

Large-scale esophageal cancer screening through noncontrast computed tomography and artificial intelligence

In the format provided by the
authors and unedited

1 Supplementary Methods for: Large-Scale
2 Esophageal Cancer Screening via Non-contrast CT
3 and AI

4 **Contents**

5	1 Supplementary Methods	4
6	1.1 Training, internal and external validation	4
7	1.1.1 Internal training and test	4
8	1.1.2 External multicenter test for regular dose non-contrast CT . . .	4
9	1.1.3 Multicenter test for low-dose CT	5
10	1.1.4 External multicenter test for endoscopic screening program . .	5
11	1.2 Deployment of AI model	5
12	1.3 Details of real-world model calibration and validation	6
13	1.4 Details of prospective validation in hospital	7
14	1.5 Details of real-world retrospective LDCT validation	9
15	1.6 Details of validation on prospectively enrolled endoscopic screening cohort	9
16	1.7 Comparison to other methods	10

17 List of Figures

18	1	ROC curves and operating point of EAGLE on centers that contain	
19		both EC and negative control cases.	12
20	2	Reader study with DAMO Med system.	13
21	3	EAGLE and EAGLE-Plus performances on RW1 and RW2 cohort. . .	14
22	4	Esophageal malignancy examples found by EAGLE-Plus in the prospec-	
23		tive validation.	15
24	5	Flowchart illustrating the potential benefits for an individual without	
25		symptoms.	16
26	6	Endoscopic findings of EAGLE predictions in the LDCT test cohort. .	17
27	7	Non-contrast CT informed consent form for SuiNing participants. . .	18
28	8	Analysis of EAGLE intepratability with positive endoscopic findings. .	19
29	9	Comparison between the Original EAGLE and the EGJ Enhanced	
30		EAGLE.	20

31

List of Tables

32	1	Reader experiences. Experience of CT reading and expertise of each	
33		reader.	21
34	2	Definition of esophageal lesions and diseases and national guidelines to	
35		manage endoscopic findings	22
36	3	Standard of truth for all training and validation cohorts.	23
37	4	Region specific economic cost estimates for endoscopy and non-	
38		contrast/low-dose CT.	24

1 Supplementary Methods

1.1 Training, internal and external validation

Positive cases in this study included patients with histopathologically confirmed esophageal malignancies, defined as high-grade intraepithelial neoplasia (HGIN) or esophageal cancer (EC), based on biopsy or surgical pathology. Lesion staging was determined according to the 8th edition of the AJCC TNM classification. For biopsy-based diagnoses, definitive pathological evidence was required. Negative controls in retrospective test cohorts were selected through a two-stage process at each participating center: First, radiologists identified patients with chest CT reports documenting no suspicious esophageal abnormalities. We established the at least 2-year follow-up as one of inclusion criteria for the negative controls to confirm the absence of malignancy at baseline excluding cases with ambiguous or short-term data. Then, patients were included if they had either at least two years of follow-up data, or endoscopic results within one year, with no clinical evidence of malignant esophageal lesions during follow-up. The standard of truth for all cohorts are provided in Supplementary Table 3.

1.1.1 Internal training and test

Center A. Between 2008 and 2021, a total of 5,737 patients were enrolled at Sun Yat-sen University Cancer Center (SYSUCC), with 4,255 patients (2008–2018) assigned to the training cohort and 1,482 patients enrolled in the internal test cohort.

Center B. Between 2009 and 2023, 3,576 patients were enrolled at Sichuan Cancer Hospital (SCCH). 2,558 patients (2009–2015) were assigned to the training cohort, and 1,018 patients were assigned to the internal test cohort.

The internal training cohort comprised 6,813 patients in total, where 3,744 patients were diagnosed with esophageal cancer or HGIN, and 3,069 patients were negative controls. The internal test cohort included 2,500 patients, with 1,281 esophageal malignant cases and 1,219 negative controls. The internal test cohort also included some challenging cases of early-stage EC or HGIN from the concurrent training cohort for which radiologists could not provide lesion annotations.

1.1.2 External multicenter test for regular dose non-contrast CT

The external test cohorts for opportunistic screening were collected from six centers across China, one center in the Czech Republic, and one center in Australia. For six centers in China, they are distributed widely in geographical areas, with three centers in the east, one center in the central region, one center in the south, and one center in the northwest region of China.

Center C (n=4,904). Shanghai Institution of Pancreatic Diseases (SIPD) is a general medical center in the east of China. We collected 1,113 esophageal malignant patients from 2014 to 2024 and enrolled 3,791 negative controls from 2014 to 2022.

Center D (n=4,313). First Affiliated Hospital of Zhejiang University (FAHZU) is a tertiary hospital in the east of China. We collected 363 esophageal malignant patients from 2018 to 2021 and enrolled 3,950 negative controls from 2018 to 2021.

Center E (n=761). Hunan Cancer Hospital (HCH) is a tertiary hospital in the central region of China. We collected 139 esophageal malignant cases and 622 negative controls between 2016 and 2020.

Center F (n=731). Shantou Central Hospital (STCH) is a tertiary hospital in the south of China. We collected 731 esophageal malignant patients between 2009 and 2020.

Center G (n=325). Xinjiang Medical University Affiliated Cancer Hospital (XMUACH) is a tertiary hospital in the northwest of China. We collected 175 esophageal malignant patients and 150 negative controls between 2014 and 2020.

Center H (n=32). Fujian Cancer Hospital (FCH) is a tertiary hospital in the east of China. We collected 32 esophageal malignant patients from 2013 to 2018.

Center I (n=195). General University Hospital in Prague (GUHP) is a hospital in the Czech Republic. We collected 54 esophageal malignant patients and 141 negative controls.

Center J (n=205). GenesisCare is a leading provider of radiation therapy services in Australia. This dataset contains de-identified clinical data collected from 205 oncology patients treated for esophageal cancer between 2015 and 2025.

1.1.3 Multicenter test for low-dose CT

To evaluate the generalizability of EAGLE on LDCT, we enrolled 1,607 subjects from SCCH (Center B) and SIPD (Center C). SCCH contributed 1,522 cases, comprising 62 esophageal malignant patients and 1,460 negative controls, while SIPD provided 85 esophageal malignant cases. This cohort included a total of 147 esophageal malignant cases and 1,460 negative controls across both institutions. TNM staging data were incomplete for a significant proportion of patients (54/147, 36.7%), mainly due to treatment at external institutions or loss to follow-up.

1.1.4 External multicenter test for endoscopic screening program

The external validation cohorts comprised retrospectively enrolled patients who had undergone non-contrast CT before the endoscopic screening program.

Center K (n=599). Suining Central Hospital (SNCH) is a tertiary hospital in China. We collected 449 esophageal malignant patients from 2016 to 2023 and 150 negative controls confirmed by endoscopy from 2020 to 2023.

Center L (n=103). Lishui Central Hospital (LCH) is a tertiary hospital in China. We collected 103 esophageal malignant patients from 2019 to 2024.

1.2 Deployment of AI model

Our model was integrated into routine clinical workflows through the DAMO MED system. This platform interfaces with Radiology Information System (RIS) databases to identify eligible examinations and automatically retrieves corresponding CT scans from Picture Archiving and Communication Systems (PACS) to local servers. For each DICOM-formatted dataset, appropriate scans are selected using DICOM metadata, such as anatomical region, slice thickness and reconstruction kernel, in combination with RIS entries (e.g. exam descriptions containing “chest non-contrast CT” or “chest

low-dose CT”). Verified non-contrast CT volumes are then processed by our model. Model’s output, including lesion probabilities and segmentation masks, are presented in the DAMO MED user interface, saved as structured Excel files and transmitted for clinical review and evaluation.

1.3 Details of real-world model calibration and validation

We assembled real-world cohorts from three medical centers, and patient reference standards were determined in a multi-step process. At each centre, two experienced radiologists first reviewed electronic medical records to establish standard-of-care (SOC) clinical diagnoses. Cases with diagnostic uncertainty were adjudicated by a multidisciplinary team (MDT) using clinical history, imaging findings and pathological results to reach a definitive diagnosis. All positive cases were confirmed by biopsy or surgical pathology. To minimise potential “double standard” bias for negative cases, approximately 1% of patients with concordant negative results from both SOC and AI assessment were randomly selected for further review. For these cases, two radiologists at each center independently re-evaluated CT images and clinical records to verify the absence of disease.

Real-world, multi-scenario cohorts were assembled to refine and evaluate EAGLE model using clinical data beyond the scope of its internal training set. The evaluation covered four distinct clinical settings, each with characteristic workflows: *physical examination* perform routine health screening, typically using non-contrast chest CT for lung nodule detection and general wellness assessment; *emergency departments* use non-contrast CT for rapid diagnosis in acute presentations, requiring expedited triage and immediate clinical decisions; *outpatient clinics* encompass initial diagnosis, treatment monitoring and follow-up for patients with visits under 24 hours; and *inpatient wards* manage patients requiring overnight or prolonged hospitalisation, often involving complex management. A total of 35,402 patients were enrolled from three centres—SYSUCC (Center A), SCCH (Center B), and SIPD (Center C). At each site, EAGLE model was deployed and executed locally. Patients were divided into two real-world cohorts to enable temporal validation of model performance.

The first real-world cohorts (RW1) for model calibration. RW1 included 20,758 patients from three centers. At SYSUCC, 4,568 patients from the outpatient and physical examination departments were collected in March-April 2024. At SCCH, 6,875 patients from the outpatient department were collected from December 2019 to March 2021. At SIPD, 9,315 patients from all four clinical settings were enrolled in December 2021. In RW1, EAGLE achieved a sensitivity of 89.8% (95% CI 86.1-93.5%), specificity of 95.5% (95% CI 95.3-95.8%), and positive predictive value (PPV) of 21.9% (95% CI 20.6-23.1%).

For the SYSUCC-calibration cohort in RW1 (n=4,568), EAGLE generated 231 positive predictions, including 49 esophageal malignant cases and 182 negative cases. Among the 55 patients with esophageal malignancies in this cohort, EAGLE achieved a sensitivity of 89.1% (95% CI 80.9-97.3%) and specificity of 96.0% (95% CI 95.4-96.5%). The MDT selected 155 of 182 as hard negative examples.

For the SCCH-calibration cohort in RW1 (n=6,875), EAGLE generated 517 positive predictions, including 173 esophageal malignant cases and 344 negative cases.

Among the 192 esophageal malignancies in this cohort, EAGLE achieved an overall sensitivity of 90.1% (95% CI 85.7-94.5%) and specificity of 94.9% (95% CI 94.3-95.4%). The MDT selected 83 of 344 as hard negative examples, as most patients were from the outpatient department who sought cancer treatments.

For the SIPD-calibration cohort in RW1 (n=9,315), EAGLE generated 422 positive predictions, including 34 esophageal malignant cases and 388 negative cases. Among the 38 esophageal malignancies in this cohort, EAGLE achieved a sensitivity of 89.5% (95% CI 80.3-98.7%) and specificity of 95.8% (95% CI 95.4-96.2%). The MDT selected 361 of 388 as hard negative examples.

The second real-world cohorts (RW2) for validation. RW2 included 14,644 patients from three centers. At SYSUCC, 5,577 patients from the outpatient and physical examination departments were collected from May 2024 to June 2024. At SCCH, 5,350 patients from the outpatient department were collected from April 2021 to June 2024. At SIPD, 3,717 patients from all four clinical settings were collected from 1 and 10 February 2022.

For the SYSUCC-validation cohort in RW2 (n=5,577) that included 51 esophageal malignant cases, EAGLE-Plus achieved a sensitivity of 90.2% (95% CI 82.4-98.0%) and specificity of 98.9% (95% CI 98.7-99.2%). Compared to the original EAGLE model that achieved a sensitivity of 88.2% (95% CI 79.4-97.1%) and specificity of 97.1% (95% CI 96.7-97.6%), PPV was increased from 22.2% to 43.8%, with false positives reduced by more than 88% (158 to 59).

For the SCCH-validation cohort in RW2 (n=5,350) that included 103 esophageal malignant cases, EAGLE-Plus achieved an overall sensitivity of 86.4% (95% CI 79.6-93.2%) and specificity of 99.1% (95% CI 98.9-99.4%). Compared to the original EAGLE model, false positives were reduced by over 66% (from 116 to 45), leading to an improved PPV from 44.2% to 66.4%.

For the SIPD-validation cohort in RW2 (n=3,717) that included 9 esophageal malignant cases, both EAGLE and EAGLE-Plus could detect all of them. Compared to the original EAGLE model, EAGLE-Plus achieved an overall specificity of 99.6% (95% CI 99.4-99.8%), reducing false positives by more than 90% (from 158 to 14).

1.4 Details of prospective validation in hospital

We conducted a prospective validation to evaluate the real-world performance of EAGLE-Plus for detecting esophageal malignancies, including lesions missed by the initial SOC pathway, in a hospital-based clinical setting. The primary endpoint was the PPV among AI-positive participants, and secondary endpoints included sensitivity and the AI-triggered referral rate. Overall study oversight was provided by the MDT comprising radiologists, endoscopists, and surgeons, who conducted case review and guided clinical triage.

Throughout the study, radiologists operating within SIPD's routine double-reading SOC pathway remained blinded to both AI outputs and follow-up/outcome information. As this SOC pathway was intended to reflect real-world clinical practice, radiologists had access to available clinical information, including patient symptoms, rather than interpreting CT images in isolation. Independent of the SOC pathway, a separate MDT reviewed AI-positive alerts but did not communicate AI findings to

the radiologists performing SOC interpretation during the study period. The MDT communicated directly with treating clinicians or patients only when a patient-safety concern arose because a clinically important finding flagged by the AI appeared to have been missed by the initial SOC pathway. No feedback was provided to the radiologists performing SOC interpretation, thereby preserving blinding throughout the study.

Between 1 January 2025 and 30 April 2025, 17,526 participants undergoing non-contrast chest CT at SIPD (Center C) across four clinical settings—physical examination, emergency, inpatient, and outpatient—were enrolled in a prospective pilot validation. Of these, 80 individuals were excluded due to prior esophageal cancer treatment, resulting in a final cohort of 17,446 participants across four main clinical settings: physical examination (n=9,484), emergency (n=5,393), outpatient (n=1,433), and inpatient (n=1,136). EAGLE-Plus was integrated into the hospital’s PACS to enable analysis of routine non-contrast CT scans. The system automatically identified eligible examinations using DICOM metadata (e.g., body region, slice thickness, convolution kernel) and RIS records (e.g., study descriptions such as “non-contrast CT”). Relevant data were transferred from PACS to a local server for AI processing. Model’s outputs were displayed in the DAMO MED user interface, facilitating immediate review by clinical staff. All positive predictions of AI were re-evaluated by an MDT in conjunction with available clinical records. Based on this review, the MDT determined whether to recall patients who had been missed by the hospital’s initial SOC assessment, thereby evaluating the incremental diagnostic value of EAGLE-Plus in a real-world setting. Systematic follow-up for the entire prospective cohort was conducted via electronic medical records from the hospital information system until 31 July 2026, ensuring a minimum follow-up duration of 15 months. Two cases initially missed by both the SOC and AI was subsequently identified as EC during follow-up clinical evaluation.

During the study, there were 90 cases predicted as positive by EAGLE-Plus, representing a AI positive rate of 0.5% among all 17,446 participants. Of these, 38 were true positives, including one cancer case initially missed in SOC, and recalled by the MDT. Among the remaining 52 cases, 18 had esophageal findings and 10 had peri-esophageal abnormalities. Furthermore, 19 were false positives, where 3 were confirmed as false positives via endoscopy, 16 were documented as having no esophageal symptoms in clinical records at the latest follow-up. 5 other cases remained lost to follow-up due to unavailable diagnoses by the follow-up cutoff on 31 July 2026.

The initial SOC identified 85 cases with the positive rate of 0.49%, 48 of which overlapped with prediction of EAGLE-Plus. 37 cases were identified by SOC but not flagged by EAGLE-Plus, 27 underwent endoscopic follow-up by the cutoff date, with 7 confirmed as esophageal malignancies (4 HGIN, 3 EC), and 3 cases were lost to follow-up.

Among all 17,446 participants, 47 esophageal malignancies (HGIN and EC) were detected, 36 cases were lost to follow-up by the cutoff date on 31 July 2026. EAGLE-Plus detected 38 of these 47 cases (80.9% sensitivity), including 36 of the 41 EC cases (87.8% sensitivity).

1.5 Details of real-world retrospective LDCT validation

The real-world LDCT cohort included individuals aged 45–75 years who underwent lung cancer screening at the physical examination department of SIPD between July 2024 and December 2024. The study follow-up period ended on 31 July 2026, which served as the primary outcome assessment date. Radiology reports were available at the time of LDCT imaging. Two experienced radiologists first reviewed radiology reports to determine SOC diagnoses of esophageal malignancy. Follow-ups were conducted on 30 June 2025 for two cases with SOC-detected only esophageal malignancies. Additionally, approximately 1% of negative cases were randomly selected from those with concordant negative findings between SOC and AI. Two independent radiologists performed blinded re-evaluations of CT scans and clinical records for these cases to confirm the absence of malignancy.

Special attention was given to eight cases that AI detected as esophageal malignant but not reported in SOC assessments. Follow-up calls were also made on 31 July 2026 for these cases if no clinical records were available. Among eight AI positive predictions, one was confirmed as cancer by the biopsy pathology; Three were diagnosed with non-malignant esophageal conditions: one with benign esophageal tumor (subsequently confirmed by endoscopy), one with reflux esophagitis (based on endoscopy within the past year), and one with esophagotracheal fistula, a serious condition requiring urgent intervention. Four cases had no available clinical records regarding esophageal findings by 31 July 2026.

1.6 Details of validation on prospectively enrolled endoscopic screening cohort

We collected non-contrast CT data from a endoscopic screening program from the SNCH. From 7 July 2023 to 26 July 2023, 530 patients participating in this program were invited to undergo the non-contrast CT examination. In total, 530 patients signed the informed consent (Supplementary Fig. 7). They were first scanned for non-contrast CT and then offered a standard upper gastrointestinal endoscopy with iodine staining (1.2% Lugol's solution) of the esophagus. All focal lesions identified during upper gastrointestinal endoscopy were sampled via biopsy. In the absence of visually detectable abnormalities, standard biopsy specimens were obtained from predetermined anatomic sites located at 28 cm and 33 cm distal to the incisors in the 6 o'clock position. Two board-certified pathologists independently evaluated all biopsy specimens, while maintaining blinded status to endoscopic findings; any inter-observer variability in histopathological interpretations was resolved through expert consultation.

In each hybrid simulation cohort ($n=533$) that includes three bootstrapped EC cases, EAGLE classified participants into positive ($n=131$) or negative ($n=402$) groups. In the positive predictions, there were 3 EC cases, 4 with HGIN, 14 with low-grade intraepithelial neoplasia (LGIN), 45 with precancerous diseases, 1 with other esophageal diseases, and 64 with no abnormalities. In the negative predictions, 300 had no esophageal disease, 77 had precancerous diseases, 23 had LGIN, and 2 had HGIN; no EC cases were contained. According to the Chinese national guidelines for the diagnosis and treatment of esophageal malignancies (Supplementary Table 2),

endoscopic surveillance is recommended every 1-3 years for patients with LGIN and every 5 years for individuals with esophageal precancerous diseases in high-incidence regions. Overall, EAGLE detected 4 of 6 HGIN cases and all 3 EC cases in most simulation analyses. Considering that non-malignant precancerous lesions (LGIN) and diseases may represent early malignant changes, EAGLE identified 14 of 37 (37.8%) LGIN lesions and 45 of 122 (36.9%) esophageal precancerous diseases. The specificity for a normal esophagus (no esophageal precancerous lesions and diseases) was 82.2% (95% CI: 77.7-85.7%).

1.7 Comparison to other methods

A comparative experiment was conducted on the training dataset using five-fold cross-validation to evaluate the proposed method against other models. The following approaches were included in the comparison:

Swin UNETR is a transformer-based architecture specifically designed for medical image segmentation tasks [1]. It integrates the Swin Transformer with a U-Net-like structure, combining the powerful feature representation capabilities of Transformers with the hierarchical, encoder-decoder design of U-Nets. This hybrid architecture enables the model to effectively capture both global context and fine-grained spatial details, making it particularly suitable for segmenting complex anatomical structures such as organs or tumors in medical images. Swin UNETR has demonstrated strong performance in various medical imaging applications, supporting accurate diagnosis and research. In this study, multi-level features extracted from the decoder of Swin UNETR are further fed into a classification head for downstream tasks.

Mask2Former is a transformer-based segmentation framework tailored for both semantic and instance segmentation tasks [2]. Built upon a powerful multi-scale deformable attention transformer, Mask2Former unifies the treatment of panoptic, instance, and semantic segmentation in a single architecture. It employs a mask prediction strategy that iteratively refines segmentation masks via set prediction, facilitating high accuracy for complex boundaries and overlapping structures. Mask2Former has shown competitive performance in both natural and medical imaging benchmarks. In this study, we utilize the mask decoder outputs of Mask2Former for both segmentation of esophageal lesions and classification of esophageal malignancy.

U-Mamba is a general-purpose neural network designed for both 2D and 3D biomedical image segmentation tasks [3]. It is built upon an innovative hybrid CNN-SSM (State Space Model) architecture, enabling the network to effectively capture both localized fine-grained features and long-range dependencies within medical images. Unlike conventional Transformer-based models that are often associated with quadratic complexity with respect to feature size, U-Mamba achieves linear computational scaling, making it more efficient for large medical data. In this study, we evaluate two architectural variants of U-Mamba:

- **U-MambaBot**: In this configuration, only the bottleneck block employs the U-Mamba mechanism, while the remaining network maintains standard convolutional operations.

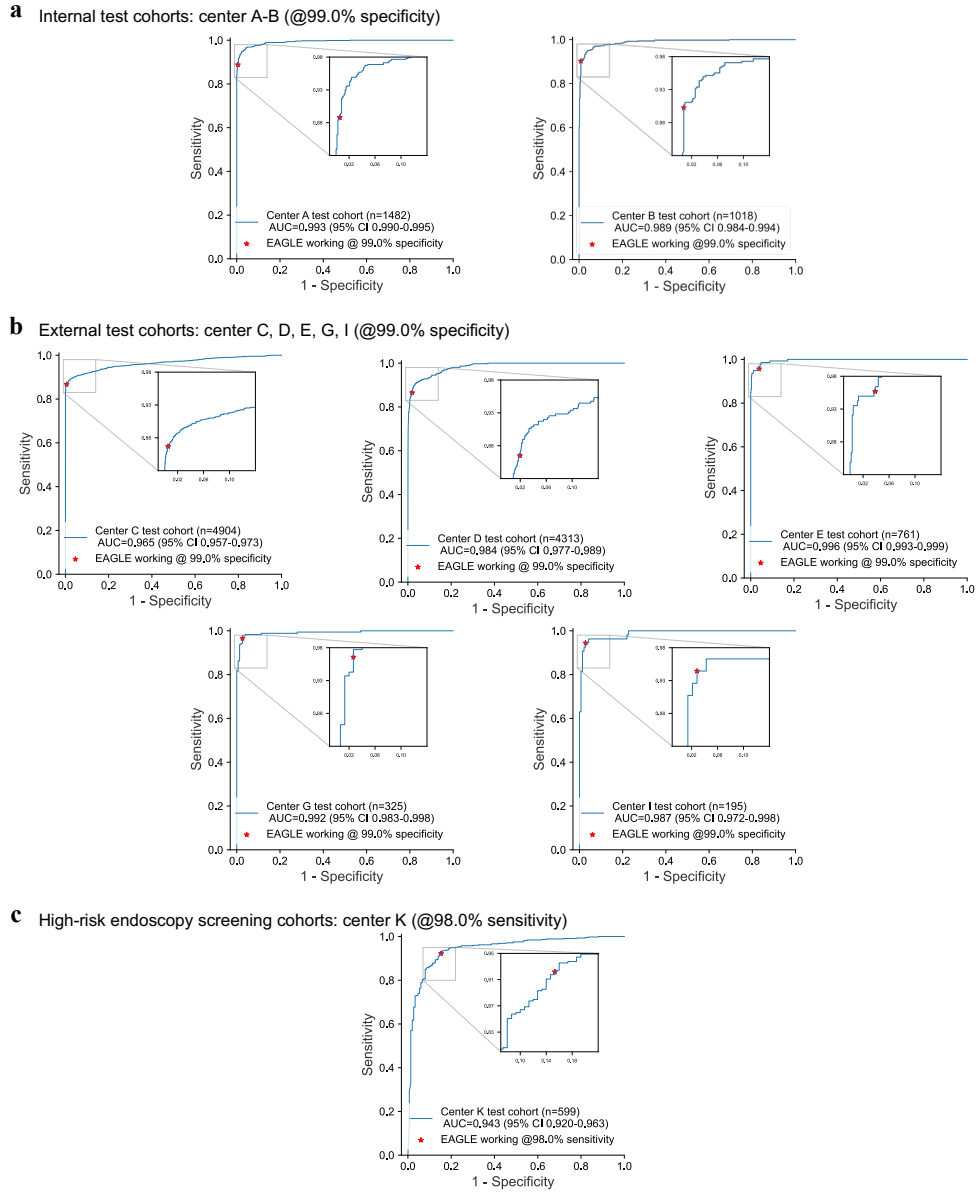
- U-MambaEnc: Here, all encoder blocks are replaced by U-Mamba blocks, enhancing the model's ability to extract global contextual information throughout the encoding pathway.

Both configurations leverage the strengths of the hybrid architecture, providing flexibility for a wide range of clinical segmentation scenarios. The extracted feature representations from these network variants are utilized for segmentation and classification.

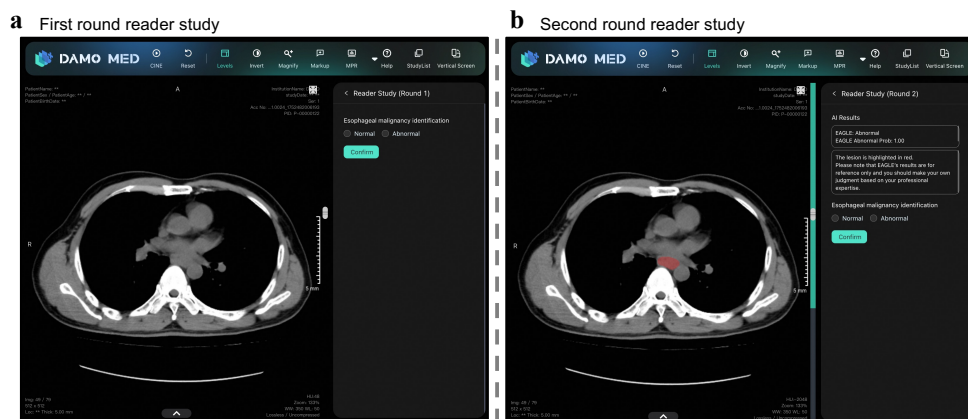
nnDetection is a deep learning framework designed for lesion detection in medical imaging, and optimized for both 2D and 3D data [4]. It provides a modular and flexible pipeline for training and evaluating detectors tailored to medical applications such as identifying anatomical structures, tumors, or pathological regions in CT. nnDetection achieves results that are on par with or even superior to other methods across various public medical detection tasks. In this study, the final patient-level classification results will be determined by the highest anomaly score among all boxes detected by the nnDetection model.

References

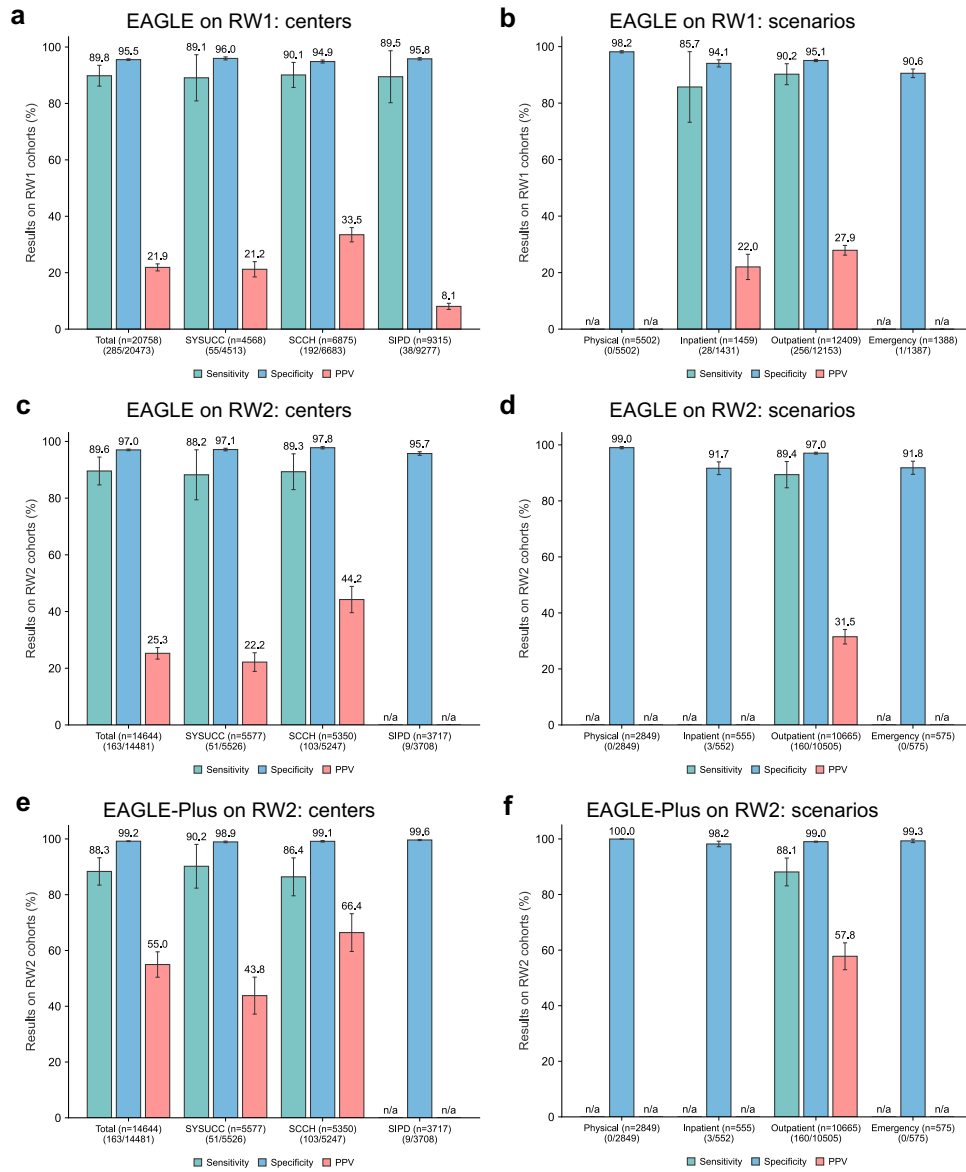
- [1] Tang, Y., Yang, D., Li, W., Roth, H.R., Landman, B., Xu, D., Nath, V., Hatamizadeh, A.: Self-supervised pre-training of swin transformers for 3d medical image analysis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 20730–20740 (2022)
- [2] Cheng, B., Misra, I., Schwing, A.G., Kirillov, A., Girdhar, R.: Masked-attention mask transformer for universal image segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1290–1299 (2022)
- [3] Ma, J., Li, F., Wang, B.: U-mamba: Enhancing long-range dependency for biomedical image segmentation. arXiv preprint arXiv:2401.04722 (2024)
- [4] Baumgartner, M., Jäger, P.F., Isensee, F., Maier-Hein, K.H.: nndetection: a self-configuring method for medical object detection. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24, pp. 530–539 (2021). Springer



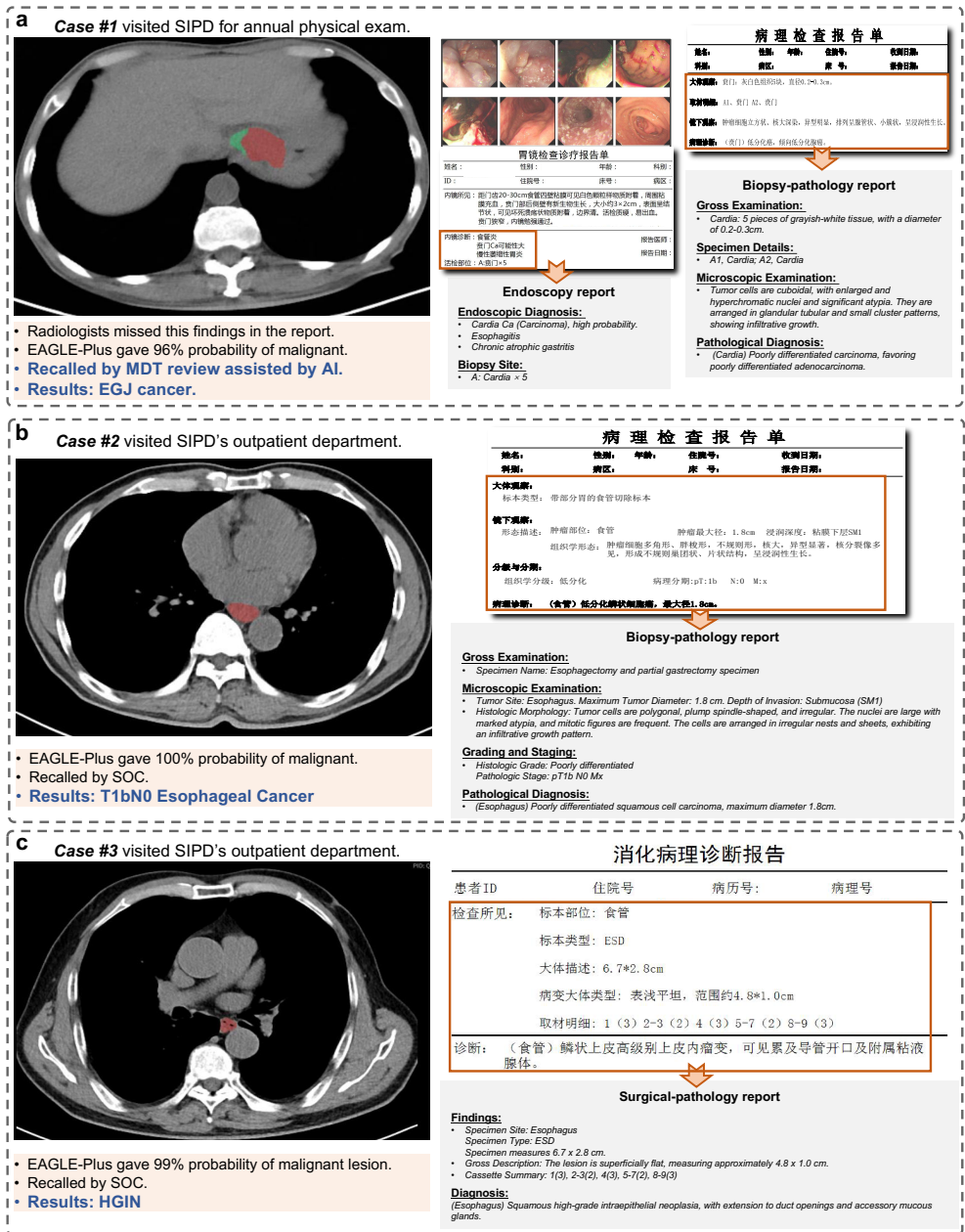
Supplementary Fig. 1: ROC curves and operating point of EAGLE on centers that contain both positive and negative control cases. a, two centers A and B of the internal test cohort, working at 99% specificity. **b,** five centers in external test cohort, working at 99% specificity. **c,** center K in external cohorts for endoscopic group, working at 98% sensitivity.



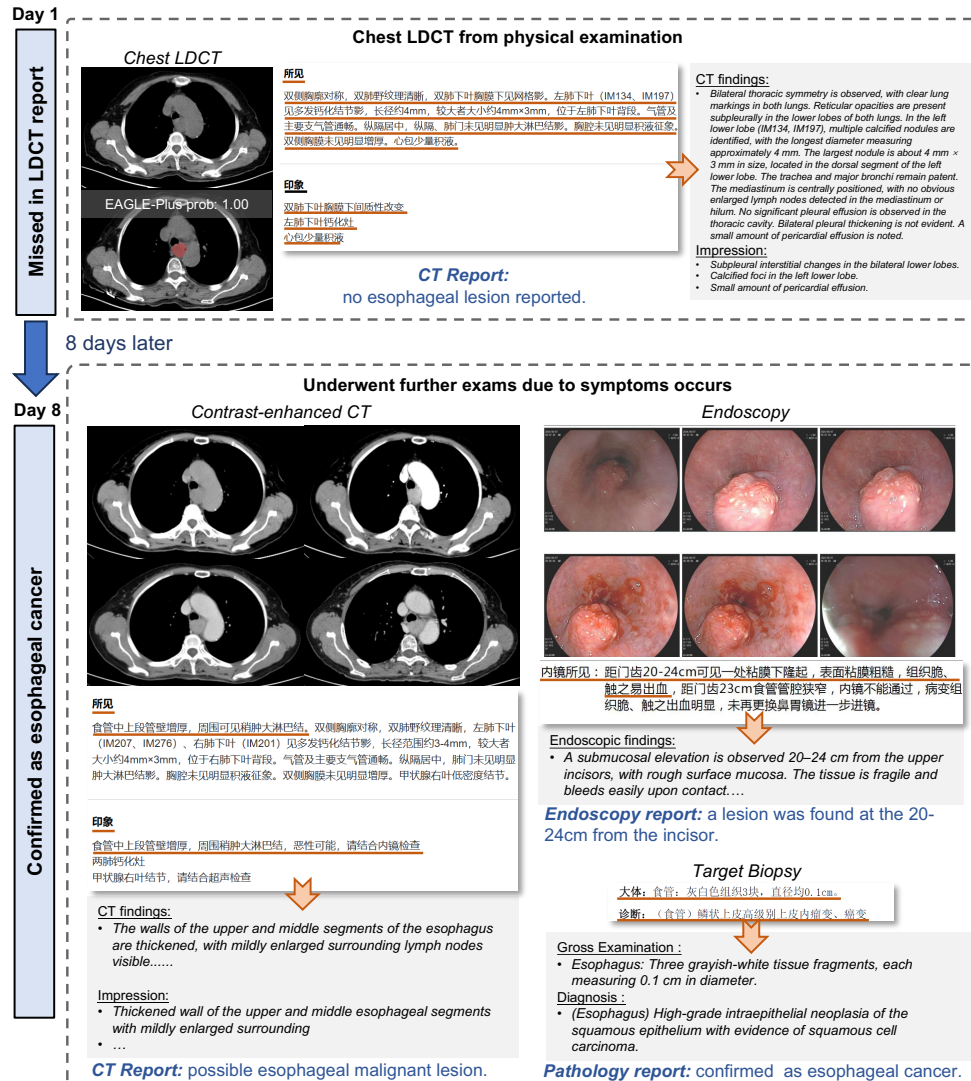
Supplementary Fig. 2: Reader study with DAMO Med system. **a**, The first reader study surface. **b**, The second reader study surface after three months. EAGLE tumor segmentation mask (red color) and positive probability were given in the second reader study, and they are summarized as text shown on the right.



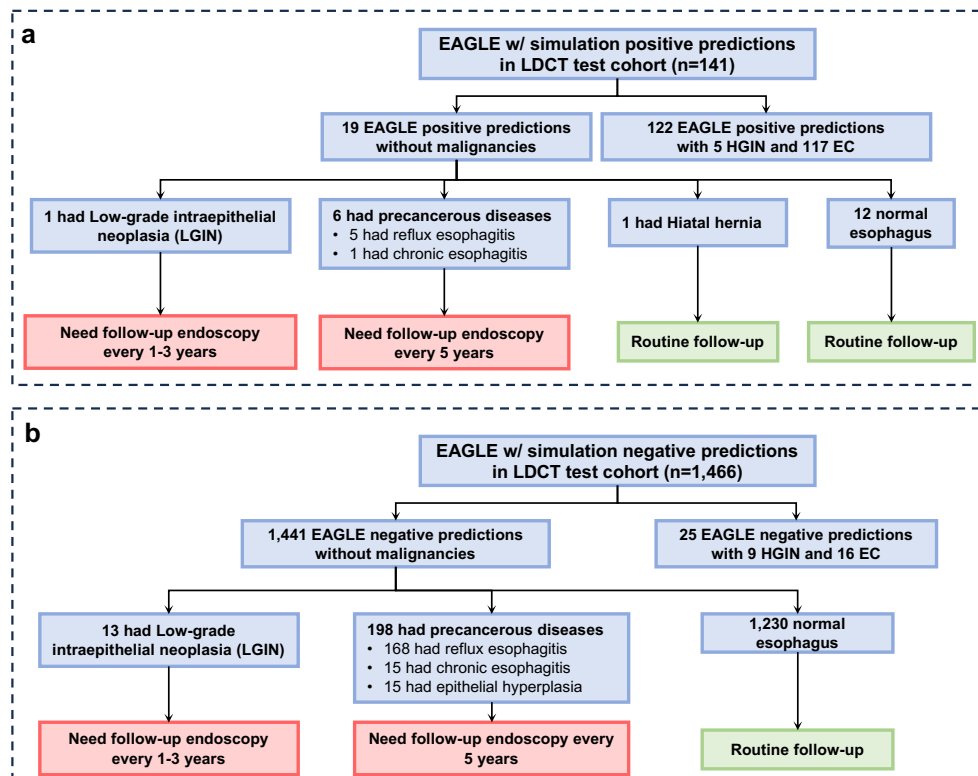
Supplementary Fig. 3: EAGLE and EAGLE-Plus performances on RW1 and RW2 cohort. **a**, EAGLE performances on RW1 across different medical centers. **b**, EAGLE performances on RW1 across different scenarios. **c**, EAGLE performances on RW2 across different medical centers. **d**, EAGLE performances on RW2 across different scenarios. **e**, EAGLE-Plus performances on RW2 across different medical centers. **f**, EAGLE-Plus performances on RW2 across different scenarios. “n/a” (not applicable) is used to indicate patient counts below 10. Positive and negative case numbers are displayed under each center’s name on the x-axis as a stacked bar labeled “positive/negative”. Error bars indicate two-sided 95% CI computed from 1,000 bootstrapping iterations; no adjustment was made for multiple comparisons.



Supplementary Fig. 4: Esophageal malignancy examples found by EAGLE-Plus in the prospective validation. **a**, A patient visited for an annual physical examination but was missed by the initial SOC. Guided by our AI model's prediction, the MDT recalled the patient for endoscopic evaluation; subsequent endoscopy and biopsy confirmed cancer in the esophagogastric junction (EGJ), and the patient subsequently underwent definitive interventional therapy. **b**, A patient visited the outpatient department and was recalled by SOC. EAGLE-Plus could also detect this patient. **c**, A patient visited the outpatient department and was recalled by SOC. EAGLE-Plus could also detect this patient. The English translation is provided for key results (in orange boxes) in reports.



Supplementary Fig. 5: Flowchart illustrating the potential benefits for the individual without symptoms who initially visited the SIPD for a physical exam. The English translation is provided for key results (with orange lines) in reports.



Supplementary Fig. 6: Endoscopic findings of EAGLE predictions in the LDCT test cohort. **a**, Endoscopic findings of 141 EAGLE high-risk participants. 122 of 141 were esophageal malignancies. 12 of 141 were normal esophagus under endoscopy. **b**, Endoscopic findings of 1466 EAGLE low-risk participants. 25 of 1466 were esophageal malignancies. 1441 of 1466 were negative without malignancies.

a

遂宁市中心医院临床试验

平扫 CT 知情同意书

研究名称：四川省科技计划项目（重点研发项目）川渝地区食管癌精准防控及诊疗体系建立与临床推广

受试者姓名 性别：男 ☐ 女 ☐ 年龄：

检查时间： 门诊/住院号：

平扫 CT 检查能清楚显示病灶的范围和性质，有利于明确诊断。平扫 CT 的危害主要来自 X 射线产生的电离辐射，其辐射剂量小，在人体正常辐射安全范围之内，一般不会对受检者造成明显辐射伤害。

以上文本由医院提供，医务人员已详细告知上诉内容，我已知晓该内容，并同意使用该文本。对医师告知的平扫 CT 风险完全知情，经慎重考虑，我 _____（同意/不同意）做平扫 CT。

我知道在本次检查开始之前，我可以随时签署拒绝检查的意见，以取消本知情同意书的决定。

受试者签名(或法定代理人签名): 电话:

研究者签名: 电话:

年 月 日

胸部平扫 CT 检查注意事项:

1、 请勿携带金属制品，会出现伪影，影响检查结果；

2、 保持呼吸节律稳定；

3、 孕妇慎做；

b

Clinical Trial at Suining Central Hospital

Non-contrast CT Informed Consent Form

Research name: Sichuan Provincial Science and Technology Plan Project (Key R&D Project) Establishment and clinical promotion of precise prevention and control and diagnosis and treatment system for esophageal cancer in Sichuan and Chongqing

Subject's name Gender: Male ☐ Female ☐ Age:

Examination time: Outpatient/inpatient number:

Non-contrast CT examination can clearly show the scope and nature of the lesion, which is conducive to a clear diagnosis. The harm of non-contrast CT mainly comes from the ionizing radiation generated by X-rays. Its radiation dose is small and is within the normal radiation safety range of the human body. Generally, it will not cause obvious radiation damage to the examinee.

The above text is provided by the hospital the medical staff has informed me of the content in detail, I have known the content and agree to use the text. I am fully aware of the risks of plain scan CT informed by the doctor. After careful consideration, I _____(agree/disagree) to do non-contrast CT.

I know that before the start of this examination, I can sign the opinion of refusing the examination at any time to cancel the decision of this informed consent.

Signature of subject (or legal representative): Phone:

Signature of researcher: Phone:

Year Month Day

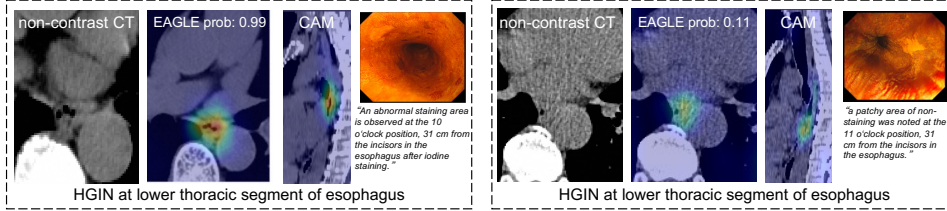
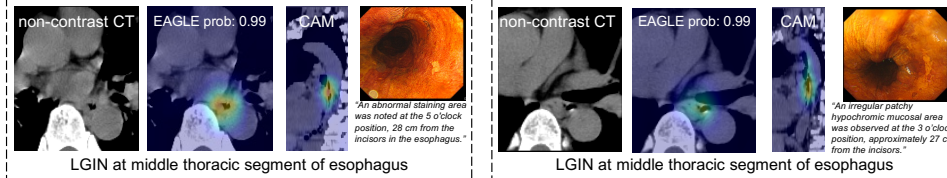
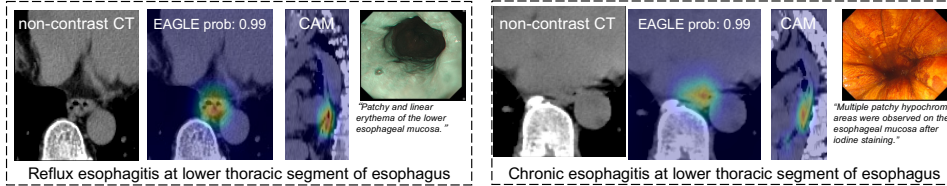
Precautions for chest non-contrast CT examination:

1. Do not carry metal products, as artifacts will appear and affect the examination results;

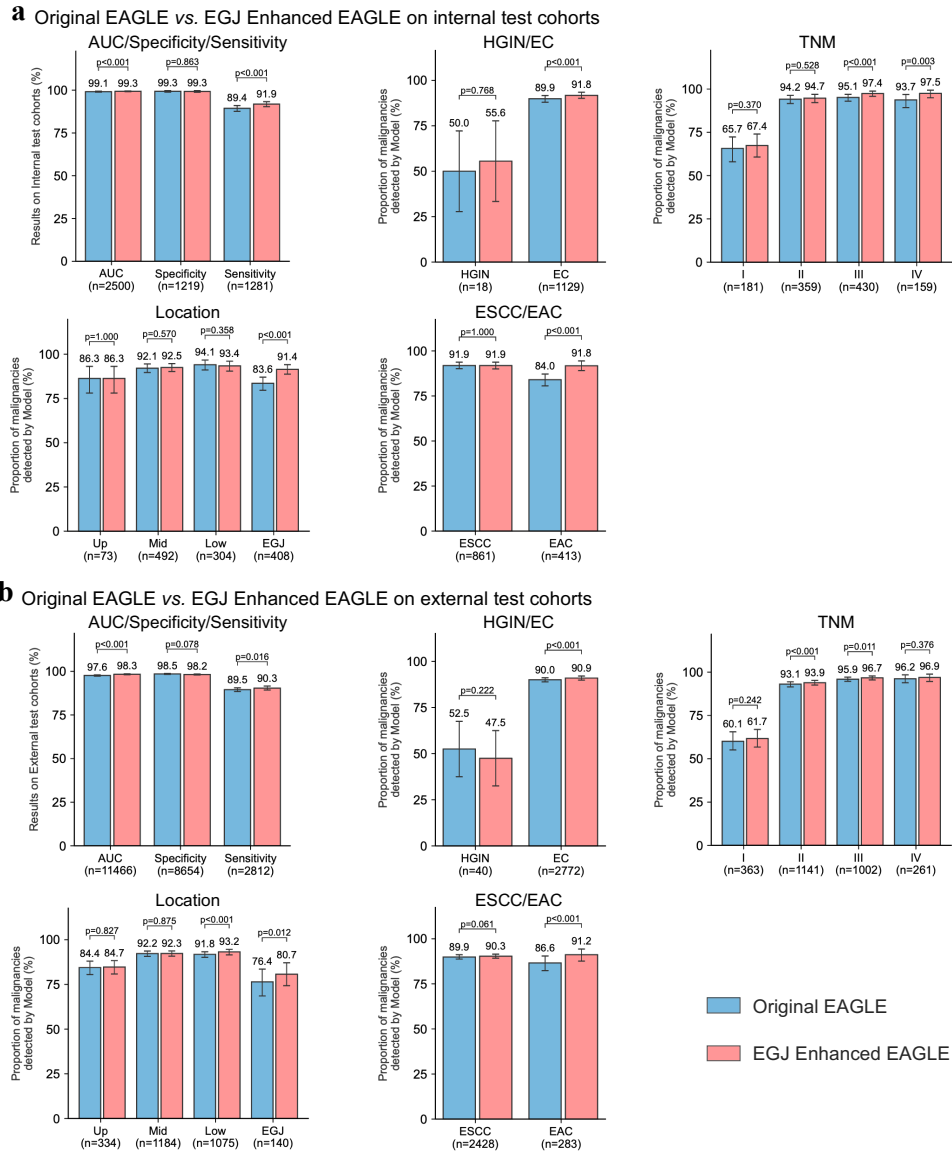
2. Keep a stable breathing rhythm;

3. Pregnant women should be cautious;

Supplementary Fig. 7: Non-contrast CT informed consent form for SuiNing participants. a, Original Chinese version. b, Translated English version.

a Malignant lesions (HGIN)**b** Precancerous lesions (LGIN)**c** Precancerous diseases

Supplementary Fig. 8: Analysis of EAGLE interpretability with positive endoscopic findings. We present the non-contrast CT, EAGLE's probability, EAGLE's CAM heatmap, and the corresponding endoscopy image and report for malignant lesions, precancerous lesions (LGIN), and precancerous diseases, respectively. The endoscopy screening method is endoscopy with esophageal mucosa iodine solution staining with targeted biopsy. All CAM results were matched with endoscopy findings confirmed by doctors. **a**, Two malignant lesions (HGIN) examples. The right example presents EAGLE made classification based on the local features of this lesion at the lower thoracic segment of the esophagus (CAM heatmap), which was matched with endoscopic findings. **b**, Two precancerous lesions (LGIN) examples. **c**, Two precancerous diseases examples.



Supplementary Fig. 9: Comparison between the Original EAGLE and the EGJ Enhanced EAGLE. The Original EAGLE refers to the frozen model utilized in the primary analysis and the prospective validation. The EGJ Enhanced EAGLE is an optimized version with an additional 81 EGJ training cases to demonstrate the learnability of EGJ and EAC on non-contrast CT. Note that the performance of the EGJ Enhanced EAGLE was not reported in the main results to maintain the integrity of the frozen model used for prospective validation; these data are presented here solely to characterize the potential for sensitivity improvement through data expansion. Error bars indicate two-sided 95% CI computed from 1,000 bootstrapping iterations. p values computed via two-sided paired permutation tests; no adjustment was made for multiple comparisons.

Reader ID	Experience (year)	Training/Expertise	CT read per year	Esophagus CT read per year
Specialist 1 (S1)	20	Esophagus radiology	2000	500
Specialist 2 (S2)	17	Esophagus radiology	3000	500
Specialist 3 (S3)	15	Esophagus radiology	3000	1000
General 1 (G1)	12	General radiology	1500	400
General 2 (G2)	12	General radiology	1500	400
General 3 (G3)	8	General radiology	1300	300
General 4 (G4)	5	General radiology	1300	300
General 5 (G5)	10	General radiology	1300	300
General 6 (G6)	3	General radiology	500	100
Resident 1 (R1)	4	General radiology	500	100
Resident 2 (R2)	3	General radiology	500	100
Resident 3 (R3)	4	General radiology	500	100
Resident 4 (R4)	4	General radiology	500	100
Resident 5 (R5)	3	General radiology	500	100
Resident 6 (R6)	3	General radiology	500	100
Resident 7 (R7)	3	General radiology	500	100
Resident 8 (R8)	3	General radiology	500	100

Supplementary Table 1: Reader experiences. Experience of CT reading and expertise of each reader.

Categories	Types	Recommendations for Individuals aged 45 years or older
Precancerous lesion	high-grade intraepithelial neoplasia (HGIN)	Endoscopic Submucosal Dissection (ESD) or Endoscopic Mucosal Resection (EMR)
Precancerous lesion	Low-grade intraepithelial neoplasia (LGIN)	Follow-up endoscopy every 1-3 years for LGIN
Precancerous diseases	Barrett's esophagus, chronic esophagitis, reflux esophagitis, epithelial hyperplasia, esophageal diverticula, or benign esophageal stricture	Follow-up endoscopy every 5 years
Other diseases	Esophageal-gastric varices, hiatal hernia	Rountinue follow-up
Normal esophagus	no findings in endoscopy	Rountinue follow-up

Supplementary Table 2: Definition of esophageal lesions and diseases and national guidelines to manage endoscopic findings (2024 Edition) by National Health Commission of the People's Republic of China. <https://www.nhc.gov.cn/ylyjs/gzdt/202406/2bb28e7a39ce4734bee4e31ff5442e0f.shtml>

Dataset	Malignancy	Standards
Training Set	Positive	Biopsy or surgical pathology
	Negative	Confirmed by at least four years of clinical follow-up
<i>Validation cohorts for opportunistic screening in hospital:</i>		
Internal & External Test	Positive	Biopsy or surgical pathology
	Negative	Confirmed by at least two years of clinical follow-up
Real-world Retrospective Hospital Validation	Positive	Biopsy or surgical pathology
	Negative	Clinical follow-up between the initial SOC and the follow-up SOC before model evaluation
Prospective Validation	Positive	Biopsy or surgical pathology
	Negative	Clinical follow-up between the initial SOC and the follow-up SOC before the cutoff
<i>Validation cohorts for opportunistic screening in established LDCT programs:</i>		
Low-dose CT Test	Positive	Endoscopic biopsy pathology
	Negative	Negative endoscopic results within one year
Real-world Retrospective LDCT Validation	Positive	Biopsy or surgical pathology
	Negative	Clinical follow-up between the initial SOC and the follow-up SOC before model evaluation
<i>Validation cohorts for population-based endoscopic screening:</i>		
External Test	Positive	Biopsy or surgical pathology
	Negative	Negative endoscopic results within one year
Prospectively Enrolled Validation	Positive	Endoscopic biopsy pathology
	Negative	Endoscopic biopsy pathology

Supplementary Table 3: Standard of truth for all training and validation cohorts, including both positive and negative cases.

Region	Examination	Link	Source
USA	Endoscopy	www.newchoicehealth.com/endoscopy/cost	New Choice Health: website for comparing prices for common medical procedures across the US. National Public Radio.
	Low-dose CT	www.npr.org/sections/health-shots/2017/01/30/512468452/lung-cancer-screening-program-finds-a-lot-thats-not-cancer	
Singapore	Endoscopy	colorectalsurgery.com.sg/areas-of-expertise/gastroscopy-singapore/	A colorectal/general surgery clinic in Singapore. A healthcare provider in Singapore.
	Low-dose CT	mediwaymedical.com/ct-scan/	
China	Endoscopy	-	Suining Central Hospital, our external test cohort, center J.
UK	Endoscopy	www.england.nhs.uk/wp-content/uploads/2025/04/25-26NHSPS-Annex-A-Prices-workbook.xlsx	NHS Payment Scheme 2025/26 guide prices for UK.
	Non-contrast CT	-	
Netherlands	Endoscopy	www.umcg.nl/documents/d/umcg.nl/tarieven-onderlinge-dienstverlening-2024	University Medical Center Groningen official price list for Netherlands 2024.
	Non-contrast CT	-	
Australia	Endoscopy	medicalcostsfinder.health.gov.au	Australian Government Medical Costs Finder.
South Korea	Endoscopy	www.nhis.or.kr/nhis/together/retrieveForumUserView.do?seq=1222	National Health Insurance Service of South Korea. Health Insurance Review & Assessment Service of South Korea.
	Non-contrast CT	www.hira.or.kr/ebooksc/ebook.560/ebook.560_201912300924568680.pdf	
Japan	Endoscopy	www.mhlw.go.jp/content/12404000/000907834.pdf	Japan's Ministry of Health, Labour and Welfare, medical fee points table.
	Non-contrast CT	-	

Supplementary Table 4: Region specific economic cost estimates for endoscopy and non-contrast/low-dose CT. The estimated cost of endoscopic cancer screening encompasses the procedure itself, sedation/anesthesia, and chromoendoscopy. Costs associated with biopsy and histopathology were excluded from the baseline estimates, as these interventions are not universally required for all screened individuals.