Automated grading of urothelial cell carcinoma of the bladder

Ilaria Jansen, MD, PhD Candidate
Amsterdam UMC, location AMC
Disclosures

• No conflicts of interest
Bladder cancer

• 9th most common form of cancer
• 90% of cases urothelial cell carcinoma (UCC)
  • Non-muscle invasive bladder cancer (NMIBC)
    • 80% of cases

• High recurrence rate: 60-70%
  • Most expensive type of cancer
Histology

- Gold standard: treatment decision making
- Grading
  - Low interobserver agreement
    - WHO’73: 38-89%
    - WHO’04: 43-100%
Aim

• Automatic detection and grading of urothelial cell carcinoma
• Compare results with three experienced pathologists
Data selection

• Transurethral resection between 2000-2016, three hospitals
  • FFPE, H&E stained
  • Digitized using a Philips InstelliSite UltraFast scanner

• 328 Tissue samples from 232 patients
Grading

• Three pathologists, 8-11 years of experience
• Grading: WHO’73 & WHO’04
  • Disagreement: consensus meeting
Manual annotations

- Manual annotations:
  - UCC
  - Healthy urothelium
  - Fibro-vascular tissue
  - Exclude
- Checked by uropathologist

- 32 billion annotated pixels, only 51 million pixels non-anaplastic urothelium
Class imbalance

- Only WHO’04 grading system
  - Low-grade
  - High-grade

<table>
<thead>
<tr>
<th>Category</th>
<th>% of annotated pixels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthy</td>
<td>0.16</td>
</tr>
<tr>
<td>1 - low</td>
<td>7.14</td>
</tr>
<tr>
<td>2 - low</td>
<td>48.14</td>
</tr>
<tr>
<td>2 - high</td>
<td>21.06</td>
</tr>
<tr>
<td>3 - high</td>
<td>23.50</td>
</tr>
</tbody>
</table>
Classification of urothelium

• Two-step approach:
  • Segmentation network
  • Classification network
• Dataset divided on patient-level
  • 60% training  → train the network
  • 20% validation → select the best network
  • 20% test      → assess the accuracy
Annotations → U-Net → Classification → VGG 16 → Results
Segmentation network

Urothelium detector

• U-Net
  • Urothelium detection
  • Weighted cross-entropy
Classification network
Grading

- Undefined
- WHO’04 Low
- WHO’04 High

Classes:
- WHO’04 Low-grade
- WHO’04 High-grade
- Undefined
  - No weight assigned to patches in the undefined class
U-Net Classification VGG 16 Results

ESP NICE | September 2019

Annotations → U-Net → Classification → VGG 16 → Results

- undefined
- low-grade
- high-grade
Accuracy analysis

• Performance measures
• Concordance using linear weighted Kappa
  • Automated vs. consensus
  • Pathologist vs. pathologist

Segmentation network
Urothelium detector

- Urothelium accurately detected in 93% of slides
- U-Net detected more urothelium than annotated
  - Von Brunn’s nests
  - Mechnical artefacts (sectioning or cauterisation)
- 13% False negative/positive regions
  - Extensive color loss
  - Inflammation
Classification network
Grading

• Correct classification
  • Low-grade 76%
  • High-grade 71%

• Automated vs. Consensus \( \kappa = 0.48 \pm 0.14 \) se
• Pathologist vs. Pathologist \( \kappa = 0.35 \pm 0.13 \) se to \( \kappa = 0.52 \pm 0.13 \) se

<table>
<thead>
<tr>
<th>Kappa’s levels of agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.21 - 0.40</td>
</tr>
<tr>
<td>0.41-0.60</td>
</tr>
<tr>
<td>0.61-0.80</td>
</tr>
<tr>
<td>0.81-1.00</td>
</tr>
</tbody>
</table>
Conclusion

• Automated grading method within range of agreement between pathologists

• Future: patient outcome as output
Team

Amsterdam UMC, location AMC
Department of Biomedical Engineering & Physics
• Marit Lucas, MSc
• Henk A. Marquering
• Ton G. van Leeuwen

Amsterdam UMC, location VUMc
Department of Urology
• Judith Bosschieter
• Jakko A. Nieuwenhuijzen

Department of Urology
• D. Martijn de Bruin

Department of Pathology
• C. Dilara Savci-Heijink
• Sybren. L. Meijer
• Onno J. de Boer
<table>
<thead>
<tr>
<th></th>
<th>Kappa (se)</th>
<th>Accuracy (%)</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automated vs. consensus</td>
<td>0.48 (0.14)</td>
<td>74</td>
<td>71</td>
<td>76</td>
</tr>
<tr>
<td>Observer 1 vs. consensus</td>
<td>0.38 (0.12)</td>
<td>69</td>
<td>100</td>
<td>38</td>
</tr>
<tr>
<td>Observer 2 vs. consensus</td>
<td>0.81 (0.09)</td>
<td>91</td>
<td>91</td>
<td>91</td>
</tr>
<tr>
<td>Observer 3 vs. consensus</td>
<td>0.62 (0.12)</td>
<td>81</td>
<td>86</td>
<td>76</td>
</tr>
<tr>
<td>Automated vs. Observer 1</td>
<td>0.35 (0.11)</td>
<td>67</td>
<td>100</td>
<td>36</td>
</tr>
<tr>
<td>Automated vs. Observer 2</td>
<td>0.38 (0.14)</td>
<td>69</td>
<td>70</td>
<td>68</td>
</tr>
<tr>
<td>Automated vs. Observer 3</td>
<td>0.48 (0.13)</td>
<td>74</td>
<td>80</td>
<td>68</td>
</tr>
<tr>
<td>Observer 1 vs. Observer 2</td>
<td>0.38 (0.11)</td>
<td>69</td>
<td>63</td>
<td>100</td>
</tr>
<tr>
<td>Observer 1 vs. Observer 3</td>
<td>0.35 (0.13)</td>
<td>69</td>
<td>65</td>
<td>88</td>
</tr>
<tr>
<td>Observer 2 vs. Observer 3</td>
<td>0.52 (0.13)</td>
<td>76</td>
<td>81</td>
<td>71</td>
</tr>
</tbody>
</table>
Limitations

- Grading assessed on specimen level
- Relative small number of patients
- Use of one grading system (WHO’04)
Network selection

- Network design
  - ResNet
    - 50
    - 101
    - 50V2
    - 101V2
  - DenseNet
    - 121
    - 169
- VGG
  - VGG 16
  - VGG 19
- Position of transfer learning
  - High-level
  - Mid-level
  - All levels
- Optimizers, learning rates etc.