Article

Effect of computer aided detection system on esophageal neoplasm diagnosis in varied levels of endoscopists

Check for updates

Bing Li[1,6], Yan-Yun Du[1,6], Wei-Min Tan[2,6], Dong-Li He[3,6], Zhi-Peng Qi[1,6], Hon-Ho Yu[4], Qiang Shi[1], Zhong Ren[1], Ming-Yan Cai[1], Bo Yan [2]✉, Shi-Lun Cai[1,3]✉ & Yun-Shi Zhong [1,3,5]✉

A computer-aided detection (CAD) system for early esophagus carcinoma identification during endoscopy with narrow-band imaging (NBI) was evaluated in a large-scale, prospective, tandem, randomized controlled trial to assess its effectiveness. The study was registered at the Chinese Clinical Trial Registry (ChiCTR2100050654, 2021/09/01). Involving 3400 patients were randomly assigned to either routine (routine-first) or CAD-assisted (CAD-first) NBI endoscopy, followed by the other procedure, with targeted biopsies taken at the end of the second examination. The primary outcome was the diagnosis of 1 or more neoplastic lesion of esophagus during the first examination. The CAD-first group demonstrated a significantly higher neoplastic lesion detection rate (3.12%) compared to the routine-first group (1.59%) with a relative detection ratio of 1.96 ($P = 0.0047$). Subgroup analysis revealed a higher detection rate in junior endoscopists using CAD-first, while no significant difference was observed for senior endoscopists. The CAD system significantly improved esophageal neoplasm detection, particularly benefiting junior endoscopists.

Esophageal cancer (EC) is one of the most common cancers in the world a leading cause of cancer death, with >600,000 new cases and >540,000 deaths in 2020[1]. Eastern Asia, including China and other developing countries, exhibits the highest regional incidence rates, and esophageal squamous cell carcinoma (ESCC) is the predominant histological type. Esophageal squamous dysplasia is considered to be the premalignant precursor lesion for ESCC and harbors a high risk for progression into invasive cancer[2,3]. Detection and resection of lesions in the precursor or early stage is an effective way to decrease the incidence and mortality of ESCC[4].

Endoscopic screening has been widely accepted as an optimal strategy for the diagnosis of ESCC and its precursor lesions[5]. However, it is not always easy to identify early-stage ESCC and dysplasia without typical features under endoscopy, which can easily pass unnoticed. To improve early ESCC detection, endoscopy with narrow-band imaging (NBI) system has been used frequently in routine screening examinations[6]. However, the diagnostic ability of NBI for early ESCC and dysplasia among variable levels of endoscopists is highly heterogeneous[7]. The sensitivity of NBI for screening of squamous mucosal high-grade neoplasia of the esophagus has been reported to be only 53% in the less experienced endoscopist group[8]. There is clearly a need to effectively assist endoscopists, especially for junior

group, to improve diagnostic accuracy. Then, more esophagus lesions can be diagnosed early and cured.

Recently, artificial intelligence (AI)-based technologies have been broadly applied in diagnostic medicine[9,10]. Many studies have shown that computer-aided diagnosis (CAD) in endoscopic analysis of early EC demonstrates excellent performance, including sensitivity and specificity[11–13]. In previous research, we reported a novel CAD system with a deep neural network (DNN) to identify early ESCC with an accuracy over 91% under conventional white-light endoscopy and over 94% under NBI[14,15]. Thereafter, we further modified and improved the model to identify esophageal neoplasms[16]. In fact, most of the reported CAD systems from different teams still used images for early ESCC diagnosis, not for real-time detection. Evidence on the ability of this technology to recognize and trace esophageal lesions in clinical practice during live endoscopy is lacking.

Here, we conducted a large-scale, randomized, comparative study in our center to evaluate the performance of CAD-NBI system compared with NBI alone in detecting esophageal neoplastic lesions. Based on the results, we aimed to verify the performance of our CAD system in the real-time clinical diagnosis to increase esophageal neoplasm detection rate. We also

[1]Endoscopy Center, Zhongshan Hospital of Fudan University, Shanghai, China. [2]School of Computer Science, Shanghai Key Laboratory of Intelligent Information Processing, Fudan University, Shanghai, China. [3]Endoscopy Center, Xuhui Hospital, Zhongshan Hospital of Fudan University, Shanghai, China. [4]Department of Gastroenterology, Kiang Wu Hospital, Macau SAR, China. [5]Endoscopy Center, Shanghai Geriatric Medical Center, Shanghai, China. [6]These authors contributed equally: Bing Li, Yan-Yun Du, Wei-Min Tan, Dong-Li He, Zhi-Peng Qi ✉e-mail: byan@fudan.edu.cn; cai.shilun@zs-hospital.sh.cn; zhongyunshi@yahoo.com

evaluated the applicability of the CAD system across endoscopists with different levels of expertise.

Results

The real-time working of CAD-NBI system

The CAD-NBI system monitor was positioned fixedly adjacent and parallel to the original endoscopy monitor. Figure 1 and Supplementary Movie 1 shows the real-time working of the CAD-NBI system. Supplementary Fig. 1 shows the schematic diagram of neoplastic lesions detection.

Study participants

Between September 3, 2021 and December 16, 2022, 3441 patients were invited to participate in this study, of whom 41 were judged ineligible and excluded before randomization (Fig. 2). Therefore, 3400 patients agreed to participate and were allocated to either endoscopy with assistance from the CAD-NBI system ($n = 1700$) or routine endoscopy with NBI ($n = 1700$). Four patients (2 in the routine-first group and 2 in the CAD-first group) were excluded during endoscopy because they met exclusion criteria. A total of 3396 eligible patients were included in the primary analysis, with 1698 patients allocated to the routine-first group and 1698 to the CAD-first group. Baseline characteristics of patients are presented in Table 1. There were no significant differences in baseline characteristics between the two groups. Of the randomly assigned participants, 91(5.36%) in the routine-first group and 101(5.95%) in the CAD-first group were awake, and others underwent endoscopy with intravenous anaesthesia.

The detection rate of lesions in each group as the main outcome

The percentage of patients diagnosed with a neoplastic lesion in the first examination by CAD-NBI (53 [57 lesions] of 1698 patients or 3.12% [95% CI, 2.35–4.06%]) was higher than that by NBI alone (27 [29 lesions] of 1698 patients or 1.59% [CI, 1.05–2.31%]) with a relative detection ratio of 1.96 (CI, 1.24–3.10; $P = 0.0047$). The combined number of patients diagnosed with neoplastic lesions in the first and second examinations was 58 (66 lesions) in the CAD-first group and 56 (64 lesions) in the routine-first group, with no difference between the groups (Table 2). To eliminate the potential interference of confounding variable, we performed a CMH test stratified by the level of operator experience. The common odds ratio of detection rate between CAD-first and routine-first is 2.02 (CI, 1.26–3.23; $P = 0.0042$), demonstrating the existence of difference between two groups.

In the initial examination using CAD-NBI, 15 out of 1698 patients (0.88% [95% CI, 0.50% to 1.45%]) were diagnosed with 19 non-neoplastic lesions. Specifically, when using NBI alone, 28 out of 1698 patients (1.65% [CI, 1.10–2.37%]) were diagnosed with 31 non-neoplastic lesions. The relative detection ratio was 0.54 (CI, 0.29–1.00; $P = 0.07$). The combined number of patients diagnosed with non-neoplastic lesions in the first and second examinations was 41 (45 lesions) in the CAD-first group and 49 (54 lesions) in the routine-first group, with no difference between the groups (Supplementary Table 1).

The characteristics of the neoplastic lesions detected in the first examination of the respective comparison groups are shown in Table 3. Lesions less than 10 mm accounted for 41.38% of the lesions detected in the first examination in the routine-first group and 45.61% in the CAD-first group. The proportion of lesions located in the middle thoracic and lower thoracic esophagus was 31.03% and 62.07% in the routine-first group, and 26.3% and 57.89% in the CAD-first group. Among the detected neoplastic lesions, one case was adenocarcinoma in the CAD-first group. As for squamous cell lesions, LGIN, HGIN and carcinoma in the routine-first group were 6.90%, 34.48%, 58.62%, respectively, whereas in the CAD-first group were 14.04%, 43.86% and 40.35%, respectively. In the routine-first group, 12 precancerous lesions (LGIN and HGIN) and 17 carcinomas were detected, while 33 precancerous lesions and 24 carcinomas were detected in the CAD-first group. The percentage of precancerous lesions diagnosed in the first examination by CAD-NBI was higher than that by NBI alone (1.94% vs. 0.71%, $P = 0.0016$). However, the detection rate of carcinoma was similar with no difference between the groups (1.41% vs. 1.00%, $P = 0.2714$).

The secondary outcomes

Supplementary Table 2 compares the secondary outcomes in the CAD-first group and the routine-first group. The FNR of CAD-first group was 8.62%, and it was significantly lower than that in the routine-first group (vs 51.79%, $P < 0.0001$). In terms of precision and accuracy, the values of the CAD-first group were significantly higher than that in the NBI group.

There was a total of 19 false alarms (false positives) in the CAD-first group, including 17 cases of squamous epithelium, 1 case of inflammation and 1 case of heterotopic gastric mucosa. Of all the detected neoplastic lesions in CAD-first group, 7 lesions, including 3 cases of LGIN and 4 cases of HGIN, were missed by the automatic detection system. No carcinoma was missed by the CAD system. Supplementary Table 3 shows the detailed clinical characteristics of the missed lesions. Among the 7 missed lesions, 6 lesions were less than 10 mm in size, and the circumference was less than 1/4. Six lesions were type 0–IIb macroscopic type, and one lesion was type 0–IIc. Five cases were examined by junior endoscopists and 2 cases by midlevel endoscopists.

The median procedure times of routine-first and CAD-first for the first esophageal examination were 25 s (IQR, 19–31 s) and 28 seconds (IQR, 20–40 s), respectively. These differences were statistically significant ($P < 0.0001$). We analyzed the procedure time for the first esophageal

Fig. 1 | The real-time working of CAD-NBI system. The system monitor was positioned fixedly adjacent and parallel to the original endoscopy monitor, achieving real-time assistance during examination of esophagus. NBI narrow-band imaging, CAD computer-aided detection.
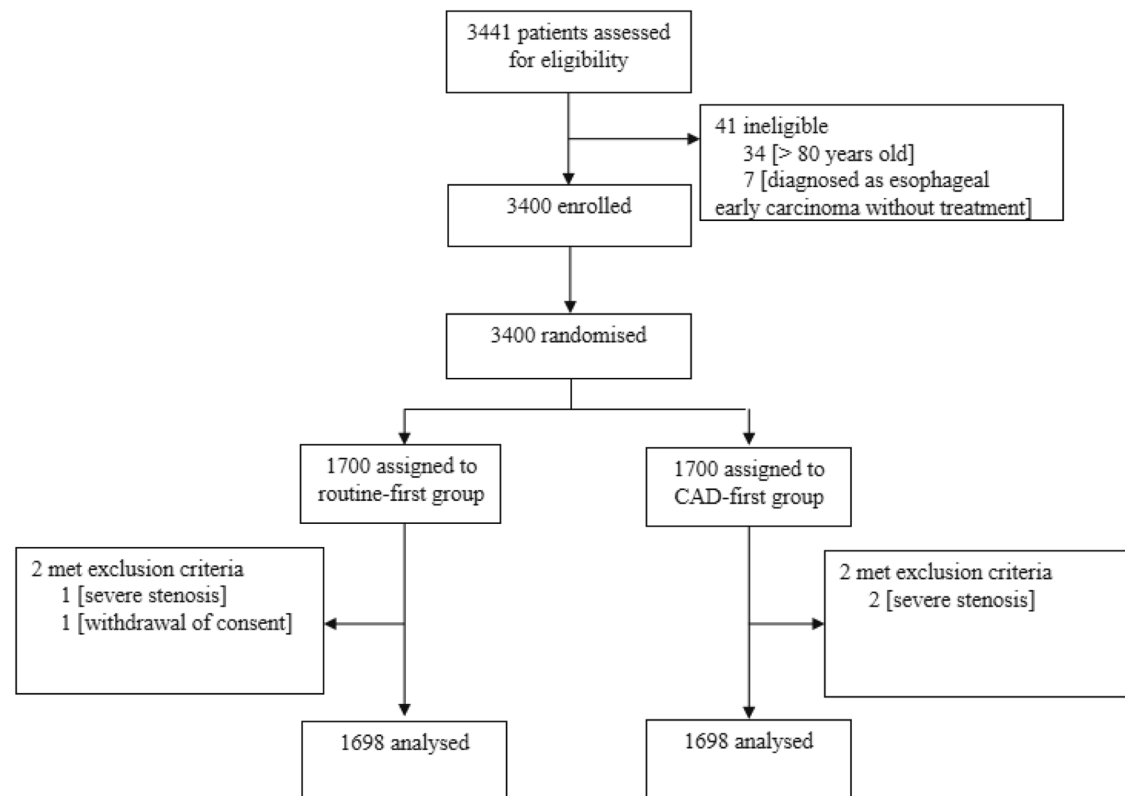
Fig. 2 | Trial profile. A total of 3400 patients were randomly assigned to the routine-first group or CAD-first group, of whom 3396 (routine-first group: 1698; CAD-first group: 1698) were eligible for analysis. NBI narrow-band imaging, CAD computer-aided detection.

Table 1 | Baseline characteristics of patients

Characteristic	Routine-first group (n = 1698)	CAD-first group (n = 1698)
Age, mean (SD), years	54.61 (14.17)	54.41 (14.07)
BMI, mean (SD), kg/m²	23.23 (3.24)	23.32 (3.49)
Gender		
Female, n (%)	933 (54.95)	897 (52.83)
Male, n (%)	765 (45.05)	801 (47.17)
Surgical history		
Upper GI tract, n (%)	167 (9.84)	172 (10.13)
Others, n (%)	510 (30.04)	506 (29.80)
Malignant tumor history		
Esophagus, n (%)	24 (1.41)	20 (1.18)
Head and neck, n (%)	1 (0.06)	0 (0)
Others, n (%)	141 (8.30)	148 (8.72)
Family history of malignant tumor		
Esophageal cancer, n (%)	17 (1.00)	13 (0.77)
Head and neck, n (%)	4 (0.24)	1 (0.06)
Indication		
Screening, n (%)	828 (48.76)	813 (47.88)
Symptomatic, n (%)	870 (51.24)	885 (52.12)

Data are mean (SD) or n (%).
CAD computer-aided detection, *BMI* body-mass index.

examination among different groups of endoscopists, as shown in Supplementary Table 4. In the routine-first group, the procedure time for the first esophageal examination performed by junior endoscopists was significantly longer than that of midlevel and senior endoscopists. In the CAD-first group, there was no significant difference in examination time among the three groups of endoscopists.

There were no serious adverse events related to examination in both groups.

Outcomes in endoscopists with different levels of expertise

Considering the experience of the operator plays a critical role in lesions detecting, a subgroup analysis was performed to evaluate the esophageal neoplastic lesions detected rate during endoscopy according to the experience level of the endoscopists (Table 4). The chance of detecting neoplastic lesions was clearly higher with NBI-CAD than NBI alone for junior endoscopists (RR = 3.58, 95% CI, 1.63–7.87). However, the detection rate of NBI-CAD with senior endoscopists was similar to that of NBI alone (RR = 0.97, 95% CI, 0.43–2.16).

Discussion

Endoscopic screening has become a key factor in early diagnosis and treatment of esophageal cancer, as well as in improving patient survival[17]. Despite the significant improvement in diagnostic techniques under endoscopy, such as narrow-band imaging, magnifying endoscopy, and chromoendoscopy, there are still several issues in early recognition of esophageal cancer[18–20]. First, early-stage esophageal cancer and precancerous lesions are often subtle and have unclear morphological differences from surrounding normal tissues, resulting in a challenging diagnosis[21]. Second, the diagnostic ability of endoscopists varies significantly. Junior physicians may not have enough experience in distinguishing normal esophageal mucosa from early lesions, leading to inaccurate judgments on lesion nature and range, influenced by various factors such as physician's operation skills, clinical experience, and knowledge reserves[8,22]. Finally, visual errors, visual and brain fatigue during endoscopic procedures, may lead to misdiagnosis or missed diagnosis[23].

AI provides a possible way to solve the above problems. By combining AI with digestive endoscopy imaging technology, it is possible to help

Table 2 | Neoplastic lesions detected in the primary and secondary modes

Variable	Routine-first group (*n* = 1698)	CAD-first group (*n* = 1698)	Risk ratio (95% CI)	*P*
Detected by primary mode (primary endpoint)				
Patients, *n* (% [95% CI])	27 (1.59 [1.05–2.31])	53 (3.12 [2.35–4.06])	1.96 (1.24–3.10)	0.0047
Lesions, *n*	29 (1.71 [1.15–2.44])	57 (3.36 [2.55–4.33])	–	–
Detected by secondary mode				
Patients, *n* (% [95% CI])	29 (1.71 [1.15–2.44])	5 (0.29 [0.10–0.69])	0.17 (0.07–0.44)	<0.0001
Lesions, *n*	35 (2.06 [1.44–2.86])	7 (0.41 [0.17–0.85])	–	–
Total				
Patients, *n* (% [95% CI])	56 (3.30 [2.50–4.26])	58 (3.42 [2.60–4.39])	1.04 (0.72–1.49)	0.92
Lesions, *n*	64 (3.77 [2.91–4.79])	66 (3.89 [3.02–4.92])	–	–

Data are *n* (%).
CAD computer-aided detection.

Table 3 | Clinicopathologic features of detected neoplastic lesions (only the first procedure in each group)

Variable	Routine-first group (27 patients/29 lesions)	CAD-first group (53 patients/57 lesions)
Size, mm, *n* (%)		
10 mm	12 (41.38)	26 (45.61)
>10–19 mm	11 (37.93)	19 (33.33)
>19 mm	6 (20.69)	12 (21.05)
Location, *n* (%)		
Ce	1 (3.45)	5 (8.77)
Ut	1 (3.45)	4 (7.02)
Mt	9 (31.03)	15 (26.32)
Lt	18 (62.07)	33 (57.89)
Ae	0 (0)	0 (0)
Morphology, *n* (%)		
Unidentified	1 (3.45)	0 (0)
Protruded	14 (48.28)	13 (22.81)
Flat	5 (17.24)	18 (31.58)
Depressed	9 (31.03)	26 (45.61)
Pathological diagnosis, *n* (%)		
LGIN	2 (6.90)	8 (14.04)
HGIN	10 (34.48)	25 (43.86)
Squamous carcinoma	17 (58.62)	23 (40.35)
Adenocarcinoma	0 (0)	1 (1.75)

Data are *n* (%).
CAD computer-aided detection, *Ut* upper thoracic esophagus, *Mt* middle thoracic esophagus, *Lt* lower thoracic esophagus, *Ae* abdominal esophagus, *Ce* cervical esophagus, *LGIN* low grade intraepithelial neoplasia, *HGIN* high grade intraepithelial neoplasia.

physicians to achieve fast and accurate diagnosis. Studies have demonstrated the ability of AI to meet or exceed the performance of human experts as a triage or screening tool for gastrointestinal diseases with endoscopy[24–26]. In this large-scale, randomized, comparative study, we found that, CAD-NBI system from our team can detect neoplastic lesions in the esophagus 1.96 times more frequently than NBI alone. Moreover, the CAD-NBI system had higher specificity for distinguish neoplastic lesions from non-neoplastic ones. Several CAD systems from other teams have been constructed in endoscopic analysis of early EC based on endoscopy images and videos[27,28]. Most of the studies at present are mainly focused on the establishment of AI algorithms based on the still images from retrospective studies of single-center data[29–33]. Although these AI systems have been demonstrated equivalent or superior diagnostic accuracy, sensitivity, and specificity compared to endoscopic experts, their clinical applicability in the actual endoscopic working environment was unknown. Moreover, these AI systems are still in the laboratory stage and have not yet been connected to endoscopic devices and evaluated in real clinical trials. Compared with these published studies, the advantages of our study are as follows: 1) This study was a large-scale prospective randomized controlled trial, where the CAD system was applied to recognize esophageal lesions in clinical practice during live endoscopy. The results confirmed that the real-time CAD-NBI system utilized during upper GI endoscopy led to a substantial improvement in the detection rate of esophageal neoplasms. Applying AI systems to clinical practice by conducting prospective clinical studies is an important step in the verification of this system and laying the foundation for future large-scale clinical applications. 2) Our CAD system is developed based on an end-to-end trainable multi-task neural network. By using a multi-threaded processing system, the system achieved a maximum processing rate of 135 frames per second in real-time video analysis, achieving real-time lesion detection. 3) Further analysis of the results in this study showed junior endoscopists, compared with senior endoscopists, could benefit more from the assistance of the CAD system. This finding suggests that the CAD system can help to homogenize the performance of endoscopists by overcoming the differences among endoscopists with varying experience levels. This is a new finding that has not been shown in a recently published clinical trial from Sichuan, China[34]. Compared with another similar clinical study from Zhejiang, China[35], our study design has some advantages. Our study was designed as a tandem randomized controlled trial. A same-day tandem upper gastrointestinal endoscopy including two examinations was done for each eligible patient. The lesions found in the first examination were the primary result, and the overlooked lesions in the first examination could be recognized in the second examination. Moreover, all participants in both groups could benefit from the CAD system.

According to the findings of this study, the CAD group exhibited a significant increase in detected neoplastic lesions in the esophagus compared to the control group. Through the analysis of the pathological results, we discovered that the enhanced detection rate of neoplastic lesions was primarily due to the increased identification of both LGIN and HGIN. The CAD system predominantly detected most cases of intraepithelial neoplasia, supporting the conventional understanding that identifying intraepithelial neoplasia, particularly in flat lesions, is often challenging during endoscopy. Although LGIN may regress or disappear, they can progress to HGIN and ESCC[36,37]. Therefore, the overall increase in the detection rate of both LGIN and HGINs may ultimately contribute to a reduced risk of esophageal cancer, which represents the primary objective of any screening endoscopy[38].

In the CAD-first group, there were 7 missed lesions including 3 cases of LGIN and 4 cases of HGIN, although no cancer lesions were missed. Based

Table 4 | Detected outcomes by the primary modes in different groups of operators

Variable	Routine-first group (n = 1698)	CAD-first group (n = 1698)	Risk ratio (95% CI)	P
Junior, n (%)	815 (48.00)	830 (48.88)	–	0.61
Patients, n (%)	8 (0.98)	26 (3.13)	3.58 (1.63–7.87)	0.0012
Lesions, n	8 (0.98)	27 (3.25)	–	–
Midlevel, n (%)	572 (33.69)	574 (33.80)	–	0.94
Patients, n (%)	7 (1.22)	16 (2.79)	2.28 (0.94–5.49)	0.09
Lesions, n	9 (1.57)	17 (2.96)	–	–
Senior, n (%)	311 (18.32)	294 (17.31)	–	0.45
Patients, n (%)	12 (3.86)	11 (3.74)	0.97 (0.43–2.16)	1
Lesions, n	12 (3.86)	13 (4.42)	–	–

Data are n (%).
CAD computer-aided detection.

on the characteristics of the missed cases, we identified two factors contributing to the missed detections of LGIN and HGIN lesions: 1) subtlety of lesion features including small size (less than 10 mm) and type 0–IIb; 2) low-quality operation of junior endoscopists. The subtle and atypical features of lesions pose challenges for CAD systems, especially in distinguishing them from benign conditions. Low-quality operation can hinder the AI's ability to identify lesions. Meticulous examination is a fundamental requirement for endoscopic screening and serves as a prerequisite for accurate diagnosis of lesions. In light of these findings, we are actively working on reducing the missed diagnosis rate of the CAD system by: 1) incorporating additional training data that includes a wider variety of small and atypical cases to better train the CAD system to recognize these lesions; 2) adopting more advanced AI algorithms to improve the feature extraction capabilities of subtle lesions; 3) adding endoscopic quality control functions to existing CAD systems. Although the CAD system can provide valuable assistance to endoscopists, the knowledge and meticulous observation skills of endoscopists remain crucial in the diagnostic process.

For the non-neoplastic lesions, the detection rate was lower in the CAD group than that in the control group, which showed the specificity of the CAD group was superior to that of the control group. The advantage of high specificity is to avoid excessive endoscopic biopsy while improving diagnostic accuracy. Unnecessary biopsies would lead to an extremely high volume of work, resulting in higher medical costs, increased risk of complications, and longer procedure times[39]. As we reported before, CAD-NBI showed a specificity of 96.7% to early ESCC using images. There were 19 non-neoplastic lesions recognized by CAD-NBI system in this study, and the false positive rate was acceptable.

To assess the different clinical application value of the CAD system, endoscopists of varying experience participated in this study, and half of them were junior endoscopists. In most endoscopy centers in China, the number of experienced endoscopists is smaller than that of junior endoscopists. This is also the case in the center where this study was conducted. Thus, we invited two senior endoscopists, three mid-level endoscopists, and five junior endoscopists to participate in the study. Several studies on the validation of AI-assisted diagnosis systems have also similar study designs, where the number of junior doctors is higher than that of senior doctors[40–42]. The distribution of endoscopists in this study, with a higher proportion of junior and less experienced endoscopists, is more aligned with the purpose of this study and reflects the real-world composition of endoscopists in clinical settings.

The subgroup analysis was performed to evaluate the detected rate of esophageal neoplastic lesions during endoscopy according to the experience level of the endoscopists. We conducted a subgroup analysis according to the experience level of the endoscopists for the following reasons: 1) Based on previous research[14,15,43], CAD systems may provide significant assistance to less experienced endoscopists. 2) Subgroup analysis is widely used in clinical trials to identify specific populations that are more likely to benefit from the intervention. This approach allows us to maximize information from clinical trials, generate new hypotheses based on the results, and inform the design of subsequent trials. 3) The subgroup analysis results in this study showed that the chance of detecting neoplastic lesions was clearly higher with NBI-CAD than with NBI alone for junior endoscopists, while the detection rate with NBI-CAD for senior endoscopists was similar to that of NBI alone. This confirms that the CAD system provides most useful references for non-expert endoscopists in real-time clinical diagnosis and facilitates further inspection of potentially malignant esophageal lesions. We believe that the main significance of using the CAD system is to homogenize the performance of endoscopists by overcoming the differences among endoscopists with varying experience levels. In clinical practice, the effectiveness of high-definition and imaging-enhanced endoscopy may be compromised by the insufficient number of well-trained endoscopists. In contrast, the CAD system does not require additional training and has been shown to enhance the performance of non-expert endoscopists, approximating the level of expertise. Therefore, the CAD system may offer particular advantages in areas with limited resources, improving the quality of endoscopic services and patient outcomes.

It is generally believed that a longer endoscopic examination time leads to a higher detection rate of lesions. The median esophageal examination time in both groups was shorter than the recommended value, as we did not impose restrictions on the endoscopy time in this study. Chinese guideline recommends a minimum upper gastrointestinal endoscopy duration of 7 min, with at least 3 min allocated for esophageal observation[44]. However, currently in China, there is a shortage of endoscopists and a large number of patients requiring examination. Under these circumstances, it is challenging to adhere to the recommended observation time. In a similar study recently published from Sichuan, China, the esophageal examination time was also shorter than the recommended[34]. Nevertheless, we believe that this factor does not impact our main findings. First, the detection rates of esophageal neoplasm in both groups were not lower than previously reported[45], indicating that esophageal examination time did not affect the detection rate significantly, lending credibility to our findings. Moreover, the actual time difference for the observation of the esophagus between the CAD-first group and the routine-first group was only 3–5 s. This difference was not clinically significant. We believe that the high detection rate of the CAD-first group is not due to the extra few seconds, but the contribution of the CAD system. Therefore, the main finding of this study is that our CAD system assists endoscopists in detecting lesions even with a relatively short esophageal observation time. In the future, we will further validate the performance of this CAD system in other centers at standard observation times and continue to improve it.

This study has several limitations. Firstly, the endoscopists were aware of the group allocation in this study. Future studies could incorporate a double-blind design to precisely determine the contribution of this system to the increased detection rate of esophageal neoplasms. Secondly, the study's limited external validity stems from its single-center nature. Further investigations should be conducted to explore the adaptability and effectiveness of this system in multiple centers across China and other countries.
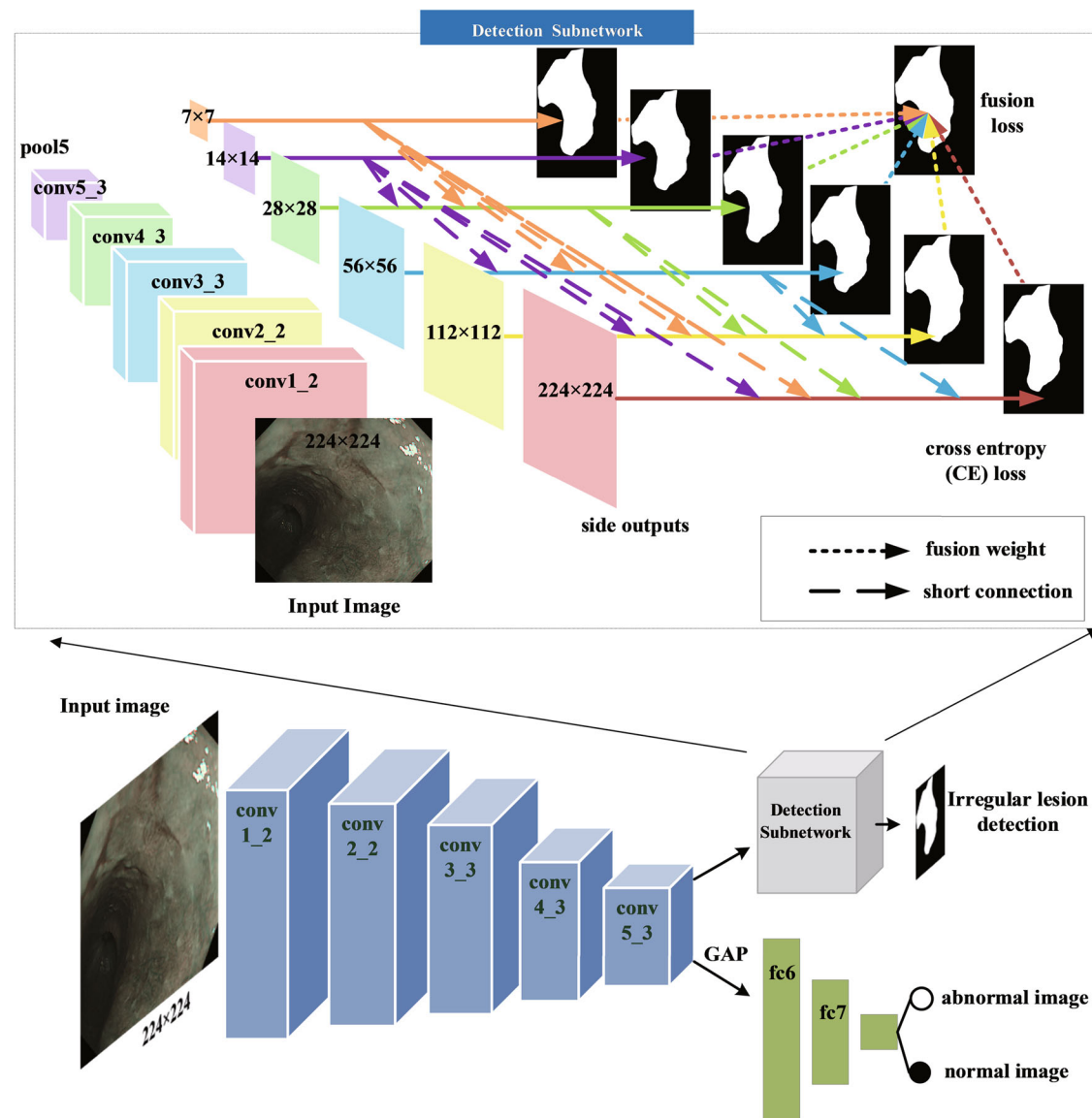
Fig. 3 | Overview of the proposed end-to-end trainable multi-task architecture based on deep learning. The input image is processed through two branches, producing recognition and detection results. GAP, or global average pooling, is used as a standard deep learning operation. The detection subnetwork's architecture is illustrated above. An endoscopic image passes through a six-layer deep network, generating six prediction results at different scales, which are then fused to produce the final prediction.

Thirdly, this study did not control for the fatigue levels of the participating endoscopists, which may independently affect the primary outcome. Future studies should examine the effectiveness of this CAD system across different levels of fatigue.

In conclusion, this study demonstrates that the real-time CAD-NBI system utilized during upper GI endoscopy led to a substantial improvement in the detection rate of esophageal neoplasms. The CAD-NBI system, known for its high accuracy, has the potential for practical application in current clinical practice for esophageal cancer screening. Further investigation is required to determine the efficacy of this technique in diverse populations from other centers.

Methods
Real-time computer-aided detection system for esophageal neoplasm
The real-time computer-aided detection system for esophageal neoplasm was developed on a deep learning architecture. In a preliminary study from our cooperative group, the algorithm was validated and found to have a per-image sensitivity of 91.0%, per-image specificity of 96.7% under the narrow-band imaging model. Moreover, the area under receiver operating characteristic curve was 0.976. We try to improve the diagnostic level of endoscopists by designing an advanced AI algorithm that is developed and evaluated on large-scale multi-centers endoscopy images[16]. Specifically, the CAD system is developed based on an end-to-end trainable multi-task neural network. The network allows precancerous lesion in esophagus to be detected and recognized simultaneously by only feeding an esophagoscope image, as shown in Fig. 3. The processing procedure is that the input image is inputted into two branches from left to right, and then outputs recognition and detection results. The GAP represents global average pooling, which is a common operation in deep learning. The detailed architecture of the detection subnetwork is shown above. An endoscopic image is inputted into the deep network with six layers. Then, these six layers output six prediction results with different scales. Finally, the final prediction result is obtained by fusing these six prediction results. Because both recognition and detection tasks are required to understand the lesion characteristics in endoscopic images, these two tasks have a close correlation. By incorporating these two tasks into the same network, these two tasks promote each other and help to achieve better recognition and detection performance. We also developed a

hardware system based on the proposed algorithm in order to provide real-time video assistance during clinical screening. By using a multithreaded processing system, the system achieved a maximum processing rate of 135 frames per second in real-time video analysis. Additionally, the detection delay was hardly noticeable for endoscopists.

Study design and participants

This prospective study was designed as a single-center, parallel (1:1), randomized, controlled, tandem trial to investigate the impact of an automatic computer-aided detection system acting as an assistant to the endoscopist on the diagnosis of esophageal neoplasm. This study was conducted at the Endoscopy Center of Zhongshan Hospital (Xuhui Hospital) of Fudan University, Shanghai, China.

This study enrolled patients aged 18–80 years who presented for diagnostic or screening upper GI endoscopy and agreed to participate in the study. The exclusion criteria were as follows: (i) confirmed diagnosis of esophageal neoplastic lesion; (ii) needing emergency upper GI endoscopy; (iii) high ASA grade greater than 3; (iv) during pregnancy or lactation; (v) esophageal or pharyngeal stricture; (vi) patients and families unable to understand enough to give their informed consent; and (vii) other reasons that made the subject ineligible to participate in this study, at the discretion of the chief investigator.

The study protocol was approved by the Institutional Review Board (approval no. 2021-077) of Zhongshan Hospital (Xuhui Hospital) of Fudan University. All eligible patients provided written informed consent to participate before the endoscopy examination. It was clearly stated in the informed consent document that the participating patients were free to withdraw their consent at any time. We followed the Consolidated Standards of Reporting Trials (CONSORT) reporting guideline[46]. This trial is registered with Chinese Clinical Trial Registry (http://www.chictr.org.cn), ChiCTR2100050654. The registration date is September 1, 2021.

Randomization

Recruited eligible subjects were randomized in a 1:1 ratio to undergo upper GI endoscopy examination with either CAD-NBI first (followed immediately by routine NBI endoscopy; CAD-first group) or routine NBI first (followed immediately by CAD-NBI endoscopy; routine-first group) without stratification or dynamic allocation. An independent member is responsible for the generation of the randomization scheme using a predefined seed from the statistical software R. Endoscopists and patients were not blinded to the randomization status and group allocation, whereas pathologists were masked to group allocation. All study analyses were done in a blinded fashion.

Procedures

Ten physicians from the endoscopy center participated in the study, including two senior endoscopists (>10,000 NBI endoscopy examinations), three mid-level endoscopists (between 2500 and 5000 NBI endoscopy examinations) and five junior endoscopists (between 1000 and 2500 NBI endoscopy examinations). High-resolution upper gastrointestinal endoscopy systems (GIF-HQ290, Olympus Medical Systems, Tokyo, Japan) were used in this study. We used an NBI system according to the standard methods.

A same-day tandem upper gastrointestinal endoscopy was done for each eligible patient by the same endoscopists. Regarding the examination procedures, the esophagus was examined for the presence or absence of lesions, using routine NBI followed by CAD-NBI (the routine-first group) or CAD-NBI followed by routine NBI (the CAD-first group), according to randomization. Endoscopists were assisted by a research assistant, who turned on or off the CAD-NBI system between the two examinations. The specific examination procedure for the esophagus in the CAD-first group was to 1) examine the esophagus using CAD-NBI model by endoscopists; 2) record the findings and detailed information for target lesions by a research assistant immediately; 3) close the CAD system, examine the esophagus with routine NBI by endoscopists; 4) record the findings and target lesions

unidentified in the first examination; 5) perform a biopsy of any targeted neoplastic lesions in the esophagus detected steps 1 and 3. Examinations in the routine-first group were performed by reversing the order of CAD-NBI and routine NBI. In CAD-assisted examinations, targeted lesions were defined as focal lesions marked by the CAD-NBI system. When the CAD-NBI system did not stably mark lesions and generally were present for less than 1 s, endoscopists could choose to ignore them on the basis of their judgment. In routine NBI examinations, targeted lesions were recognized by the endoscopists according to the characteristics of suspected areas. The lesions detected by secondary mode did not include the lesions detected by the primary mode. Other locations were examined according to the standard endoscopy examination procedures. There was no restriction on inspection time in both examinations. During the procedure, we measured the observation time from start to finish of examination at esophagus by the first method. These procedure times included the evaluation of the lesion but not the biopsy procedure.

Three experienced GI pathologists who were blinded to the recorded endoscopic assessment conducted histologic assessments in the pathology departments. Two pathologists first evaluated the histologic images independently. The type of lesion was determined when the initial diagnosis of the 2 pathologists was the same. In the case of disagreement, the third pathologist conducted further evaluation, and together they made the final decision. Histologic diagnoses were made according to WHO criteria and were classified into two groups: neoplastic lesions and non-neoplastic lesions[47]. The neoplastic lesions group included low-grade dysplasia (LGD), high-grade dysplasia (HGD), or EC; the non-neoplastic lesions group included parakeratosis, inflammation, squamous epithelium, heterotopic gastric mucosa, gastric metaplasia, intestinal metaplasia without dysplasia, and others according to published standards.

Outcomes

The primary endpoint was the diagnosis of one or more neoplastic lesions in the esophagus by the primary method in the tandem examination. The secondary outcomes were the rate of false negatives (FNR) and the rate of false positives (FPR). False negatives were neoplastic lesions overlooked by the first method (that is, neoplastic lesions detected by the second examination but not by the first). False positives were non-neoplastic lesions recognized by the first method (that is, lesions detected by the first examination but confirmed as non-neoplastic by the histologic diagnoses). We used the histologic diagnosis from a biopsy specimen as the gold standard diagnosis.

Statistical analysis

We calculated the minimum sample size using a traditional 2-sample proportion test. The detection rate in endoscopic screening for esophageal neoplastic lesions (squamous dysplasia and carcinoma) is approximately 4.5–5.1% in a high-incidence area (not including Shanghai) of China[45,48]. We estimated that the CAD system would at least double the detection yield for esophageal neoplastic lesions compared with NBI alone in once examination. After the research group discussion, we considered 1.5% to be the detection rate by endoscopy under routine NBI model in once examination, and the detection rates in the CAD-first group were assumed to be 3%. With a significance level of 0.05 (2-sided) and power of 80%, the sample size adequate to detect significant differences was planned to be 1700 patients per group, or 3400 in total.

Statistical analysis was done with R version 4.2.2. Comparison of baseline clinical and demographic characteristics between the CAD-first and the routine-first group was performed using the χ2 test for categorical variables and using the two-sample t-test for continuous variables. The primary analysis was per-protocol, which included all randomized patients excluding those who dropped out or who were subsequently found to meet exclusion criteria. The primary analysis compared the percentage of patients diagnosed with a neoplastic lesion in esophagus in the first examination using either routine NBI or CAD-NBI system (that is, incidence of neoplastic lesion). The denominator is the number of patients in each group,

and the numerator is the number of patients diagnosed with 1 or more neoplastic lesions. The risk ratio was obtained by division of the incidence in the 2 groups, and the 95% CI values were calculated. The comparison of the primary endpoint between two groups was also made using the Cochrane–Mantel–Haenszel $\chi2$ test, stratified by the factor for the experience of the operators. A two-sided *P* value of 0.05 was used as the threshold for significance.

Data availability

Code availability

References

1. Sung, H. et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
2. Mwachiro, M. M. et al. Esophageal Squamous Dysplasia is Common in Asymptomatic Kenyans: A Prospective, Community-Based, Cross-Sectional Study. *Am. J. Gastroenterol.* **111**, 500–507 (2016).
3. Taylor, P. R., Abnet, C. C. & Dawsey, S. M. Squamous dysplasia-the precursor lesion for esophageal squamous cell carcinoma. *Cancer Epidemiol. Biomark. Prev.* **22**, 540–552 (2013).
4. Muto, M. Endoscopic diagnostic strategy of superficial esophageal squamous cell carcinoma. *Dig. Endosc.* **25**, 1–6 (2013).
5. di Pietro, M., Canto, M. I. & Fitzgerald, R. C. Endoscopic Management of Early Adenocarcinoma and Squamous Cell Carcinoma of the Esophagus: Screening, Diagnosis, and Therapy. *Gastroenterology* **154**, 421–436 (2018).
6. Ohashi, S. et al. Recent Advances From Basic and Clinical Studies of Esophageal Squamous Cell Carcinoma. *Gastroenterology* **149**, 1700–1715 (2015).
7. Wang, W. L. et al. A training program of a new simplified classification of magnified narrow band imaging for superficial esophageal squamous cell carcinoma. *J. Gastroenterol. Hepatol.* **33**, 1248–1255 (2018).
8. Ishihara, R. et al. Prospective evaluation of narrow-band imaging endoscopy for screening of esophageal squamous mucosal high-grade neoplasia in experienced and less experienced endoscopists. *Dis. Esophagus* **23**, 480–486 (2010).
9. Uche-Anya, E., Anyane-Yeboa, A., Berzin, T. M., Ghassemi, M. & May, F. P. Artificial intelligence in gastroenterology and hepatology: how to advance clinical practice while ensuring health equity. *Gut* **71**, 1909–1915 (2022).
10. Reyna, M. A., Nsoesie, E. O. & Clifford, G. D. Rethinking Algorithm Performance Metrics for Artificial Intelligence in Diagnostic Medicine. *JAMA* **328**, 329–330 (2022).
11. de Groof, A. J. et al. Deep-Learning System Detects Neoplasia in Patients With Barrett's Esophagus With Higher Accuracy Than Endoscopists in a Multistep Training and Validation Study With Benchmarking. *Gastroenterology* **158**, 915–929 e914 (2020).
12. Luo, H. et al. Real-time artificial intelligence for detection of upper gastrointestinal cancer by endoscopy: a multicentre, case-control, diagnostic study. *Lancet Oncol.* **20**, 1645–1654 (2019).
13. Sharma, P. & Hassan, C. Artificial Intelligence and Deep Learning for Upper Gastrointestinal Neoplasia. *Gastroenterology* **162**, 1056–1066 (2022).
14. Cai, S. L. et al. Using a deep learning system in endoscopy for screening of early esophageal squamous cell carcinoma (with video). *Gastrointest. Endosc.* **90**, 745–753 e742 (2019).
15. Li, B. et al. Comparative study on artificial intelligence systems for detecting early esophageal squamous cell carcinoma between narrow-band and white-light imaging. *World J. Gastroenterol.* **27**, 281–293 (2021).
16. Lin, Q. et al. Lesion-Decoupling-Based Segmentation With Large-Scale Colon and Esophageal Datasets for Early Cancer Diagnosis. *IEEE Trans. Neural Netw. Learn. Syst.* (IEEE, 2023).
17. 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).
18. Gruner, M. et al. Narrow-band imaging versus Lugol chromoendoscopy for esophageal squamous cell cancer screening in normal endoscopic practice: randomized controlled trial. *Endoscopy* **53**, 674–682 (2021).
19. Mizumoto, T. et al. Magnifying Endoscopy with Narrow Band Imaging in Estimating the Invasion Depth of Superficial Esophageal Squamous Cell Carcinomas. *Digestion* **98**, 249–256 (2018).
20. Safatle-Ribeiro, A. V. et al. Diagnostic accuracy of probe-based confocal laser endomicroscopy in Lugol-unstained esophageal superficial lesions of patients with head and neck cancer. *Gastrointest. Endosc.* **85**, 1195–1207 (2017).
21. Shimizu, M., Zaninotto, G., Nagata, K., Graham, D. Y. & Lauwers, G. Y. Esophageal squamous cell carcinoma with special reference to its early stage. *Best. Pr. Res. Clin. Gastroenterol.* **27**, 171–186 (2013).
22. Rodriguez de Santiago, E. et al. Rate of missed oesophageal cancer at routine endoscopy and survival outcomes: A multicentric cohort study. *U. Eur. Gastroenterol. J.* **7**, 189–198 (2019).
23. Chan, M. Y., Cohen, H. & Spiegel, B. M. Fewer polyps detected by colonoscopy as the day progresses at a Veteran's Administration teaching hospital. *Clin. Gastroenterol. Hepatol.* **7**, 1217–1223 (2009).
24. Barua, I. et al. Artificial intelligence for polyp detection during colonoscopy: a systematic review and meta-analysis. *Endoscopy* **53**, 277–284 (2021).
25. Wu, L. et al. Effect of a deep learning-based system on the miss rate of gastric neoplasms during upper gastrointestinal endoscopy: a single-centre, tandem, randomised controlled trial. *Lancet Gastroenterol. Hepatol.* **6**, 700–708 (2021).
26. Wang, P. et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. *Gut* **68**, 1813–1819 (2019).
27. Guo, L. et al. Real-time automated diagnosis of precancerous lesions and early esophageal squamous cell carcinoma using a deep learning model (with videos). *Gastrointest. Endosc.* **91**, 41–51 (2020).
28. Ebigbo, A. et al. Real-time use of artificial intelligence in the evaluation of cancer in Barrett's oesophagus. *Gut* **69**, 615–616 (2020).
29. Yang, X. X. et al. Real-time artificial intelligence for endoscopic diagnosis of early esophageal squamous cell cancer (with video). *Dig. Endosc.* **33**, 1075–1084 (2021).
30. Shiroma, S. et al. Ability of artificial intelligence to detect T1 esophageal squamous cell carcinoma from endoscopic videos and the effects of real-time assistance. *Sci. Rep.* **11**, 7759 (2021).
31. Yuan, X. L. et al. Artificial intelligence for detecting superficial esophageal squamous cell carcinoma under multiple endoscopic imaging modalities: A multicenter study. *J. Gastroenterol. Hepatol.* **37**, 169–178 (2022).
32. Liu, W. et al. Artificial Intelligence for Detecting and Delineating Margins of Early ESCC Under WLI Endoscopy. *Clin. Transl. Gastroenterol.* **13**, e00433 (2022).
33. Yuan, X. L. et al. Artificial intelligence for detecting and delineating the extent of superficial esophageal squamous cell carcinoma and precancerous lesions under narrow-band imaging (with video). *Gastrointest. Endosc.* **97**, 664–672 e664 (2023).

34. Yuan, X. L. et al. Effect of an artificial intelligence-assisted system on endoscopic diagnosis of superficial oesophageal squamous cell carcinoma and precancerous lesions: a multicentre, tandem, double-blind, randomised controlled trial. *Lancet Gastroenterol. Hepatol.* **9**, 34–44 (2024).

35. Li, S. W. et al. Deep learning assists detection of esophageal cancer and precursor lesions in a prospective, randomized controlled study. *Sci. Transl. Med.* **16**, eadk5395 (2024).

36. Shimizu, M., Nagata, K., Yamaguchi, H. & Kita, H. Squamous intraepithelial neoplasia of the esophagus: past, present, and future. *J. Gastroenterol.* **44**, 103–112 (2009).

37. Li, H. et al. Long-term Incidence Rates of Esophageal Squamous Cell Carcinoma in Chinese Patients With Low-grade Intraepithelial Neoplasia and Association of Surveillance Endoscopy With Incidence. *JAMA Netw. Open* **5**, e2247415 (2022).

38. Wei, W. Q. et al. Esophageal Histological Precursor Lesions and Subsequent 8.5-Year Cancer Risk in a Population-Based Prospective Study in China. *Am. J. Gastroenterol.* **115**, 1036–1044 (2020).

39. Oude Nijhuis, R. A. B. et al. Utility of Routine Esophageal Biopsies in Patients With Refractory Reflux Symptoms. *Am. J. Gastroenterol.* **116**, 816–820 (2021).

40. Yao, L. et al. Development and validation of an artificial intelligence-based system for predicting colorectal cancer invasion depth using multi-modal data. *Dig. Endosc.* **35**, 625–635 (2023).

41. Yuan, X. L. et al. Artificial intelligence for diagnosing microvessels of precancerous lesions and superficial esophageal squamous cell carcinomas: a multicenter study. *Surg. Endosc.* **36**, 8651–8662 (2022).

42. Zhou, W. et al. Interpretable artificial intelligence-based app assists inexperienced radiologists in diagnosing biliary atresia from sonographic gallbladder images. *BMC Med.* **22**, 29 (2024).

43. Zhao, Y. Y. et al. Computer-assisted diagnosis of early esophageal squamous cell carcinoma using narrow-band imaging magnifying endoscopy. *Endoscopy* **51**, 333–341 (2019).

44. He, J. et al. [China guideline for the screening, early detection and early treatment of esophageal cancer (2022, Beijing)]. *Zhonghua Zhong Liu Za Zhi* **44**, 491–522 (2022).

45. He, Z. et al. Efficacy of endoscopic screening for esophageal cancer in China (ESECC): design and preliminary results of a population-based randomised controlled trial. *Gut* **68**, 198–206 (2019).

46. Schulz, K. F., Altman, D. G., Moher, D. & Group, C. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Ann. Intern Med* **152**, 726–732 (2010).

47. Nagtegaal, I. D. et al. The 2019 WHO classification of tumours of the digestive system. *Histopathology* **76**, 182–188 (2020).

48. Niu, C. et al. Risk factors for esophageal squamous cell carcinoma and its histological precursor lesions in China: a multicenter cross-sectional study. *BMC Cancer* **21**, 1034 (2021).

Acknowledgements

Author contributions

Y.-S.Z., B.Y. and S.-L.C. had the idea for the study and contributed to study design. B.L., Y.-Y.D., W.-M.T., D.-L.H., Z.-P.Q., Z.R. and Q.S. contributed to acquisition of data. B.L., Y.-Y.D., W.-M.T., H.-H.Y., M.-Y.C., and Y.-S.Z. contributed to analysis and interpretation of the data. B.L., Y.-Y.D., W.-M.T. and D.-L.H. contributed to drafting of the article. Y.-S.Z. and S.-L.C. contributed to critical revision of the article for important intellectual content. All authors read and approved the final report.

Competing interests

The authors declare no competing interests.

Additional information