ICPR 2010
August 2010 Istanbul Turkey

Automated Gland Segmentation and Classification for Prostate Cancer Grading

Kien Nguyen and Anil K. Jain
Michigan State University
East Lansing, Michigan

Ronald L. Allen
Biolimagene
Sunnyvale, California
# United States Mortality Data

## US Mortality, 2006

<table>
<thead>
<tr>
<th>Rank</th>
<th>Cause of Death</th>
<th>No. of deaths</th>
<th>% of all deaths</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Heart Diseases</td>
<td>631,636</td>
<td>26.0</td>
</tr>
<tr>
<td>2.</td>
<td>Cancer</td>
<td>559,888</td>
<td>23.1</td>
</tr>
<tr>
<td>3.</td>
<td>Cerebrovascular diseases</td>
<td>137,119</td>
<td>5.7</td>
</tr>
<tr>
<td>4.</td>
<td>Chronic lower respiratory diseases</td>
<td>124,583</td>
<td>5.1</td>
</tr>
<tr>
<td>5.</td>
<td>Accidents (unintentional injuries)</td>
<td>121,599</td>
<td>5.0</td>
</tr>
<tr>
<td>6.</td>
<td>Diabetes mellitus</td>
<td>72,449</td>
<td>3.0</td>
</tr>
<tr>
<td>7.</td>
<td>Alzheimer disease</td>
<td>72,432</td>
<td>3.0</td>
</tr>
<tr>
<td>8.</td>
<td>Influenza &amp; pneumonia</td>
<td>56,326</td>
<td>2.3</td>
</tr>
<tr>
<td>9.</td>
<td>Nephritis*</td>
<td>45,344</td>
<td>1.9</td>
</tr>
<tr>
<td>10.</td>
<td>Septicemia</td>
<td>34,234</td>
<td>1.4</td>
</tr>
</tbody>
</table>

Some facts and figures about cancer

2009 Estimated US Cancer Deaths*

**Men (292,540)**

- Lung & bronchus: 30%
- Prostate: 9%
- Colon & rectum: 9%
- Pancreas: 6%
- Leukemia: 4%
- Liver & intrahepatic bile duct: 4%
- Esophagus: 4%
- Urinary bladder: 3%
- Non-Hodgkin lymphoma: 3%
- Kidney & renal pelvis: 3%
- All other sites: 25%

**Women (269,800)**

- 26% Lung & bronchus
- 15% Breast
- 9% Colon & rectum
- 6% Pancreas
- 5% Ovary
- 4% Non-Hodgkin lymphoma
- 3% Leukemia
- 3% Uterine corpus
- 2% Liver & intrahepatic bile duct
- 2% Brain/ONS
- 25% All other sites

Source: American Cancer Society, 2009.
Prostate cancer diagnosis

- Prostate biopsy: a procedure to obtain a sample of the prostate tissue
- Pathologists look at a tissue slide under a microscope to detect abnormal regions
  - Tedious and time-consuming job
  - Variation in diagnosis depending on the experience & skill of the pathologist

Need for an automatic diagnostic process
Image Acquisition

Image acquisition system

An image of a prostate tissue stained by H&E (Hematoxylin and eosin) method

(H&E is used in this work)
Prostate tissue structure

The tissue includes glands surrounded by stroma. Each gland consists of epithelial nuclei on the boundary and lumen in the center.

In cancer tissue, blue mucin may invade the lumen.
Gleason grading system for prostate cancer

Tissue grading is based on architectural pattern of the glands

Ref: Gleason Grading of Prostatic Carcinoma, Johns Hopkins Pathology
Classification problem

Benign (normal, grades 1 & 2) pattern       Grade 3 pattern       Grade 4 pattern

Not easy to find data for grade 5 cancer
• **Non-segmentation based approach:**
  – Diamond et al. classify regions of interest into stroma, carcinoma and normal tissue based on texture features; 79.3% accuracy on 8 images
  – Doyle et al. extracted 594 features (first-order statistics, texture, wavelet) at 3 different image scales for classifying a tissue pattern as cancerous or non-cancerous; 88% accuracy on 22 test images

*Limitation:*
Did not analyze the tissue structure which is the grading criteria

• **Segmentation-based approach:**
  – Naik et al. segment glands from the tissue. Eight shape features for lumen and for gland inner boundary are computed; 86.35% (grade 3 vs benign), 92.9% (grade 4 vs benign), 95.19% (grade 3 vs grade 4) on a dataset of 44 images

*Limitation:*
Did not utilize the nuclei and blue mucin information in the tissue
Gland segmentation based approach

Given an input pattern:

1. Segment glands from background
2. Extract gland features (from lumen, nuclei, mucin and gland morphology)
3. Classify the pattern using multilayer perceptron, SVM, fusion of classifiers
Overview of the gland segmentation

Three stages of segmentation

1. Pixel classification
2. Gland boundary construction
3. Complete gland segmentation

Input image  Output of step 1  Output of step 2  Output of step 3
1. Pixel classification

- **Make use of:** color information (La*b* color system)
- **Observation:** Each pixel belongs to one of five classes: **stroma, cytoplasm, lumen, nuclei** or **blue mucin**
- **Goal:** A labeled image in which each pixel is assigned one of five labels
- **Method:**
  - Nearest neighbor classifier
  - Training samples are typical pixels from each class
- **Result:**
  - Result of step 1
2. Gland boundary construction

- **Make use of**: nuclei and cytoplasm pixels on the gland boundary
- **Observation**: Nuclei are denser at the gland boundary than other areas; Cytoplasm can be the bridge to connect nuclei
- **Goal**: Components of Nuclei-Cytoplasm (NC)
- **Method**:

  Neighborhood $N_i$ of each nuclei (size $S_1 \times S_1$)

  If (nuclei pixels in $N_i > T_1$) →
  Unify cytoplasm and nuclei pixels in $N_i$

  Result of step 2 (Nuclei-cytoplasm components)
3. Complete gland segmentation

- **Make use of**: lumen component
- **Observation**: lumen is surrounded by Nuclei-Cytoplasm (NC) components
- **Goal**: complete gland regions
- **Method**: lumen expanding

Binary image of lumen

NC components (from step 2) are separated

The lumen (left) can be unified with the boundary (right) to create a complete gland
Discard pixels which exceed a distance DMAX computed using the estimated gland size.

Estimate gland size by looking at the points that pixels meet the boundary.

Lumen boundary is expanding.

For each expanding lumen pixel, check its neighborhood. Merge the cytoplasm (purple region) in its neighborhood to the lumen component.

The NC components surrounding glands.
Complete Segmentation Result

Segmented glands are highlighted

Segmented gland regions only
Gland feature extraction

15 features of each gland are extracted

<table>
<thead>
<tr>
<th>Feature type</th>
<th>Feature description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ten lumen features</td>
<td>Average, variance, max of lumen area; average, variance, max of lumen perimeter; average and variance of lumen roundedness; number of lumen and ratio of lumen area to total segment area.</td>
</tr>
<tr>
<td>Two nuclei features</td>
<td>Nuclei density and ratio of nuclei area to total segment area</td>
</tr>
<tr>
<td>Two gland size features</td>
<td>Average and variance of the distances from the lumen center to the nuclei boundary</td>
</tr>
<tr>
<td>Blue mucin feature</td>
<td>Ratio of blue mucin area to total segment area</td>
</tr>
</tbody>
</table>

Why these features?

- Large, less uniform lumen area
- Thick gland boundary (rich of nuclei)
- Small, uniform lumen area
- Thin gland boundary (fewer nuclei)
- Small lumen area
- Nuclei are more uniformly distributed

Features are averaged for all the glands within a pattern to create a 15-dim feature vector for the pattern.
**Dataset**

**Original data:** 52 10x whole-slide images with an average size of ~90,000 x 45,000

**Testing data:** 30 patterns of benign, 28 of grade 3, 20 of grade 4 cases. Average size: 501 x 526
Experimental Setup:
- Classifiers: SVM, multilayer perceptron, classifier fusion
- Validation: 10 runs of 10-fold cross validation

Results:

### Three-class classification

<table>
<thead>
<tr>
<th>Test case</th>
<th>Best Classifier</th>
<th>Accuracy (variance)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full feature set (15)</td>
<td>SVM (linear kernel, C = 10)</td>
<td>87.5% (1.11%)</td>
</tr>
<tr>
<td>Best subset of features (9)</td>
<td>Multilayer perceptron (one hidden layer, 16 hidden nodes)</td>
<td>88.4% (6.2%)</td>
</tr>
<tr>
<td>Classifier fusion with best subset</td>
<td>Max of probability of SVM &amp; Multilayer perceptron (same parameters as above)</td>
<td>88.8% (1.8%)</td>
</tr>
</tbody>
</table>

Classifier fusion reduces the variance of cross-validation

### Two-class classification

<table>
<thead>
<tr>
<th>Test case</th>
<th>Best Classifier</th>
<th>Accuracy (variance)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benign vs. grade 3</td>
<td>SVM (linear kernel, C = 10)</td>
<td>97.75% (1.35%)</td>
</tr>
<tr>
<td>Benign vs. grade 4</td>
<td>Multilayer perceptron (one hidden layer, 16 hidden nodes)</td>
<td>94% (3.55%)</td>
</tr>
<tr>
<td>Grade 3 vs. grade 4</td>
<td>SVM (linear kernel, C = 10)</td>
<td>87.3% (0.43%)</td>
</tr>
<tr>
<td>Benign vs. Carcinoma (grades 3 &amp; 4)</td>
<td>Majority voting of SVM &amp; Multilayer perceptron</td>
<td>98.58% (0.16%)</td>
</tr>
</tbody>
</table>
Hierarchical classification scheme

Input image

Majority voting of SVM & Multilayer perceptron (98.58%)

Benign

Carcinoma

SVM (87.3%)

Grade 3

Grade 4

Overall accuracy:

\[ p(\text{correct}) = p(\omega_B)p(I \in \omega_B|\omega_B) + p(\omega_{G3})p(I \in \omega_{G3}|\omega_{G3}) + p(\omega_{G4})p(I \in \omega_{G4}|\omega_{G4}) \]

\[ = \frac{30}{78} \times 98.5\% + \frac{28}{78} \times 98.5\% \times 87.3\% + \frac{20}{78} \times 98.5\% \times 87.3\% = 91\% \]

Where:

- \( \omega_B, \omega_{G3}, \omega_{G4} \) : benign, grade 3, grade 4 class
- \( p(\omega_B) \) : percent of benign samples in the dataset
- \( p(I \in \omega_B|\omega_B) \) : probability that the image is classified as benign while the true label is also benign (accuracy of benign case)
Misclassifications: (a) Grade 3 is classified as grade 4; (b) grade 4 is classified as grade 3.
Correct classifications: (c) Benign; (d) grade 3 carcinoma; (e) grade 4 carcinoma
A method to automate the prostate cancer grading is presented.

Compared to non-segmentation based method, proposed method has two advantages:

- Extracted glands not only provide grading information but also facilitate other tasks such as gland retrieval.
- Segmented glands can serve as landmarks to register different images (by different staining methods) of a tissue to improve grading results.

Future work:

- Improve the discrimination between grade 3 and grade 4 carcinoma.
- Search for carcinoma patterns in a whole slide of tissue image (90,000 x 45,000 pixels).
- Improve the computational efficiency (current MATLAB algorithm takes about 2.2 minutes to grade a 500 x 500 image).
Thank you