Current status and future perspectives of polyp detection and characterization

Daniel von Renteln, Yuichi Mori
COI (Yuichi Mori)

➢ Olympus Corp. (consulting and speaking honorarium)
➢ Cybernet Corp. (ownership interest)

COI (Daniel von Renteln)

➢ Boston Scientific (research grant, consulting and speaking honorarium)
➢ ERBE (research grant and speaking honorarium)
➢ Pendopharm (research grant, consulting and speaking honorarium)
➢ Pentax (research grant and speaking honorarium)
➢ Imagia (Research collaboration/grant)
OBJECTIVES

- AI Background
- Current state and future of AI in colonoscopy
  - CADe (detection), CADx (characterisation), Quality auditing
  - Available technology, studies, results
- Challenges and barriers of AI in colonoscopy
- Conclusion
EMERGENCE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence
Computer programs that mimic human cognitive functions such as learning and problem solving.

Machine learning
Computer-based methods for analyzing data and learning descriptive or predictive models.

Deep learning

Time line of artificial intelligence concepts, (le Berre et al. (2020))
Three roles of AI in colonoscopy

1. Computer-aided detection (CADe)

2. Computer-aided diagnosis (CADx)

3. Computer-aided quality assurance (CAQ)
Increased ADR: 1.47 risk ratio

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>CAD Events</th>
<th>CAD Total</th>
<th>WL Events</th>
<th>WL Total</th>
<th>Weight</th>
<th>Risk Ratio M-H, Random, 95% CI</th>
<th>Risk Ratio M-H, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gong et al., 2020</td>
<td>54</td>
<td>324</td>
<td>26</td>
<td>318</td>
<td>6.6%</td>
<td>2.04 [1.31, 3.17]</td>
<td></td>
</tr>
<tr>
<td>Su et al., 2020</td>
<td>89</td>
<td>303</td>
<td>52</td>
<td>315</td>
<td>11.7%</td>
<td>1.75 [1.29, 2.37]</td>
<td></td>
</tr>
<tr>
<td>Wang et al., 2019</td>
<td>151</td>
<td>522</td>
<td>109</td>
<td>536</td>
<td>17.9%</td>
<td>1.42 [1.15, 1.75]</td>
<td></td>
</tr>
<tr>
<td>Wang et al., 2020</td>
<td>165</td>
<td>484</td>
<td>134</td>
<td>478</td>
<td>20.2%</td>
<td>1.22 [1.01, 1.47]</td>
<td></td>
</tr>
<tr>
<td>Liu et al., 2020</td>
<td>198</td>
<td>508</td>
<td>124</td>
<td>518</td>
<td>20.4%</td>
<td>1.64 [1.38, 1.97]</td>
<td></td>
</tr>
<tr>
<td>Repici et al., 2020</td>
<td>187</td>
<td>341</td>
<td>139</td>
<td>344</td>
<td>23.2%</td>
<td>1.38 [1.16, 1.62]</td>
<td></td>
</tr>
<tr>
<td><strong>Total (95% CI)</strong></td>
<td><strong>2487</strong></td>
<td><strong>2509</strong></td>
<td><strong>100.0%</strong></td>
<td></td>
<td></td>
<td><strong>1.47 [1.30, 1.67]</strong></td>
<td></td>
</tr>
</tbody>
</table>

Total events: 845 584

Heterogeneity: $\tau^2 = 0.01$, $\chi^2 = 9.41$, df = 5 ($P = 0.09$), $I^2 = 47\%$

Test for overall effect: $Z = 6.04$ ($P < 0.00001$)

ADR: 25% $\rightarrow$ 37%
<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Study design</th>
<th>System</th>
<th>Image modality</th>
<th>Number of patients/colonoscopies used for training/test datasets (total)</th>
<th>Number of colonoscopy/polyp images/videos used for training/test datasets</th>
<th>Diagnostic properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Park and Sargent</td>
<td>2016</td>
<td>Retrospective</td>
<td>CADe based on DCNN using a conditional random field model</td>
<td>Still images</td>
<td>35 (colonoscopy videos)</td>
<td>562/562 (colonoscopy still images)</td>
<td>Sensitivity=86%; specificity=85%; AUC=0.8585</td>
</tr>
<tr>
<td>Fernández-Esparrach et al</td>
<td>2016</td>
<td>Retrospective</td>
<td>CADe based on energy map</td>
<td>Still images</td>
<td>NA/24 colonoscopy videos containing 31 different polyps</td>
<td>NA/Experiment A: 612 polyp images from all 24 videos Experiment B: 47,886 frames from the 24 videos</td>
<td>Experiment A: accuracy= small vs all polyps=77.5%, 95% CI=71.5%–82.6% vs. 66.2%, 95% CI=61.4%–70.7%; P &lt;0.01 Experiment B: the AUC=high quality frames vs all Frames= 0.79, 95% CI=0.70–0.87 vs 0.75, 95% CI=0.86–0.83</td>
</tr>
<tr>
<td>Yu et al</td>
<td>2017</td>
<td>Retrospective</td>
<td>CADe based on three-dimensional (3-D) deep learning integration framework by leveraging the 3-D fully CNN (3D-FCN)</td>
<td>Videos</td>
<td>20/18 (colonoscopy videos)</td>
<td>3,799 frames with polyps in total</td>
<td>Sensitivity=71%; PPV=88%; precision=88.1%</td>
</tr>
<tr>
<td>Billah et al</td>
<td>2017</td>
<td>Retrospective</td>
<td>CADe based on CNN and color wavelet features using a linear support vector machine</td>
<td>Still images</td>
<td>100 (colonoscopy videos for combined training and test datasets)</td>
<td>14,000 still images (combined for training and test datasets)</td>
<td>Accuracy=98.65%; sensitivity=98.79%; specificity=98.52%</td>
</tr>
<tr>
<td>Zhang et al</td>
<td>2017</td>
<td>Retrospective</td>
<td>CADe based on DCNN</td>
<td>Still images</td>
<td>NA</td>
<td>2262/150 random, 30 NBI (colonoscopy still images)</td>
<td>Accuracy=85.9%; sensitivity=98%; PPV=99%; precision=87.3%; recall rate=87.6%; AUC=1.0</td>
</tr>
<tr>
<td>Wang et al</td>
<td>2018</td>
<td>Retrospective</td>
<td>CADe based on DNN</td>
<td>Still images</td>
<td>1,290/1,138 (2428) patients</td>
<td>27,113/5,545 (colonoscopy images)</td>
<td>Sensitivity=94.38%, 95% CI=93.80%–94.96% in images with polyp; AUC=0.984</td>
</tr>
<tr>
<td>Misawa et al</td>
<td>2018</td>
<td>Retrospective</td>
<td>CADe based on CNN</td>
<td>Videos</td>
<td>59/14 (73)</td>
<td>411/135 (colonoscopy videos containing 150 polyps)</td>
<td>Per-polyp Sensitivity=94%; per-frame sensitivity=90%; specificity=63.3%; accuracy=76.5%; False positive rate=60%; AUC=0.87</td>
</tr>
<tr>
<td>Yamada et al</td>
<td>2018</td>
<td>Retrospective</td>
<td>CADe based on DNN</td>
<td>Videos</td>
<td>NA/77 (number of videos)</td>
<td>13,983/4,840 (colonoscopy videos)</td>
<td>Sensitivity=97.2%, 95% CI=95.9%–98.4%; specificity=99.0%, 95% CI=98.6%–99.2%; AUC=0.975, 95% CI=0.964–0.986</td>
</tr>
<tr>
<td>Urban et al</td>
<td>2018</td>
<td>Retrospective</td>
<td>CADe based on deep learning CNN</td>
<td>Videos</td>
<td>Several training and validation sets: 1) Cross-validation on the 8,641 images; 2) Training on the 8,641 images and testing on the 9 videos, 11 videos, and independent dataset; 3) Training on the 8,641 images and 9 videos and testing on the 11 videos and independent dataset</td>
<td>411/135 (colonoscopy videos containing 150 polyps)</td>
<td>Sensitivity=96.9%; specificity=95%; AUC=0.991; accuracy=96.4%; false positive rate=7%</td>
</tr>
<tr>
<td>Klare et al</td>
<td>2019</td>
<td>Prospective</td>
<td>automated polyp detection software (&quot;KoloPol,&quot; Fraunhofer IIS, Erlangen, Germany) based on CNN</td>
<td>Live colonoscopy videos</td>
<td>NA</td>
<td>NA/55 (colonoscopy videos)</td>
<td>Per-polyp sensitivity=75.3%, 95% CI=62.3%–84.9%, PDR=50.9%, 95% CI=37.1%–64.4%, ADR=29.1%, 95% CI=17.6%–42.9%</td>
</tr>
<tr>
<td>Ozawa et al</td>
<td>2020</td>
<td>Retrospective</td>
<td>CADe based on DCNN</td>
<td>Still images</td>
<td>12,895 patients</td>
<td>16,418/7,077</td>
<td>Sensitivity=92%; positive predictive value=86%; precision=83%; identified adenomas=97%</td>
</tr>
</tbody>
</table>
Problems

• Some studies with longer withdrawal time in CADe group
• Overall modest ADR performance (25% as comparator vs 37%)
• What is the effect for high vs low performer?
Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model

<table>
<thead>
<tr>
<th>Model based prediction</th>
<th>Nice type I</th>
<th>NICE type II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pathology</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hyperplastic</td>
<td>33</td>
<td>7</td>
</tr>
<tr>
<td>Adenoma</td>
<td>1</td>
<td>65</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Diagnostic property</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>98% (95% CI: 92% to 100%)</td>
</tr>
<tr>
<td>Specificity</td>
<td>98% (95% CI: 92% to 100%)</td>
</tr>
<tr>
<td>NPV</td>
<td>97%</td>
</tr>
<tr>
<td>PPV</td>
<td>90%</td>
</tr>
<tr>
<td>Accuracy</td>
<td>94% (95% CI: 86% to 97%),</td>
</tr>
</tbody>
</table>
CADx Problems

- Sens, spec, accuracy high but level of data preparation influences outcomes (Preselected frames, sequences)
- NICE 1 and 2, for most CADx studies removal of SSL from training and test datasets, cleaned up dataset that does not reflect clinical practise and is not granular enough to predict HGD, SSL, NICE3/Cancer (yet)
- Prospective clinical data sparse
### Author Year Study design System Number of patients/colonoscopies used for training/test datasets (total) Number of colonoscopy/polyp images/videos used in training/test datasets Diagnostic properties

**Tischendorf et al 2010** Prospective pilot CADx based on SVMs NA/128 Colonoscopy videos NA/209 polyps containing 160 neoplastic and 49 non-neoplastic polyps in the test dataset **CADx**: sensitivity=90%, specificity=91%, correct classification rate=93%. 
**Consensus decision between the human Observers: sensitivity=93.8%, specificity=85.7%, correct classification rate=91.9%.**

**Gross et al 2011** Prospective CADx based on SVMs NA/214 patients in the test dataset NA/434 polyp images with the maximum size of 10 mm containing 258 neoplastic and 176 non-neoplastic polyps in the test dataset **Accuracy=93.1%; sensitivity=95.0%; specificity=93.3%; correct classification rate=95.3% for non-neoplastic and 95% for adenomatous polyps**

**Takeamura et al 2012** Retrospective CADx based on support vector machines (SVMs) NA NA/137 neoplastic and 39 non-neoplastic polyps in the test dataset **Accuracy=97.8%; sensitivity=97.8% for type B-C3 lesions for a diagnosis of neoplastic lesion; specificity=97.9%**

**Aihara et al 2013** Prospective CADx based on numerical color analysis of autofluorescence endoscopy as an Adobe AIR application NA/32 patients in the test dataset NA/102 lesions containing 75 neoplastic lesions in the test dataset **Sensitivity=94.2%; specificity=88.8%; PPV=95.6%; NPV=85.2%**

**Mori et al 2015** Retrospective pilot CADx (EC-CAD) based on CNN NA/152 patients in the test dataset NA/176 small polyps in the test dataset containing 137 neoplastic and 39 non-neoplastic polyps for the test dataset **Accuracy=98.2%, 95% CI=83.7%–93.4%; Sensitivity= 92.0%, 95% CI=80.3%–93.5%; Specificity=96.1%–95.9%; specificity of 78.4%, 95% CI=0.65%–90.7%**

**Kuiper et al 2015** Retrospective CADx (WavSTAT) based on CNN NA/87 patients in the test dataset NA/207 small lesions in the test dataset **Accuracy=74.4%, 95% CI=68.1%–79.7%; sensitivity=85.3%, 95% CI=0.78–0.90; specificity=58.8%, 95% CI=0.48–0.69; PPV=74.8%, 95% CI=0.67–0.81; NPV=73.5%; accuracy of on-site recommended surveillance interval=73.7%**

**Kominami et al 2016** Prospective CADx based on support vector machines (SVMs) NA/41 patients 1262/118 containing 45 non-neoplastic lesions and 73 neoplastic lesions **Accuracy=93.2%; sensitivity=93.8%; specificity=93.3%; positive predictive value (PPV)=93.0%; negative predictive value=93.3%; concordance between the endoscopic diagnosis and diagnosis by CADx=97.5%**

**Misawa et al 2016** Retrospective CADx based on support vector machines (SVMs) NA 979 images containing 381 non-neoplasms and 598 neoplasms in the training dataset/100 images containing 50 non-neoplasms and 50 neoplasms in the test dataset **Accuracy=90.0%, 95% CI=82.4%–95.1%; sensitivity=84.5%, 95% CI=72.6%–92.7%; specificity=97.6%, 95% CI=87.4%–99.9%; PPV=98.0%, 95% CI=96.8%–99.4%; NPV=92.9%, 95% CI=82.8%–98.4%**

**Byrne et al 2018** Retrospective CADx + CADe based on an improved DCNN model using NBI NA/21,804 unseen frames in the test dataset **Accuracy=99.94%; sensitivity=95.95%; specificity=91.66%; NPV=93.6%; prediction of polyp videos= 97.6%**

**Mori et al 2018** Prospective CADx based on support vector machines (SVMs) used with NBI and endocytoscope NA/791 patients in the test dataset 61,925/466 polyps from 325 patients in the test dataset **CADx NBI : sensitivity=92.7%, 95% CI=89.1%–95.4; sensitivity=89.8%, 95% CI=84.4%–89.9; PPV=93.7%, 95% CI=90.2%–96.2%; specificity=98.3%, 95% CI=92.7%–95.6%; CADx-endocytoscope: sensitivity=91.3%, 95% CI=87.5%–94.3; specificity=88.7%, 95% CI=83.1%–93.0; PPV=92.9%, 95% CI=89.3%–95.9; NPV=96.3%, 95% CI=80.4%–96.0**

**Byrne et al 2019** CADx based on CNN Training dataset: 60098 frames from 223 polyp videos (29% NICE type 1, 53% NICE type 2 and 18% of normal mucosa with no polyp)/validation dataset: 40 videos (NICE type 1, NICE type 2 and two videos of normal mucosa)/test dataset: 125 consecutively identified diminutive polyps, comprising 51 hyperplastic polyps and 74 adenomas **Accuracy=94%, 95% CI=86%–97%; sensitivity=98%, 95% CI=92%–100%; Specificity=83%, 95% CI=67%–93%; NPV=97%; PPV=90%**

**Song et al 2020** Retrospective CADx based on CNN NA 12480 image patches of 624 polyps/two test datasets of 545 polyp **Agreement between the true polyp history CADx=0.614–0.642; accuracy=81.3–82.4%; sensitivity=81.2%; specificity=93.7%; PPV=78%; NPV=95%; the AUC=93.0–96.4, 0.86–0.89, and 0.89–0.91 for serrated polyps, benign adenoma/mucosal or superficial submucosal cancer, and deep submucosal cancer, respectively**

**Kudo et al 2020** Retrospective The EndoBRAIN system (CADx + CADe based on CNN) NA/89 patients test set 69,142 images taken at 520-fold magnification and 2,000 polyp images/100 lesions (≤ 10 mm) in the test dataset **CADe: accuracy=98.9%, 95% CI=97.3%–99.6%; sensitivity=96.5%, 95% CI=95.8%–97.8%; specificity=99.6%, 95% CI=99.9%–100%; PPV=99.9%, 95% CI=99.8%–100%; NPV=94.6%, 95% CI=92.7%–96.1%; CADx: accuracy=96%, 95% CI=95.1%–96.8%; sensitivity=96.9%, 95% CI=95.8%–97.8%; specificity=94.3%, 95% CI=92.7%–95.6%**
Guideline

Role of artificial intelligence in detection and characterization of colorectal polyps

RECOMMENDATION

2019 statement:
ESGE suggests the possible incorporation of computer-aided diagnosis (detection and characterization of lesions) into colonoscopy, if acceptable and reproducible accuracy for colorectal neoplasia is demonstrated in high quality multicenter in vivo clinical studies. Possible significant risks with implementation, specifically endoscopist deskilling and over-reliance on artificial intelligence (AI), unrepresentative training datasets, and hacking, need be considered.*

Weak recommendation, low quality evidence.

“Recommendations from professional endoscopy societies and local institutional policies should change from “you can do resect and discard” to “you should do resect and discard.”

D Rex, Endoscopy 2021
„The fear that a cancer will be resected and thrown away [...] defies evidence-based rational decision-making. Based on literature review and [...] data, the risk of invasive cancer in lesions ≤5mm is 0.009%“

D Rex, Endocopy 2021
KNOWLEDGE GAPS

-Topics to be discussed in the EWG between 2021 and 2024-

1. Unknown cancer prevention effect of CADe.
2. Unknown performance of CADe in the cancer screening programmes (e.g., ADR, FIT based screening programs etc.).
3. Unknown cost-effectiveness of CADe +/- CADx.
4. Unknown role of CAQ.
5. Lack of reimbursement of AI in colonoscopy.
6. Imperfect CADx in surveillance interval prediction (e.g., no SSL prediction).
„what is essential is invisible to the eye“
Antoine de Saint-Exupery

„what is essential is visible because of AI“
Endoscopist 2021
Questions?

My ability to come up with a great answer to a question.

In the moment vs. Four hours later.
ENDO 2022
Connecting the World of Endoscopy
3rd World Congress of GI Endoscopy

May 13-15, 2022
Kyoto, Japan
Kyoto International Conference Center & Grand Prince Hotel

www.worldendo.org

In cooperation with
103rd Congress of the
Japan Gastroenterological Endoscopy Society

SAVE THE DATE!