Mapping Suspected Prostatic Carcinoma Using Smoothed Co-occurrence Texture Features

Matthew D. DiFranco¹, Gillian O’Hurley², Elaine W. Kay³, R. William G. Watson⁴, and Pádraig Cunningham⁴

¹ Department of Computer Science and Informatics, University College Dublin, Ireland.
² Royal College of Surgeons in Ireland, Dublin, Ireland.
³ Department of Pathology, Beaumont Hospital, Dublin, Ireland.
⁴ Conway Institute of Biomolecular & Biomedical Research, University College Dublin, Ireland.

Abstract. In this paper we present a system to assist in the scoring of prostatic carcinoma in histopathology images. Cancer and non-cancer regions of interest, annotated by an expert pathologist on whole-mount slides, are used to build a training set for support vector machine classification. Our strategy entails extraction of co-occurrence texture features from image tiles within these regions of interest. These features are mapped to the underlying image, and 2D smoothing is applied to these maps using mean, median, and Gaussian filters to produce new texture features. Inclusion of smoothed features in the classifier improves both sensitivity and specificity when testing on regions of interest from unseen images. Heat maps generated from tile classification probabilities highlight suspect regions on full section images for further inspection by a pathologist or CAD system. Improvements in training data and inclusion of additional texture features should improve the overall performance and robustness of the underlying methodology.

1 Introduction

Histopathology utilizes colour staining to highlight cellular structures in tissue samples and aid microscopic analysis by a pathologist. Diagnosis of prostate cancer relies in part on histological evaluation of biopsy samples to estimate disease progression. Many patients subsequently undergo radical prostatectomy (whole organ removal), and following this procedure histology is again used to assess the extent of disease in the resected organ.

This paper demonstrates a process for highlighting regions of suspected prostatic carcinoma (PCa) in histology samples taken from resected prostates. In daily practice pathologists utilize the Gleason scoring system, which is a unique prostate cancer grading system (see section 2.1 for more detail) based solely on tissue architecture as seen under a microscope in histological samples.

* Corresponding author: Matthew D. DiFranco, email: matthew.difranco@ucd.ie.
Our strategy uses global textural features to build heat maps of PCa in whole-section images. These heat maps can be used to address a number of key metrics which pathologists require in prostate cancer assessment, including overall disease volume, evidence of extracapsular extension (PCa at the edge of the organ), and Gleason grade. Small foci of carcinoma, often missed by pathologists, also show up clearly in our heat maps.

Our method uses no segmentation or localization techniques. We instead extract textural features from regularly spaced image tiles and build a support vector machine (SVM) classifier from training examples using these features. Our heat maps are generated from assignment of probability scores (based on a sigmoid transformed distance from the classifier margin) for each image tile. We also show improved results through smoothing of our texture features with mean, median, and Gaussian filters, and subsequent inclusion of the new features for training the SVM classifier.

In section 2 we present a brief background on prostate cancer histopathology and the role of image analysis in Gleason scoring. Section 3 describes our methodology, including data acquisition and image annotation (sections 3.1 and 3.2), random forest feature selection (section 3.3), and descriptions of our classification (section 3.4) and feature map smoothing (section 3.5) approaches. Results seen in section 4 show the impact of feature smoothing on classification accuracy and present examples of heat maps generated from histopathology slides. A brief discussion of results is found in section 5.

2 Background

2.1 Prostate Cancer

As of 2002, prostate cancer was found to be the fifth most common cancer in the world as measured by number of new cases, and the second most common cancer among men, with an estimated 679,023 cases globally [1]. Diagnosis of prostate cancer requires the appearance of adenocarcinoma in prostate needle biopsy cores, which are histopathology samples extracted directly from the prostate and assigned a Gleason score.

Gleason scoring is a unique prostatic carcinoma grading system created by Donald F. Gleason in 1966 which is based solely on the architectural pattern of the tumor [2–4]. Prostatic carcinoma often displays morphological heterogeneity [5]. In the majority of prostatectomy specimens, more than one histological pattern is present, so the Gleason score is defined as the sum of the two most common Gleason grade patterns in a sample. Grades range from 1-5, with scores ranging from 2-10. Cancer of a single grade is not commonly seen [6, 7].

Despite the value of Gleason score in prediction of patient outcomes, studies have consistently shown that pathologists are susceptible to errors, including difficulty with diagnosis of high-grade prostatic intraepithelial neoplasia (PIN) which can precede PCa spread, and failure to identify small foci of carcinoma [8].
2.2 Texture as Architecture

The Gleason scoring system relies on the architecture of prostatic tissue patterns, and the majority of work to date has relied in part on architectural or morphological features for computer aided detection [9–11]. Extraction of such features generally requires the segmentation or localization of tissue structures such as glandular lumen and cell nuclei [12–14]. The approach of explicitly identifying tissue structures before extracting image features seeks to translate the tissue patterns seen in digital images into a representation which fits the language of the Gleason guidelines. However, the generalization and reproducibility of such approaches proves difficult when the heterogeneity and variety of PCa patterns in histopathology images are considered.

Our approach seeks to avoid segmentation or localization techniques, instead relying on texture features which provide a sufficient representation of the underlying tissue architecture. Haralick et al. state that textural information for an image can be found by analyzing the spatial relationship of the gray-tones within the image [15]. Since it is ultimately spatial information that pathologists seek when assigning a Gleason score to prostate tissue, textural features, and the manipulation of such features, should prove effective in providing the building blocks for an automated Gleason scoring system.

In the work presented here we use co-occurrence texture features extracted from high-resolution image tiles as input to an SVM classifier in order to predict individual tiles in histopathology images as belonging to cancerous or non-cancerous regions. We are then able to improve classification results by applying smoothing filters to the original textural features and reclassifying the tiles. Heat maps of these tile scores provide a compelling way of reporting these results to the pathologist for follow-up examination.

Other texture features have already proved successful in prostate histopathology classification, including Gabor filters [16], multiwavelets [17] and fractal dimension [18]. Our feature smoothing technique could be applied to such features, but is beyond the scope of this work.

3 Methodology

3.1 Data Acquisition and Feature Extraction

Histopathology samples used in this study have been obtained through the Prostate Cancer Research Consortium (PCRC), a program under the Dublin Molecular Medicine Centre (DMMC). Slides have been prepared from paraffin-embedded prostates removed surgically from patients during radical prostatectomy. The slides are stained with hematoxylin-and-eosin (H&E) and have already been assigned a Gleason score by a clinician. Digital images are produced using an Aperio ScanScope XT System under 40x magnification.

Our approach begins with subdivision of each image into 512x512 pixel tiles at 40x magnification. Each tile within an annotated region of interest (see section 3.2) then becomes a training example. 90 colour and texture features are
Fig. 1. Features are extracted at the tile level from regions of interest in the training images. A subset of training tiles are randomly selected to provide a balance between cancer and non-cancer for model building. Tile key: G3 (yellow), G4 (blue), G5 (light blue), PIN (Red), BPH (green), BS (pink), Inflammation (Black)

extracted from each tile. Each tile is then resized to 256x256 pixels using bicubic interpolation (equivalent to 20x magnification), and the same 90 features are again extracted.

Features include histogram statistics (mean, range, variance, standard deviation, skewness, and kurtosis) for each channel of the RGB and HGV color spaces for a total of 36 features. In addition, a gray-level co-occurrence matrix (GLCM) is constructed for each channel of the RGB and CIEL*a*b* colour spaces. For each GLCM, nine Haralick co-occurrence features are extracted: energy, contrast, homogeneity, inverse difference moment, entropy, correlation, variance, and information measure of correlation I and II. In total there are 54 GLCM features.

3.2 Annotation

Annotations are performed using Aperio’s ImageScope software. Regions of interest (ROIs) classed as Cancer include Gleason 3 (G3), Gleason 4 (G4), Gleason 5 (G5), and Prostatic Intraepithelial Neoplasia (PIN). Noncancer annotations include Benign Prostatic Hyperplasia (BPH), Benign Stroma (BS), and Inflammation (INF). Figure 1 shows example ROIs. All annotations are carried out by an expert pathologist, and ROIs are annotated to be unambiguous (i.e., the pathologist is confident that an ROI contains a single class of tissue pattern).

Because we impose a regular tile grid, ROI boundaries inevitably bisect tiles. We considered a tile to be within an ROI if 15/16ths of the tile area was contained within the ROI. This threshold is applied by creating a binary mask for each ROI and removing tiles from the ROI that do not meet the criteria. Such a strict
boundary condition is chosen to ensure that training data is representative of the expert annotations.

3.3 Random Forest Feature Selection

Random Forest classification is an ensemble technique which utilises weak learners (many classification trees) to form one strong learner [19]. In addition to providing an accurate classifier, the process of building a forest of diverse decision trees yields extra information about the data. In particular it provides a ranking of the importance of the features in the classification process [20]. This is achieved using a perturbation principle. If a feature is important in the classification process then perturbing the values of that feature will have a significant influence on classification performance. By contrast, perturbing a feature that is not useful will have little impact on generalisation accuracy. Evaluations have shown [19, 20] that this is a very effective mechanism for assessing the contribution of features in the context of other features in classification.

Feature selection is performed on a training set of annotated regions from seven whole-section images, comprising 5687 cancer and 2521 non-cancer tiles, using the 90 features described in section 3.1 at 40x magnification, then at 20x magnification, and finally on a combined set of 40x and 20x.

The 30 most important features are captured from the Random Forest algorithm for each training set. These feature sets are further subdivided by removing features which depend on illumination (RGB, Value from HSV, and Luminance in CIEL*a*b*). Each permutation of features is then used to build an SVM training model.

3.4 Support Vector Machine Classification

We classify the image tiles using a support vector machine (SVM) algorithm implemented with LibSVM[21]. A radial basis function kernel is chosen, and a grid search is performed to optimise the cost and gamma parameters of the SVM. We train the SVM using the feature sets described in section 3.3 over the entire training set. For comparison, we also train an SVM with a linear kernel. Evaluation of SVM performance is carried out using 9856 tiles from annotated regions of interest (70% cancer and 30% non-cancer tiles) in three unseen images.

Our training data comes from annotated regions from eight images (the seven images used in feature selection, plus an additional image). However, we choose to subsample the tiles from these training images in order to train the classifiers with a balanced set of cancer and non-cancer data randomly sampled from the entire training set. The training set included 1582 cancer (700 G3, 700 G4, 170 G5, 12 PIN) and 1481 non-cancer (700 BPH, 700 BS, 81 INF) (see figure 1 for a breakdown of the training data). We balance the classes in order to minimize bias towards a given class in the training phase. Subsampling also improves the efficiency of training without sacrificing performance, as we found little improvement in classification accuracy above the subsampled training set size specified here.
3.5 Feature Map Smoothing

Because each feature can be assigned to a tile of the original image, feature maps of the tiled image can be created. Each tile in a feature map is independent of the neighboring tile owing to the fact that no tile overlap was used in the feature extraction process.

Benign tissue, particularly BPH, is often defined by relatively large glands separated by regions of stroma. When the gland lumen area and the interstitial stroma area exceeds the tile size, the tiles no longer characterize the texture of the underlying tissue. To account for this problem, we produce a new set of feature maps by smoothing the original feature maps with a 2D filter. For this analysis, we chose three types of smoothing filters: a mean filter, a median filter, and a symmetrical 2D Gaussian filter. Feature smoothing was applied to the training set of 3063 tiles sampled as described at the end of section 3.4.

4 Results

4.1 Feature Set Evaluation

Feature set evaluation was performed by predicting tiles from annotated regions of images unseen by the SVM training model. A total of 9856 tiles were included in the test set, taken from three unseen images. Classification accuracies are shown in Table 1.

We find nearly identical classification results between features from 20x, 40x, or both 20x and 40x, with the best accuracy, 87%, found using the RBF kernel and including illumination-based color channels. There is, however, no significant change in classification accuracy when using only texture features based on illumination-independent color channels. However, inclusion of convolution features results in accuracies over 90% for all feature sets. Inspection of classification errors reveals that BPH produces a high false positive rate, and that the inclusion of convolution features reduces these false positives, as can be seen in Section 4.2.

4.2 Impact of Feature Map Smoothing

The addition of smoothed features in the training data results in an increase in classifier accuracy, including an increase in both sensitivity and specificity, as shown in figure 2. Using Area under ROC curves (AUC) in figure 2, we show that, for each filter type, classification errors decrease with increasing filter width up to a maximum AUC. The red ROC curves correspond to those optimum widths ($\sigma = 1.0$ for the Gaussian filter, 9x9 for the median filter, and 5x5 for the mean filter).
Fig. 2. The left column shows the effect of feature convolution using (a) Gaussian, (c) median, and (e) mean filters. The right column shows corresponding ROC curves for each filter up to a filter width corresponding to maximum AUC.
Fig. 3. Heatmap results for High-grade PCa. (a) Annotated snapshot. (b) Heatmap from texture features alone. (c) After Gaussian feature convolution. (d) After median filter convolution. Annotation key: G3 (yellow), G4 (blue), G5 (light blue), BPH (green), BS (pink)
Fig. 4. Heatmap results for low-grade PCa (a-d) and both high- and low-grade PCa (e-h). (a,e) Annotated snapshot. (b,f) Heatmap from texture features alone. (c,g) After Gaussian feature convolution. (d,h) After median filter convolution. See figure 3 for annotation legend.
Table 1. Classification accuracies for different feature combinations, with and without convolution features. Results are reported for SVM classifiers using a radial basis kernel function (RBF), as well as using a linear SVM kernel.

<table>
<thead>
<tr>
<th>Resolution</th>
<th>Features</th>
<th>Count</th>
<th>No Convolution (%)</th>
<th>Convolution (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>RBF</td>
<td>Linear</td>
</tr>
<tr>
<td>20x</td>
<td>All</td>
<td>30</td>
<td>86.62</td>
<td>83.20</td>
</tr>
<tr>
<td>20x</td>
<td>Illum. Inv.</td>
<td>8</td>
<td>86.73</td>
<td>86.52</td>
</tr>
<tr>
<td>40x</td>
<td>All</td>
<td>30</td>
<td>87.02</td>
<td>86.33</td>
</tr>
<tr>
<td>40x</td>
<td>Illum. Inv.</td>
<td>9</td>
<td>86.23</td>
<td>85.58</td>
</tr>
<tr>
<td>20x,40x</td>
<td>All</td>
<td>30</td>
<td>86.89</td>
<td>86.16</td>
</tr>
<tr>
<td>20x,40x</td>
<td>Illum. Inv.</td>
<td>15</td>
<td>86.53</td>
<td>85.41</td>
</tr>
</tbody>
</table>

* Feature counts are doubled for convolution sets (i.e., 30 becomes 60, 8 becomes 16, etc.)

4.3 Heat Map Visualization

Heat maps produced for the three test images are shown in figures 3 and 4. In each figure, the original annotated snapshot is shown along with three heatmaps: no convolution features, Gaussian filter features, and Median filter features (mean filter maps are not shown). It should be noted that the classification accuracies reported in figure 2 come from the annotated regions of these three images, but not the whole images.

The heat maps are produced by mapping the color of each tile to the probability score output by the SVM classifier. This probability is based on the distance from the SVM margin between cancer and non-cancer tiles in the feature space. Cancer tiles are labelled red, while non-cancer are labelled blue. Tiles which fall near the margin are labelled green.

The heat maps with no filtering are noisy, making it hard to isolate areas of suspected carcinoma for the pathologist. Inclusion of the convolution features improves classification accuracy on a per-tile basis, while also producing a more visually meaningful heat map. Note that the annotated cancer and non-cancer regions are well-discriminated in the heat maps.

5 Discussion and Conclusions

Histopathology is defined by the use of color stains to highlight and differentiate important structures and features in tissue samples. Texture features alone can be used to identify cancerous tissue in resected prostate specimens. The challenge of minimizing false positives (seen in many other CAD domains, not just prostate cancer detection) has been addressed here using the technique of feature convolution. Furthermore, the aggregation of tile-level classification results into visually meaningful heat maps using only basic textural features suggests
a need for further analysis of global features for use in prostate histopathology CAD systems.

Our results from the inclusion of smoothing features indicate a potential method for exploiting texture features to build more robust classifiers for automated Gleason scoring. We identify optimum filter widths for these smoothing filters, but we believe that these values may be dictated by the test images given the heterogeneity of tissue samples, particularly BPH. Future work should focus on finding ways to identify these characteristic filter widths for given test inputs, and also to explore the impact of combining filters.

The heat maps shown in this paper demonstrate a potentially valuable tool for aiding the pathologist in Gleason grading. Further, rigorous studies of the accuracy and utility of such heat maps could lead to more accurate reporting of extracapsular extension and tumour volume while also pin-pointing regions of uncertain grading to the pathologist and identifying small foci of carcinoma which can be easily missed during visual inspection. The heat map format can also be explored for visualization of tumour severity and other key histopathological indicators in future studies.

References