**Supplementary materials**

**Methods**

*Inclusion and exclusion criteria*

The inclusion criteria were as follows: a diagnosis of early ESCC or reflux esophagitis or normal esophageal mucosa; for early ESCC, patients underwent endoscopic submucosal dissection (ESD) with histological proof at three hospitals; for reflux esophagitis or normal esophageal mucosa, all the cases were confirmed by two endoscopists with > 10 years of endoscopic experience. The exclusion criteria were as follows: a history of surgery or chemotherapy, or radiation to the esophagus.

*Construction of deep learning model*

The PMG architecture had two characteristics. The training strategy added new layers in each training step to exploit information based on the smaller granularity information found at the last step and the previous stage. With such operation, the information cross multi-granularity was fused. Second, a simple jigsaw generator was used to form images contain information about different granularity levels to allow the network to focus on different scales of features. In this study, the Resnet50 is employed as the backbone of the PMG architecture as feature extractors. After the training process of the PMG model, the Gradient-weighted Class Activation Mapping (Grad-CAM) method was employed for the visualization of ESCC and reflux esophagitis prediction (1).

*Outcome calculation*

Accuracy = number of true diagnostic images/total number of images, Sensitivity = number of true positive images/total number of positive images, Specificity = number of true negative images/total number of negative images, PPV = number of true positive images/(number of true positive images + number of false-positive images), NPV = number of true negative images/(number of true negative images + number of false-negative images). The accuracy of the deep learning model prediction was defined as predictive value > cut-off value. The optimum cut-off point was determined in the training dataset using the Youden index (sensitivity + specificity - 1).

**Reference**

1. Selvaraju RR, Cogswell M, Das A, et al. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. International Journal of Computer Vision 2020;128:336-59.

**Figure legends**

Figure S1. The architecture of the DCNN system.

Figure S2. Training curves of the DCNN system. (A & B) The DCNN model in identifying early ESCC.

Figure S3. Website of the open-access DCNN model (<http://112.74.182.39/esophagus>).

Table S1 Baseline characteristics of training and validation datasets

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Training | Internal validation | External validation | |
| WXPH | TZPH |
| Normal esophagus (N, %) | 328 (18.7%) | 25 (20.2%) | 25 (17.8%) | 26 (18.9%) |
| Reflux esophagitis |  |  |  |  |
| A (N, %) | 47 (2.7%) | 3 (2.4%) | 3 (2.7%) | 4 (2.3%) |
| B (N, %) | 230 (13.1%) | 13 (10.5%) | 14 (10.3%) | 15 (10.6%) |
| C (N, %) | 118 (6.7%) | 6 (4.8%) | 7 (4.1%) | 6 (5.3%) |
| D (N, %) | 18 (1.0%) | 1 (0.8%) | 2 (1.4%) | 2 (1.5%) |
| Early esophageal cancer |  |  |  |  |
| Site |  |  |  |  |
| Upper (N, %) | 45 (2.6%) | 5 (4.0%) | 2 (0.7%) | 1 (1.5%) |
| Middle (N, %) | 138 (7.9%) | 12 (9.7%) | 10 (8.2%) | 12 (7.6%) |
| Lower (N, %) | 154 (8.8%) | 16 (12.9%) | 15 (12.3%) | 18 (11.4%) |
| Size |  |  |  |  |
| < 2 cm (N, %) | 94 (5.4%) | 12 (9.7%) | 19 (17.1%) | 25 (14.4%) |
| ≥ 2 cm (N, %) | 243 (13.9%) | 21 (16.9%) | 8 (4.1%) | 6 (6.1%) |
| Depth |  |  |  |  |
| M (N, %) | 316 (18.0%) | 29 (23.4%) | 25 (19.9%) | 29 (18.9%) |
| SM (N, %) | 21 (1.2%) | 4 (3.2%) | 2 (1.4%) | 2 (1.5%) |

M, Intramucosal lesions; SM, Submucosal lesions.

Table S2 The number of patients and images of Q260 endoscopes

|  |  |  |  |
| --- | --- | --- | --- |
|  | Q260 endoscopes | | |
|  | Normal esophagus | Reflux esophagitis | Early esophageal cancer |
| Patient/images | 24/117 | 30/126 | 16/145 |

Table S3 The performance of the DCNN model in different tumor sizes

|  |  |  |
| --- | --- | --- |
|  | Internal validation | |
|  | Size < 2 cm | Size ≥ 2 cm |
| Accuracy (95% CI) | 0.932 (0.851–0.971) | 0.966 (0.916–0.987) |
| Sensitivity (95% CI) | 1.000 (0.906–0.999) | 1.000 (0.939–1.000) |
| Specificity (95% CI) | 0.865 (0.720–0.941) | 0.932 (0.838–0.973) |
| PPV (95% CI) | 0.881 (0.750–0.948) | 0.937 (0.848–0.975) |
| NPV (95% CI) | 1.000 (0.893–1.000) | 1.000 (0.935–1.000) |
| AUC | 0.956 (0.904–1.000) | 0.975 (0.942–1.000) |

PPV, Positive predictive value; NPV, Negative predictive value

Table S4 The performance of the DCNN model in different locations

|  |  |  |  |
| --- | --- | --- | --- |
|  | Internal validation | | |
|  | Upper | Middle | Lower |
| Accuracy (95% CI) | 0.950 (0.764–0.997) | 0.933 (0.841–0.974) | 0.946 (0.888–0.975) |
| Sensitivity (95% CI) | 1.000 (0.722–1.000) | 1.000 (0.886–1.000) | 1.000 (0.936–1.000) |
| Specificity (95% CI) | 0.900 (0.596–0.995) | 0.867 (0.703–0.947) | 0.893 (0.785–0.950) |
| PPV (95% CI) | 0.909 (0.623–0.995) | 0.882 (0.734–0.953) | 0.903 (0.805–0.955) |
| NPV (95% CI) | 1.000 (0.701–1.000) | 1.000 (0.871–1.000) | 1.000 (0.929–1.000) |
| AUC | 0.900 (0.704–1.000) | 0.933 (0.866–1.000) | 0.964 (0.933–0.996) |

PPV, Positive predictive value; NPV, Negative predictive value

Table S5 The performance of the DCNN model in different invasion depths

|  |  |  |
| --- | --- | --- |
|  | Internal validation | |
|  | Intramucosal lesions | Submucosal lesions |
| Accuracy (95% CI) | 0.940 (0.894–0.967) | 0.917 (0.742–0.985) |
| Sensitivity (95% CI) | 0.976 (0.917–0.996) | 1.000 (0.758–1.000) |
| Specificity (95% CI) | 0.905 (0.823–0.951) | 0.833 (0.552–0.970) |
| PPV (95% CI) | 0.911 (0.834–0.954) | 0.857 (0.601–0.975) |
| NPV (95% CI) | 0.974 (0.911–0.995) | 1.000 (0.722–1.000) |
| AUC | 0.957 (0.925–0.989) | 0.889 (0.720–1.000) |

PPV, Positive predictive value; NPV, Negative predictive value

Table S6 The performance of the DCNN model in different endoscopes

|  |  |  |
| --- | --- | --- |
|  | H260 & H290 | Q260 |
| Accuracy (95% CI) | 0.913 (0.878–0.939) | 0.912 (0.880–0.937) |
| Sensitivity (95% CI) | 0.979 (0.927–0.996) | 1.000 (0.974–1.000) |
| Specificity (95% CI) | 0.886 (0.839-0.921) | 0.860 (0.811–0.898) |
| PPV (95% CI) | 0.777 (0.695–0.842) | 0.810 (0.746–0.861) |
| NPV (95% CI) | 0.991 (0.966–0.998) | 1.000 (0.982–1.000) |
| AUC | 0.954 (0.934–0.974) | 0.921 (0.893–0.949) |

PPV, Positive predictive value; NPV, Negative predictive value

Table S7 Results of a significant statistical test

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Experts | |  | Novices | |
|  | Experts vs. DCNN model | With DCNN vs. without DCNN |  | Novices vs. DCNN model | With DCNN vs. without DCNN |
| Accuracy (95% CI) | < 0.001 | < 0.001 |  | <0.001 | < 0.001 |
| Sensitivity (95% CI) | < 0.001 | < 0.001 |  | <0.001 | < 0.001 |
| Specificity (95% CI) | 0.704 | 0.010 |  | 0.575 | 0.052 |
| PPV (95% CI) | 0.113 | < 0.001 |  | 0.004 | < 0.001 |
| NPV (95% CI) | <0.001 | < 0.001 |  | <0.001 | < 0.001 |

PPV, Positive predictive value; NPV, Negative predictive value