# Machine learning approaches to analyze histological images of tissues from radical prostatectomies

Arkadiusz Gertych [a,b,*], Nathan Ing [a], Zhaoxuan Ma [c], Thomas J. Fuchs [d,e,f], Sadri Salman [a], Sambit Mohanty [b], Sanica Bhele [b], Adriana Velásquez-Vacca [a], Mahul B. Amin [b], Beatrice S. Knudsen [b,c]

[a] Department of Surgery, Cedars-Sinai Medical Center, 8700 Beverly Blvd, Los Angeles, CA 90048, USA
[b] Department of Pathology and Laboratory Medicine, Cedars-Sinai Medical Center, 8700 Beverly Blvd, Los Angeles, CA 90048, USA
[c] Department of Biomedical Sciences Cedars-Sinai Medical Center, 8700 Beverly Blvd, Los Angeles, CA 90048, USA
[d] Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA
[e] Department of Medical Physics, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, NY 10065, USA
[f] Department of Pathology, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, NY 10065, USA

## ARTICLE INFO

## ABSTRACT

Computerized evaluation of histological preparations of prostate tissues involves identification of tissue components such as stroma (ST), benign/normal epithelium (BN) and prostate cancer (PCa). Image classification approaches have been developed to identify and classify glandular regions in digital images of prostate tissues; however their success has been limited by difficulties in cellular segmentation and tissue heterogeneity. We hypothesized that utilizing image pixels to generate intensity histograms of hematoxylin (H) and eosin (E) stains deconvoluted from H&E images numerically captures the architectural difference between glands and stroma. In addition, we postulated that joint histograms of local binary patterns and local variance (LBPxVAR) can be used as sensitive textural features to differentiate benign/normal tissue from cancer. Here we utilized a machine learning approach comprising of a support vector machine (SVM) followed by a random forest (RF) classifier to digitally stratify prostate tissue into ST, BN and PCa areas. Two pathologists manually annotated 210 images of low- and high-grade tumors from slides that were selected from 20 radical prostatectomies and digitized at high-resolution. The 210 images were split into the training ($n = 19$) and test ($n = 191$) sets. Local intensity histograms of H and E were used to train a SVM classifier to separate ST from epithelium (BN + PCa). The performance of SVM prediction was evaluated by measuring the accuracy of delineating epithelial areas. The Jaccard $J = 59.5 \pm 14.6$ and Rand $Ri = 62.0 \pm 7.5$ indices reported a significantly better prediction when compared to a reference method (Chen et al., Clinical Proteomics 2013, 10:18) based on the averaged values from the test set. To distinguish BN from PCa we trained a RF classifier with LBPxVAR and local intensity histograms and obtained separate performance values for BN and PCa: $J_{BN} = 35.2 \pm 24.9$, $O_{BN} = 49.6 \pm 32$, $J_{PCa} = 49.5 \pm 18.5$, $O_{PCa} = 72.7 \pm 14.8$ and $Ri = 60.6 \pm 7.6$ in the test set. Our pixel-based classification does not rely on the detection of lumens, which is prone to errors and has limitations in high-grade cancers and has the potential to aid in clinical studies in which the quantification of tumor content is necessary to prognosticate the course of the disease. The image data set with ground truth annotation is available for public use to stimulate further research in this area.

© 2015 Elsevier Ltd. All rights reserved.

## 1. Introduction

Prostate cancer (PCa) remains the most commonly diagnosed cancer in men in developed countries. Fortunately, cancer deaths are steadily declining despite a fairly steady rate of new incidences per year [1]. Microscopic evaluation of prostate needle biopsies is the gold standard for PCa diagnosis and criteria have been established to manage patients based on histopathologic observations in the biopsy and radical prostatectomies. While normal glands are

* Correspondence to: BioImage Informatics Laboratory, Department of Surgery, 116 N Robertson Blvd. Suite #903, Los Angeles CA 90048, USA. Tel.: +1 310 423 2964; fax: +1 310 423 7707.
*E-mail addresses:* arkadiusz.gertych@cshs.org, agertych@gmail.com (A. Gertych).

organized into ducts and acini and well separated by stroma, as PCa develops, the malignant acinar structures undergo excessive branching morphogenesis. This is the reason for the histological appearance of small and tightly packed glands with little or no intervening stroma that has become a diagnostic hallmark of low-grade PCa. The architecture in high-grade cancer is different. Cancer cells form glands within glands (Gleason grade 4 (G4) cribriform) loose their ability to form glands that possess a lumen (G4 non-cribriform) or grow in sheets (G4 or G5) [2]. The association between the severity and growth pattern of the prostate cancer provides the basis for the Gleason grading scheme [2,3], which is used clinically. Accurate grading by pathologists requires extensive experience and is occasionally associated with disagreement about low- versus high-grade diagnostic interpretation. In fact, in the early days of the Gleason grading scheme, the inter-observer reproducibility to distinguish low-grade (Gleason grade 3 (G3)) from high-grade tumor growth patterns (G4) ranged between 25% and 47% depending on the grade distribution in the study cohort [4–6].

One way to potentially improve the reproducibility and accuracy of tumor grading is through a computer-assisted approach. Tools for recognition and quantification of morphological characteristics, which correlate with individual Gleason grades have been under intense development by computational pathology researchers. The vast majority of software algorithms for image analysis employs context-based gland quantification to distinguish benign/normal tissue from low- and high-grade areas of cancer [7–12]. As a starting point for image analysis, a typical scenario involves the generation of image tiles with several areas of tumor cells, which receive a grade annotation by an experienced pathologist. A set of descriptors that reflect the cellular organization, inflammation or various secretions is first extracted from the image tiles and then classified according to the cancer grade annotation of the image. Typically, the image content of the entire tile is predefined as stroma (ST), benign/normal (BN), low-grade (G3) or high-grade (G4) cancer [13,14].

Numerous approaches have been developed to capture the growth pattern of prostate cancers. Existing techniques involve various kinds of image features to capture growth patterns that are related to color, texture and nuclear topology [7–11,14].

Yet, the performance of these classification methods varies greatly. The uniformity of the image content, ideally with only one tissue component in each tile, has a major impact on the accuracy of the tissue classification. For images with heterogeneous content, such as those from cancer tissues, which contain admixtures of benign structures and cancer, the performances of the classifiers decline. While the speed of evaluating the entire slide constitutes a major advantage of the tile-based analysis, the approach has several shortcomings. The impact of the tile size, which varies among studies on the performance of the classifier, is unknown. Moreover, the predictive power of tissue descriptors and classifiers can be artificially high, if training and validation is performed on small sets of tiles (usually <100). This problem is particularly grave in prostate cancer since large tiles with homogenous tissue content cannot be generated in sufficient quantity.

To overcome the problems that are caused by tissue heterogeneity, tissue classes can be manually delineated by a pathologist for algorithm training and validation [13,15–17]. The training set consists of similar manually annotated tissue regions. While this approach is more laborious, it provides additional opportunities for computational image classification [15–17]. Low-power image magnification (<20×) is often employed for prostate cancer image analysis mainly, because it is the most efficient way to manually grade prostate cancer [7,8,10] [15]. However, a recent study shows the improved performance of high-resolution imaging. In the work published by Kwak et al. [16], 21 intensity and 42 texture features

(including local binary patterns) were utilized to segment stroma from the epithelium and the analysis was conducted on 4 different resolution scales. The training was performed manually by a pathologist and ROC curves showed high concordance rates with final algorithm output. However, since the robustness of image analysis is a complex product of the quality of manual ground truth, image resolution and tissue heterogeneity, it is important to determine the effect of each component on the accuracy of segmentation

Recently, machine learning approaches have become popular to quantify tumor areas in histopathological preparations. When expression of protein biomarkers in breast cancer specimens was visualized by staining with antibodies and quantified by image analysis, human and software-derived annotations showed strong agreement in the classification of cancer areas. Furthermore, software developed by academic or commercial groups efficiently separated cancer from stroma within image tiles on a sub-tile resolution [18,19]. However, quantitative analysis of specimens stained with hematoxylin and eosin (H&E), which is routinely used for histopathologic evaluations, is much more challenging and the development of software for analysis of H&E stained slides is the main determinant of the pace at which the image analysis field advances.

Since benign prostate glands and G3 and G4 cancer areas are morphologically distinct from stroma in H&E stained tissue sections, the differences can be numerically captured by image analysis. These slide preparations provide an ideal starting point to demonstrate the power of machine learning tools for classification of H&E images on a sub-tile resolution. In other words, instead of classifying the whole tile content into one class, a machine learning tool can classify individual pixels and deliver pixel-based tissue quantifications. Towards the development of such tools, our team designed three separate classifiers to identify and quantify areas of stroma in images with benign glands, or G3, or G4 prostate cancer with excellent performance [17]. In contrast to other methods, the histograms composed of pixels from H&E intensity measurements that we utilized to describe and classify images were superior to histograms of oriented gradients and provided the highest tissue classification rates. Encouraged by the preliminary results, we continued to improve the approach through the employment of intensity features combined with a more complex texture features for the capture of patterns within areas of glandular architecture. The tissue segmentation results were compared to manual annotations by a pathologist in a large set of high-resolution images of radical prostatectomies. Overall, this approach combines improvements in classification performance and speed and is ready for preliminary testing in prognostic and predictive biomarker studies.

## 2. Materials

Radical prostatectomy specimens from 20 patients with a diagnosis of G3 or G4 prostate cancer according to the contemporary grading criteria [2,3] were retrieved from archives in the Pathology Department at our institution under an Institutional Review Board approval no. Pro00029960. Slides were digitized by a high resolution whole slide scanner SCN400F (Leica Biosystems, Buffalo Grove, IL) dedicated for pathology research. The scanning objective was set to 20× and the focusing was automatically adjusted by the scanner. The output was a color RGB image with the pixel size of 0.5 μm × 0.5 μm and 8 bit intensity depth for each color channel. We utilized freely available libraries from OpenSlide.org [20] to import Leica (.scn) images, select histopathologically important fields of view (FOV) that were converted to TIFFs for methods development. Furthermore, the FOVs were split into tiles of 1200 × 1200 pixels for image analysis. Of the total 5000 tiles, a subset of 210 images was selected by pathologists (SM, SB) who identified ST,