Automated classification of gastric neoplasms in endoscopic images using a convolutional neural network

Cho et al.

Supplementary material

Table 1s Data composition of the training and test datasets for gastric lesions.

Lesion type	Whole dataset, no. of images	Training dataset		Test dataset	
		No. of images (%)	No. of patients	No. of images (%)	No. of patients
Overall	5017	4205 (83.8)	1066	812 (16.2)	212
AGC	679	570 (83.9)	155	109 (16.1)	30
EGC	1104	919 (83.2)	297	185 (16.8)	59
HGD	380	322 (84.7)	104	58 (15.3)	20
LGD	628	529 (84.2)	180	99 (15.8)	35
Non-neoplasm	2226	1865 (83.8)	360	361 (16.2)	71

AGC, advanced gastric cancer; EGC, early gastric cancer; HGD, high grade dysplasia; LGD, low grade dysplasia.

% is proportion of each class in the overall dataset.

Table 2s Per-category diagnostic performance of convolutional neural network models in the five-class classification on endoscopic images in the test dataset.

Model	Performance parameters, % (95%CI)					AUC (95%CI)
	Accuracy	Sensitivity	Specificity	PPV	NPV	
Inception-v4						
AGC	90.2 (87.9–92.1)	77.1 (68.0–84.6)	92.2 (89.9–94.1)	60.4 (53.7–66.8)	96.3 (94.8–97.3)	0.846 (0.820–0.870)
EGC	74.8 (71.6–77.7)	51.4 (43.9–58.8)	81.7 (78.4–84.6)	45.2 (39.9–50.6)	85.1 (83.0–86.9)	0.665 (0.631–0.697)
HGD	88.1 (85.6–90.2)	10.3 (3.9–21.2)	94.0 (92.1–95.6)	11.8 (5.6–23.0)	93.2 (92.6–93.7)	0.522 (0.487–0.557)
LGD	83.9 (81.2–86.3)	32.3 (23.3–42.5)	91.0 (88.7–93.0)	33.3 (25.7–42.0)	90.6 (89.4–91.8)	0.617 (0.582–0.650)
Non-neoplasm	82.6 (79.9–85.2)	74.2 (69.4–78.7)	89.4 (86.1–92.0)	84.8 (80.9–88.0)	81.3 (78.4–83.8)	0.818 (0.790–0.844)
Resnet-152						
AGC	89.2 (86.8–91.2)	67.9 (58.3–76.5)	92.5 (90.3–94.3)	58.3 (51.1–65.1)	94.9 (93.4–96.1)	0.802 (0.773–0.829)
EGC	77.6 (74.6–80.4)	33.0 (26.3–40.3)	90.8 (88.2–92.9)	51.3 (43.3–59.2)	82.1 (80.3–83.6)	0.619 (0.584–0.652)
HGD	91.3 (89.1–93.1)	3.5 (0.4–12.1)	97.9 (96.6–98.8)	11.1 (2.9–34.7)	93.1 (92.7–93.4)	0.507 (0.472–0.542)
LGD	84.0 (81.3–86.5)	21.2 (13.6–30.6)	92.7 (90.5–94.5)	28.8 (20.3–39.0)	89.4 (88.4–90.4)	0.570 (0.535–0.604)
Non-neoplasm	77.3 (74.3–80.2)	90.3 (86.8–93.2)	67.0 (62.4–71.3)	68.6 (65.6–71.5)	89.6 (86.2–92.2)	0.786 (0.757–0.814)
Inception-Resnet-v2						
AGC	90.4 (88.2–92.3)	78.9 (70.0–86.1)	92.2 (89.9–94.1)	61.0 (54.4–67.2)	96.6 (95.1–97.6)	0.855 (0.829–0.879)
EGC	80.8 (77.9–83.4)	52.4 (45.0–59.8)	89.2 (86.5–91.5)	58.8 (52.3–65.0)	86.4 (84.5–88.1)	0.708 (0.675–0.739)
HGD	91.1 (89.0–93.0)	0 (0.0–6.2)	98.1 (96.9–99.0)	0	92.7 (92.7–92.8)	0.491 (0.456–0.526)
LGD	87.1 (84.6–89.3)	22.2 (14.5–31.7)	96.1 (94.4–97.4)	44.0 (31.9–56.9)	89.9 (88.9–90.8)	0.591 (0.557–0.626)
Non-neoplasm	83.1 (80.4–85.6)	92.2 (89.0–94.8)	75.8 (71.6–79.7)	75.3 (72.1–78.3)	92.4 (89.5–94.6)	0.840 (0.813–0.865)

CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value; AUC, area under the curve; AGC, advanced gastric cancer; EGC, early gastric cancer; HGD, high grade dysplasia; LGD, low grade dysplasia.

Each value is described with 95% confidence interval.

Table 3s Diagnostic performance of convolutional neural network models in classifying gastric cancer or neoplasm on endoscopic images in the test dataset.

Model	Performance parameters, % (95%CI)					AUC (95%CI)
	Accuracy	Sensitivity	Specificity	PPV	NPV	
Cancer vs. non-cancer						
Inception-v4	81.0 (78.2–83.9)	71.8 (66.6–77.1)	86.3 (83.3–89.3)	74.8 (69.2–80.1)	84.3 (81.2–87.5)	0.875 (0.848–0.898)
Resnet-152	79.8 (77.1–82.6)	66.3 (60.9–71.5)	87.5 (84.5–90.3)	75.0 (69.8–80)	82.1 (79–85.5)	0.862 (0.838–0.886),
Inception-Resnet-v2	81.9 (79.3–84.6)	75.9 (79.3–84.6)	85.3 (79.3–84.6)	74.6 (69.5–79.6)	86.2 (83.2–89)	0.877 (0.851–0.901),
Neoplasm vs. non-neoplasm						
Inception-v4	82.9 (80.5–85.7)	79.6 (75.7–83.3)	87.0 (83.5–90.4)	88.4 (85.4–91.4)	77.3 (73.6–81.3)	0.912 (0.890–0.931)
Resnet-152	83.4 (80.5–85.8)	86.7 (83.9–90)	79.2 (75.1–83.7)	83.9 (80.5–87.2)	82.7 (78.4–86.6)	0.908 (0.889–0.928)
Inception-Resnet-v2	85.5 (83.3–87.8)	84.0 (80.5–87.3)	87.3 (83.8–90.5)	89.2 (86.2–92)	81.4 (77.5–85.3)	0.927 (0.908–0.944)

CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value; AUC, area under the curve.

Each value is described with 95% confidence interval.

Fig. 1s Validation accuracy by training epochs of CNN using Inception-resnet-v2. CNN, convolutional neural network. Accuracy is for the five-class classification.