Artificial Intelligence for polyp detection

Raf Bisschops M.D., PhD
Associate professor, head of endoscopy
University Hospitals, KU Leuven
Belgium
VOTING SLIDE

• I will be using AI in my daily endoscopy

1. Tomorrow
2. Within 1-2 years
3. Within 5-10 years
4. Never
Challenges for polyp detection where AI can help

• Quality of bowel preparation
• Quality of endoscopic equipment
• Operator dependent factors
  – Fatigue: better performance in the morning
  – Successive repetitive procedures
  – Distraction: phone calls, people asking stuff, holiday conversations, ...

→ 26% miss rate for polyps reported in back to back and CT Colonography studies.
AI = Learning from data

So performance of the system will depend on the training data and validation study.

Deep learning

Predictive algorithm → Data → Output

Courtesy C. Hassan
Training: risk of overfitting

- The algorithm performs almost perfect in the training set but not well in the validation set.
COMPUTER SAYS “No”

No neoplasia
## TRAINING (CADe)

<table>
<thead>
<tr>
<th>Regulatory approved</th>
<th>Company</th>
<th>Approved</th>
<th>N° cases</th>
<th>N° centres</th>
</tr>
</thead>
<tbody>
<tr>
<td>GI-Genius</td>
<td>Medtronic</td>
<td>CE</td>
<td>2 684</td>
<td>20</td>
</tr>
<tr>
<td>Discovery AI</td>
<td>Pentax</td>
<td>CE</td>
<td>788</td>
<td>3</td>
</tr>
<tr>
<td>CAD-EYE Detection</td>
<td>Fujifilm</td>
<td>CE</td>
<td>2 001</td>
<td>4</td>
</tr>
<tr>
<td>CAD-EYE Characterization</td>
<td>Fujifilm</td>
<td>CE</td>
<td>1 864</td>
<td>4</td>
</tr>
<tr>
<td>Endobrain-EYE</td>
<td>Cybernet</td>
<td>PMDA</td>
<td>NA</td>
<td>5</td>
</tr>
<tr>
<td><strong>Not regulatory approved</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yamada et al.</td>
<td>NEC</td>
<td>-</td>
<td>2 116</td>
<td>1</td>
</tr>
<tr>
<td>Endoscreener</td>
<td>Wision AI</td>
<td>-</td>
<td>NA</td>
<td>1</td>
</tr>
</tbody>
</table>

Courtesy C. Hassan
Does AI work for polyp detection?

- Is it as good as the human ground truth?
- No need for randomization
- Can be tested benchtop (Multicase-multireader setting)
## AI standalone performance

### Regulatory approved

<table>
<thead>
<tr>
<th>System</th>
<th>Manufacturer</th>
<th>N° cases</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>GI-Genius</td>
<td>Medtronic</td>
<td>338</td>
<td>99.70%</td>
<td>0.9% (FP)</td>
</tr>
<tr>
<td>Discovery AI</td>
<td>Pentax</td>
<td>UK</td>
<td>90%</td>
<td>80% (spec)</td>
</tr>
<tr>
<td>CAD-EYE Detection</td>
<td>Fujifilm</td>
<td>UK</td>
<td>92.90%</td>
<td>90.6% (spec)</td>
</tr>
<tr>
<td>CAD-EYE Characterization</td>
<td>Fujifilm</td>
<td>UK</td>
<td>85%</td>
<td>79.4% (spec)</td>
</tr>
<tr>
<td>Endobrain-EYE</td>
<td>Cybernet</td>
<td>UK</td>
<td>95%</td>
<td>89% (spec)</td>
</tr>
</tbody>
</table>

### Not regulatory approved

<table>
<thead>
<tr>
<th>System</th>
<th>Manufacturer</th>
<th>N° cases</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yamada et al.</td>
<td>NEC</td>
<td>808</td>
<td>97.30%</td>
<td>99% (spec)</td>
</tr>
<tr>
<td>Wang et al.</td>
<td>Wision AI</td>
<td>1 138</td>
<td>94%</td>
<td>96% (spec)</td>
</tr>
</tbody>
</table>

Courtesy C. Hassan
Does AI have a clinical benefit?

Detection

Randomized setting

- Endoscopist
- Endoscopist + AI
- RCT
- Endoscopist
- Endoscopist + AI
- Endoscopist
- Endoscopist + AI
- RCT
GI Genius: RCT shows increase in ADR


EXCLUSION POOR BOWEL PREP: BBPS 0 or 1 in a segment

AI CAD e group

- 700 patients
- 350 patients
- 341 patients
- ADR: 54.8%
- Adenoma per colonoscopy: 1.07

standard group

- 350 patients
- 344 patients
- ADR: 40.4%
- Adenoma per colonoscopy: 0.71
GI Genius: RCT shows increase in ADR

Limitations:
No non-experts or low detectors
Positive
No difference in clean withdrawal time
No increase in unnecessary resections

Back to back randomization

Wang et al Gastroenterology 2020 online
<table>
<thead>
<tr>
<th>Pt with adenomas</th>
<th>Routine-CADe (n-185)</th>
<th>CADe-Routine (n-185)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>First pass</td>
<td>49</td>
<td>64</td>
<td></td>
</tr>
<tr>
<td>Second pass</td>
<td>34</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Additional detection</td>
<td>17</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>2nd pass</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADR first pass</td>
<td>26.5 %</td>
<td>34.8%</td>
<td>0.0846</td>
</tr>
<tr>
<td>Miss rate</td>
<td>25.8%</td>
<td>17.9%</td>
<td>0.258</td>
</tr>
</tbody>
</table>

Wang et al Gastroenterology 2020 online
Back to back randomization

<table>
<thead>
<tr>
<th>Per adenoma analysis</th>
<th>Routine-CADe (n-185)</th>
<th>CADe-Routine (n-185)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>First pass</td>
<td>72</td>
<td>124</td>
<td></td>
</tr>
<tr>
<td>Second pass</td>
<td>48</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Miss rate</td>
<td>40%</td>
<td>13.9%</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Wang et al Gastroenterology 2020 online
Pooled data from 5 RCTs: 36.6% ADR versus 25.2% ADR in favour of CADe
Does AI have any clinical benefit?

- Does the endoscopist reject the TP?
- Does the endoscopist accept the FP as TP?
- Does the endoscopist detect the FN?
- Does the endoscopist confirm it as TN?

Neoplasia | No neoplasia
--- | ---
+ | TP | FP
- | FN | TN

Courtesy C. Hassan
Study design: real time unblinding to assess additional value of CADe

Sinonquel P, Eelbode T, Hassan C, ... Bisschops R. UEGW 2020 and Gut 2020 accepted
Study design: real time unblinding to assess additional value of CADe

Sinonquel P, Eelbode T, Hassan C,... Bisschops R. UEGW 2020 and Gut 2020 accepted
We defined four different situations:

(1) **Obvious false positive**: the system gives an obviously false detection (stool, air bubbles, ...).

(2) **Other positive**: after its location disappears from the image, the endoscopist is asked to return. If there is a polyp, this is an additional detection.

(3) **False negative**: the endoscopist found a polyp, but the system missed it.

(4) **True positive**: the system and endoscopist found the polyp.

Sinonquel P, Eelbode T, Hassan C,... Bisschops R. UEGW 2020 and Gut 2020 accepted
In 295 patients a total of 606 polyps were detected

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity CADe</td>
<td>96.5%</td>
</tr>
<tr>
<td>Sensitivity endoscopists</td>
<td>98.2%</td>
</tr>
<tr>
<td>Miss rate CADe, %</td>
<td>3.5</td>
</tr>
<tr>
<td>Miss rate endoscopists, %</td>
<td>1.8</td>
</tr>
<tr>
<td>False positive rate per minute clean withdrawal time (n=1663)</td>
<td>1.4</td>
</tr>
</tbody>
</table>

CAD e was non inferior to high detectors.
AID STUDY: False Positives (GI GENIUS)

- 0.7% of the mean WT
- 27.3 FP per colonoscopy
- 4.8±6.2 sec mean time spent per FP
- 0.7% of the mean WT
Conclusion

• CADe works and is here to stay
• First trials show good performance of system with increase in ADR
Future prospects

• How does quality of endoscopist affect performance of AI?
• How does AI affect the technical performance of endoscopists?
• Postmarketing follow-up of Q will be imperative after implementation