Assessment of mitosis detection algorithms

2013

AMIDA13 | MICCAI Grand challenge
September 22nd, Nagoya, Japan
Outline

- Introduction;
  - Digital pathology at UMC Utrecht;
  - Importance of assessment of tumor proliferation;
- AMIDA13 data set overview;
  - Observer agreement;
- AMIDA13 statistics;
- Results from the challenge;
  - After the presentation of the methods;
Introduction

- Digital pathology at UMC Utrecht\(^1\);
- Medium size pathology laboratory handling more than 144,000 surgical pathology slide year;
- Since 2007 there is full digital slide archiving (all slides are scanned);
- We are interested mainly in automatic image analysis methods that work with H&E stained slides;

\(^1\) Stathonikos et. al JPI 2013
Introduction

- Tumor proliferation is routinely assessed by mitosis counting;
  - Part of the Bloom and Richardson (B&R) system for grading of invasive breast carcinoma;

- The mitotic activity index (MAI) is arguably the strongest prognosticator among the 3 components of the B&R grading system;

Kaplan-Meier survival curves for node–negative patients younger than 55 years by mitotic activity index (MAI).¹

¹Baak et. al JCO 2005
Introduction

- Mitosis counting is notorious for the observer variability;
- It is also tedious and time consuming task;
- There are alternative methods for assessing tumor proliferation;
  - Immunohistochemistry;
  - Ki67, PPH3;
- The alternative methods are more expensive and require more time to perform;

PPH3 and H&E staining of the same section.
AMIDA13 data set

- 23 cases admitted at UMC Utrecht between July 2009 and January 2010;
  - Routinely prepared;
  - Selected based solely on the availability;
- One expert pathologist selected one slide per patient and one representative region per slide;
  - Following standard guidelines;
  - Representative regions vary in size from 7 mm$^2$ to 27 mm$^2$;

Example of a marked region on the glass slide.
AMIDA13 data set

- Slide digitization;
  - Aperio ScanScope XT scanner;
  - 40× magnification, 0.25 μm/pixel;
  - Manually selected focus points;
  - JPEG 2000 compression (quality factor 85);
AMIDA13 data set

- Initial experiment: mitosis counting by light microscopy and on digital slides;
- One expert pathologist at UMC Utrecht;
  - Fully traversed the marked regions;
  - Aperio ImageScope viewer;
  - Standard computer monitor;
  - Marked the locations on the digital slide;

Scatter and Bland-Altman plots of MAI by light microscopy and on digital slides
AMIDA13 data set

- Second observer;
  - Fully traversed the marked regions;
  - Pathoconsult.nl, an online slide viewing platform;
  - Standard computer monitor;
  - Marked the locations on the digital slide;
- Objects annotated by both observers were taken as ground truth;
- Discordant annotations were reviewed by a panel of two additional observers;
  - Only looked at the discordant cases (did not traverse the slide);
  - Accepted or rejected the object as ground truth;
- In this way, all ground truth objects have been agreed upon by at least two observers;
AMIDA13 data set

- Observer 1: 1088 annotations;
- Observer 2: 1599 annotations;
- Total: 2687 annotations;
- Overlap/agreement: 649 annotations;

\[
\text{Dice overlap}(\text{Observer 1, Observer 2}) = \frac{2 \times 649}{1088 + 1599} = 0.483
\]

\[\text{Dice overlap} \equiv F_1\text{-Score}\]

- Needed to be resolved: 1389 annotations;
- Total number after consensus annotation: 1157;
AMIDA13 data set

- The annotated regions were split into individual high power fields (HPF);
  - 1 HPF = 0.5×0.5 mm² = 2000×2000 pixels;
  - Only HPFs that contain mitotic figures were included;
  - For cases with < 10 HPFs with a mitotic figure, additional “empty” fields were included;

- The 23 cases were split onto 2 groups;
  - 12 for training;
  - 11 for testing;

Marked region split into HPFs. HPFs outside of the marked region or intersecting the marker lines were not included.
AMIDA13 statistics

- Dates;
  - March 28th – Training set available for download;
  - May 27th – Testing set available for download;
  - September 8th – Deadline for submission of results that will be presented at the workshop;
- 1300 unique visits to the website;
- More than 110 registered users from more than 30 countries;
- 14 teams submitted results for evaluation;
- 8 will be presented today;
Assessment of mitosis detection algorithms
2013

Part 2: Results from the challenge
Challenging cases

- Faintly stained mitotic figures and/or mitotic figures with atypical appearance;
  - Cases #1, 2, 5 and 7 from the testing set;
Challenging cases

- Many dark nuclei that are not mitotic figures;
  - Case #3 from the testing set;
Challenging cases

- Many dark nuclei that are not mitotic figures;
  - Case #9 from the testing set;
Less challenging cases

- Rich in mitotic figures, most of the dark objects are mitotic figures;
  - Case #6 from the testing set;
Results

- Two rankings were produced;
  - According to the overall $F_1$-score (all ground truth objects are considered as single data set);
  - According to the $F_1$-score computed for each case separately;
    - The proposed methods are individually ranked for each case and an average rank is computer;
    - Cases with different number of mitotic figures are weighted equally;
    - More forgiving in case of “catastrophic” failure in small number of cases;

\[ F_1 = 2 \frac{P \times R}{P + R} \quad P = \frac{TP}{TP + FP} \quad R = \frac{TP}{TP + FN} \]
Results

- Ranking according to the overall $F_1$-score;

<table>
<thead>
<tr>
<th>Team name</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDSIA</td>
<td>0.610</td>
<td>0.612</td>
<td>0.611</td>
</tr>
<tr>
<td>DTU</td>
<td>0.427</td>
<td>0.555</td>
<td>0.483</td>
</tr>
<tr>
<td>SURREY</td>
<td>0.357</td>
<td>0.332</td>
<td>0.344</td>
</tr>
<tr>
<td>ISIK</td>
<td>0.306</td>
<td>0.351</td>
<td>0.327</td>
</tr>
<tr>
<td>PANASONIC</td>
<td>0.336</td>
<td>0.310</td>
<td>0.322</td>
</tr>
<tr>
<td>CCIPD/MINDLAB</td>
<td>0.353</td>
<td>0.291</td>
<td>0.319</td>
</tr>
<tr>
<td>WARWICK</td>
<td>0.171</td>
<td>0.552</td>
<td>0.261</td>
</tr>
<tr>
<td>POLYTECH/UCLAN</td>
<td>0.186</td>
<td>0.263</td>
<td>0.218</td>
</tr>
<tr>
<td>MINES</td>
<td>0.139</td>
<td>0.490</td>
<td>0.217</td>
</tr>
<tr>
<td>SHEFFIELD/SURREY</td>
<td>0.119</td>
<td>0.107</td>
<td>0.113</td>
</tr>
<tr>
<td>SEOUL</td>
<td>0.032</td>
<td>0.630</td>
<td>0.061</td>
</tr>
<tr>
<td>NTUST</td>
<td>0.011</td>
<td>0.685</td>
<td>0.022</td>
</tr>
<tr>
<td>UNI-JENA</td>
<td>0.007</td>
<td>0.077</td>
<td>0.013</td>
</tr>
<tr>
<td>NIH</td>
<td>0.002</td>
<td>0.049</td>
<td>0.003</td>
</tr>
</tbody>
</table>
Results

- Ranking according to the average rank;
Results

- Ranking according to the average rank;
  - F1-scores per case;

<table>
<thead>
<tr>
<th>Team name</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8*</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>Average rank</th>
<th>Average F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>IDSIA</td>
<td>.50</td>
<td>.00</td>
<td>.66</td>
<td>.57</td>
<td>.29</td>
<td>.73</td>
<td>.00</td>
<td>.46</td>
<td>.60</td>
<td>.64</td>
<td>2.364</td>
<td>0.445</td>
</tr>
<tr>
<td>2</td>
<td>DTU</td>
<td>.00</td>
<td>.11</td>
<td>.41</td>
<td>.56</td>
<td>.20</td>
<td>.63</td>
<td>.00</td>
<td>.35</td>
<td>.61</td>
<td>.68</td>
<td>2.818</td>
<td>0.352</td>
</tr>
<tr>
<td>3</td>
<td>WARWICK</td>
<td>.13</td>
<td>.34</td>
<td>.14</td>
<td>.59</td>
<td>.40</td>
<td>.62</td>
<td>.00</td>
<td>.3</td>
<td>.38</td>
<td>.34</td>
<td>4.000</td>
<td>0.302</td>
</tr>
<tr>
<td>4</td>
<td>SEOUL</td>
<td>.07</td>
<td>.18</td>
<td>.02</td>
<td>.50</td>
<td>.20</td>
<td>.54</td>
<td>.00</td>
<td>.3</td>
<td>.41</td>
<td>.35</td>
<td>5.273</td>
<td>0.231</td>
</tr>
<tr>
<td>5</td>
<td>ISIK</td>
<td>.00</td>
<td>.10</td>
<td>.37</td>
<td>.44</td>
<td>.00</td>
<td>.47</td>
<td>.00</td>
<td>.16</td>
<td>.26</td>
<td>.46</td>
<td>5.545</td>
<td>0.226</td>
</tr>
<tr>
<td>6</td>
<td>PANASONIC</td>
<td>.00</td>
<td>.00</td>
<td>.28</td>
<td>.20</td>
<td>.40</td>
<td>.52</td>
<td>.00</td>
<td>.10</td>
<td>.30</td>
<td>.33</td>
<td>5.818</td>
<td>0.213</td>
</tr>
<tr>
<td>7</td>
<td>CCIPD/ MINDLAB</td>
<td>.00</td>
<td>.15</td>
<td>.36</td>
<td>.38</td>
<td>.00</td>
<td>.42</td>
<td>.00</td>
<td>.07</td>
<td>.38</td>
<td>.33</td>
<td>6.091</td>
<td>0.208</td>
</tr>
<tr>
<td>8</td>
<td>SURREY</td>
<td>.00</td>
<td>.10</td>
<td>.48</td>
<td>.30</td>
<td>.00</td>
<td>.47</td>
<td>.00</td>
<td>.14</td>
<td>.19</td>
<td>.38</td>
<td>6.273</td>
<td>0.205</td>
</tr>
<tr>
<td>9</td>
<td>MINES</td>
<td>.00</td>
<td>.14</td>
<td>.11</td>
<td>.40</td>
<td>.16</td>
<td>.49</td>
<td>.00</td>
<td>.07</td>
<td>.32</td>
<td>.34</td>
<td>6.273</td>
<td>0.203</td>
</tr>
<tr>
<td>10</td>
<td>POLYTECH/ UCLAN</td>
<td>.00</td>
<td>.00</td>
<td>.07</td>
<td>.18</td>
<td>.00</td>
<td>.50</td>
<td>.00</td>
<td>.04</td>
<td>.30</td>
<td>.39</td>
<td>6.909</td>
<td>0.148</td>
</tr>
<tr>
<td>11</td>
<td>SHEFFIELD/ SURREY</td>
<td>.03</td>
<td>.07</td>
<td>.35</td>
<td>.15</td>
<td>.00</td>
<td>.05</td>
<td>.00</td>
<td>.09</td>
<td>.08</td>
<td>.18</td>
<td>8.636</td>
<td>0.099</td>
</tr>
<tr>
<td>12</td>
<td>NTUST</td>
<td>.01</td>
<td>.03</td>
<td>.01</td>
<td>.09</td>
<td>.03</td>
<td>.23</td>
<td>.02</td>
<td>.29</td>
<td>.01</td>
<td>.05</td>
<td>.21</td>
<td>9.455</td>
</tr>
<tr>
<td>13</td>
<td>UNI-JENA</td>
<td>.00</td>
<td>.00</td>
<td>.01</td>
<td>.02</td>
<td>.00</td>
<td>.06</td>
<td>.00</td>
<td>.10</td>
<td>.01</td>
<td>.01</td>
<td>.07</td>
<td>10.455</td>
</tr>
<tr>
<td>14</td>
<td>NIH</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>632</td>
<td>.01</td>
<td>.00</td>
<td>.01</td>
<td>11.273</td>
<td>0.002</td>
</tr>
</tbody>
</table>

* Number of false positives
Results

- Ranking according to the average rank;
  - Rankings per case;

<table>
<thead>
<tr>
<th>Team name</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>Average rank</th>
<th>Average F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDSIA</td>
<td>1</td>
<td>10</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2.364</td>
<td>0.445</td>
</tr>
<tr>
<td>DTU</td>
<td>7</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2.818</td>
<td>0.352</td>
</tr>
<tr>
<td>WARWICK</td>
<td>2</td>
<td>1</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>9</td>
<td>6</td>
<td>4</td>
<td>7</td>
<td>4.000</td>
<td>0.302</td>
</tr>
<tr>
<td>SEOUl</td>
<td>3</td>
<td>2</td>
<td>11</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>9</td>
<td>10</td>
<td>3</td>
<td>6</td>
<td>5.273</td>
<td>0.231</td>
<td></td>
</tr>
<tr>
<td>ISIK</td>
<td>7</td>
<td>6</td>
<td>4</td>
<td>5</td>
<td>8</td>
<td>8</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>9</td>
<td>3</td>
<td>5.545</td>
<td>0.226</td>
</tr>
<tr>
<td>PANASONIC</td>
<td>7</td>
<td>10</td>
<td>7</td>
<td>9</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>8</td>
<td>9</td>
<td>5.818</td>
<td>0.213</td>
</tr>
<tr>
<td>CCIPD/MINDLAB</td>
<td>7</td>
<td>3</td>
<td>5</td>
<td>7</td>
<td>8</td>
<td>10</td>
<td>2</td>
<td>1</td>
<td>9</td>
<td>5</td>
<td>10</td>
<td>6.091</td>
<td>0.208</td>
</tr>
<tr>
<td>SURREY</td>
<td>7</td>
<td>6</td>
<td>2</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>2</td>
<td>8</td>
<td>4</td>
<td>10</td>
<td>5</td>
<td>6.273</td>
<td>0.205</td>
</tr>
<tr>
<td>MINES</td>
<td>7</td>
<td>4</td>
<td>9</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>2</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>8</td>
<td>6.273</td>
<td>0.203</td>
</tr>
<tr>
<td>POLYTECH/UCLAN</td>
<td>7</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>8</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td>11</td>
<td>7</td>
<td>4</td>
<td>6.909</td>
<td>0.148</td>
</tr>
<tr>
<td>SHEFFIELD/SURREY</td>
<td>4</td>
<td>8</td>
<td>6</td>
<td>11</td>
<td>8</td>
<td>13</td>
<td>2</td>
<td>13</td>
<td>7</td>
<td>11</td>
<td>9</td>
<td>8.636</td>
<td>0.099</td>
</tr>
<tr>
<td>NTUST</td>
<td>5</td>
<td>9</td>
<td>12</td>
<td>12</td>
<td>7</td>
<td>11</td>
<td>1</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>9.455</td>
<td>0.068</td>
</tr>
<tr>
<td>UNI-JENA</td>
<td>7</td>
<td>10</td>
<td>13</td>
<td>13</td>
<td>8</td>
<td>12</td>
<td>2</td>
<td>11</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>10.455</td>
<td>0.016</td>
</tr>
<tr>
<td>NIH</td>
<td>6</td>
<td>10</td>
<td>14</td>
<td>14</td>
<td>8</td>
<td>14</td>
<td>2</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>11.273</td>
<td>0.002</td>
</tr>
</tbody>
</table>
## Results

- Ranking according to the average rank;
  - Rankings per case;

<table>
<thead>
<tr>
<th>Team name</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>Average rank</th>
<th>Average F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDSIA</td>
<td>1</td>
<td>10</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2.364</td>
<td>0.445</td>
</tr>
<tr>
<td>DTU</td>
<td>7</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2.818</td>
<td>0.352</td>
</tr>
<tr>
<td>WARWICK</td>
<td>2</td>
<td>1</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>9</td>
<td>6</td>
<td>4</td>
<td>7</td>
<td>4.000</td>
<td>0.302</td>
</tr>
<tr>
<td>SEOUL</td>
<td>3</td>
<td>2</td>
<td>11</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>9</td>
<td>10</td>
<td>3</td>
<td>6</td>
<td></td>
<td>5.273</td>
<td>0.231</td>
</tr>
<tr>
<td>ISIK</td>
<td>7</td>
<td>6</td>
<td>4</td>
<td>5</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>9</td>
<td>3</td>
<td></td>
<td>5.545</td>
<td>0.226</td>
</tr>
<tr>
<td>PANASONIC</td>
<td>7</td>
<td>10</td>
<td>7</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>8</td>
<td>9</td>
<td></td>
<td>5.818</td>
<td>0.213</td>
</tr>
<tr>
<td>CCIPD/MINDLAB</td>
<td>7</td>
<td>3</td>
<td>5</td>
<td>7</td>
<td>8</td>
<td>10</td>
<td>2</td>
<td>1</td>
<td>9</td>
<td>5</td>
<td>10</td>
<td>6.091</td>
<td>0.208</td>
</tr>
<tr>
<td>SURREY</td>
<td>7</td>
<td>6</td>
<td>2</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>2</td>
<td>8</td>
<td>4</td>
<td>10</td>
<td>5</td>
<td>6.273</td>
<td>0.205</td>
</tr>
<tr>
<td>MINES</td>
<td>7</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>2</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>8</td>
<td></td>
<td>6.273</td>
<td>0.203</td>
</tr>
<tr>
<td>POLYTECH/UCLAN</td>
<td>7</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>8</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td>11</td>
<td>7</td>
<td>4</td>
<td>6.909</td>
<td>0.148</td>
</tr>
<tr>
<td>SHEFFIELD/SURREY</td>
<td>4</td>
<td>8</td>
<td>6</td>
<td>11</td>
<td>8</td>
<td>13</td>
<td>2</td>
<td>13</td>
<td>7</td>
<td>11</td>
<td>12</td>
<td>8.636</td>
<td>0.099</td>
</tr>
<tr>
<td>NTUST</td>
<td>5</td>
<td>9</td>
<td>12</td>
<td>12</td>
<td>7</td>
<td>11</td>
<td>1</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>9.455</td>
<td>0.068</td>
</tr>
<tr>
<td>UNI-JENA</td>
<td>7</td>
<td>10</td>
<td>13</td>
<td>13</td>
<td>8</td>
<td>12</td>
<td>2</td>
<td>11</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>10.455</td>
<td>0.016</td>
</tr>
<tr>
<td>NIH</td>
<td>6</td>
<td>10</td>
<td>14</td>
<td>14</td>
<td>8</td>
<td>14</td>
<td>2</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>11.273</td>
<td>0.002</td>
</tr>
</tbody>
</table>
Conclusions

- Some of the methods are very sensitive to the staining variability;
  - Reflected in the results for cases #3 and #6;
- Most of the proposed methods depend heavily on the intensity as a feature;
  - This does not work well when there are many dark non-mitotic objects;
  - There are many false positives that are clearly apoptotic nuclei, lymphocytes or compressed nuclei;
- The top scoring method seems to detect some mitotic figures with faint staining and typical appearance;
In the future...

- The challenge website will soon be reopened for new submissions;
  - We will evaluate them on a regular basis;
- Expand the data set;
  - Use immunohistochemistry to obtain better ground truth;
- Examine if a combination of methods or ideas can provide better performance;
- Write an overview paper;
Acknowledgements

Image Sciences Institute, UMC Utrecht
Max A. Viergever
Josien P.W. Pluim

Pathology Department, UMC Utrecht
Paul J. van Diest
Nikolaos Stathonikos
Assessment of mitosis detection algorithms
2013

Questions?