Supplementary material

Supplementary Table 1 Overview of the prospectively, multicenter, multivendor collected training and testing dataset used for AI4CRP (Artificial Intelligence for ColoRectal Polyps).

	Training set I	**Training set II**	**Training set III**	**Training set IV**	**Training set I-IV**	**Internal validation test set I**	**Real-time validation test set I[a]**
Population	CZE (NL)	MUMC+ (NL)	QA (UK)	Multicenter[b]	Multicenter	CZE (NL)	CZE (NL)
Endoscopy vendor	Fujifilm	Pentax	Fujifilm	Fujifilm	Multivendor	Fujifilm	Fujifilm
Adenomas	199	230	30	32	491	58	33
Hyperplastic polyps	58	30	19	35	142	19	12
SSLs	29	20	3	12	64	5	6
CRCs	12	9	0	16	37	4	0
Total polyps (images)	298 (889)	289 (785)	52 (107)	95 (248)	734 (2029)	86 (258)	51 (153)

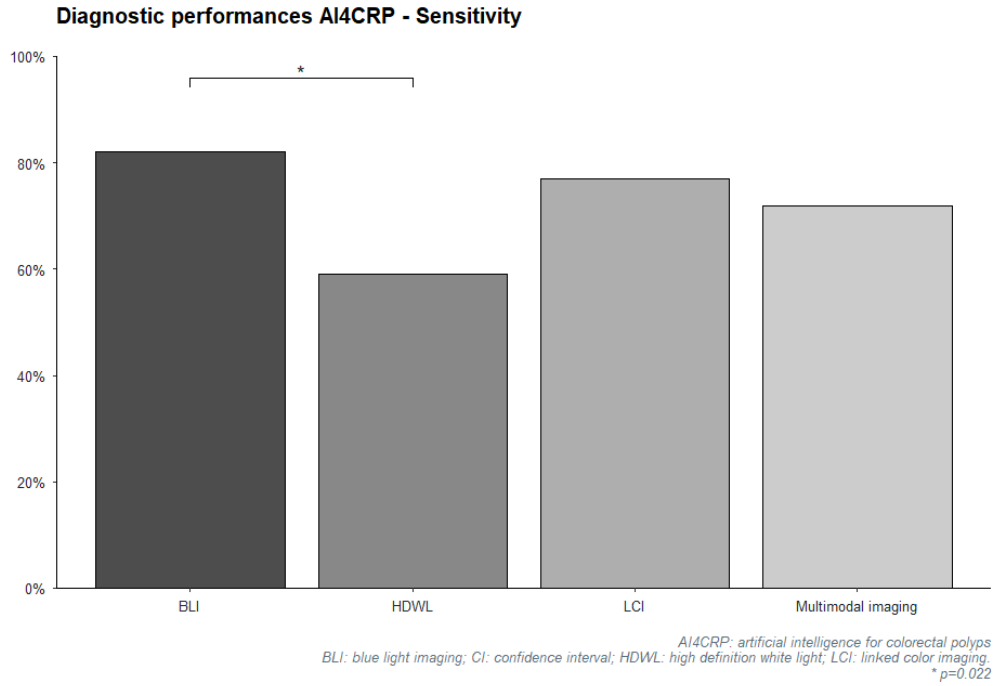
[a] The external validation test set I was collected prospectively and real-time. The images of the colorectal polyps in the external validation set were not included in one of the training sets.

[b] Multicenter data collection consisting of Diagnostični Center Bled (Ljubljana, Slovenia), King's College Hospital (London, UK), QA (Portsmouth, UK), and CZE (Eindhoven, NL).

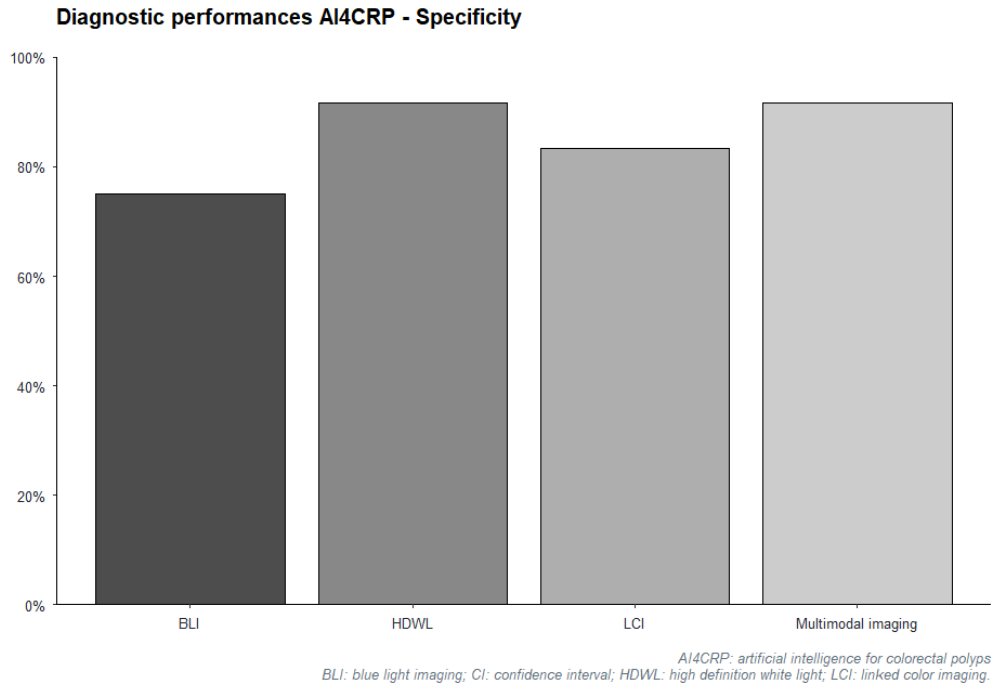
CRCs: colorectal cancers; CZE: Catharina Hospital Eindhoven; MUMC+: Maastricht University Medical Center+; NL: the Netherlands; QA: Queen Alexandra Hospital; SSLs: sessile serrated lesions; UK: United Kingdom.

Supplementary Figure 1. Statistical testing of the diagnostic performances of AI4CRP (Artificial Intelligence for ColoRectal Polyps) in different image enhancement modes.

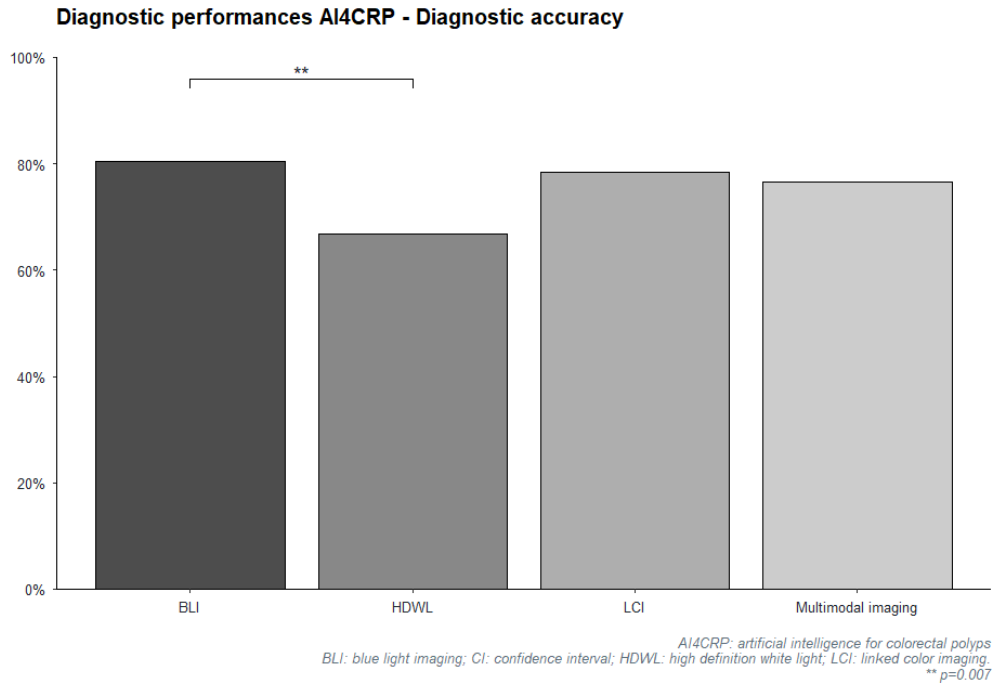
A)



B)



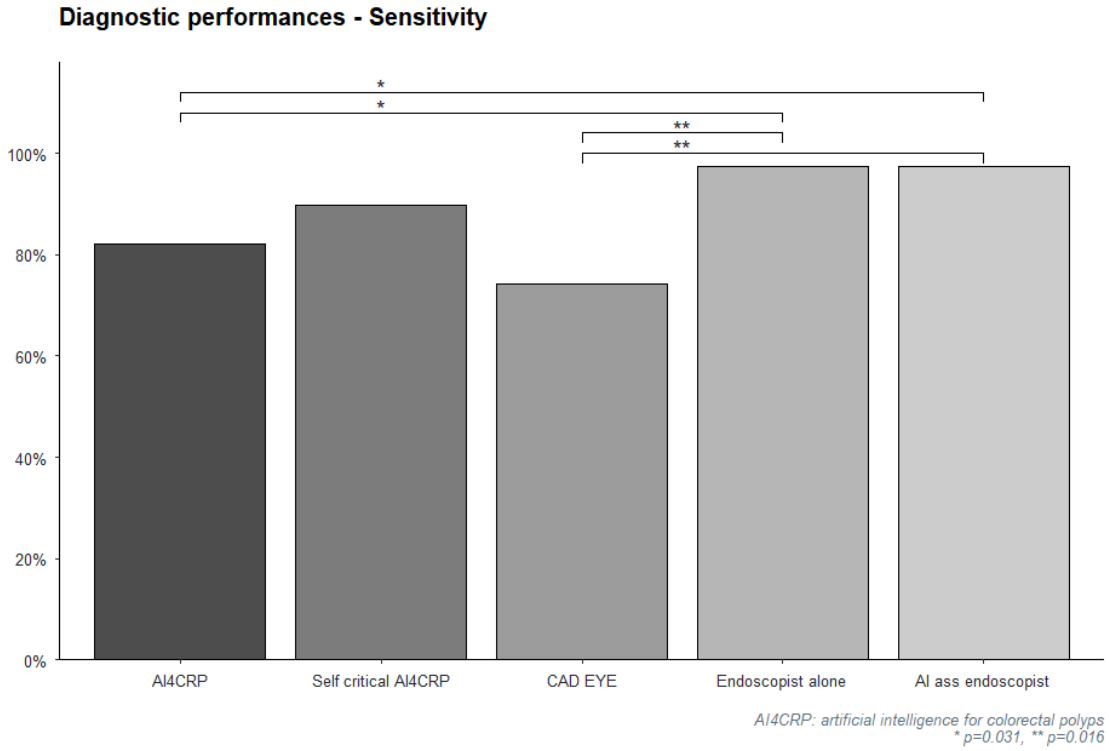
C)



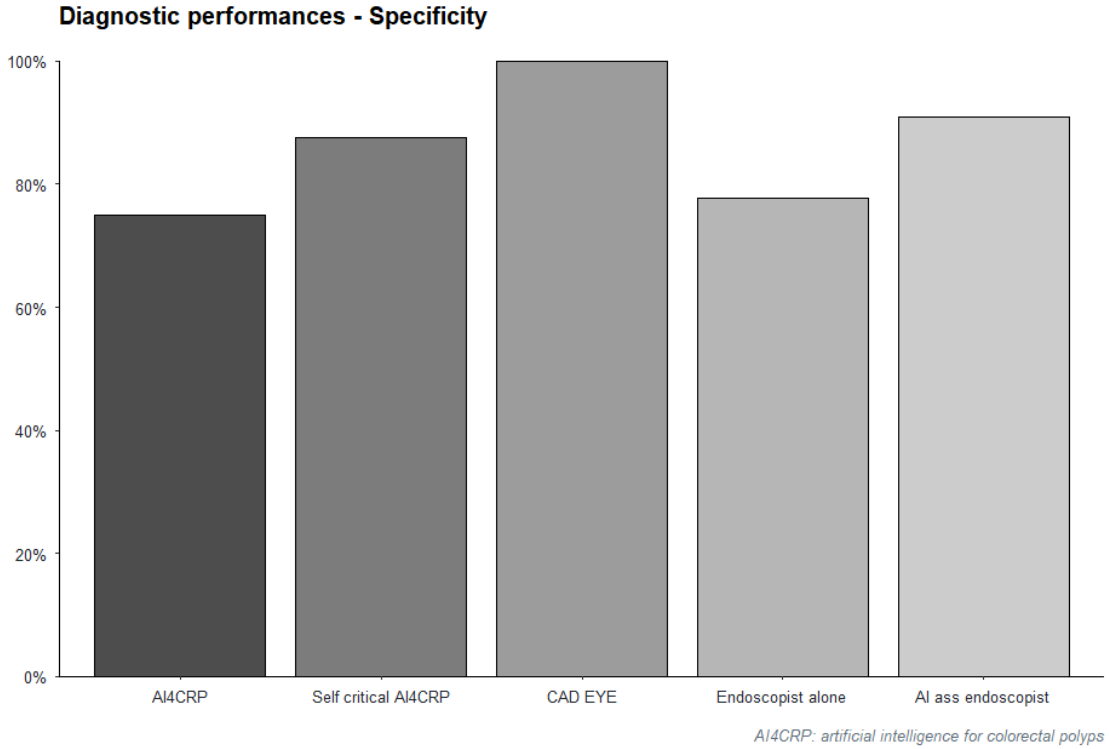
Statistical testing of the diagnostic performances of AI4CRP (Artificial Intelligence for ColoRectal Polyps) in different image enhancement modes. A) Sensitivity, B) specificity, and C) diagnostic accuracy. AI4CRP and self-critical AI4CRP both in BLI mode.

Supplementary Figure 2. Statistical testing of the diagnostic performances of AI4CRP (Artificial Intelligence for ColoRectal Polyps), self-critical AI4CRP, CAD EYE, the endoscopist alone, and the AI-assisted endoscopist.

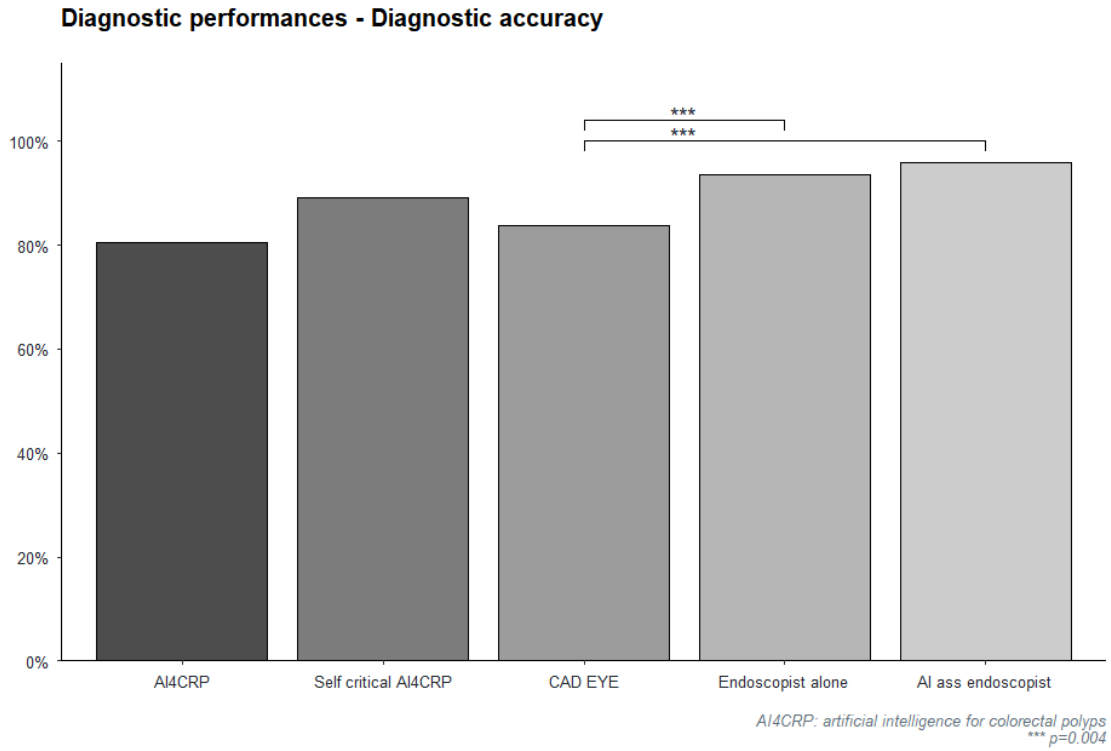
A)



B)


Diagnostic performances - Specificity
100%
80%
60%
40%
20%
0%
AI4CRP
Self critical AI4CRP
CAD EYE
Endoscopist alone
AI ass endoscopist
AI4CRP: artificial intelligence for colorectal polyps

C)



Statistical testing of the diagnostic performances of AI4CRP (Artificial Intelligence for ColoRectal Polyps) in different image enhancement modes. A) Sensitivity, B) specificity, and C) diagnostic accuracy. AI4CRP and self-critical AI4CRP both in BLI mode.

Supplementary Table 2. Confidence level by the expert endoscopist before and after reviewing the predictions of both computer-aided diagnosis systems.

	Endoscopist alone confidence level		**AI-assisted endoscopist confidence level[a]**		**p value[b]**
	High, n (%)	Low, n (%)	High, n (%)	Low, n (%)	
Tubular adenoma	31 (60.8)	1 (2.0))	31 (60.8)	1 (2.0)	
Tubulovillous adenoma	1 (2.0)	0 (0.0)	1 (2.0)	0 (0.0)	
Sessile serrated lesion	6 (11.8)	0 (0.0)	6 (11.8)	0 (0.0)	
Hyperplastic polyp	9 (17.6)	3 (5.9)	11 (21.6)	1 (2.0)	
Total	**47 (92.2)**	**4 (7.8)**	**49 (96.1)**	**2 (3.9)**	**0.500**

[a] Optical diagnosis by the endoscopist after reviewing predictions of both AI4CRP and CAD EYE.

[b] McNemar test.

AI: artificial intelligence

Supplementary Table 3. Diagnostic performance of AI4CRP (Artificial Intelligence for ColoRectal Polyps), self-critical AI4CRP, CAD EYE, and the endoscopist with cluster bootstrapping.

	AI4CRP[a], % (95% CI) n=51	**Self-critical AI4CRP**[a], % (95% CI) n=37	**CAD EYE**, % (95% CI) n=49	**Endoscopist alone**[b], % (95% CI) n=47	**AI-assisted endoscopist**[b,c], % (95% CI) n=49
Sensitivity	82.1 (0.71-0.93)	89.5 (0.79-1.00)	74.0 (0.57-0.91)	97.5 (0.93-1.00)	97.4 (0.92-1.00)
Specificity	75.3 (0.52-0.99)	87.6 (0.64-1.00)	100.0 (1.00-1.00)	77.8 (0.49-1.00)	90.9 (0.74-1.00)
PPV	91.6 (0.83-1.00)	96.4 (0.90-1.00)	100.0 (1.00-1.00)	94.9 (0.88-1.00)	97.3 (0.92-1.00)
NPV	56.2 (0.31-0.81)	69.7 (0.40-0.99)	69.1 (0.50-0.88)	87.9 (0.64-1.00)	90.9 (0.73-1.00)
Diagnostic accuracy	80.4 (0.71-0.90)	89.1 (0.79-0.99)	83.0 (0.71-0.95)	93.7 (0.87-1.00)	95.9 (0.90-1.00)

[a] AI4CRP and self-critical AI4CRP both in BLI mode.

[b] Optical diagnosis by the endoscopist only taking into account diagnosis made with high confidence.

[c] Optical diagnosis by the endoscopist after reviewing predictions of both AI4CRP and CAD EYE.

AI: artificial intelligence; CI: confidence interval; NPV: negative predictive value; PPV: positive predictive value.

Supplementary Table 4. Diagnostic performance of AI4CRP (Artificial Intelligence for ColoRectal Polyps), self-critical AI4CRP, CAD EYE, and the endoscopist according to location.

	AI4CRP[a], % (95% CI) n=16	**Self-critical AI4CRP**[a], % (95% CI) n=11	**CAD EYE**, % (95% CI) n=16	**Endoscopist alone**[b], % (95% CI) n=14	**AI-assisted endoscopist**[b,c], % (95% CI) n=15
Rectosigmoid					
Sensitivity	75.0 (0.36-0.96)	100.0 (0.46-1.00)	100.0 (0.56-1.00)	87.5 (0.47-0.99)	87.5 (0.47-0.99)
Specificity	75.0 (0.36-0.96)	83.3 (0.36-0.99)	100.0 (0.63-1.00)	100.0 (0.52-1.00)	100.0 (0.56-1.00)
PPV	75.0 (0.36-0.96)	83.3 (0.36-0.99)	100.0 (0.56-1.00)	100.0 (0.56-1.00)	100.0 (0.56-1.00)
NPV	75.0 (0.36-0.96)	100.0 (0.46-1.00)	100.0 (0.63-1.00)	85.7 (0.42-0.99)	87.5 (0.47-0.99)
Diagnostic accuracy	75.0 (0.47-0.92)	90.9 (0.57-1.00)	100.0 (0.76-1.00)	92.9 (0.64-1.00)	93.3 (0.66-1.00)
Proximal to rectosigmoid	n=35	n=26	n=33	n=33	n=34
Sensitivity	83.9 (0.66-0.94)	87.5 (0.67-0.97)	66.7 (0.45-0.84)	100.0 (0.86-1.00)	100.0 (0.86-1.00)
Specificity	75.0 (0.22-0.99)	100.0 (0.20-1.00)	100.0 (0.63-1.00)	33.3 (0.02-0.87)	75.0 (0.22-0.99)
PPV	96.3 (0.79-1.00)	100.0 (0.81-1.00)	100.0 (0.76-1.00)	93.8 (0.78-0.99)	96.8 (0.81-1.00)
NPV	37.5 (0.10-0.74)	40.0 (0.07-0.83)	52.9 (0.29-0.76)	100.0 (0.05-1.00)	100.0 (0.31-1.00)
Diagnostic accuracy	82.9 (0.66-0.93)	88.5 (0.69-0.97)	75.8 (0.57-0.88)	93.9 (0.78-0.99)	97.1 (0.83-1.00)

[a] AI4CRP and self-critical AI4CRP both in BLI mode.

[b] Optical diagnosis by the endoscopist only taking into account diagnosis made with high confidence.

[c] Optical diagnosis by the endoscopist after reviewing predictions of both AI4CRP and CAD EYE.

AI: artificial intelligence; CI: confidence interval; NPV: negative predictive value; PPV: positive predictive value