation in h&e stained breast cancer histopathology images," *PLoS One* **8**, e70221 (2013).

[13] Marcelpoil, R., Davoine, F., and Robert-Nicoud, M., "Cellular sociology: parametrization of spatial relationships based on voronoi diagram and ulam trees," in [*Fractals in biology and medicine*], 201–209, Springer (1994).