nature medicine

Article

https://doi.org/10.1038/s41591-026-04656-4

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

Received: 22 July 2026

Accepted: 17 August 2026

Published online: 22 September 2026

Check for updates

A list of authors and their affiliations appears at the end of the paper

The absence of accurate, noninvasive, scalable screening tools keeps early esophageal cancer (EC) detection a global health challenge. Although noncontrast computed tomography (NC CT) is widely accessible, the esophagus is a hollow tubular structure prone to collapse and motion artifacts, making small early malignant lesions difficult to distinguish from normal tissue. Here we developed the Esophageal AI-Guided malignant Lesion Evaluation (EAGLE) model to detect precancerous lesions and cancer from chest NC CT, a task historically considered impossible. EAGLE was trained on 6,813 patients from two centers and validated across 12 centers in three countries involving 80,612 patients in opportunistic and population-based screening settings. For opportunistic screening on existing CT scans, multicenter external test cohorts (eight centers, $n = 11{,}466$) achieved 98.5% specificity, with 90.0% sensitivity for cancer and 52.5% for precancerous lesions; low-dose CT (LDCT) validation (two centers, $n = 1{,}607$) showed comparable performance, supporting EC screening through lung-cancer screening programs. Calibration in a real-world cohort (three centers, $n = 35{,}402$) reduced false positives by 72.7% while preserving sensitivity; prospective hospital validation ($n = 17{,}446$) achieved a 42.2% PPV, and real-world low-dose screening ($n = 10{,}959$) reached 99.94% specificity. EAGLE also detected precancerous lesions—in paired CT–endoscopy cohorts (two centers, $n = 702$), sensitivities were 65.0% for precancerous lesions and 78.4% for stage I EC at a higher-sensitivity operating point. Exploratory analyses of a prospectively enrolled cohort suggest that referring high-risk individuals for endoscopy could improve screening efficiency. In conclusion, EAGLE has the potential to serve as a scalable tool for early EC screening. Chictr.org.cn identifier: ChiCTR2300074806.

Esophageal cancer (EC) is one of the most common digestive tract cancers worldwide, with estimated 511,000 new cases and 445,000 deaths in 2022 (ref. 1). EC comprises two major histological subtypes, esophageal squamous cell carcinoma (ESCC) and esophageal adenocarcinoma (EAC), which differ substantially in their risk factors, anatomical locations and geographic distributions. It remains a critical global health issue, characterized by high mortality rates due to its frequent diagnosis at advanced stages and the absence of formalized screening programs[2,3]. Although the prognosis of EC has improved in the past decades, survival remains unsatisfactory, as the 5-year overall survival rates are 36.9% in China and 18.5% in the United States[4,5]. While high-risk population-based endoscopic screening has made a substantial contribution to the early detection and intervention of EC in China[6–8], its invasive nature and low compliance make it impractical for large-scale screening of asymptomatic populations globally, particularly in countries and regions lacking adequate endoscopy

e-mail: zhangdingwen2006yyy@gmail.com; jijiansong@zju.edu.cn; 1500921245@qq.com; xchuanm@sysucc.org.cn; mdkaicao163@163.com; zhaoyuqian1211@scszlyy.org.cn; ling.z@alibaba-inc.com; wangqifeng@scszlyy.org.cn

https://doi.org/10.1038/s41591-026-04656-4

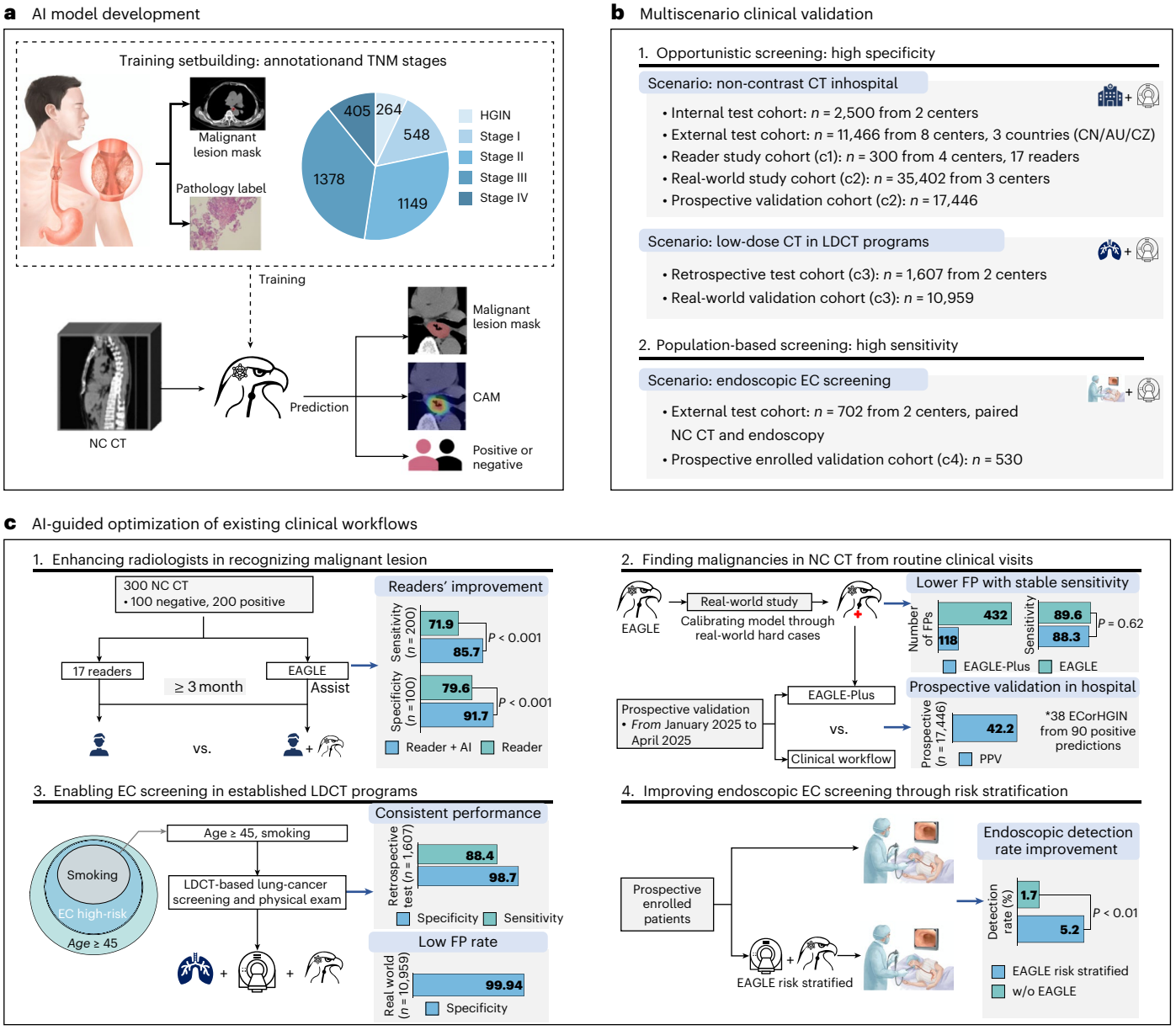



**Fig. 1 | Overview of the EAGLE development, validation and clinical integration framework. a**, AI model development. EAGLE was designed to process NC CT scans for EC diagnosis, output binary classification (positive/negative), malignant lesion masks and heatmaps that highlight regions contributing to diagnostic decisions. EAGLE was trained on CT datasets from two medical centers, and each malignant case included histopathology-confirmed labels and expert-annotated lesion masks. **b**, Multiscenario evaluation. Validation of opportunistic screening in the hospital, including an internal validation cohort, an external validation cohort across eight centers from three countries (China, Czech Republic and Australia), a reader study, a large-scale real-world study and a prospective validation. Validation in existing LDCT-based programs, including a retrospective assessment on two centers, and a real-world LDCT cohort. **c**, AI-guided optimization of existing clinical workflows. Four strategic implementation scenarios demonstrate EAGLE's potential to enhance existing EC care pathways. CAM, class activation mapping; w/o, without.

infrastructure. Nonendoscopic methods, such as liquid biopsy[9–13] and sponge cytology[14,15], have been evaluated for EC screening, but the moderate sensitivity in early-stage cancer, the substantial costs and the requirement of several visits to the clinics pose challenges for their wide implementation. An accurate, noninvasive, cost-effective and rapid tool for worldwide EC screening remains an unsolved clinical problem.

Noncontrast computed tomography (NC CT) is a widely used medical imaging protocol in routine clinical visits and physical examinations, showing characteristics that are noninvasive, cost-effective and rapid[16]. However, identifying esophageal malignancies on NC CT is challenging even for experienced radiologists, largely because early-stage esophageal malignant lesions can be extremely small, often confined to the epithelial layer (high-grade intraepithelial neoplasia, HGIN) or invading only the mucosa or submucosa (stage I). In addition, the esophagus is a thoracic organ with a long, narrow tubular structure and is therefore frequently affected by physiological collapse and susceptible to motion from the heart and great vessels, making subtle lesions difficult to distinguish from normal tissue and limiting the utility of CT for EC screening. Consistent with this, a recent study revealed that the majority of individuals who died from EC had no suspicious esophageal findings documented in their chest NC CT reports within the preceding 5 years[17]. Recent advancements in artificial intelligence (AI) have shown that data-driven models can detect subtle lesions on medical images[18–23], and have highlighted the potential of

AI approaches to improve cancer screening[24–27], offering an avenue for early EC detection on NC CT. Enabling EC screening could open new possibilities for chest NC CT scans (40% of all CT examinations[28]) that already cover the full esophagus, including both regular-dose CT obtained during clinical visits and LDCT used in lung-cancer screening programs. Particularly, given that LDCT lung-cancer screening is now performed worldwide, and smoking is a shared risk factor for lung and ECs[29,30], integrating EC detection into LDCT workflows may enable a one-scan, multicancer screening paradigm without requiring additional imaging procedures. At the same time, LDCT introduces an additional level of difficulty for esophageal lesion detection because it has lower image quality and increased noise.

In this study, we present Esophageal AI-Guided malignant Lesion Evaluation (EAGLE; Fig. 1), a new AI model that not only identifies EC but also detects malignant precancerous lesions (HGIN). This capability supports the potential expansion of AI from opportunistic finding to active population-based prevention—it not only achieves high accuracy in opportunistically identifying malignancies missed by the standard of care (SOC), but also improves efficiency in population-based endoscopic screening. For 'opportunistic screening', where EAGLE operates at high specificity, multicenter reader studies showed that it significantly enhances radiologists' ability to identify early cancers while lowering the false-positive (FP) rate; large-scale multicenter external validation across three countries confirmed its generalizability in hospital scenarios; a simulation training strategy demonstrated comparable specificity and sensitivity to those achieved with regular-dose scans in an LDCT cohort; and large-scale real-world studies recalibrated the model to markedly lower the FP rate without sacrificing sensitivity, confirmed exceptionally high specificity in a consecutive physical examination LDCT cohort and demonstrated a clinically desirable positive predictive value (PPV) in a prospective hospital validation.

EAGLE also showed potential to boost established 'population-based endoscopic screening programs', where it operates at a high sensitivity threshold to minimize the risk of missed cancers—we retrospectively evaluated performance on NC CT scans from individuals who underwent endoscopic screening at two external centers, and integrated EAGLE into a standard endoscopic screening workflow for a prospectively enrolled high-risk cohort, where simulation analysis suggested that using EAGLE for pre-endoscopy risk stratification could markedly reduce unnecessary endoscopic procedures while preserving cancer detection, thereby improving the detection rate and coverage of current endoscopic screening programs.

## Results

### The EAGLE model

EAGLE identifies high-risk patients with malignant esophageal lesions using NC CT scans, outputting binary classification (positive or negative), malignant lesion segmentation masks and classification heatmaps that highlight regions contributing to diagnostic decisions (Fig. 1a). Positive cases were defined as malignant esophageal lesions, including histology-confirmed esophageal carcinoma (EC) and HGIN. Negative controls were patients without EC and HGIN, who were confirmed by at least 2 years of clinical follow-up or negative endoscopic screening within 1 year in the retrospective datasets. The standard of truth for all cohorts is provided in Supplementary Table 3. EAGLE was trained on a two-center cohort of NC CT scans from 6,813 patients (264 = HGIN, 548 = stage I, 2,932 = stages II–IV and 3,069 = negative controls) from Sun Yat-sen University Cancer Center (SYSUCC) and Sichuan Cancer Hospital (SCCH). The patient characteristics are shown in Extended Data Table 1. Besides patient-level classification labels, EAGLE was also supervised by voxel-wise annotations, including both the esophagus and malignant lesions. We recruited an annotation team to perform a rigorous annotation procedure using the cloud-based DAMO MED annotation system (Extended Data Fig. 1 and Methods).

EAGLE is a two-stage approach—the first stage localizes the esophagus within the entire three-dimensional (3D) CT scan and the second stage simultaneously segments lesions and predicts the esophageal malignant probability for the patient. The second stage model consists of dual branches, which share multilevel features between the segmentation and classification branches. The patient-level classification output is used for risk prediction, whereas the segmentation output and classification heatmap localize suspicious lesions and provide spatial evidence to support interpretability (Extended Data Fig. 2).

The decision threshold of our model can be adjusted to meet specific application requirements. For the opportunistic screening scenario, which involves a general clinical population at the average risk of esophageal malignancy, the threshold was set to achieve 99% specificity during cross-validation. When EAGLE was used as a risk-stratification tool in the endoscopic screening program for high-risk individuals, the threshold was set to achieve 98% sensitivity during cross-validation. Detailed information on operation point and threshold selection is provided in Methods.

### Internal and external validation for opportunistic screening

**Regular-dose NC CT validation.** EAGLE's performance on regular-dose NC CT was first evaluated in two internal cohorts comprising 2,500 patients (18 = HGIN, 181 = stage I, 948 = stages II–IV, 134 positive cases with unavailable staging information, and 1,219 = negative controls), followed by testing in eight external cohorts across China, the Czech Republic and Australia, comprising 11,466 patients (40 = HGIN, 363 = stage I, 2,404 = stages II–IV and 8,654 = negative controls). In internal testing, EAGLE achieved an area under the curve (AUC) of 0.991 (95% confidence interval (CI) = 0.989–0.994), a sensitivity of 89.4% (95% CI = 87.7–91.1%) and a specificity of 99.3% (95% CI = 98.8–99.8%; Fig. 2b). External multicenter testing demonstrated consistent performance across institutions and countries, with AUC of 0.976 (95% CI = 0.972–0.980), sensitivity of 89.5% (95% CI = 88.3–90.6%) and specificity of 98.5% (95% CI = 98.2–98.8%; Fig. 2b). In the international cohorts, sensitivity was 94.4% in the Czech Republic cohort and 90.2% in the Australian cohort, with a specificity of 97.2% in the Czech Republic cohort. AUCs of each internal and external cohort are shown in Extended Data Table 2.

We also analyzed the proportion of esophageal malignancies detected by EAGLE in subgroups (Fig. 2d). First, EAGLE could detect malignant precancerous lesions and showed high performance for cancerous lesions (Fig. 2d; HGIN/EC)—internal validation showed sensitivities of 50.0% (95% CI = 27.8–72.2%) for HGIN and 89.9% (95% CI = 88.0–91.6%) for EC; external validation further yielded sensitivities of 52.5% (95% CI = 37.5–67.5%) for HGIN and 90.0% (95% CI = 88.9–91.1%) for EC. Second, for TNM stages of EC, sensitivity increased with stage advancement (Fig. 2d; TNM stage). Particularly, for early-stage EC (stage I), EAGLE achieved a sensitivity of 65.7% (95% CI = 58.6–72.4%) in the internal test and 60.1% (95% CI = 54.8–65.0%) in the external test. Example cases of HGIN and stage I EC are shown in Fig. 2h. Third, EAGLE demonstrated higher sensitivity for lesions in the middle esophagus than for those at the esophagogastric junction (EGJ). Histologic subtype-specific analyses showed performance across both ESCC and EAC, and detection of EAC can be improved after the incorporation of additional EAC training cases (Supplementary Fig. 9). In both the internal and external test cohorts, sensitivity was higher in males than in females. Performance was also higher in patients with body mass index ≤22.

**LDCT validation.** To assess EAGLE's performance on chest LDCT, we collected 147 LDCT scans from patients with esophageal malignancy and 1,460 LDCT scans from negative controls at Shanghai Institution of Pancreatic Diseases (SIPD; center C) and SCCH (center B). A low-dose simulation tool[31] was used to generate simulated LDCT data from the training set, and we trained a corresponding model, termed EAGLE with

https://doi.org/10.1038/s41591-026-04656-4

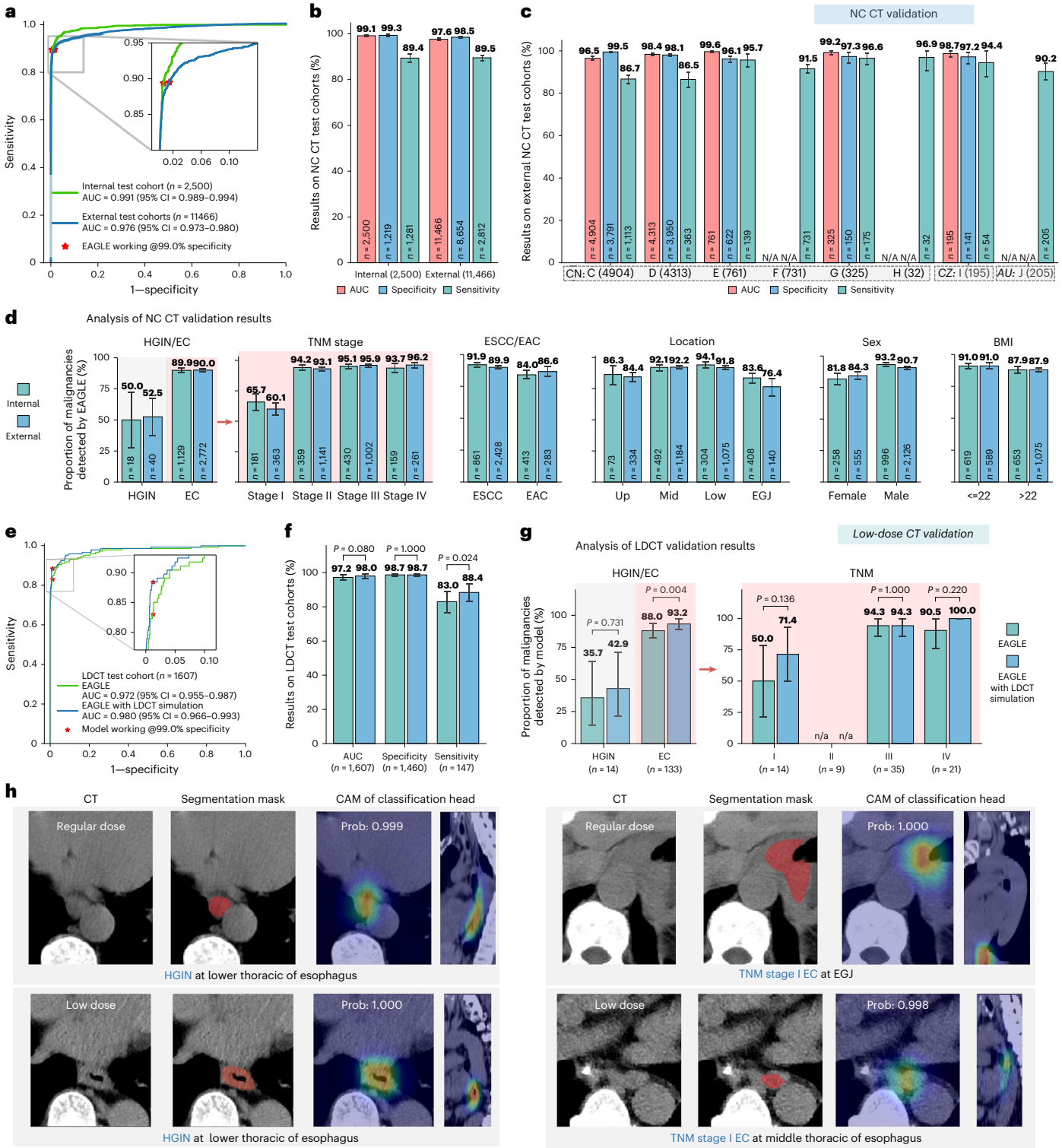



**Fig. 2 | Internal and external validation for opportunistic EC screening.** **a**–**d**, Results from NC CT test cohorts. **a**, ROC curves of EAGLE in internal ($n = 2{,}500$) and external NC CT test cohorts ($n = 11{,}466$). **b**, The sensitivity, specificity and AUC. **c**, The sensitivity, specificity and AUC of EAGLE in each center of external NC CT test cohorts (centers C–J). **d**, The proportion of esophageal malignancies detected by EAGLE for HGIN and EC, TNM stages, histology subtype, locations of lesion, sex and BMI of patients. **e**–**g**, Results from LDCT test cohorts (centers B and C, $n = 1{,}607$). We report the performance of EAGLE with or without LDCT simulation strategy. **e**, ROC curves. **f**, The sensitivity, specificity and AUC. **g**, The proportion of esophageal malignancies detected for HGIN and EC, and TNM stages. **h**, Four malignant lesion examples detected by EAGLE. Lesion segmentation masks and heatmap visualizations are provided. In **b**–**d**, **f** and **g**, error bars indicate two-sided 95% CI computed from 1,000 bootstrapping iterations. In **f** and **g**, *P* values were calculated using two-sided paired permutation tests of 1,000 permutations, with AUC compared by the two-sided DeLong test; no adjustment was made for multiple comparisons. Results for groups with too few samples ($n \leq 10$) are marked as not applicable (N/A). ROC, receiver operating characteristic; BMI, body mass index.

LDCT simulation (Fig. 2e,f and Methods). At the operating threshold corresponding to 99.0% specificity, EAGLE with LDCT simulation model achieved an overall sensitivity of 88.4% (95% CI = 83.2–88.7%), significantly outperforming the original EAGLE's 83.0% (95% CI = 76.9–88.7%; $P = 0.024$; Fig. 2f). The same trend was observed when stratified by TNM stage (Fig. 2g), particularly for HGIN, where the sensitivity reached 42.9% (95% CI = 21.4–71.4%). Example cases are shown in Fig. 2h. Endoscopic findings for both EAGLE-positive and EAGLE-negative predictions are provided in Supplementary Fig. 6.

Ablation studies about different sizes of training data and baseline comparisons are provided in the Methods (Extended Data Table 3). Increasing the training data led to consistently improved performance, particularly in terms of sensitivity (Extended Data Fig. 3). In the internal test set of 515 NC CT scans with lesion annotations (Extended Data Fig. 1), the mean lesion segmentation of Dice similarity coefficient (DSC) was 0.744 (95% CI = 0.722–0.766). Interpretability through activation maps was provided in Extended Data Fig. 2c.

### Reader studies

We conducted two rounds of reader studies with 17 radiologists to evaluate EAGLE's ability to assist clinicians, using 300 NC CT scans collected from four centers (two internal and two external). To mitigate potential recall bias, a washout period of at least 3 months was implemented between two reading rounds. The reader studies were conducted using the online DAMO MED system. In the first round, readers classified each case as esophageal malignancy or not (Supplementary Fig. 2a). For lesion detection, EAGLE outperformed all 17 readers in identifying HGIN and EC (Fig. 3a). Its performance was also superior to the reader average, with a 19.1% higher sensitivity and an 18.4% higher specificity (Fig. 3c; $P < 0.001$). For early-stage EC, the average sensitivity of the 17 readers was 36.3% (95% CI = 27.4–46.1%) for HGIN and 59.8% (95% CI = 49.3–69.1%) for stage I (Fig. 3d). In contrast, EAGLE achieved significantly higher sensitivities than the reader average ($P < 0.05$), reaching 65.7% (95% CI = 48.6–80.0%) for HGIN and 87.1% (95% CI = 74.2–96.8%) for stage I (Fig. 3d).

In the second round, the readers evaluated the same cases with EAGLE's assistance, including predicted malignant lesion segmentation and malignancy probabilities (Supplementary Fig. 2b). EAGLE assistance improved readers' accuracy in identifying EC (Fig. 3b)—sensitivity improved from 71.9% (95% CI = 67.7–76.4%) to 85.7% (95% CI = 82.2–89.2%; Fig. 3c; $P < 0.001$). Also, EAGLE can help to reduce FPs—specificity improved from 79.6% (95% CI = 76.6–82.6%) to 91.7% (95% CI = 87.6–95.1%; Fig. 3c; $P < 0.001$). Notably, EAGLE could help readers in identifying early-stage lesions—the mean reader sensitivity was improved by 16.6% ($P < 0.05$) for HGIN and by 22.7% ($P < 0.001$; Fig. 3d) for stage I. With the help of EAGLE, the residents' performance with AI could approach that of esophageal specialists (Fig. 3e). Representative cases were shown in Fig. 3f, where most readers failed to identify them in the first round. With the assistance of EAGLE, more readers could identify those malignant esophageal lesions in the second round.

### Real-world validation for opportunistic screening

**Model calibration and real-world retrospective validation.** To control the workload of radiologists and reduce downstream evaluation of patients in real-world opportunistic screening, a lower FP rate is particularly important. Therefore, we conducted a large-scale ($n = 35{,}402$) multiscenario real-world study across three large-volume medical centers (SIPD, SYSUCC and SCCH). The model was calibrated through hard cases from the first cohort of each center (RW1), and then the enhanced model (termed EAGLE-Plus) was evaluated on the second cohort of each center (RW2). Two time points were used to define the standard of truth for each patient—the initial SOC, referring to the diagnosis made at the initial clinical visit when the NC CT was acquired; and the follow-up SOC, referring to the diagnosis established during subsequent follow-up before EAGLE evaluation.

The RW1 cohort enrolled 20,758 consecutive patients from three centers, of whom 285 were diagnosed with esophageal malignancies. EAGLE demonstrated a sensitivity (89.6%; 95% CI = 86.1–93.5%), comparable to that in the external test cohort, but the specificity decreased to 95.5% (95% CI = 95.3–95.8%; Supplementary Fig. 3a). This decrease in specificity was primarily attributed to the inclusion of patients from outpatient, inpatient and emergency settings, whose clinical profiles differed substantially from the curated cohorts used for model development (Supplementary Fig. 3b). Notably, EAGLE identified three esophageal malignant cases that were missed by the initial SOC (Fig. 4a). One lesion was detected 21 months earlier than standard clinical diagnosis.

To reduce FPs, we implemented an iterative refinement using 'hard cases' identified during deployment across three clinical centers (RW1; Methods). In the RW2 cohort ($n = 14{,}644$), which included 163 cases of HGIN or EC, EAGLE-Plus led to a 72.7% reduction in FPs (from 432 to 118; Fig. 4a). Consequently, the PPV more than doubled, improving from 25.3% (95% CI = 23.2–27.3%) to 55.0% (95% CI = 50.4–59.5%; $P < 0.001$). Regarding sensitivity, EAGLE-Plus (88.3%; 95% CI = 83.4–93.3%) maintained comparable performance to EAGLE (89.8%; 95% CI = 86.1–93.5%; $P = 0.6186$; Supplementary Fig. 3c,e). Furthermore, EAGLE-Plus demonstrated high specificity in four clinical scenarios, notably reaching nearly 100% specificity in the physical examination scenario (Supplementary Fig. 3f).

**Prospective validation in hospital setting.** To assess whether EAGLE-Plus could maintain performance and identify unrecognized esophageal malignancies in real-world clinical applications, EAGLE-Plus was deployed in SIPD for a prospective validation from 1 January to 30 April 2025. Follow-up of AI-positive cases was completed by 31 July 2026. In total, EAGLE-Plus processed 17,446 patients during the study period (Extended Data Fig. 4 and Methods).

By the end of follow-up, 41 ECs and 6 HGIN lesions were confirmed among all enrolled individuals. EAGLE-Plus generated 90 positive predictions, corresponding to a positive rate of approximately 0.5%. Of these, 38 AI-positive cases (36 = ECs, 2 = HGIN) were confirmed as true positives, yielding a sensitivity of 87.8% (95% CI = 77.4–97.3%) for EC and 80.9% (95% CI = 69.2–91.3%) for esophageal malignancy, with a PPV of 42.2% (95% CI = 31.8–52.1%). Of the remaining 52 AI-positive cases, 28 had esophageal and peri-esophageal findings, 19 were considered negative or reported no esophageal symptoms on follow-up and 5 were lost to follow-up.

Of the 90 AI-positive individuals, 48 had already been managed by the initial SOC pathway (Extended Data Fig. 4). Among the remaining 42 individuals, the multidisciplinary team (MDT) referred 16 for endoscopy, corresponding to an AI-triggered referral rate of 0.09% (16/17,446). Of these referred individuals, three completed endoscopy, yielding one EC (T3N0M0) and two negative cases. The patient with EC subsequently underwent definitive interventional therapy (Fig. 4b and Supplementary Fig. 4a). Among the other 13 referred individuals, 9 underwent structured telephone follow-up and reported no esophageal symptoms and 4 were lost to follow-up. The remaining 26 were not referred because the findings were considered nonmalignant or negative after the review of available clinical records. Detailed information on each AI-positive prediction is provided in Extended Data Table 4, and representative examples of HGIN and EC cases are illustrated in Supplementary Fig. 4.

**Real-world retrospective validation in LDCT program.** To further evaluate EAGLE-Plus within established LDCT programs for asymptomatic populations, we conducted a retrospective validation from 1 July 2024 to 31 December 2024, comprising 10,959 consecutive participants aged 45 to 75 years. For EAGLE-Plus-predicted positive cases, MDT reviewed initial radiology reports and available clinical records at SIPD.

https://doi.org/10.1038/s41591-026-04656-4

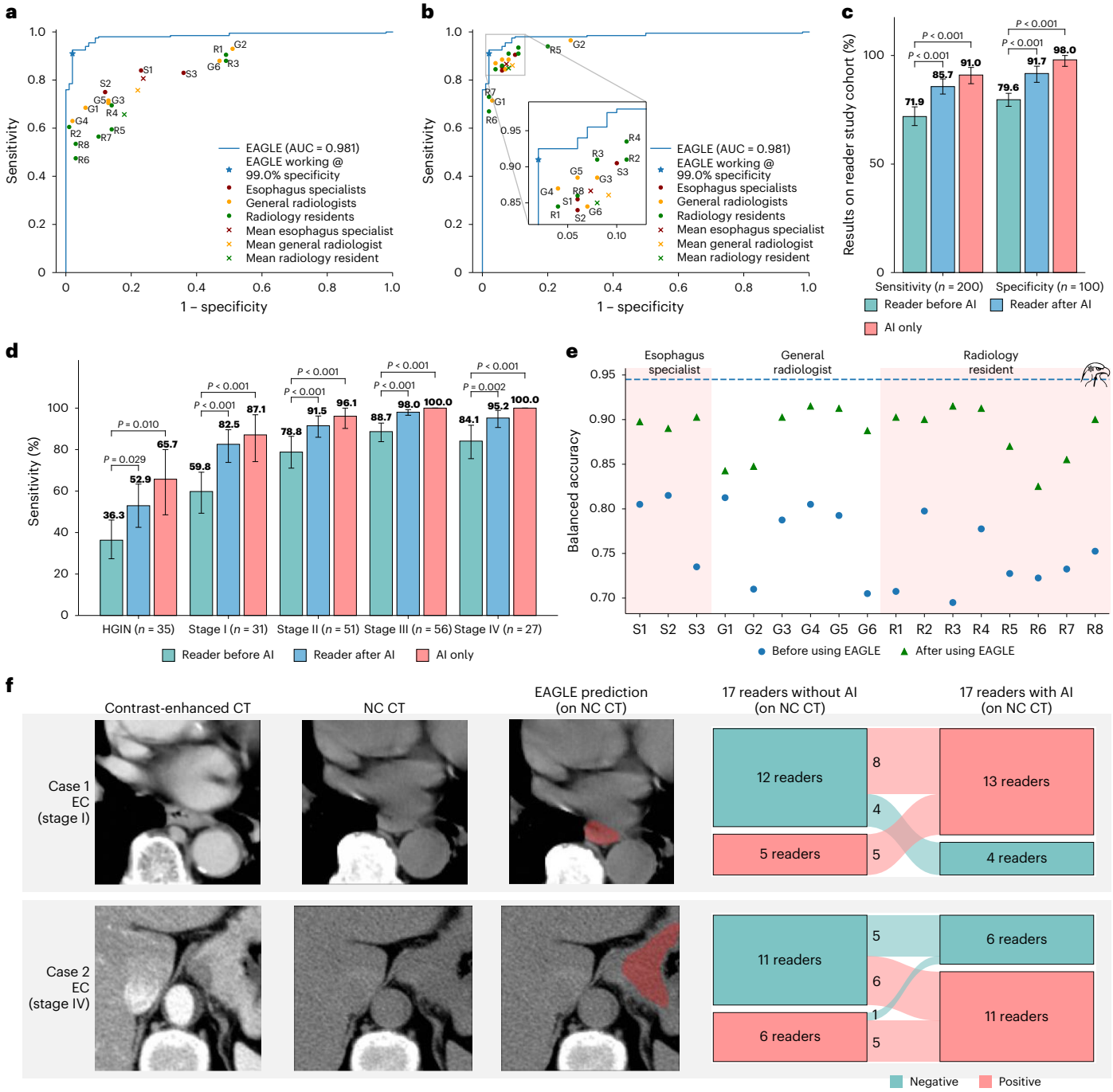



**Fig. 3 | Comparative analysis of EAGLE and radiologists. a**, The comparison between EAGLE and 17 readers with different levels of expertise on NC CT for esophageal malignancy detection. **b**, The performance in esophageal malignancy detection of the same set of 17 readers with the assistance of EAGLE on NC CT. **c**, The sensitivity, specificity and AUC of readers alone, EAGLE alone and readers with the assistance of EAGLE in the reader study cohort. **d**, The sensitivity of HGIN and EC in different TNM stages by readers alone, EAGLE alone and readers with the assistance of EAGLE. **e**, The balanced accuracy improvements in readers with different levels of expertise for EC detection. **f**, Examples of ECs initially missed by readers on NC CT, and identified with the assistance of EAGLE; contrast-enhanced CT scans were only used for reference purposes in this figure. In **c** and **d**, error bars indicate two-sided 95% CI computed from 1,000 bootstrapping iterations. *P* values computed through two-sided paired permutation tests; no adjustment was made for multiple comparisons.

EAGLE-Plus generated eight positive predictions from all participants, corresponding to a positive rate of 0.07%. One of these cases had not been reported as suspicious in the initial radiology report but was confirmed as EC 8 days later (Fig. 4c and Supplementary Fig. 5). Five patients did not undergo the MDT-recommended endoscopic examination. Three patients were diagnosed with esophageal disease—one with benign esophageal tumor, one with reflux esophagitis and another with esophagotracheal fistula, a serious disease that requires immediate care. In addition, SOC recalled two participants due to suspicious findings for further endoscopic examinations, both were predicted as negatives by EAGLE-Plus, and neither showed esophageal abnormalities on endoscopy. In total, EAGLE-Plus model achieved a PPV of 12.5% (95% CI = 7.1–25.0%) and a specificity of 99.94% (95% CI = 99.88–99.97%).

https://doi.org/10.1038/s41591-026-04656-4

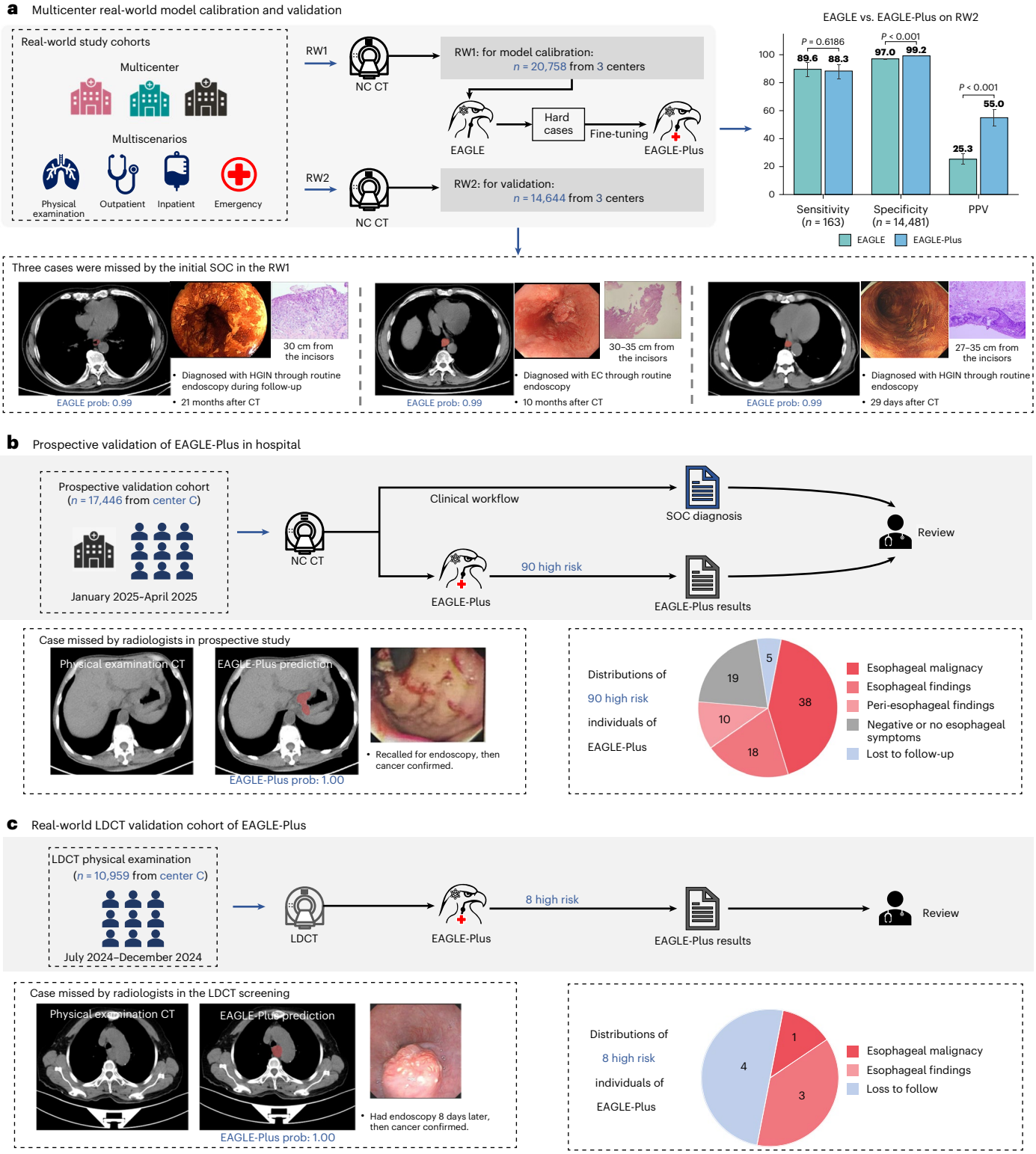



**Fig. 4 | Real-world validation for opportunistic screening. a**, Real-world model calibration and validation. Top-left, the study design stratified a multicenter, multiscenario dataset into two cohorts (RW1 and RW2). RW1 was used to reduce the FP rate, resulting in the calibrated EAGLE-Plus model. Top-right, EAGLE versus EAGLE-Plus on RW2. Bottom, three malignancies that were not detected by the initial SOC. **b**, Prospective validation of EAGLE-Plus. Top, pipeline of validation. Bottom-left, an EC initially missed by radiologists but detected by EAGLE-Plus. Bottom-right, categories of 90 EAGLE-Plus predictions. **c**, Real-world LDCT performance assessment. Top, pipeline of the real-world LDCT validation. Bottom-left, an EC undetected in initial LDCT interpretation but identified by EAGLE-Plus. Bottom-right, categories of eight EAGLE-Plus predictions. 'Loss to follow' refers to patients who could not be reached during follow-up. In the top-right bar chart of **a**, error bars indicate two-sided 95% CI computed from 1,000 bootstrapping iterations.

### EAGLE risk-stratified endoscopic screening

**External validation on paired NC CT–endoscopy data.** China has implemented nationwide EC screening programs by endoscopy in high-incidence areas. To evaluate EAGLE's performance in this high-risk population, we collected NC CT scans from two designated EC screening centers—Suining Central Hospital (SNCH) and Lishui Central Hospital (LCH). Given the requirement for high sensitivity in population-based screening, EAGLE's operating threshold was set to achieve 98% sensitivity during cross-validation to support its utility as a risk-stratification tool before the endoscopic examination.

This dual-center, external validation cohort ($n$ = 702) was established through retrospective medical record review of patients who had undergone both endoscopic examination and NC CT. Among 552 malignant cases, 68.6% were HGIN or stages I–II EC. EAGLE achieved an overall AUC of 0.932 (95% CI = 0.909–0.954), sensitivity of 90.2% (95% CI = 87.8–92.7%) and specificity of 84.7% (95% CI = 79.1–90.3%). Stage-specific sensitivities are shown in Fig. 5c—EAGLE detected EC with 92.2% (95% CI = 89.8–94.5%) sensitivity, HGIN with 65.0% (95% CI = 50.0–80.0%) sensitivity and stage I EC with 78.4% (95% CI = 71.9–84.3%) sensitivity. Sensitivity for stage IV cases was lower (89.5%) than for earlier stages, owing to two false-negative metastatic (M1) cases that were T1 lesions.

**Validation of prediagnostic detection ability.** To assess whether EAGLE could identify EC patients on CT scans obtained before clinical diagnosis, we retrospectively collected available NC CT scans for 28 EC patients at LCH. EAGLE classified 18 of 28 patients (64.3%) as positive based on their prediagnostic CT scans. Notably, EAGLE detected four patients from CT scans obtained ≥9 months before EC diagnosis, including two stage I, one stage II and one stage III cases (Fig. 5d).

**EAGLE as a risk-stratification tool before endoscopic screening.** The ability of EAGLE to detect precancerous malignant lesions suggests its potential to have a role in proactive prevention settings, as progression from HGIN to stage I EC is relatively slow (annual progression rate = 8–18%[32]), providing a meaningful window for earlier malignant lesion identification through repeated screening. To explore the potential role of EAGLE in an endoscopy-based screening program, we invited the patients from the program in SNCH to have an NC CT scan before the endoscopy. Between 7 July and 26 July 2023, 530 participants who agreed to the CT examination were prospectively enrolled (Chictr.org.cn identifier: ChiCTR2300074806). Among them, six had precancerous lesions (HGIN), and no invasive cancers were found in this cohort. To enable an exploratory analysis under a prevalence structure closer to that expected in high-risk populations undergoing endoscopy, that is, an HGIN-to-EC incidence ratio of 2:1 (ref. 2), we generated a hybrid dataset by adding three EC cases through resampling from the SNCH test cohort (Fig. 5e and Methods). This yielded a hybrid cohort of 533 patients (Extended Data Fig. 5), comprising nine esophageal malignant cases (six HGINs and three ECs) and 524 negative controls, corresponding to a detection rate of 1.7% (9/533). In this generated hybrid cohort, EAGLE achieved a sensitivity of 76.0% (95% CI = 73.8–78.2%) and a specificity of 76.3% (95% CI = 72.9–80.1%). The negative predictive value is 99.46% (95% CI = 99.42–99.51%), indicating a low missed-cancer risk in negative results in this setting.

When EAGLE was used as a risk-stratification tool to triage only EAGLE-identified high-risk patients for endoscopic examination (Fig. 5e), the simulations indicated that approximately 131 patients (130.8; 95% CI = 130.6–131.0) would be classified as positive. In the simulation, EAGLE detected 6.8 (95% CI = 6.6–7.0) true positives with four of six HGIN cases and 2.8 cancers (95% CI = 2.6–3.0), missed two HGIN cases. This EAGLE risk-stratification strategy increased the malignant lesion detection rate of endoscopy from 1.7% (9/533) to 5.2% (6.8/130.8; 95% CI = 5.1–5.4%; $P < 0.01$). The number of participants needed to be screened to detect one esophageal malignant patient would decrease from 59.2 (533/9) to 19.2 (130.8/6.8; 95% CI = 18.6–19.8). Consequently, 67.6% endoscopic examinations could be avoided, representing approximately a threefold improvement in endoscopy efficiency.

To further evaluate the impact of the EAGLE risk-stratified strategy on efficient use of resources and safety, we analyzed the distribution of esophageal lesions in both positive and negative prediction groups. Among the positive predictions without malignancy ($n$ = 124), nearly half ($n$ = 59, 47.6%) had esophageal abnormalities requiring dedicated follow-up endoscopy, including 14 LGIN (a nonmalignant precancerous lesion) and 45 precancerous esophageal diseases (Extended Data Fig. 5). These findings suggest that EAGLE could reduce unnecessary procedures while identifying clinically relevant lesions among its FP predictions. In the EAGLE-negative group ($n$ = 402), 99.5% participants ($n$ = 400) showed no evidence of esophageal malignancy, and the remaining two cases were HGIN (a malignant precancerous lesion), indicating that EAGLE maintains a high level of safety in its negative predictions (Extended Data Fig. 5). Precancerous lesions and diseases that require dedicated follow-up endoscopy were determined according to the Chinese national guideline (Supplementary Table 2)[33]. Representative examples were shown in Supplementary Fig. 8.

## Discussion

We present EAGLE, an imaging AI-based approach to detect esophageal malignancy from routine chest NC CT scans. Through comprehensive real-world validation, EAGLE showed robust performance and potential translation relevance across diverse clinical applications for EC screening. EAGLE's clinical values manifest in the following scenarios: (1) EAGLE demonstrated high sensitivity and exceptional specificity in hospital settings, while also establishing the capability of AI to detect precancerous lesions on NC CT scans. (2) Radiologists showed significantly enhanced diagnostic accuracy through EAGLE's interpretable outputs, establishing its role as a decision-support tool. (3) EAGLE could repurpose established lung-cancer screening programs for EC screening, expanding the significance of established large-scale LDCT examinations. (4) EAGLE demonstrated prescreening capabilities for population-based endoscopic programs by prioritizing high-risk individuals and optimizing endoscopy resource allocation.

EAGLE supports early detection of EC on NC CT and LDCT, even at precancerous stages[19,22]. The operating point of EAGLE can be adjusted by varying the threshold. At a 98.5% specificity, external validation across eight centers showed approximately 60% sensitivity for stage I EC and 52.5% for HGIN. At a matched specificity of 91.0%, corresponding to that of a noninvasive liquid biopsy[34], EAGLE's sensitivity for stage I EC increased to 73.4%. This capability is supported by a large training set containing 264 HGIN and 548 stage I cancers (Fig. 1a), each with voxel-level lesion masks that provided strong supervision and enabled the model to learn subtle signatures on NC CT (Extended Data Fig. 1). Notably, reliable voxel-level annotation of HGIN and early cancers stemmed from an annotation strategy that integrated information across the clinical workflow rather than relying solely on lesion visibility on CT—pathology reports from endoscopic surgery provided centimeter-level localization, which was cross-referenced with abnormalities identified by esophageal-specialized radiologists on contrast-enhanced CT through a standardized, expert-adjudicated two-annotator workflow (Methods). This framework may inform early-detection approaches for other cancers long considered impractical for NC CT, as integrating complementary clinical information across the diagnostic workflow can enable reliable annotation.

EAGLE enhances radiologists' diagnostic accuracy for esophageal malignancies. In clinical practice, radiologists encounter challenges in distinguishing between normal esophageal constrictions (for example, physiological narrowing) and early-stage EC. In the reader study, EAGLE outperformed human radiologists on early-stage esophageal malignancies on NC CT and, when used as assistance, significantly improved their sensitivity for HGIN and stage I EC. The reader study

https://doi.org/10.1038/s41591-026-04656-4

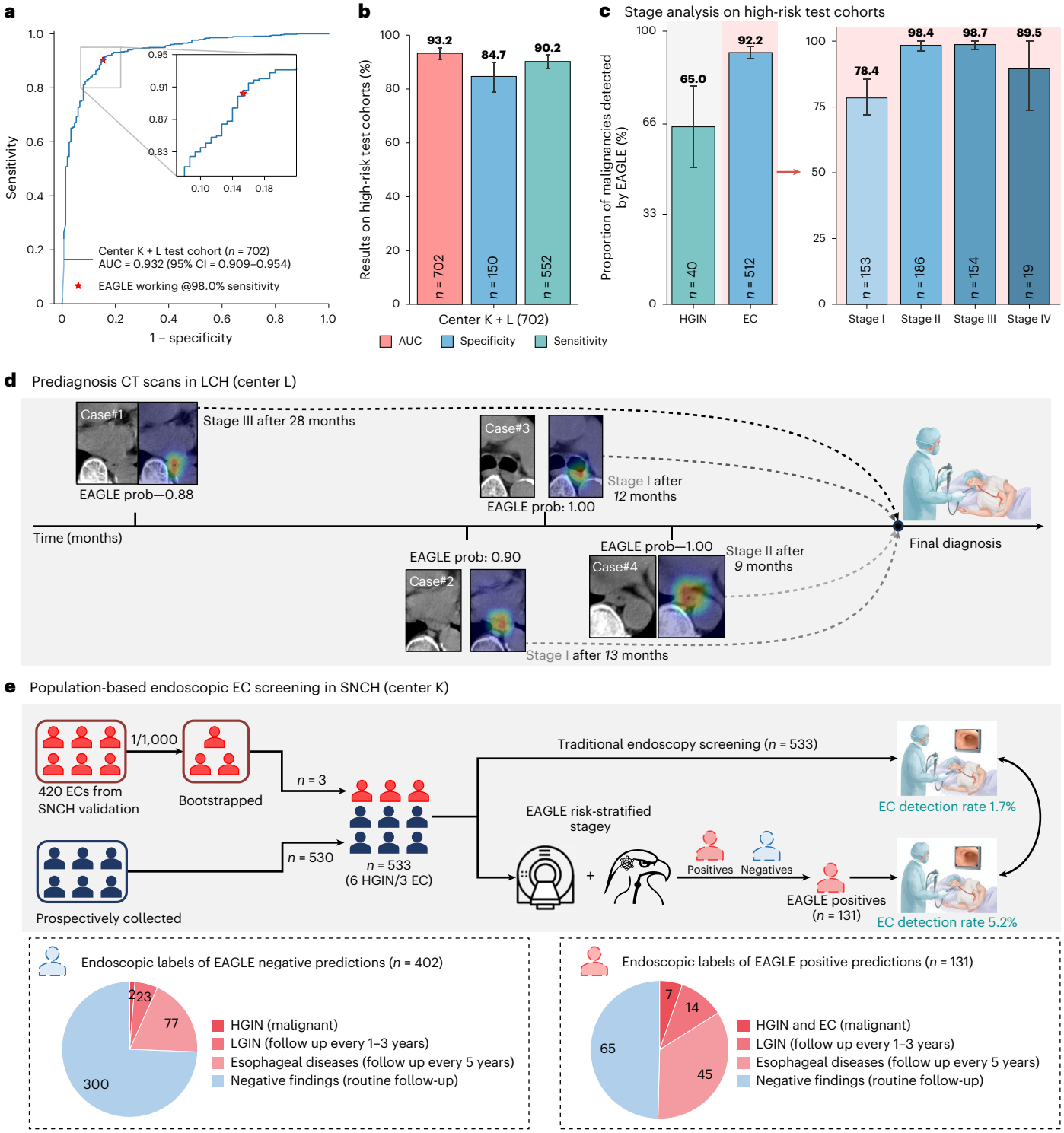



**Fig. 5 | Endoscopic EC screening. a–c**, Results on external test cohort in the endoscopic screening program ($n = 702$, centers K and L). **a**, ROC curves. **b**, The sensitivity, specificity and AUC. **c**, The proportion of esophageal malignancies detected by EAGLE in different TNM stages. **d**, Four EC cases from center L could be detected by EAGLE on their prediagnostic CT examinations at least 9 months earlier than their final diagnosis date by endoscopy. **e**, The validation of EAGLE risk-stratified strategy in a prospectively enrolled population-based EC endoscopic screening cohort from center K. Top, the pipeline of our endoscopic detection rate comparison. Bottom-left, the distribution of 402 negative predictions, where two HGINs are missed by EAGLE. Bottom-right, the distribution of 131 EAGLE-positive predictions, where seven are HGIN or EC. In **b** and **c**, error bars indicate two-sided 95% CI computed from 1,000 bootstrapping iterations. In **e**, the *P* value was calculated using a two-sided McNemar test.

also indicated that the AI model's standalone performance was not fully translated into human–AI collaboration. This may be because evaluation of early-stage esophageal malignancy on NC CT is not currently part of standard clinical practice, radiologists generally lack specific training for this task, and they may show bias when interpreting AI alerts for nonroutine indications. In addition, EAGLE's activation maps and lesion segmentation indicate the most suspicious areas of the esophagus (Extended Data Fig. 2), which would enhance radiologists'

prioritization of esophageal malignant lesions. Given the widespread use of chest NC CT for clinical indications and routine physical examination in China, EAGLE can be deployed in hospitals with limited radiology resources to bridge diagnostic gaps in EC detection, thereby increasing the rate of subsequent endoscopic confirmation and redefining the standard pathway for early EC intervention.

EAGLE supports integrating EC screening into established LDCT lung-cancer screening and general physical examination workflows, transforming these platforms into dual-cancer screening systems without compromising operational efficiency. Annual LDCT screening is recommended by the US Preventive Services Task Force for individuals at high risk of lung cancer (aged 50–80 years with ≥20 pack-year smoking history)[30,35]. Notably, smoking represents a shared risk factor for both ESCC and EAC[29], while LDCT scans for lung-cancer screening inherently include full visualization of the esophagus. This overlap in risk factors and anatomical coverage suggests that integrating EC screening into established LDCT lung-cancer screening programs potentially represents a pragmatic strategy for identifying esophageal malignancy. In our two-center cohort, the model with LDCT simulation strategy achieved sensitivity and specificity for detecting EC comparable to its performance on NC CT, including the detection of precancerous lesions, demonstrating robust generalization to LDCT imaging despite having been originally trained on NC CT scans. In subsequent large-scale, real-world LDCT screening of asymptomatic populations, the model maintained a low FP rate (<0.1%), thereby minimizing unnecessary follow-ups and patient anxiety—key requirements for practical, population-level screening[36].

EAGLE could serve as a pre-endoscopic risk-stratification tool combining noninvasive operation, rapid processing and operator-independent performance. EC is primarily diagnosed through endoscopy, which directly visualizes suspicious lesions. However, endoscopic screening shows low detection rates (~0.80%) even in high-risk populations[37]. Endoscopy's high cost, invasiveness and suboptimal patient adherence further constrain its effectiveness in population screening, underscoring the urgent need for risk-stratification tools to prioritize high-risk candidates for endoscopic validation. When implemented at two Chinese endoscopic screening centers, EAGLE demonstrated 65.0% sensitivity for HGIN and 78.4% sensitivity for stage I EC, with >98% sensitivity across stages II–III EC. The majority of missed stage I cases (>90%) were classified as Paris type 0-II (superficial and flat)[38]. This morphology is known to be the most challenging type during standard endoscopy. A simulation using prospectively collected data from one screening site suggested that integrating EAGLE as a pre-endoscopy triage tool could triple the detection rate of subsequent endoscopic procedures and achieve considerably lower time cost—EAGLE risk-stratification tool reduced the required time by 70.4% (from 355.33 to 105.10 h), lowered costs in seven of eight countries by 37.5–64.5% compared to universal endoscopy (Extended Data Fig. 6). Moreover, the cumulative early-detection rate of a screening strategy depends on the combined effects of test sensitivity, screening frequency, patient adherence and the natural history of disease progression. The high acceptability of CT screening combined with the indolent progression of esophageal precancerous lesions[8,32,39] creates an opportunity to minimize diagnostic gaps. By improving screening frequency (for example, annual intervals), this approach may detect early-stage premalignant transformations, particularly among high-risk cohorts where preventive measures greatly improve prognostic outcomes. A similar paradigm has been demonstrated in colorectal cancer screening using the fecal immunochemical test, which achieves meaningful population-level mortality reduction despite only moderate sensitivity for early-stage disease[40–44].

Our study has several limitations. First, the diversity of the training and validation populations requires further expansion. Although preliminary validation was conducted in two countries outside China, broader international validation remains essential, particularly in populations with a higher prevalence of distal EC and adenocarcinoma; in pilot experiments, adding EGJ/EAC cases to the training set improved EAC sensitivity from 86.6% to 91.2% ($P < 0.001$; Supplementary Fig. 9). The observed sex-based differences in sensitivity may also reflect an imbalance in the training data due to the markedly higher incidence of EC in men[4], highlighting the need for more well-annotated data from female patients. Second, the value of EAGLE as a pre-endoscopy screening test requires larger and more mature prospective data. The present evidence rests on retrospective paired CT–endoscopy cohorts and a simulation from a single prospectively enrolled screening site, in which positive cases were limited, as also in the real-world LDCT cohort. Third, the follow-up durations in both hospital prospective validation and the real-world LDCT validation were less than 2 years, and compliance with MDT-recommended endoscopic examinations was suboptimal, consistent with the low adherence reported in China[13], so the model's true sensitivity may not be fully reflected. Finally, although a 3-month washout period and randomized ordering of 300 CT scans were used in the reader study, the absence of a full cross-over design means that potential recall bias cannot be completely excluded.

In conclusion, as an accurate, noninvasive and rapid tool, EAGLE demonstrates the feasibility of identifying esophageal malignant and precancerous lesions from NC CT and LDCT. EAGLE has the potential to extend the clinical applicability of AI-based CT analysis beyond opportunistic screening settings to risk stratification in population-based prevention.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41591-026-04656-4.

## References


1. Bray, F. et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **74**, 229–263 (2024).
2. Lao-Sirieix, P. & Fitzgerald, R. C. Screening for oesophageal cancer. *Nat. Rev. Clin. Oncol.* **9**, 278–287 (2012).
3. Yang, H., Wang, F., Hallemeier, C. L., Lerut, T. & Fu, J. Oesophageal cancer. *Lancet* **404**, 1991–2005 (2024).
4. He, S. et al. Cancer profiles in China and comparisons with the USA: a comprehensive analysis in the incidence, mortality, survival, staging, and attribution to risk factors. *Sci. China Life Sci.* **67**, 122–131 (2024).
5. An, L. et al. The survival of esophageal cancer by subtype in China with comparison to the United States. *Int. J. Cancer* **152**, 151–161 (2023).
6. Wei, W.-Q. et al. Long-term follow-up of a community assignment, one-time endoscopic screening study of esophageal cancer in China. *J. Clin. Oncol.* **33**, 1951–1957 (2015).
7. Chen, R. et al. Effectiveness of one-time endoscopic screening programme in prevention of upper gastrointestinal cancer in China: a multicentre population-based cohort study. *Gut* **70**, 251–260 (2021).
8. Liu, M. et al. Effectiveness of endoscopic screening on esophageal cancer incidence and mortality: a 9-year report of the endoscopic screening for esophageal cancer in China (ESECC) randomized trial. *J. Clin. Oncol.* **42**, 1655–1664 (2024).
9. Klein, E. et al. Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. *Ann. Oncol.* **32**, 1167–1177 (2021).
10. Liu, M. C. et al. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Ann. Oncol.* **31**, 745–759 (2020).

11. Gao, Q. et al. Unintrusive multi-cancer detection by circulating cell-free DNA methylation sequencing (THUNDER): development and independent validation studies. *Ann. Oncol.* **34**, 486–495 (2023).
12. Wang, Y. et al. Highly sensitive detection platform-based diagnosis of oesophageal squamous cell carcinoma in China: a multicentre, case-control, diagnostic study. *Lancet Digit. Health* **6**, 705–717 (2024).
13. Bao, H. et al. Early detection of multiple cancer types using multidimensional cell-free DNA fragmentomics. *Nat. Med.* **31**, 2737–2745 (2025).
14. Gao, Y. et al. Machine learning-based automated sponge cytology for screening of oesophageal squamous cell carcinoma and adenocarcinoma of the oesophagogastric junction: a nationwide, multicohort, prospective study. *Lancet Gastroenterol. Hepatol.* **8**, 432–445 (2023).
15. Falk, G. W. et al. Surveillance of patients with Barrett's esophagus for dysplasia and cancer with balloon cytology. *Gastroenterology* **112**, 1787–1797 (1997).
16. Aberle, D. R. et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N. Engl. J. Med.* **365**, 395–409 (2011).
17. Gros, L. et al. GI cancer mortality in participants in low dose CT screening for lung cancer with a focus on pancreatic cancer. *Sci. Rep.* **14**, 29851 (2024).
18. Lotter, W. et al. Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach. *Nat. Med.* **27**, 244–249 (2021).
19. Cao, K. et al. Large-scale pancreatic cancer detection via non-contrast CT and deep learning. *Nat. Med.* **29**, 3033–3043 (2023).
20. Ying, H. et al. A multicenter clinical AI system study for detection and diagnosis of focal liver lesions. *Nat. Commun.* **15**, 1131 (2024).
21. Wang, C. et al. Data-driven risk stratification and precision management of pulmonary nodules detected on chest computed tomography. *Nat. Med.* **30**, 3184–3195 (2024).
22. Hu, C. et al. AI-based large-scale screening of gastric cancer from noncontrast CT imaging. *Nat. Med.* **31**, 3011–3019 (2025).
23. Chen, X. et al. Colorectal cancer detection using non-contrast CT and deep learning: a multicenter and international cohort study. *Ann. Oncol.* **37**, 1144–1156 (2026).
24. Ardila, D. et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat. Med.* **25**, 954–961 (2019).
25. McKinney, S. M. et al. International evaluation of an AI system for breast cancer screening. *Nature* **577**, 89–94 (2020).
26. Chang, Y.-W. et al. Artificial intelligence for breast cancer screening in mammography (AI-STREAM): preliminary analysis of a prospective multicenter cohort study. *Nat. Commun.* **16**, 2248 (2025).
27. Eisemann, N. et al. Nationwide real-world implementation of AI for cancer detection in population-based mammography screening. *Nat. Med.* **31**, 917–924 (2025).
28. Sodickson, A. et al. Recurrent CT, cumulative radiation exposure, and associated radiation-induced cancer risks from CT of adults. *Radiology* **251**, 175–184 (2009).
29. Morgan, E. et al. The global landscape of esophageal squamous cell carcinoma and esophageal adenocarcinoma incidence and mortality in 2020 and projections to 2040: new estimates from GLOBOCAN 2020. *Gastroenterology* **163**, 649–658 (2022).
30. Dai, X. et al. Health effects associated with smoking: a Burden of Proof study. *Nat. Med.* **28**, 2045–2055 (2022).
31. Yu, L., Shiung, M., Jondal, D. & McCollough, C. H. Development and validation of a practical lower-dose-simulation tool for optimizing computed tomography scan protocols. *J. Comput. Assist. Tomogr.* **36**, 477–487 (2012).
32. Xia, R. et al. Estimated cost-effectiveness of endoscopic screening for upper gastrointestinal tract cancer in high-risk areas in China. *JAMA Netw. Open* **4**, 2121403 (2021).
33. National Health Commission of the People's Republic of China. National guidelines for diagnosis and treatment of esophageal carcinoma 2022 in China (English version). *Chin. J. Cancer Res.* **34**, 309–334 (2022).
34. Qin, Y. et al. Discovery, validation, and application of novel methylated DNA markers for detection of esophageal cancer in plasma. *Clin. Cancer Res.* **25**, 7396–7404 (2019).
35. Krist, A. H. et al. Screening for lung cancer: US Preventive Services Task Force recommendation statement. *JAMA* **325**, 962–970 (2021).
36. Cohen, J. D. et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* **359**, 926–930 (2018).
37. Li, H. et al. Profiles and findings of population-based esophageal cancer screening with endoscopy in China: systematic review and meta-analysis. *JMIR Public Health Surveill.* **9**, e45360 (2023).
38. Endoscopic Classification Review Group. Update on the Paris classification of superficial neoplastic lesions in the digestive tract. *Endoscopy* **37**, 570–578 (2005).
39. Xia, R. et al. Cost-effectiveness of risk-stratified endoscopic screening for esophageal cancer in high-risk areas of China: a modeling study. *Gastrointest. Endosc.* **95**, 225–235 (2022).
40. Quintero, E. et al. Colonoscopy versus fecal immunochemical testing in colorectal-cancer screening. *N. Engl. J. Med.* **366**, 697–706 (2012).
41. Levin, T. R. et al. Effects of organized colorectal cancer screening on cancer incidence and mortality in a large community-based population. *Gastroenterology* **155**, 1383–1391 (2018).
42. Shaukat, A. & Levin, T. R. Current and future colorectal cancer screening strategies. *Nat. Rev. Gastroenterol. Hepatol.* **19**, 521–531 (2022).
43. Bretthauer, M. & Kalager, M. First head-to-head trial of colonoscopy versus faecal testing for colorectal cancer screening. *Lancet* **405**, 1204–1206 (2025).
44. Westerberg, M. et al. Colonoscopy and fecal immunochemical testing versus usual care in diagnostic colorectal cancer screening: the SCREESCO randomized controlled trial. *Nat. Med.* **32**, 1278–1285 (2026).


**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2026

**Jian Zhou[1,27], Guangyu Guo[2,3,27], Jiawen Yao[2,3,27], Qinji Yu[2,3,27], Chunpeng Zheng[4,27], Xiaochen Feng[5,27], Xianghua Ye[6,27], Yongjin Zhou[7,27], Pei Yang[8,27], Lukas Lambert[9,27], Shaojun Zheng[10], Qiong Li[1], Yingda Xia[3,11], Wenchao Guo[2,3], Yi Chen[2,3], Mengdi Zhang[1], Yu Jiang[1], Dandan Zheng[6], Jia Ge[6], Haomiao Qing[12], Wusong Liu[13], Peng Zhou[12], Mei Lan[14], Lei Wu[14], Yong Li[15], Haifeng Wang[16], Yong Zhou[17], Junqiang Chen[18], Liyan Xu[19], Chenying Lu[7], Lin Chen[20], Junwei Han[21], Chengwei Shao[5], Wenqiang Wei[22], Kai Zhang[23], Jianfeng Zhang[2,3], Dingwen Zhang[24] ✉, Jiansong Ji[7] ✉, Na Li[25] ✉, Chuanmiao Xie[1] ✉, Kai Cao[5] ✉, Yuqian Zhao[26] ✉, Ling Zhang[3,11] ✉ & Qifeng Wang[14] ✉**

[1]Department of Radiology, State Key Laboratory of Oncology in South China, Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center, Guangzhou, China. [2]DAMO Academy, Alibaba Group, Hangzhou, China. [3]Hupan Lab, Hangzhou, China. [4]Department of Surgical Oncology, Shantou Central Hospital, Shantou, China. [5]Department of Radiology, Shanghai Institution of Pancreatic Disease, Shanghai, China. [6]Department of Radiation Oncology, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China. [7]Zhejiang Key Laboratory of Imaging and Interventional Medicine, Department of Radiology, The Fifth Affiliated Hospital of Wenzhou Medical University, Lishui Central Hospital, Lishui, China. [8]Department of Radiotherapy, Cancer Hospital Affiliated to Xiangya Medical College, Central South University, Hunan Cancer Hospital, Changsha, China. [9]Department of Radiology, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic. [10]Department of radio-oncology, People's Hospital of Chenghai Shantou, Shantou, China. [11]DAMO Academy, Alibaba Group, Washington, DC, USA. [12]Department of Radiology, Sichuan Clinical Research Center for Cancer, Sichuan Cancer Hospital & Institute, Sichuan Cancer Center, Chengdu, China. [13]Department of Endoscopy Center, Sichuan Clinical Research Center for Cancer, Sichuan Cancer Hospital & Institute, Sichuan Cancer Center, Chengdu, China. [14]Department of Radiation Oncology, Precision Radiation in Oncology Key Laboratory of Sichuan Province, Sichuan Clinical Research Center for Cancer, Sichuan Cancer Hospital & Institute, Sichuan Cancer Center, Chengdu, China. [15]Department of Radiological Imaging, Suining Central Hospital, Suining, China. [16]Department of Radiotherapy, The Third Affiliated Teaching Hospital of Xinjiang Medical University (Affiliated Cancer Hospital), Urumqi, China. [17]Medical imaging center, The Third Affiliated Teaching Hospital of Xinjiang Medical University (Affiliated Cancer Hospital), Urumqi, China. [18]Department of Radiation Oncology, The Teaching Hospital of Fujian Medical University, Fujian Provincial Cancer Hospital, Fuzhou, China. [19]Guangdong Provincial Key Laboratory of Infectious Diseases and Molecular Immunopathology, Institute of Oncologic Pathology, Cancer Research Center, Shantou University Medical College, Shantou, China. [20]International Health Management Center, Shanghai Institution of Pancreatic Disease, Shanghai, China. [21]School of Artificial Intelligence, Chongqing University of Posts and Telecommunications, Chongqing, China. [22]National Central Cancer Registry, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. [23]Department of Cancer Prevention, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. [24]Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei, China. [25]Department of oncology, Suining Central Hospital, Suining, China. [26]Sichuan Provincial Engineering Research Center of Tumor Organoids and Clinical Transformation, Sichuan Clinical Research Center for Cancer, Sichuan Cancer Hospital & Institute, Sichuan Cancer Center, Chengdu, China. [27]These authors contributed equally: Jian Zhou, Guangyu Guo, Jiawen Yao, Qinji Yu, Chunpeng Zheng, Xiaochen Feng, Xianghua Ye, Yongjin Zhou, Pei Yang, Lukas Lambert. ✉e-mail: zhangdingwen2006yyy@gmail.com; jijiansong@zju.edu.cn; 1500921245@qq.com; xchuanm@sysucc.org.cn; mdkaicao163@163.com; zhaoyuqian1211@scszlyy.org.cn; ling.z@alibaba-inc.com; wangqifeng@scszlyy.org.cn

## Methods

### Ethics approval

For the retrospective study, ethical approval was obtained individually from the institutional review board of SYSUCC, SCCH, SIPD, First Affiliated Hospital of Zhejiang University, Hunan Cancer Hospital (HCH), Shantou Central Hospital, Xinjiang Medical University Affiliated Cancer Hospital (XMUACH), Fujian Cancer Hospital, General University Hospital in Prague, GenesisCare, SNCH and LCH. The study was registered at https://www.chictr.org.cn/ (Chictr.org.cn identifier: ChiCTR2300074806).

### Dataset

This study aims to classify patients into positive and negative categories, where the 'positive' category includes patients with esophageal malignancy, including HGIN and EC. Data were used for model training and for validation in three clinical scenarios—opportunistic screening in hospitals, opportunistic screening in established LDCT programs and the population-based endoscopic screening program. For the retrospective test cohorts, patient labels were defined as follows: cases of esophageal malignancy were confirmed through surgical or biopsy histopathology, positive patients in all cohorts were staged according to the eighth edition of the American Joint Committee on Cancer pathological staging system. Negative controls were defined as individuals free of malignant esophageal lesions by at least 2 years of clinical follow-up or negative endoscopic results within 1 year. It remains possible that a minor fraction of indolent lesions remained undetected during the follow-up period, thereby introducing potential label noise. Detailed information on the dataset is provided in Supplementary Methods—Training, internal and external validation. The standard of truth for each cohort is provided in Supplementary Table 3.

### Training data

**Internal training cohort.** The internal training cohort contains 6,813 NC CT scans, including 3,744 esophageal malignant patients and 3,069 negative controls. Among these, 2,716 esophageal malignant cases and 1,539 negative controls were recruited from SYSUCC (center A), while 1,028 esophageal malignancy cases and 1,530 negative controls were sourced from SCCH (center B).

**Annotation protocol.** Early-stage esophageal malignant lesions are often extremely small, and the esophagus itself is a long, narrow tubular organ that frequently undergoes physiological collapse and is subject to motion from the heart and great vessels. These factors make subtle lesions difficult to distinguish from normal tissues on CT, even on contrast-enhanced scans, posing substantial challenges for accurate lesion identification and annotation. To address these challenges, we established a rigorous annotation protocol in which strong reference information was provided to two annotators and one expert reviewer to ensure high-quality data for EAGLE model development (Extended Data Fig. 1). For each patient, paired contrast-enhanced CT and NC CT scans with corresponding surgical pathology reports were available. We selected the most recent CT scan before the first surgery or endoscopy examination. This examination is typically a contrast-enhanced CT, and in Chinese cancer hospitals, the same session also includes an accompanying NC series. Pathology reports were obtained from that surgical or endoscopy episode. Lesion annotations were performed on contrast-enhanced CT volumes due to their superior visualization of lesion boundaries. Two radiologists with ≥3 years' experience in esophageal imaging independently annotated lesions using standardized platforms. Interannotator agreement was quantified through DSC—masks with DSC >0.6 were combined as gold-standard annotations, while lower agreement cases underwent senior radiologist review (>8 years' expertise). Reviewers integrated imaging-pathology data to refine annotations, and excluded cases from the training cohort if they could not confidently delineate the lesion. Final contrast-enhanced CT annotations were then precisely transferred to NC CT through image registration[45], applying transformation matrices to achieve voxel-wise alignment. This systematic workflow ensured accurate cross-modal ground truth data generation for model training and validation.

### Validation cohorts for opportunistic screening in hospital

**Internal test cohort.** The internal test cohort was used to evaluate the model's performance within the same institutional framework and comprised 2,500 patients, including 1,281 patients with esophageal malignant lesions, and 1,219 negative controls. Among these, 767 esophageal malignant patients and 715 negative controls were sourced from the SYSUCC; the remaining 223 esophageal malignant patients and 504 negative controls originated from the SCCH. Notably, 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 (Extended Data Fig. 1). The inclusion of these cases substantially increased the diagnostic complexity of the test dataset, allowing for a more rigorous validation of the model's classification performance.

**External multicenter test cohort.** The external test cohort was used to assess the cross-institutional generalizability and was established from six centers in China, one center in the Czech Republic (General University Hospital in Prague, center I), and one center in Australia (GenesisCare, center J). Among centers in China, four centers were located in the east (SIPD, center C; First Affiliated Hospital of Zhejiang University, center D; Shantou Central Hospital, center F; Fujian Cancer Hospital, center H), one center in the central (HCH, center E) and one center in the northwest (XMUACH, center G). The inclusion criteria were as follows: NC CT scans must fully cover the chest region, encompassing the esophagus and EGJ regions. Positive cases were confirmed through surgical or biopsy histopathology, while negative controls were confirmed based on at least 2 years of clinical follow-up. In total, the external test cohort comprised 2,812 patients with esophageal malignancy and 8,654 negative controls.

**Model calibration and real-world retrospective validation cohort.** The real-world study collected NC CT scans with continuous time series data from SYSUCC, SCCH and SIPD, consisting of 35,402 patients. For the real-world cohorts, cases of esophageal malignancy were confirmed through surgical or biopsy histopathology, while negative controls were determined by clinical follow-up. The dataset from each center was divided into two distinct continuous series—one for model calibration (RW1) and the other for model validation (RW2). The overall RW1 comprised 20,758 patients for model calibration, and the refined model was subsequently evaluated on RW2, comprising 14,644 patients. The standard of truth was defined by two time points for each patient—the initial SOC, which was the diagnosis at the first visit when the NC CT was acquired, and the follow-up SOC, which was the diagnosis during subsequent follow-up before model evaluation.

**Prospective validation cohort.** A total of 17,446 patients were enrolled from January 2025 to April 2025 at SIPD. This study aimed to validate the clinical reliability of the calibrated EAGLE-Plus model for opportunistic screening in real-world clinical settings, particularly uncontrolled, high-throughput clinical environments. Esophageal malignancy was confirmed by biopsy or surgical pathology. Follow-up of positive predictions was completed on 31 July 2026.

### Validation cohorts for opportunistic screening in established LDCT programs

Two LDCT validation cohorts were established to evaluate the model's accuracy and clinical utility in LDCT-based physical examination and lung-cancer screening programs.

**LDCT test cohort.** This cohort enrolled 147 patients with esophageal malignancy and 1,460 negative controls recruited from SCCH and SIPD, which were used to assess the generalizability of the EAGLE on LDCT. Specifically, from SCCH, we retrospectively collected 1,464 individuals who had undergone both chest LDCT and endoscopy as part of routine physical examinations, of whom 4 were confirmed to have esophageal malignancy by biopsy histopathology. To enhance the estimation of LDCT sensitivity, we additionally retrieved 85 patients with esophageal malignancy confirmed at SIPD, all of whom had undergone chest LDCT within 1 month before confirmation.

**Real-world retrospective LDCT validation cohort.** This cohort enrolled 10,959 consecutive individuals between July and December 2024 at SIPD aged 45–75 years. The original purpose of those CT scans was for a physical examination. This validation was designed to derive specificity thresholds under real-world healthcare conditions and estimate the specificity of esophageal malignancy in the LDCT-based program. The standard of truth for negative controls was established by first identifying cases with radiology reports showing no esophageal abnormalities, then including only those with follow-up records and no clinical diagnosis of EC.

### Validation cohorts for population-based endoscopic screening

**External test cohort for endoscopic screening scenario.** This cohort included 552 patients with esophageal malignancy and 150 negative controls, recruited from SNCH (center K) and LCH (center L). All patients underwent both NC CT scans and endoscopic examinations. Positive cases were confirmed through surgical or biopsy histopathology; negative controls were confirmed through endoscopic results within 1 year. The cohort was designed to evaluate EAGLE's cross-institutional generalizability in high-risk populations and ability to differentiate esophageal malignancy from benign esophageal conditions.

**Prospective enrolled cohort from endoscopic screening program.** This cohort was completed and registered at http://www.chictr.org.cn (Chictr.org.cn identifier: ChiCTR2300074806), and prospectively enrolled 530 patients between 7 and 26 July 2023 at SNCH. All patients were initially referred for endoscopic examination and provided written informed consent to undergo an additional NC CT scan (Supplementary Fig. 7). The cohort was specifically established to evaluate EAGLE's potential clinical utility in real-world endoscopic screening workflows for high-risk individuals, focusing on its role as a risk-stratification tool for endoscopy.

### Model: EAGLE

EAGLE is a two-stage framework (Extended Data Fig. 2a). The first stage focuses on esophageal localization by using a segmentation UNet. The second stage combines classification and segmentation using a joint UNet-based model, which takes the cropped region of interest as input to simultaneously predict esophageal and malignant lesion segmentation masks, diagnostic outcomes (positive or negative) and class activation mapping heatmaps[46] that highlight regions driving diagnostic decisions.

**Esophageal localization.** Stage 1 focuses on identifying the esophageal region. Given that esophageal lesions typically occupy a small volume in CT scans, isolating the esophagus expedites lesion detection and eliminates nontarget anatomical structures to enhance region-specific analysis. Here an nnU-Net V2 framework[47] was trained to segment the entire esophagus, including both healthy tissue and pathological anomalies—from NC CT inputs. Supervised learning uses voxel-level annotations of both esophageal boundaries and lesions. A 3D low-resolution architecture was adopted for computational efficiency, leveraging a downsampled UNet variant and a ResEnc backbone within the nnU-Net V2 framework[48].

**Malignant lesion detection.** Stage 2 aims to identify the malignant lesion. Given a 3D region-of-interest CT, we feed it into a 3D UNet-like backbone and obtain multiscale feature maps from the image decoder. Then, we standardize the spatial dimensions of all feature maps from the decoders to a uniform spatial size using trilinear interpolation. Subsequently, we concatenate these maps by aligning them along their depth dimension, forming an aggregated feature representation that serves as the input for the subsequent cancer screening task. The segmentation head takes the last feature map as input and outputs the malignant lesion segmentation result. The cancer screening head takes the aggregated feature map as input and outputs the probability of the patient having cancer. It consists of one 3D convolution layer with kernel size $3 \times 3 \times 3$, one global average pooling layer and one fully connected layer, as illustrated in Extended Data Fig. 2b. The overall loss of stage 2 is:

$$\mathcal{L}_{\mathrm{EAGLE}} = \mathcal{L}_{\mathrm{seg}} + \alpha \mathcal{L}_{\mathrm{cls}} \quad (1)$$

where $\mathcal{L}_{\mathrm{seg}}$ is the loss for the 3D UNet segmentation network; $\mathcal{L}_{\mathrm{cls}}$ is cross-entropy loss for the cancer screening task and $\alpha$ is its weight.

**Interpretability.** Our model's interpretability is achieved through dual mechanisms (Extended Data Fig. 2c). First, stage 2 of EAGLE generates precise segmentation masks delineating detected malignant lesions. Second, we implemented class activation mapping[46] to visualize diagnostic heatmaps derived from the classification head's convolutional feature maps. These heatmaps highlight anatomical regions that influenced the model's diagnostic decisions, providing spatial justification for classification outcomes.

**Operation point and threshold selection.** Because the model outputs continuous risk scores, the operating point can be adjusted by varying the decision threshold. For both EAGLE and EAGLE-Plus, thresholds were determined during fivefold cross-validation and were fixed after training. In each fold, the threshold was first selected on the corresponding validation split according to the predefined operating criterion, and the final threshold was fixed as the mean value across the five folds after training. In the opportunistic screening setting, the threshold for each fold was selected to achieve a specificity of 99% on its respective validation split. The final fixed thresholds were 0.93 for EAGLE and 0.88 for EAGLE-Plus. Conversely, in the population-based screening setting, EAGLE was evaluated as a risk-stratification tool for endoscopy EC screening procedures like narrow band imaging endoscopy or Lugol's chromoendoscopy[49,50], and thresholds were calibrated to achieve a sensitivity of 98% for each fold to prioritize sensitivity and minimize the risk of missed cancers. Notably, EAGLE's operating point can be flexibly adjusted according to application requirements; specifically, a lower threshold increases sensitivity while decreasing specificity.

**Performance comparison.** As illustrated in Extended Data Table 3, we first compared five state-of-the-art backbone architectures integrated into EAGLE's implementation leveraging the nnU-Net framework[47], including convolutional neural network-based backbone ResNet encoder (ResEnc[48]), two Transformer-based backbones Swin-UNETR[51] and Mask2Former[52], and two Mamba-based architectures[53]. In the internal validation cohort, ResEnc demonstrated superior performance. In the eight-center external validation cohort, although ResEnc's AUC was marginally lower (0.976 versus 0.979), it achieved the highest sensitivity (89.5%) and specificity (98.5%), outperforming other models. Given its simplicity (pure convolutional neural network design) and robust generalization across multicenter data, ResEnc[48] was selected as the default backbone for EAGLE. We further compare EAGLE (ResEnc) with another widely adopted framework, nnDetection[54], which demonstrates markedly inferior performance with an AUC of 0.954

in the eight-center external validation cohort. Detailed descriptions of the compared methods are provided in Supplementary Methods—Comparison to other methods.

**LDCT simulation strategy.** To enhance the model's performance on LDCT scans, we developed EAGLE with LDCT simulation. A low-dose simulation tool[31] was used to generate synthetic LDCT data from the NC CT scans in the training set, which were then used for model training. Validation on real-world LDCT scans demonstrates that this strategy consistently improves sensitivity while maintaining the same level of specificity.

### Reader study

This study establishes a multicenter ($n = 4$) reader study cohort to conduct blinded comparisons between EAGLE predictions and interpretations by experienced radiologists, thereby validating the potential clinical integration of AI-driven workflows. This reader study comprised 300 patients collected from four centers, including two internal centers (SYSUCC and SCCH) and two external centers (HCH and XMUACH), including 200 patients diagnosed with esophageal malignancy and 100 negative controls. NC CT scans from additional external centers were not included in this analysis because stage information or imaging data were unavailable from those institutions until the initiation of the first-round reader study.

A total of 17 readers from three institutions participated in this study, comprising 3 subspecialized esophageal radiologists, 6 general radiologists and 8 radiology residents. The cohort demonstrated a mean clinical experience of 7.6 years (range = 3–20) in radiology practice, with radiologists having interpreted an average of 1,141 esophageal CT scans (range = 100–1,000) before the year preceding the study (Supplementary Table 1).

This reader study comprised two sequential sessions designed to evaluate AI-assisted diagnostic performance. The initial phase established baseline interpretation accuracy without AI support, followed by an AI-assisted evaluation phase after at least a 3-month washout period to mitigate learning effects. During both sessions, readers independently interpreted cases presented in randomized order through our customized online platform (Supplementary Fig. 2), blinded to clinical data and required to classify each case as esophageal malignant or not.

### Opportunistic screening: real-world retrospective and prospective validations

Clinical applications of opportunistic screening impose rigorous requirements on the model's FP rate. To address this, we calibrated the model using large-scale real-world retrospective cohorts and subsequently assessed its performance changes in independent real-world cohorts, yielding the optimized EAGLE-Plus framework. The overall real-world model calibration and validation collected 35,402 NC CT scans from SYSUCC, SCCH and SIPD. Data were collected from four scenarios, that is, physical examination, emergency, outpatient and inpatient. Each center contributed two distinct continuous data series.

**Iterative training of EAGLE-Plus.** Initially, the EAGLE model was evaluated using the first data series of each center (RW1, $n = 20{,}758$). Subsequently, the model was calibrated using hard cases from RW1 and four internal and external test cohorts to fine-tune the EAGLE model. Finally, these additional training samples were annotated by the same expert radiologist to ensure consistency, and were merged with the original training data for further fine-tuning of EAGLE, with an additional 200 epochs using the same training pipeline. The upgraded EAGLE-Plus model was evaluated on the second real-world continuous data series from each center (RW2, $n = 14{,}644$). Performance on RW2 reflects its effectiveness in uncontrolled, real-world opportunistic screening settings, particularly its stable sensitivity and low FP rate.

Hard cases were selected as follows. First, all FP cases identified by the MDT were categorized as hard negatives. Second, hard positives were defined by the following two components: (1) all false-negative cases; and (2) borderline positives (that is, true-positive cases with scores just above the threshold). The latter were included to compensate for the lower frequency of false negatives and to maintain an overall ratio of hard negatives to hard positives at approximately 1:1. Detailed results are provided in Supplementary Methods—Details of real-world model calibration and validation.

**Prospective validation in hospital.** We conducted a prospective validation to evaluate the real-world clinical implementation of the EAGLE-Plus model in clinical settings. The study enrolled patients from 1 January 2025 to 30 April 2025, with a final data analysis conducted after a follow-up period ending on 31 July 2026. EAGLE-Plus was seamlessly integrated into the DAMO MED system, where it prospectively analyzed routine NC CT scans and flagged high-risk individuals for further review.

The primary endpoint was PPV among AI-positive cases. Secondary endpoints were sensitivity for HGIN and EC and the AI-triggered referral rate. PPV was defined as the proportion of confirmed true-positive cases among all AI-positive predictions. Sensitivity was defined as the proportion of confirmed HGIN and EC cases identified during the study and follow-up period that had been classified as AI-positive. The AI-triggered referral rate was defined as the proportion of screened individuals referred for further clinical evaluation on the basis of AI findings after MDT review. Study oversight was provided by the principal investigator and the MDT, which included radiologists, endoscopists and thoracic surgeons. The SOC pathway consisted of routine double reading by radiologists blinded to AI outputs and follow-up outcomes. AI-positive cases were independently reviewed by the MDT, which intervened only when suspicious findings had not been captured by the initial SOC pathway. No feedback from the AI system or MDT review was provided to SOC radiologists during image interpretation. When further clinical action was considered necessary, the MDT communicated directly with the treating clinicians or patients.

Esophageal malignancy was confirmed by histopathological examination of endoscopic biopsy or surgical specimens. For cases without pathological confirmation, outcomes were ascertained by endoscopy, electronic medical record review or telephone follow-up. Among individuals with negative findings by both AI and SOC, the MDT randomly reviewed approximately 1% of cases using available electronic medical records through 31 July 2026. Detailed methodology, results, workflow and follow-up procedures are provided in Extended Data Table 4 and Supplementary Methods—Details of prospective validation in hospital.

**Real-world validation in LDCT program.** This study aims to evaluate the performance of EAGLE-Plus in the large-scale LDCT real-world cohort. Individuals aged 45–75 who underwent LDCT for routine health examination program at SIPD between 1 July 2024 and 31 December 2024 were eligible. A total of 10,959 participants were included. Diagnostic labels were confirmed for all cases flagged by either SOC or EAGLE-Plus-positive, based on initial radiology reports, clinical records and telephone follow-up through 31 July 2026. For participants with concordant negative findings by both SOC and AI, an additional team randomly selected approximately 1% for review to confirm the absence of malignancy, provided their clinical records were available through 31 July 2026. Further details are provided in Supplementary Methods—Details of real-world retrospective LDCT validation.

### Population-based endoscopic screening: EAGLE for pre-endoscopy risk stratification

Suining is a high-incidence area for EC. As part of the National Screening Program for upper gastrointestinal cancers, SNCH provides free endoscopic screening to residents aged 40–69 with high-risk factors.

Only individuals who provided informed consent forms for NC CT were included in this study (Supplementary Fig. 7). Between 7 July 2023 and 26 July 2023, 530 participants were prospectively enrolled. Each participant first underwent NC CT, followed by standard upper gastrointestinal endoscopy with iodine staining (1.2% Lugol's solution). All focal lesions detected during endoscopy were biopsied, and two board-certified pathologists independently evaluated the specimens using established diagnostic criteria[55].

In the prospectively enrolled cohort, six cases of HGIN were enrolled, but no invasive cancers were found. Because the low number of malignant events limited direct evaluation of EAGLE as a risk-stratification tool, we performed an exploratory simulation study using a bootstrapping approach. In each iteration, and based on the reported 2:1 incidence ratio between HGIN and EC reported in high-risk populations undergoing endoscopic screening[2], three cancer cases were randomly sampled with replacement from the 420 EC cases of the SNCH retrospective test cohort. These sampled EC cases were combined with the prospectively enrolled cohort to construct a hybrid dataset for analysis. For each simulated hybrid cohort, we estimated the detection rate achievable when using EAGLE to triage only high-risk patients for endoscopic examination. This process was repeated 1,000 times to reflect the expected disease prevalence and variability in real-world screening settings. Detailed methodology and results were provided in Supplementary Methods—Details of validation on prospectively enrolled endoscopic screening cohort.

### Statistical analysis

All statistical analyses were performed in Python (v3.9.19) using scikit-learn (v1.4.2), NumPy (v1.26.4) and SciPy (v1.13.0). Binary classification performance was evaluated using the AUC, sensitivity and specificity. PPV was used to assess the precision of the model's positive predictions in clinical practice. Balanced accuracy was additionally used in the reader study to account for class imbalance. Unless otherwise stated, each metric is reported as a point estimate with a 95% CI. Bootstrap resampling with 1,000 iterations was used to estimate 95% CIs. In each iteration, patients were resampled with replacement from the cohort under analysis and the metric was recalculated; the 95% CI was defined by the 2.5th and 97.5th percentiles of the bootstrap distribution. AUCs of two models evaluated on the same patients were compared using the DeLong test for correlated receiver operating characteristic curves. Permutation tests were used for statistical comparisons of sensitivity, specificity, stage-specific sensitivity and balanced accuracy, with at least 1,000 random permutations. The McNemar test was used to compare paired binary outcomes, specifically the esophageal malignancy detection rate between the EAGLE risk-stratified pathway and direct endoscopy in the population-based screening cohort. Comparisons between models or between readers were performed on the same set of patients and were therefore treated as paired, repeated measurements, whereas all other analyses were based on measurements from distinct, independent patients. The bootstrap and permutation procedures are nonparametric and do not assume normality of the underlying data distributions. All tests were two sided, no adjustment was made for multiple comparisons and a $P$ value of <0.05 was considered statistically significant. Exact $P$ values are reported in the Figs. 1c, 2f,g, 3c,d, 4a and 5e, and $P$ values below 0.001 are reported as $P < 0.001$. A fixed random seed of 65,535 was used throughout.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

The datasets generated and/or analyzed in this study are not publicly available because public release is not allowed by the relevant institutional review boards. To facilitate accessibility, we provide a sample dataset and an interactive portal at https://eagle.damomed.com/. All shared data will be de-identified to protect participant privacy in accordance with applicable laws and regulations. Researchers seeking access to the data and supporting documentation may submit a request to the first author (J. Zhou) or corresponding author (Q. Wang). Requests will be evaluated by an independent review panel based on scientific merit and compliance with ethical and regulatory requirements. Such requests will normally be processed within 6 weeks.

## Code availability

The code for the EAGLE model is protected by patents (CN CN117853490B, US 18046385) and cannot be publicly released. Implementation details are fully documented in the Methods, enabling replication. The core code is implemented using open-source repositories—PyTorch (https://pytorch.org, v2.5.1) and nnU-Net (https://github.com/MIC-DKFZ/nnUNet, v2).

## References


45. Heinrich, M. P., Jenkinson, M., Brady, M. & Schnabel, J. A. MRF-based deformable registration and ventilation estimation of lung CT. *IEEE Trans. Med. Imaging* **32**, 1239–1248 (2013).
46. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A. & Torralba, A. Learning deep features for discriminative localization. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 2921–2929 https://doi.org/10.1109/CVPR.2016.319 (IEEE, 2016).
47. Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J. & Maier-Hein, K. H. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **18**, 203–211 (2021).
48. Isensee, F. et al. nnU-Net revisited: A call for rigorous validation in 3D medical image segmentation. In *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention* (eds Linguraru, M. G. et al.) 488–498 https://doi.org/10.1007/978-3-031-72114-4_47 (Springer, 2024).
49. Muto, M. et al. Early detection of superficial squamous cell carcinoma in the head and neck region and esophagus by narrow band imaging: a multicenter randomized controlled trial. *J. Clin. Oncol.* **28**, 1566–1572 (2010).
50. Li, J. et al. Lugol chromoendoscopy detects esophageal dysplasia with low levels of sensitivity in a high-risk region of China. *Clin. Gastroenterol. Hepatol.* **16**, 1585–1592 (2018).
51. Tang, Y. et al. Self-supervised pre-training of Swin Transformers for 3D medical image analysis. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition* 20730–20740 https://doi.org/10.1109/CVPR52688.2022.02007 (IEEE, 2022).
52. Cheng, B., Misra, I., Schwing, A. G., Kirillov, A. & Girdhar, R. Masked-attention mask transformer for universal image segmentation. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition* 1280–1289 https://doi.org/10.1109/CVPR52688.2022.00135 (IEEE, 2022).
53. Ma, J., Li, F. & Wang, B. U-Mamba: enhancing long-range dependency for biomedical image segmentation. Preprint at https://doi.org/10.48550/arXiv.2401.04722 (2024).
54. Baumgartner, M., Jäger, P. F., Isensee, F. & Maier-Hein, K. H. nnDetection: a self-configuring method for medical object detection. In *Proc. 24th International Conference on Medical Image Computing and Computer Assisted Intervention* (eds de Bruijne, M. et al.) 530–539 https://doi.org/10.1007/978-3-030-87240-3_51 (Springer, 2021).
55. Dawsey, S. M., Lewin, K. J., Liu, F.-S., Wang, G.-Q. & Shen, Q. Esophageal morphology from Linxian, China. Squamous histologic findings in 754 patients. *Cancer* **73**, 2027–2037 (1994).

## Acknowledgements
The authors acknowledge external contributions provided by Y. Qiao at the Chinese Academy of Medical Sciences and Peking Union Medical College for many valuable discussions on cancer screening. The authors acknowledge GenesisCare for its support in collecting and providing the data for external validation.

## Author contributions
L.Z., J.Y., J. Zhou, Y. Zhao and Q.W. conceptualized the study. J.Y., L.Z. and G.G. designed the study. J. Zhou, C.Z., X.F., X.Y., Yongjin Zhou, P.Y., L.L., S.Z., Q.L., Y.C., M.Z., Y.J., D. Zheng, J.G., H.Q., W.L., P.Z., M.L., L.W., Y.L., H.W., Yong Zhou, J.C., L.X., C.L., L.C., C.S., J.J., N.L., C.X., K.C., Y. Zhao and Q.W. were responsible for the acquisition of the data. J.Y., G.G. and Q.Y. contributed to data preprocessing. G.G., J.Y. and Q.Y. contributed to the AI model development. J.Y., G.G., J. Zhou, Q.Y., K.C., Y. Zhao, L.Z., Q.W., W.W., K.Z., J. Zhang, Y.X., J.H. and D. Zhang conducted the data analysis and interpretation. W.G. and J.Y. contributed to the clinical deployment. J.Y., G.G., Y. Zhao and L.Z. conducted statistical analysis. G.G., J.Y., L.Z., Q.Y. and Y. Zhao wrote and revised the manuscript.

## Funding
This work was supported by DAMO Academy (Hupan Laboratory) through the DAMO Academy (Hupan Laboratory) Innovative Research Program. J. Zhou was supported by Guangdong Esophageal Cancer Institute Science and Technology (grant Q202314) and Medical Scientific Research Foundation of Guangdong Province (grant A2021179). N.L. was supported by Sichuan Provincial Science and Technology Program (Key Research and Development Project, 2023ypfs0488). Y. Zhao was supported by the Sichuan Science and Technology Program (2026YFHZ0166).

## Competing interests
G.G., J.Y., Q.Y., Y.X., W.G., Y.C., J. Zhang and L.Z. are employees of Alibaba Group and own Alibaba stock as part of the standard compensation package. The other authors declare no competing interests.

## Additional information
**Extended data** is available for this paper at https://doi.org/10.1038/s41591-026-04656-4.

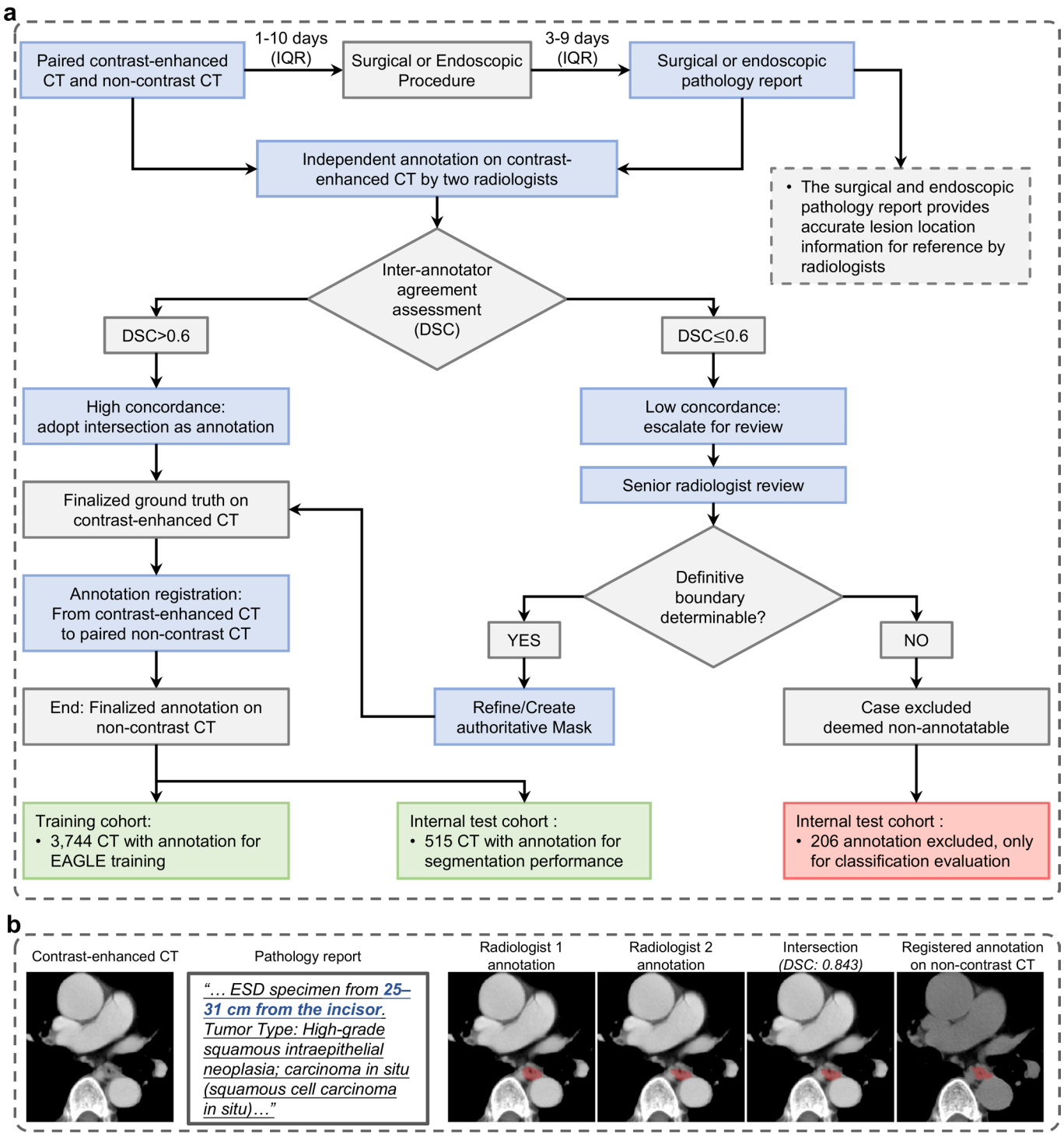
**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41591-026-04656-4.

**Correspondence and requests for materials** should be addressed to Dingwen Zhang, Jiansong Ji, Na Li, Chuanmiao Xie, Kai Cao, Yuqian Zhao, Ling Zhang or Qifeng Wang.

**Peer review information** *Nature Medicine* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: Joao Monteiro, in collaboration with the *Nature Medicine* team.

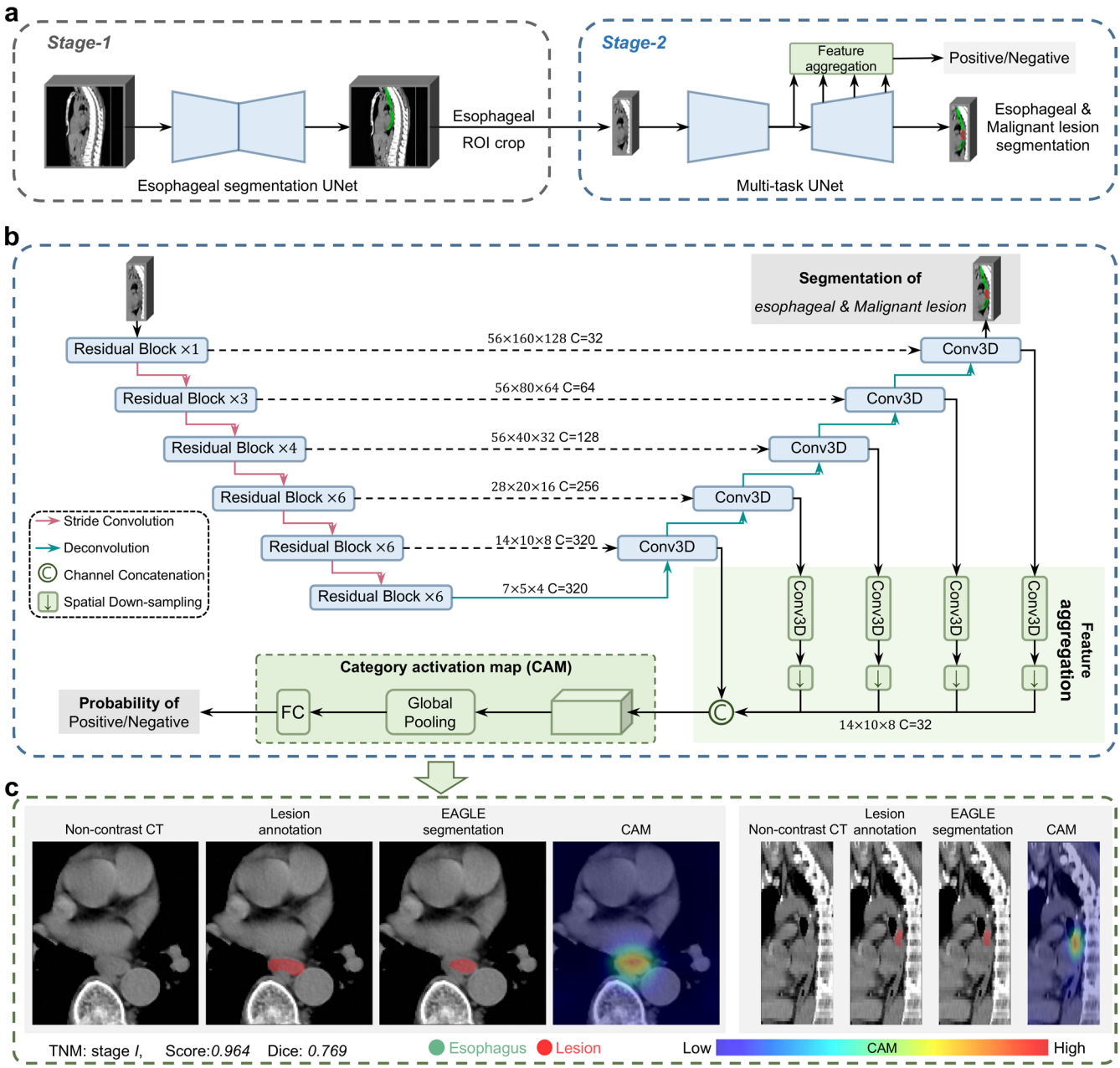
**Reprints and permissions information** is available at www.nature.com/reprints.

https://doi.org/10.1038/s41591-026-04656-4



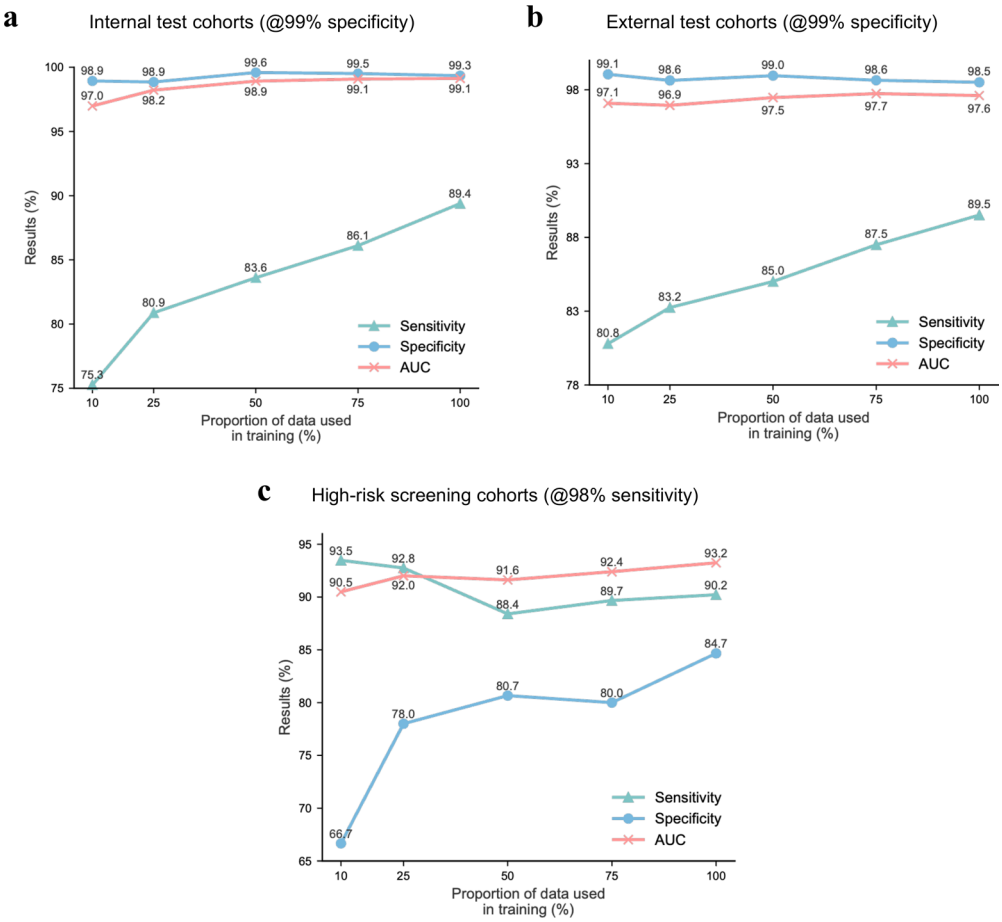

**Extended Data Fig. 1 | Voxel-level lesion annotation protocol. a**, Lesion annotation workflow: two radiologists independently annotated lesions on paired contrast-enhanced CT (CECT). High interannotator agreement (DSC > 0.6) resulted in an accepted mask intersection; discordant cases were reviewed by a senior radiologist or excluded if undeterminable. Final CECT annotations were registered to noncontrast CT for use in EAGLE model training and validation. **b**, Example of annotating a high-grade intraepithelial neoplasia (HGIN) case: both radiologists independently delineated highly consistent lesion masks (DSC = 0.843) on CECT guided by lesion location in the surgical pathology report. The intersection mask was adopted as ground truth and accurately registered onto the paired noncontrast CT scan.

https://doi.org/10.1038/s41591-026-04656-4



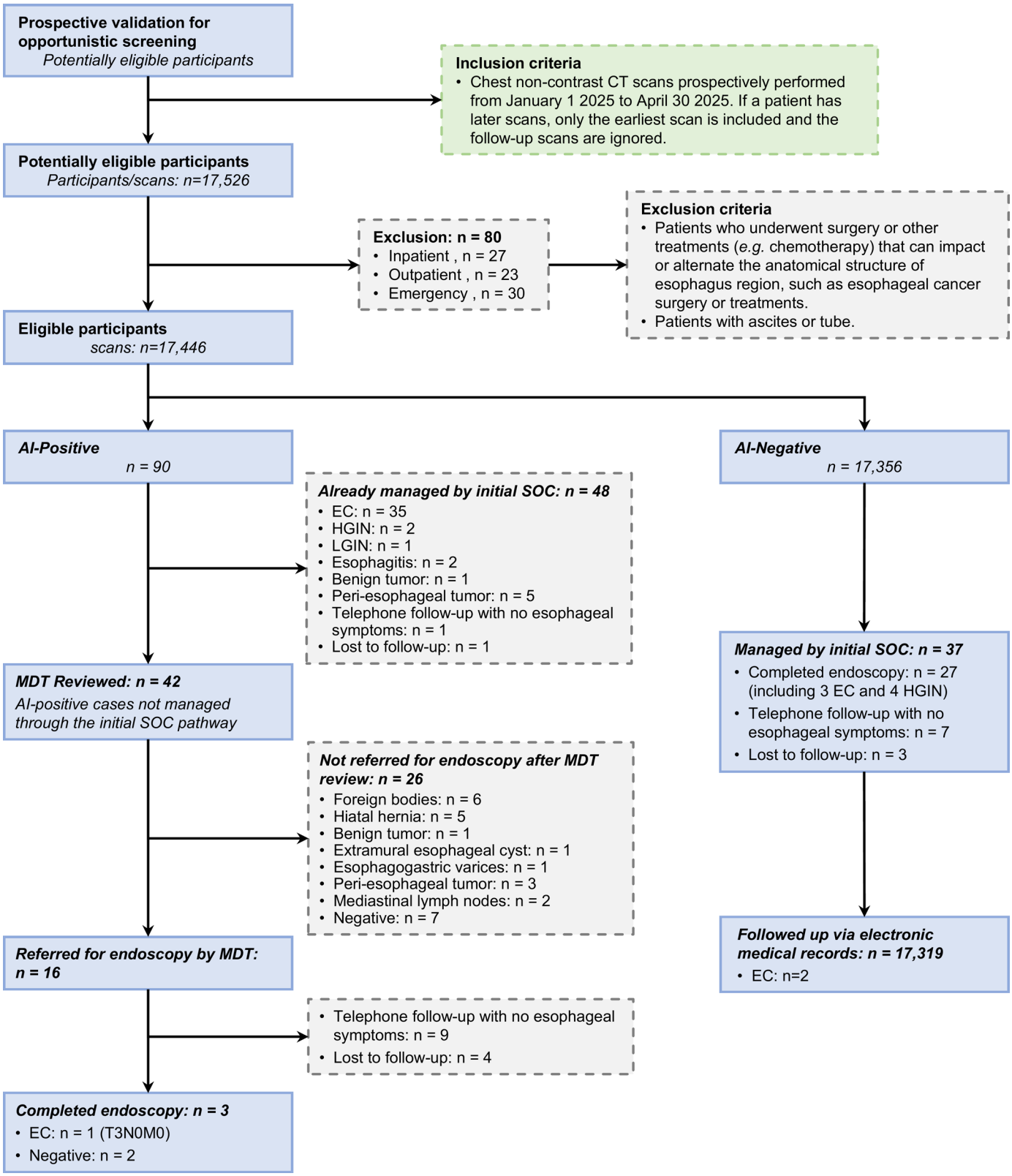

**Extended Data Fig. 2 | EAGLE's network architecture: stages, details, and interpretability analysis. a**, EAGLE consists of two stages: the esophageal localization stage that uses a segmentation UNet, and the prediction stage using a joint classification and segmentation UNet, which takes the cropped ROI region as input and predicts malignant diagnosis, and segmentation of esophageal and malignant lesion. **b**, Detailed architecture of stage 2. We extract multiscale features from the segmentation UNet, align their spatial sizes via downsampling, and concatenate all features. The classification head consists of a global pooling layer and a fully connected layer. **c**, Interpretability demonstration: a representative cancer case illustrates model explainability across axial and sagittal planes. Each plane visualizes a noncontrast CT image, malignant lesion annotation (ground truth), EAGLE segmentation mask, and class activation mapping (CAM) from stage-2 classification head, highlighting regions that contribute to the diagnostic decision-making process.

https://doi.org/10.1038/s41591-026-04656-4



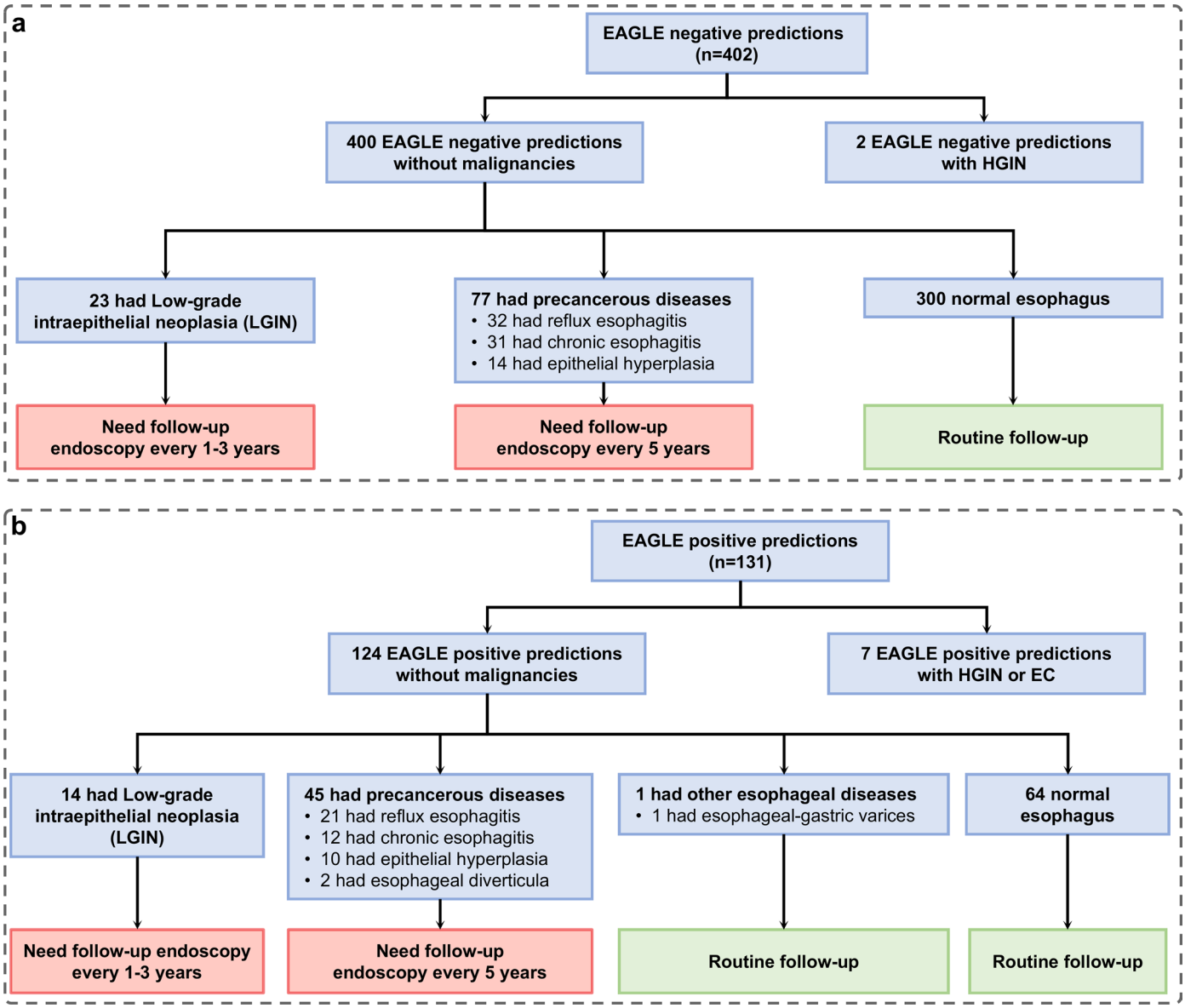

**Extended Data Fig. 3 | Influence of the proportion of training data.** We report the sensitivity, specificity, and AUC of EAGLE. **a**, Results on internal test cohorts at 99% specificity. **b**, Results on external test cohorts at 99% specificity. **c**, Results on external high-risk screening test cohorts at 98% sensitivity.

https://doi.org/10.1038/s41591-026-04656-4



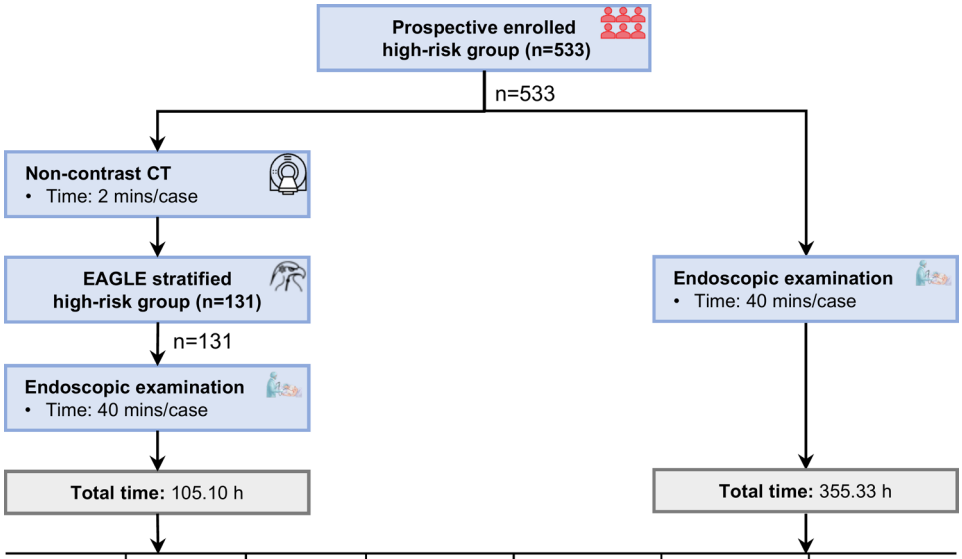

**Extended Data Fig. 4 | Workflow of the prospective validation for opportunistic screening.** Summarizing participant selection, AI deployment, multidisciplinary review, and follow-up procedures in the prospective hospital validation of EAGLE-Plus. Noncontrast CT examinations performed between 1 January and 30 April 2025 were screened for eligibility, and 17,446 eligible scans from unique participants were included after exclusions. High-risk cases identified by either EAGLE-Plus or the standard-of-care (SOC) pathway were reviewed by the multidisciplinary team (MDT). For AI-positive cases not identified through the initial SOC pathway, the MDT determined whether endoscopic evaluation or follow-up was warranted. Outcome ascertainment was based on pathology when available, and otherwise on endoscopy, telephone follow-up, or electronic medical record review.

https://doi.org/10.1038/s41591-026-04656-4




**Extended Data Fig. 5 | Endoscopic findings of EAGLE predictions in endoscopy screening cohort. a**, Endoscopic findings of 402 EAGLE low-risk participants; 2 of 402 were HGIN. **b**, Endoscopic findings of 131 EAGLE high-risk participants; 7 of 131 were esophageal malignancies.

https://doi.org/10.1038/s41591-026-04656-4




| Region | Currency | CT Cost | Endoscopy Cost | Total Cost (Endoscopy) | Total Cost (EAGLE Stratified) | Cost Reduction (%) |
|---|---|---|---|---|---|---|
| USA | USD | $300 | $2,759 | $1,470,547 | $521,329 | - 64.5% |
| Singapore | SGD | $360 | $2,500 | $1,332,500 | $519,380 | - 61.0% |
| China | CNY | ¥210 | ¥1,200 | ¥639,600 | ¥269,130 | - 57.9% |
| UK | GBP | £79 | £426 | £227,058 | £97,913 | - 56.9% |
| Netherlands | EUR | €411.18 | €1,593.87 | €849,533 | €427,955 | - 49.6% |
| Australia | AUD | $465 | $1,600 | $852,800 | $457,445 | - 46.4% |
| South Korea | KRW | ₩75,840 | ₩200,000 | ₩106,600,000 | ₩66,622,720 | - 37.5% |
| Japan | JPY | ¥9,500 | ¥10,600 | ¥5,649,800 | ¥6,452,100 | +14.2% |

**Extended Data Fig. 6 | Time and cost analysis of EAGLE risk-stratified population endoscopic screening.** The left branch illustrates the EAGLE risk-stratified screening strategy: all participants first undergo noncontrast CT, followed by EAGLE risk assessment, with only high-risk individuals proceeding to endoscopic screening. The right branch represents the conventional strategy, where all participants receive endoscopic examination. This analysis was performed in the context of government-led screening programs, in which screening-eligible individuals are typically identified in advance through questionnaire-based risk assessment. Under this setting, we assumed a baseline scenario in which the entire target cohort would otherwise undergo endoscopic examination as part of the established screening pathway. Total screening costs were estimated from a societal perspective. The sources of noncontrast CT and endoscopic examination costs are provided in Supplementary Table 4. Time required for noncontrast CT and endoscopy examinations is derived from clinical experience at Suining. Endoscopy cost includes anesthesia but excludes biopsy. The analysis suggests a substantial reduction in total screening time and cost when applying the EAGLE risk-stratified approach. These model-based estimates are context-specific and may be influenced by local screening practices.

https://doi.org/10.1038/s41591-026-04656-4

## Extended Data Table 1 | Patient characteristics of the training, internal validation and external validation cohort

| | Int. Train | Int. Test | Ext. Test Opp. | LDCT Test | Ext. Test High-risk | Int. Train SYSUCC | Int. Train SCCH | Int. Test SYSUCC | Int. Test SCCH | Ext. Test SIPD | Ext. Test FAHZU | Ext. Test HCH | Ext. Test STCH | Ext. Test XMUACH | Ext. Test FCH | Ext. Test GUHP | Ext. Test GenesisCare | Ext. Test SNCH | Ext. Test LCH |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **n** | 6813 | 2500 | 11466 | 1607 | 702 | 4255 | 2558 | 1482 | 1018 | 4904 | 4313 | 761 | 731 | 325 | 32 | 195 | 205 | 599 | 103 |
| **Lesion Types** | | | | | | | | | | | | | | | | | | | |
| EC/HGIN | 3744 | 1281 | 2812 | 147 | 552 | 2716 | 1028 | 767 | 514 | 1113 | 363 | 139 | 731 | 175 | 32 | 54 | 205 | 449 | 103 |
| Negative | 3069 | 1219 | 8654 | 1460 | 150 | 1539 | 1530 | 715 | 504 | 3791 | 3950 | 622 | 0 | 150 | 0 | 141 | 0 | 150 | 0 |
| **Age** | 59 | 60 | 54 | 57 | 65 | 60 | 57 | 60 | 60 | 50 | 56 | 53 | 59 | 58 | 62 | - | 74 | 64 | 69 |
| **(IQR)** | 52-65 | 52-67 | 44-64 | 50-62 | 58-70 | 53-66 | 50-65 | 53-67 | 52-67 | 39-61 | 47-65 | 45-62 | 54-65 | 50-65 | 59-68 | - | 67-79 | 57-69 | 69-73 |
| **Sex** | | | | | | | | | | | | | | | | | | | |
| Female | 2780 | 1021 | 5645 | 550 | 207 | 1335 | 1445 | 511 | 510 | 1670 | 3167 | 410 | 171 | 165 | 9 | 0 | 53 | 185 | 22 |
| Male | 4033 | 1452 | 5555 | 1057 | 495 | 2920 | 1113 | 950 | 502 | 3234 | 1076 | 351 | 560 | 160 | 23 | 0 | 151 | 414 | 81 |
| Missing | 0 | 27 | 266 | 0 | 0 | 0 | 0 | 21 | 6 | 0 | 70 | 0 | 0 | 0 | 0 | 195 | 1 | 0 | 0 |
| **TNM** | | | | | | | | | | | | | | | | | | | |
| HGIN | 264 | 18 | 40 | 14 | 40 | 129 | 135 | 9 | 9 | 34 | 0 | 3 | 2 | 1 | 0 | 0 | 0 | 29 | 11 |
| Stage I | 548 | 181 | 363 | 14 | 153 | 462 | 86 | 136 | 45 | 233 | 32 | 9 | 48 | 19 | 1 | 0 | 21 | 85 | 68 |
| Stage II | 1149 | 359 | 1141 | 9 | 186 | 857 | 292 | 258 | 101 | 525 | 65 | 41 | 341 | 57 | 17 | 14 | 81 | 174 | 12 |
| Stage III | 1378 | 430 | 1002 | 35 | 154 | 980 | 398 | 267 | 163 | 279 | 181 | 72 | 299 | 75 | 12 | 18 | 66 | 144 | 10 |
| Stage IV | 405 | 159 | 261 | 21 | 19 | 288 | 117 | 97 | 62 | 42 | 85 | 14 | 41 | 23 | 2 | 21 | 33 | 17 | 2 |
| Missing | 0 | 134 | 5 | 54 | 0 | 0 | 0 | 0 | 134 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 4 | 0 | 0 |
| **Location** | | | | | | | | | | | | | | | | | | | |
| Upper | 476 | 73 | 334 | 4 | 82 | 216 | 260 | 33 | 40 | 115 | 96 | 34 | 41 | 14 | 2 | 11 | 21 | 69 | 13 |
| Middle | 1849 | 492 | 1184 | 15 | 241 | 1339 | 510 | 369 | 123 | 423 | 170 | 37 | 392 | 83 | 19 | 18 | 42 | 202 | 39 |
| Lower | 1050 | 304 | 1075 | 16 | 192 | 839 | 211 | 246 | 58 | 416 | 91 | 58 | 281 | 70 | 11 | 10 | 138 | 168 | 24 |
| EGJ | 317 | 408 | 140 | 6 | 12 | 303 | 14 | 116 | 292 | 108 | 2 | 3 | 4 | 6 | 0 | 14 | 3 | 4 | 8 |
| Missing | 52 | 4 | 79 | 106 | 25 | 19 | 33 | 3 | 1 | 51 | 4 | 7 | 13 | 2 | 0 | 1 | 1 | 6 | 19 |
| **Histology** | | | | | | | | | | | | | | | | | | | |
| ESCC | 3339 | 861 | 2428 | 139 | 525 | 2379 | 960 | 639 | 222 | 950 | 359 | 134 | 705 | 158 | 32 | 30 | 60 | 429 | 96 |
| EAC | 317 | 413 | 283 | 8 | 11 | 308 | 9 | 121 | 292 | 81 | 4 | 4 | 12 | 14 | 0 | 24 | 144 | 7 | 4 |
| Others | 55 | 7 | 20 | 0 | 12 | 29 | 26 | 7 | 0 | 16 | 0 | 1 | 0 | 3 | 0 | 0 | 0 | 9 | 3 |
| Missing | 33 | 0 | 81 | 0 | 4 | 0 | 33 | 0 | 0 | 66 | 0 | 0 | 14 | 0 | 0 | 0 | 1 | 4 | 0 |
| **Thickness** | | | | | | | | | | | | | | | | | | | |
| **mean (mm)** | 5 | 5 | 4.955 | 5 | 2.014 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 3 | 2.2 | 1.5 | 5 |
| Q1 | 5 | 5 | 4.953 | 5 | 1.800 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 3 | 2 | 1.25 | 5 |
| Q3 | 5 | 5 | 4.957 | 5 | 2.014 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 3 | 2.5 | 1.5 | 5 |
| IPR | 0 | 0 | 0.005 | 0 | 0.214 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.5 | 0.25 | 0 |
| **Spacing** | | | | | | | | | | | | | | | | | | | |
| **mean (mm)** | 0.742 | 0.703 | 0.736 | 0.684 | 0.748 | 0.735 | 0.684 | 0.764 | 0.684 | 0.758 | 0.703 | 0.725 | 0.656 | 0.748 | 0.761 | 0.782 | 1.147 | 0.762 | 0.684 |
| Q1 | 0.684 | 0.677 | 0.665 | 0.684 | 0.703 | 0.686 | 0.652 | 0.718 | 0.654 | 0.707 | 0.684 | 0.684 | 0.625 | 0.703 | 0.728 | 0.732 | 0.977 | 0.703 | 0.672 |
| Q3 | 0.795 | 0.781 | 0.806 | 0.711 | 0.807 | 0.779 | 0.732 | 0.822 | 0.713 | 0.813 | 0.705 | 0.775 | 0.707 | 1.117 | 0.782 | 0.782 | 1.27 | 0.816 | 0.731 |
| IPR | 0.111 | 0.104 | 0.141 | 0.027 | 0.104 | 0.093 | 0.08 | 0.104 | 0.059 | 0.106 | 0.021 | 0.091 | 0.082 | 0.414 | 0.054 | 0.05 | 0.293 | 0.113 | 0.059 |
| **z axis extent** | | | | | | | | | | | | | | | | | | | |
| **mean (mm)** | 370 | 325 | 337 | 350 | 341 | 380 | 315 | 360 | 310 | 357 | 315 | 300 | 295 | 410 | 335 | 363 | 449 | 473 | 330 |
| Q1 | 305 | 289 | 304 | 330 | 285 | 335 | 290 | 235 | 295 | 335 | 295 | 280 | 280 | 370 | 318 | 333 | 407 | 406 | 300 |
| Q3 | 495 | 400 | 369 | 370 | 424 | 415 | 350 | 450 | 330 | 390 | 335 | 330 | 320 | 440 | 361 | 396 | 482 | 644 | 360 |
| IPR | 190 | 111 | 65 | 40 | 139 | 80 | 60 | 215 | 35 | 55 | 40 | 50 | 40 | 70 | 43 | 63 | 75 | 238 | 60 |
| **KVP mean** | 120 | 120 | 120 | 100 | 120 | 120 | 120 | 120 | 120 | 120 | 120 | 120 | 120 | 120 | 120 | 120 | 120 | 120 | 120 |
| Q1 | 120 | 120 | 120 | 100 | 120 | 120 | 120 | 120 | 120 | 120 | 120 | 120 | 120 | 120 | 120 | 120 | 120 | 120 | 120 |
| Q3 | 120 | 120 | 120 | 100 | 120 | 120 | 120 | 120 | 120 | 120 | 120 | 120 | 120 | 120 | 120 | 120 | 120 | 120 | 120 |
| IPR | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Exposure** | | | | | | | | | | | | | | | | | | | |
| **mean (mAs)** | 93.2 | 89 | 83 | 30 | 82 | 84 | 83.8 | 135 | 86.5 | 110 | 15.5 | 81.3 | 82.3 | 150 | 82 | 116 | 100 | 74 | 126 |
| Q1 | 19 | 15.8 | 14 | 20 | 4 | 65 | 19 | 72 | 15.8 | 86 | 10.8 | 53.8 | 26 | 53 | 57 | 75 | 21 | 4 | 97 |
| Q3 | 132 | 139.5 | 150 | 30 | 122 | 131 | 129 | 176 | 115.2 | 151 | 24.9 | 149 | 104 | 180 | 95 | 170 | 156 | 111 | 135 |
| IPR | 113 | 123.7 | 136 | 10 | 118 | 66 | 110 | 104 | 99.4 | 65 | 14.1 | 95.2 | 78 | 127 | 38 | 95 | 135 | 107 | 38 |
| **CTDIvol** | | | | | | | | | | | | | | | | | | | |
| **mean (mGy)** | 7.93 | 7.9 | 11.5 | 1.2 | 7.7 | 15.4 | 7 | 11.6 | 6.7 | 9.8 | 9.55 | 6.8 | 8.1 | 11.5 | 9.1 | 8.7 | 13.5 | 7.03 | 8.8 |
| Q1 | 5.51 | 6 | 7.6 | 0.8 | 5.8 | 5.5 | 5.5 | 7.5 | 5.6 | 6.9 | 7.18 | 4.9 | 7.2 | 11 | 6 | 5.7 | 9.7 | 5.7 | 7.31 |
| Q3 | 11.98 | 11.7 | 15.4 | 1.2 | 9.5 | 11.9 | 13.1 | 14.4 | 8.9 | 12.5 | 9.87 | 15.9 | 9.5 | 23.6 | 11.6 | 12.4 | 16.4 | 9.3 | 9.99 |
| IPR | 6.47 | 5.7 | 7.8 | 0.4 | 3.7 | 6.4 | 7.6 | 6.9 | 3.3 | 5.6 | 2.69 | 11 | 2.3 | 12.6 | 5.6 | 6.7 | 6.7 | 3.6 | 2.68 |
| **Tube current** | | | | | | | | | | | | | | | | | | | |
| **mean (mA)** | 234 | 295 | 240 | 91 | 250 | 238 | 279 | 258 | 294 | 258 | 228 | 225 | 200 | 265 | 116 | 233 | 263 | 263 | 187 |
| Q1 | 175 | 226 | 182 | 61 | 144 | 188 | 195 | 200 | 225 | 202 | 181 | 150 | 194 | 240 | 103 | 169 | 151 | 148 | 136 |
| Q3 | 300 | 300 | 313 | 91 | 408 | 250 | 300 | 342 | 297 | 328 | 298 | 295 | 257 | 420 | 151 | 329 | 341 | 430 | 295 |
| IPR | 125 | 74 | 131 | 30 | 264 | 62 | 105 | 142 | 72 | 126 | 117 | 145 | 63 | 180 | 48 | 160 | 190 | 282 | 159 |

TNM stage (AJCC eighth edition), lesion location and histology are displayed only for the patients with esophageal malignancy, whereas age, sex and the CT acquisition parameters refer to all patients of the corresponding column. Age is reported as median (IQR), and slice thickness, in-plane spacing, z axis extent, peak kilovoltage, exposure, CTDIvol and tube current are reported as mean together with Q1, Q3 and IPR. n, number of patients; int., internal; ext., external; opp., opportunistic screening scenario; high-risk, high-risk endoscopic screening scenario; LDCT, low-dose CT; IQR, interquartile range; Q1, first quartile; Q3, third quartile; IPR, interpercentile range, computed as Q3 minus Q1; SYUCC, Sun Yat-sen University Cancer Center; SCCH, Sichuan Cancer Hospital; SIPD, Shanghai Institution of Pancreatic Diseases; FAHZU, First Affiliated Hospital of Zhejiang University; HCH, Hunan Cancer Hospital; STCH, Shantou Central Hospital; XMUACH, Xinjiang Medical University Affiliated Cancer Hospital; FCH, Fujian Cancer Hospital; GUHP, General University Hospital in Prague, Czech Republic; GenesisCare, GenesisCare, Australia; SNCH, Suining Central Hospital; LCH, Lishui Central Hospital.

https://doi.org/10.1038/s41591-026-04656-4

**Extended Data Table 2 | The performance of EAGLE in internal and external test cohorts under different scenarios**

| Validation cohort | AUC | 95% CI of AUC | Specificity | 95% CI of Specificity | Sensitivity | 95% CI of Sensitivity |
|---|---|---|---|---|---|---|
| Internal test (n=2500) | | | | | | |
| Scenario: hospital | 0.991 | 0.989-0.994 | 99.3% | 98.8-99.8% | 89.4% | 87.7-91.0% |
| Operating point: 99% specificity | | | | | | |
| SYSUCC (n=1482) | 0.993 | 0.990-0.995 | 99.4% | 98.9-99.9% | 88.8% | 86.6-90.9% |
| SCCH (n=1018) | 0.989 | 0.984-0.994 | 99.2% | 98.3-99.8% | 90.3% | 87.6-92.8% |
| External test (n=11466) | | | | | | |
| Scenario: hospital | 0.976 | 0.973-0.980 | 98.5% | 98.3-98.8% | 89.5% | 88.4-90.6% |
| Operating point: 99% specificity | | | | | | |
| SIPD (n=4904) | 0.965 | 0.957-0.973 | 99.5% | 99.2-99.7% | 86.7% | 84.7-88.6% |
| FAHZU (n=4313) | 0.984 | 0.977-0.989 | 98.1% | 97.6-98.5% | 86.5% | 82.8-90.0% |
| HCH (n=761) | 0.996 | 0.993-0.999 | 96.1% | 94.6-97.6% | 95.7% | 92.3-98.6% |
| STCH (n=731) | - | - | - | - | 91.5% | 89.5-93.4% |
| XMUACH (n=325) | 0.992 | 0.983-0.998 | 97.3% | 94.3-99.4% | 96.6% | 93.6-98.9% |
| FCH (n=32) | - | - | - | - | 96.9% | 90.6-100.0% |
| GUHP (n=195) | 0.987 | 0.971-0.998 | 97.2% | 94.5-99.3% | 94.4% | 87.7-100.0% |
| GenesisCare (n=205) | - | - | - | - | 90.2% | 86.3-94.1% |
| LDCT test (n=1607) | | | | | | |
| Scenario: lung cancer&physical exam | | | | | | |
| EAGLE | 0.972 | 0.953-0.988 | 98.7% | 98.1-99.2% | 83.0% | 76.9-88.7% |
| EAGLE with LDCT simulation | 0.980 | 0.966-0.993 | 98.7% | 98.1-99.2% | 88.4% | 83.2-93.5% |
| Operating point: 99% specificity | | | | | | |
| External high-risk test (n=702) | | | | | | |
| Scenario: endoscopic screening | 0.932 | 0.909-0.954 | 84.7% | 79.1-90.3% | 90.2% | 87.8-92.7% |
| Operating point: 98% sensitivity | | | | | | |
| SNCH (n=599) | 0.943 | 0.921-0.964 | 84.7% | 78.8-90.2% | 92.2% | 89.6-94.6% |
| LCH (n=103) | - | - | - | - | 81.6% | 73.8-88.3% |

n/a.

https://doi.org/10.1038/s41591-026-04656-4

**Extended Data Table 3 | Comparison of backbone architectures and frameworks**

| | Internal | | | External | | |
|---|---|---|---|---|---|---|
| **Backbone** | **AUC** | **Specificity** | **Sensitivity** | **AUC** | **Specificity** | **Sensitivity** |
| EAGLE (ResEnc) | 0.991 | 99.3 | 89.4 | 0.976 | 98.5 | 89.5 |
| EAGLE (UMambaBot) | 0.990 | 99.1 | 87.7 | 0.979 | 98.5 | 89.1 |
| EAGLE (UMambaEnc) | 0.990 | 98.9 | 87.8 | 0.979 | 98.1 | 89.4 |
| EAGLE (SwinUNETR) | 0.985 | 98.7 | 84.1 | 0.969 | 96.5 | 86.3 |
| EAGLE (Mask2Former) | 0.978 | 99.5 | 71.7 | 0.925 | 98.2 | 65.8 |
| nnDetection | 0.934 | 99.5 | 50.0 | 0.954 | 99.6 | 68.1 |

We compare five state-of-the-art backbone architectures integrated into EAGLE's implementation leveraging nnU-Net, and also evaluate performance against the nnDetection framework. Detailed descriptions of the compared methods are provided in Supplementary Methods—Comparison to other methods.

https://doi.org/10.1038/s41591-026-04656-4

## Extended Data Table 4 | 90 EAGLE-Plus positive cases in the hospital prospective validation

| Category | Number | SOC recall | MDT recall | Diagnostic label | Diagnosis method |
|---|---|---|---|---|---|
| HGIN (n=2) | 2 | Y | N | HGIN | Surgical pathology |
| EC (n=36) | 35 | Y | N | EC | Surgical pathology |
| | 1 | N | Y | EC | Biopsy pathology |
| Esophageal findings (n=18) | 6 | N | N | Foreign bodies | CT |
| | 2 | Y | N | Esophagitis | Endoscopy |
| | 5 | N | N | Hiatal hernia | CT |
| | 1 | Y | N | Benign Tumor | CT |
| | 1 | N | N | Benign Tumor | Previous endoscopy within one year |
| | 1 | N | N | Extramural esophageal cyst | MRI |
| | 1 | N | N | Esophagogastric varices | Clinical records |
| | 1 | Y | N | LGIN | Endoscopy |
| peri-esophageal findings (n=10) | 5 | Y | N | Peri-esophageal tumor | Clinical records |
| | 3 | N | N | Peri-esophageal tumor | Clinical records |
| | 2 | N | N | Mediastinal lymph nodes | CT |
| No esophageal symptoms (n=10) | 9 | N | Y | No esophageal symptoms | Follow-up contact |
| | 1 | Y | N | No esophageal symptoms | Follow-up contact |
| Negative findings (n=9) | 1 | N | N | Negative | Previous endoscopy within one year |
| | 6 | N | N | Negative | Clinical records |
| | 2 | N | Y | Negative | Follow-up endoscopy |
| Lost to follow-up (n=5) | 1 | Y | N | Lost to follow-up | NA |
| | 4 | N | Y | Lost to follow-up | NA |

Among 90 cases, initial SOC recalled 48 cases due to esophageal findings and 37 were confirmed as esophageal malignancies. 16 cases were additionally recalled by MDT, and one cancer case was missed by initial SOC. Twenty-six cases were not recalled by both SOC and MDT. SOC, standard-of-care. MDT, multidisciplinary team. Y, yes; N, no. 10 cases who did not undergo complete endoscopy were categorized as no esophageal symptoms, as they remained asymptomatic according to telephone follow-ups conducted. 'Lost to follow-up' was defined based on the last recorded outcome as of 31 July, 2026.

nature portfolio

Corresponding author(s): Dingwen zhang, Jiansong Ji, Na Li, Chuanmiao Xie, Kai Cao,Yuqian Zhao, Ling Zhang, Qifeng Wang

Last updated by author(s): Aug 10, 2026

# Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☒ | ☐ | A description of all covariates tested |
| ☐ | ☒ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ | Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | Dicom files were processed with libraries such as Pydicom (https://pydicom.github.io/, version 2.3.0), SimpleITK (https://simpleitk.org/, version 2.0.2), and NiBabel (https://nipy.org/nibabel/, version 3.2.2). Custom Python (version 3.9.7) scripts were developed for data de-identification. |
| Data analysis | The code for the EAGLE model cannot be publicly released. Implementation details are fully documented in the Methods section, enabling replication. The core code is implemented using open-source repositories: PyTorch (https://pytorch.org, 2.5.1) and nnU-Net (https://github.com/MIC-DKFZ/nnUNet, v2). All statistical analyses were performed in Python (v3.9.19) using scikit-learn (v1.4.2), NumPy (v1.26.4), and SciPy (v1.13.0). |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The datasets generated and/or analyzed in this study are not publicly available because public release is not permitted by the relevant institutional review boards. To facilitate accessibility, we provide a sample dataset and an interactive portal at https://eagle.damomed.com/. All shared data will be de-identified to protect participant privacy in accordance with applicable laws and regulations. Researchers seeking access to the data and supporting documentation may submit a request to the first author (Jian Zhou: zhoujian@sysucc.org.cn) or corresponding author (Qifeng Wang: wangqifeng@scszlyy.org.cn). Requests will be evaluated by an independent review panel based on scientific merit and compliance with ethical and regulatory requirements. Such requests will normally be processed within 6 weeks.

## Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

| | |
|---|---|
| Reporting on sex and gender | Sex was collected based on the government-issued ID. The datasets used in the internal training and test cohorts, and the external multi-center test cohorts have sex distributions reported in the paper. Sex-based sensitivity of cancer was reported but sex was actually unrelated to model development. We didn't collect self-identification gender from the patients. |
| Reporting on race, ethnicity, or other socially relevant groupings | Race, ethnicity, and other socially relevant groupings were not collected from the patients and were unrelated to model implementation or deployment. |
| Population characteristics | This study involved three clinical scenarios: opportunistic screening in hospitals, opportunistic screening in established low-dose CT (LDCT) programs, and population-based high-risk endoscopic screening.<br><br>Validation of opportunistic screening in the hospital scenario was stratified across four distinct patient cohorts, each designed to address specific research objectives. Retrospective validation contained three cohorts: internal, dual-center cohorts for the model's performance within the same institutional framework (6,813 patients for training and 2,500 patients for testing), and external, multicenter (n=8, 11,466 patients) cohorts across China, the Czech Republic and Australia to assess the cross-institutional generalizability of EAGLE; Large-scale real-world, multicenter (n=3, 35,402 patients) cohorts were used to calibrate and evaluate EAGLE in the uncontrolled real-world population. Besides retrospective validation, a prospective validation cohort of 17,526 patients was used to evaluate the model in ambulatory settings.<br><br>Two LDCT validation cohorts were established to evaluate EAGLE's diagnostic performance in cost-effective clinical settings: dual-center LDCT test cohorts of 1,607 patients for evaluating EAGLE's diagnostic performance; a large-scale retrospective real-world LDCT cohort of 10,959 patients for reflecting the model's practical efficacy in high-throughput health screening environments for a target population aged between 45 to 75.<br><br>For population-based endoscopic screening scenarios, the validation involved two cohorts of patients enrolled in endoscopic screening programs: retrospective, external, dual-center test cohorts of 702 patients to assess EAGLE's cross-institutional generalizability in identifying high-risk individuals; a prospective enrolled cohort of 530 patients to evaluate the model's clinical utility in real-world endoscopic screening workflows for high-risk individuals, focusing on its role as a risk stratification tool for endoscopy. |
| Recruitment | The internal training cohort comprised 6,813 patients in total from two centers, SYSUCC and SCCH, where 3,744 patients were diagnosed with esophageal cancer or HGIN, and 3,069 patients were negative controls. From SYSUCC, 4,255 patients from 2009 to 2018 were collected. From SCCH, 2,558 patients from 2009 to 2015 were collected.<br><br>The internal test cohort included 2,500 patients with 1,281 esophageal malignant cases and 1,219 negative controls. From SYSUCC, 1,482 patients were collected. From SCCH, 1,018 patients were collected.<br><br>In the multi-center test cohorts , the non-contrast CT scans of 11,466 patients, including 2,812 esophageal malignant cases and 8,654 negative controls were collected from these centers. 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.

In the external test cohorts for endoscopic group, we collected data from two centers where endoscopic screening programs were widely implemented. Center I (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 J (n=103), Lishui Central Hospital (LCH) is a tertiary hospital in China. We collected 103 esophageal malignant patients from 2019 to 2024.

A total of 35,402 patients were enrolled from three centers for the real-world (RW) study: SYSUCC (Center A), SCCH (Center B), and SIPD (Center C). The EAGLE model was deployed and run locally at each center. Patients were divided into two real-world cohorts, RW1 and RW2. RW1 included 20,758 patients from three centers. At SYSUCC, 4,568 patients from the outpatient and physical examination departments were collected from March 2024 and 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. 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.

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 study. From 7 July 2023 to 26 July 2023, 530 patients in the endoscopic screening program at SNCH (Center I) were invited for the non-contrast CT examination free of charge, this study were completed and registered with http://www.chictr.org.cn, No. ChiCTR2300074806.

The real-world LDCT cohort included 10,959 individuals aged 45–75 years who underwent lung cancer screening at the physical examination department of SIPD between July 2024 and December 2024.

Ethics oversight

For the retrospective study, ethical approval was obtained individually from the institutional review board (IRB) of Sun Yat-Sen University Cancer Center (SYSUCC), Sichuan Cancer Hospital (SCCH), Shanghai Institution of Pancreatic Diseases (SIPD), First Affiliated Hospital of Zhejiang University (FAHZU), Hunan Cancer Hospital (HCH), Shantou Central Hospital (STCH), Xinjiang Medical University Affiliated Cancer Hospital (XMUACH), Fujian Cancer Hospital (FCH), General University Hospital in Prague (GUHP), GenesisCare, Suining Central Hospital (SNCH), and Lishui Central Hospital (LCH). The clinical trial was registered with ClinicalTrials https://www.chictr.org.cn, ChiCTR2300074806).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

The internal training cohort includes 6,813 patients. 5-fold cross validation was performed where each fold randomly selected 80% for training and 20% for the validation. This scheme follows machine learning convention for model tuning and hyperparameters selection.

The internal test cohort includes 2,500 patients and the external test cohorts include 11,466 patients across 8 medical centers.

Real-world, multicenter (n=3, 35,402 patients) cohorts were used to calibrate and evaluate EAGLE in the uncontrolled real-world population. A prospective validation cohort of 17,526 patients was used to evaluate the upgraded model. Two LDCT validation cohorts were established to evaluate EAGLE's diagnostic performance in cost-effective clinical settings: dual-center LDCT test cohorts of 1,607 patients for evaluating EAGLE's diagnostic performance; a large-scale retrospective real-world LDCT cohort of 10,959 patients for reflecting the model's practical efficacy in high-throughput health screening environments for a target population aged between 45 to 75.

For population-based endoscopic screening scenarios, two retrospective cohorts of patients enrolled in endoscopic screening programs of 702 patients to assess EAGLE's cross-institutional generalizability in identifying high-risk individuals; a prospective enrolled cohort of 530 patients to evaluate the model's clinical utility in real-world endoscopic screening workflows for high-risk individuals, focusing on its role as a risk stratification tool for endoscopy.

Data exclusions

Patients who comply one or more of the following criteria were excluded from the studies: (1) patients who underwent surgery that can impact or alternate the anatomical structure of the esophagus region, such as esophageal, gastric or endoscopic resection procedure, etc; (2) patients who underwent treatment to esophageal cancer (chemotherapy, radiotherapy, and chemoradiotherapy); (3) patients with low image quality due to artifacts caused by metal in stents or drastic motion during imaging; (4) Patients with ascites or tube.

Replication

The performance of EAGLE was consistent across the internal, external multi-center validation, real-world study cohorts and prospective cohorts, different CT scanners (GE, Philips, Siemens and UIH), scanning protocols (regular-dose non-contrast CT and low-dose non-contrast CT), and clinical scenarios (physical examination, emergency, inpatient and outpatient scenario).

| | |
|---|---|
| Randomization | In the internal training cohort, patients were randomly assigned to training and validation in the process of the cross-validation. |
| Blinding | The internal test cohort , the external multicenter test cohort, the LDCT test cohort, and the real-world clinical test cohort were not used for the development of EAGLE. The second subset of real-world clinical test cohort (RW2) were not used for the development of EAGLE-Plus. In two-round reader studies, readers were blinded to pathology results and other clinical information. Readers were also blinded to the data collection, exact proportions of positive patients, and blinded to other readers. Readers were blinded to the ground-truth labels and their performance after the study. |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ Antibodies |
| ☒ | ☐ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology and archaeology |
| ☒ | ☐ Animals and other organisms |
| ☐ | ☒ Clinical data |
| ☒ | ☐ Dual use research of concern |
| ☒ | ☐ Plants |

## Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☒ | ☐ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |

## Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

| | |
|---|---|
| Clinical trial registration | ChiCTR(chictr.org.cn): ChiCTR2300074806 |
| Study protocol | ChiCTR clinical trial protocols:<br>https://www.chictr.org.cn/showprojEN.html?proj=202090 |
| Data collection | A total of 17,446 patients were enrolled from January 2025 to April 2025 at SIPD. This prospective study aimed to validate the clinical reliability of the calibrated EAGLE-Plus model for opportunistic screening in real-world clinical settings, particularly uncontrolled, high-throughput clinical environments.<br><br>A cohort prospectively enrolled 530 patients from 1 July 2023 to 23 July 2023 at SNCH. All patients were initially referred for endoscopic examination and provided written informed consent to undergo an additional non-contrast CT scan. The cohort was specifically established to evaluate EAGLE's performance in risk stratification for esophageal malignant precancerous lesion and cancer. |
| Outcomes | In hospital settings, the primary outcome in the prospective validation study of EAGLE-Plus was the positive predictive value among AI-positive participants, defined as the proportion of AI-positive predictions confirmed as esophageal malignancy.<br><br>In the endoscopy prospective validation study, the primary outcome was the esophageal malignancy detection rate of endoscopy under an EAGLE risk-stratified pathway compared with direct endoscopy in the same participants. |

## Plants

| | |
|---|---|
| Seed stocks | *Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.* |
| Novel plant genotypes | *Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.* |
| Authentication | *Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.* |