Effective Opportunistic Esophageal Cancer Screening using Noncontrast CT Imaging

Jiawen Yao[1], Xianghua Ye[2], Yingda Xia[3], Jian Zhou[4], Yu Shi[5], Ke Yan[1], Fang Wang[2], Lili Lin[2], Haogang Yu[2], Xian-Sheng Hua[1], Le Lu[3], Dakai Jin[3], and Ling Zhang[3]

[1]DAMO Academy, Alibaba Group, Beijing, China
[2]The first Affiliated Hospital of Zhejiang University, Hangzhou, China
[3] DAMO Academy, Alibaba Group, New York, USA
[4]Sun Yat-sen University Cancer Center, Guangzhou, China
[5] Shengjing Hospital of China Medical University, Shenyang, China

Abstract. Esophageal cancer is the second most deadly cancer. Early detection of resectable/curable esophageal cancers has a great potential to reduce mortality, but no guideline-recommended screening test is available. Although some screening methods have been developed, they are expensive, might be difficult to apply to the general population, and often fail to achieve satisfactory sensitivity for identifying early-stage cancers. In this work, we investigate the feasibility of esophageal tumor detection and classification (cancer or benign) on the noncontrast CT scan, which could potentially be used for opportunistic cancer screening. To capture the global context, a novel position-sensitive self-attention is proposed to augment nnUNet with non-local interactions. Our model achieves a sensitivity of 93.0% and specificity of 97.5% for the detection of esophageal tumors on a holdout testing set with 180 patients. In comparison, the mean sensitivity and specificity of four doctors are 75.0% and 83.8%, respectively. For the classification task, our model outperforms the mean doctors by absolute margins of 17%, 31%, and 14% for cancer, benign tumor, and normal, respectively. Compared with established state-of-the-art esophageal cancer screening methods, e.g., blood testing and endoscopy AI system, our method has comparable performance and is even more sensitive for early-stage cancer and benign tumor. Our proposed method is a novel, non-invasive, low-cost, and highly accurate tool for opportunistic screening of esophageal cancer.

Keywords: Esophageal Cancer · Cancer Screening · Self-attention · Noncontrast CT

1 Introduction

Esophageal cancer (EC) is the second most deadly cancer, with a 5-year survival rate of only 20% [18], and even less than 5% in many developing countries [2]. This poor survival is mainly due to patients are usually diagnosed at advanced stages with unresectable tumors [2,5,21], because their signs and symptoms tend

to be latent and non-specific [13]. Early-stage disease, however, is associated with a substantially higher 5-year survival rate of 80%–90% [16, 21]. Therefore, the early detection of resectable/curable esophageal cancers (ideally, before symptoms) has a great potential to reduce mortality [21]. Unfortunately, EC has no guideline-recommended screening tests available [1]. Several tools have been developed, and some are implemented in high-risk areas, such as endoscopic techniques [13], Cytosponge procedure [6], and blood-based biomarkers [11, 15]. However, they are difficult to apply to the general population due to moderate sensitivity and high-cost [2, 15]. Novel screening methods that are noninvasive with low-cost, ready to distribute, and highly accurate are eagerly needed.

Routine CT imaging performed for other clinical indications offers an opportunity for opportunistic screening of diseases at no additional cost or radiation exposure to patients. Previous studies show that abdominal and chest CT with or without contrast enhancement provide values for incidental osteoporosis screening [9] and cardiovascular event prediction [14]. For cancer detection, researchers have found that pancreatic cancer could be detected with high accuracy by deep learning from noncontrast CT [22], which has long been thought to be impossible (only detectable from contrast-enhanced CT). However, detection of esophageal cancer on noncontrast CT can be extremely challenging. The early-stage esophageal carcinoma tumors can be very small, invading only lamina propria (stage I) and muscle layer (stage II) [17]. Given the extremely poor contrast between the tumor and normal esophageal tissues in the noncontrast CT (e.g., chest CT), the early-stage tumor detection task is highly challenging. Actually, even on the contrast-enhanced CT, human experts often require substantial effort and expertise to detect early-stage esophageal tumors by referring to several other clinical information, such as endoscopy, endoscopic ultrasound, and FDG-PET; even so, some tiny tumors are still hard to be detected on CT.

So far, studies on deep learning-based esophageal cancer image analysis all focus on the tumor segmentation task by improving the local image feature extraction/modeling [26, 28] or fusion of multi-modal imaging [10, 25] to improve the segmentation accuracy. In this study, we propose the first deep learning-based tool for opportunistic esophageal cancer screening using noncontrast CT – specifically, detecting the esophageal tumor if there exists and then classifying the detection as cancer or benign. As discussed above, the local image texture could be insufficient to detect esophageal tumors in noncontrast CT. In clinical practice, global context features of the esophagus, such as "asymmetric esophageal wall thickening" and "squeezed esophageal wall", are key signs to diagnose esophageal cancer, especially early-stage ones. On the other hand, for deep learning, each convolutional kernel could only attend a local-subset of voxels or local patterns rather than the global context. Therefore, we incorporate global attention layers with positional embedding to enhance the ability to model global context as well as long-range dependencies in 3D medical image segmentation. This design could improve the ability of tumor distinction especially for early stage tumors. We collect a multi-center dataset including two main esophageal tumor types (ESCC and leiomyoma) and normal esophagus

from 741 patients. On the holdout test set, our model achieves an AUC of 0.990, sensitivity of 93.0%, and specificity of 97.5% for tumor detection, surpassing the average sensitivity of 75.0% and specificity of 83.8% of four doctors. The main contributions of this paper can be summarized as follows:

- We present a deep learning method to detect and classify esophageal tumors from noncontrast CT, a novel, non-invasive, low-cost, ready-to-distribute, and highly accurate tool, for screening esophageal cancer.
- The position-sensitive full-attention layer shows its better use of positional information and long-range dependencies in 3D noncontrast CT, therefore, could improve the performance over a strong baseline nnUNet model [8].
- Compared with doctors' reading of noncontrast CT, our automated method shows substantially higher accuracy in both detection and classification. Compared with established state-of-the-art esophageal cancer screening methods, e.g., blood testing [11] and endoscopy AI system [13], our screening tool has comparable performance and is even more sensitive for early-stage cancer and benign tumor.

2 Methods

We aim at a three-class classification problem in noncontrast CT scans. We denote the whole dataset as $S = \{(\mathbf{X_i}, \mathbf{Y_i}, \mathbf{P_i})|i = 1, 2, ..N\}$, where $\mathbf{X_i} \in \mathbb{R}^{H_i \times W_i \times D_i}$ is a 3D CT volume of the i-th patient. $\mathbf{Y_i} \in \mathcal{L}^{H_i \times W_i \times D_i}$ is a voxel-wise annotated label with the same (H_i, W_i, D_i) three dimensional size as $\mathbf{X_i}$ and represents our segmentation targets, *i.e.*, background, esophagus, esophageal cancer, and benign tumor. To obtain tumor annotations in $\mathbf{Y}$ (upper panel in Fig. 1), for those patients who have radiotherapy CT, we directly use their gross tumor volume (GTV) masks. For others, manual tumor annotation is first performed on the contrast-enhanced CT phase, referring to clinical and panendoscopy reports when necessary. Then, a robust image registration method, DEEDS [7], is used to register the annotated mask from the contrast-enhanced CT to the noncontrast CT, followed by a manual correction during quality check. $\mathbf{P_i} \in \mathcal{L}$ is the patient-level label (esophageal cancer, benign, and normal), confirmed by either pathology or radiology reports with follow-up.

Segmentation for Classification with Self-Attention Segmentation for classification is the most straightforward method and has been successfully adopted for the task of tumor detection [3, 22, 24, 29]. In those methods, a localization UNet [4] is first trained to locate the ROI. However, vanilla UNet-based segmentation networks have limited receptive field and heavily rely on local textual patterns rather than the global context. It will definitely affect the later classification task if the first segmentation model fails to segment the target. Therefore, it is important to build a more robust segmentation model that is sensitive to tumors especially early stage tumors. In this paper, we propose an architectural improvement by integrating the global self-attention layer to enhance the model's ability on modeling global context. As shown in Fig. 1, we add self-attention layers after each convolutional layers in the encoder. Let us consider

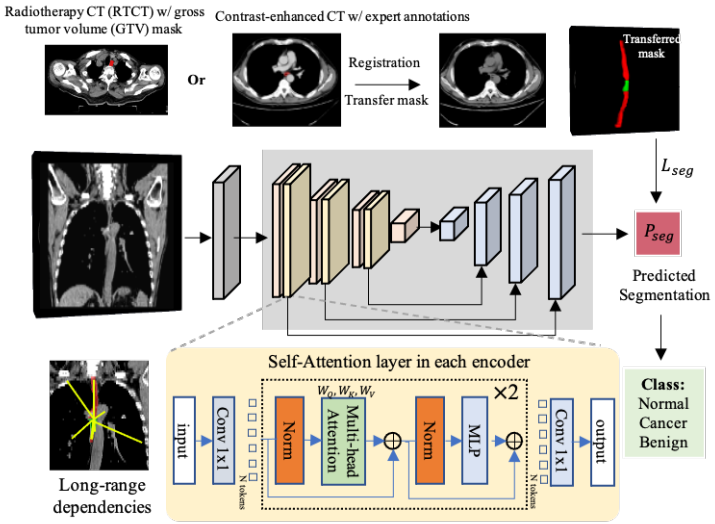

Fig. 1. The main architecture diagram of the proposed model. GTV masks and transferred masks from registration are used as labels to train the model. To capture long-rang dependencies and global context, the position-sensitive self-attention is added to augment convolutional layers with non-local interactions in the encoder block.

an input feature map $x \in \mathbb{R}^{C_{in} \times H \times W \times D}$. The output at position $o = (i, j, k)$, $y_o \in \mathbb{R}^{C_{out} \times H \times W \times D}$ of a self-attention layer is computed by pooling over the projected input as the following:

$$y_o = \sum_{p \in \mathcal{N}} softmax_p(q_o^T k_p) v_p \tag{1}$$

where $\mathcal{N}$ is the whole location lattice, and queries $q_o = W_Q x_o$, keys $k = W_K x_o$ and values $v_o = W_V x_o$ are all projections of the input x_o. W_Q, W_K, W_V are all learnable matrices. The $softmax_p$ denotes a softmax function applied to all possible positions. When applying self-attention to vision problems, one of the main obstacles is that the computational complexity of attention mechanism scales and this is even more of a problem when dealing with 3D data in medical segmentation. Another issue is that self-attention layer does not utilize any positional information when computing the non-local context. Motivated by recent advances applying attention to multi-dimensional data [19, 20], we add local constraints to serve as a memory bank for computing the output y_o which could significantly reduces the original computation of Eq.1. For each location o, a local region $\mathcal{M} = \mathcal{N}_{m_h \times m_w \times m_d}(o)$ is extracted in each self-attention layer. Additionally, a learned relative positional encoding term is introduced and the additional position embedding in query, key and values is shown to capture long-

range interaction with precise positional information. The updated self-attention with positional encoding given input feature map x can be written as

$$y_o = \sum_{p \in \mathcal{M}} softmax_p(q_o^T k_p + q_o^T r_{p-o}^q + k_p^T r_{p-o}^k)(v_p + r_{p-o}^v) \quad (2)$$

where the learnable r_{p-o}^k and r_{p-o}^v are the positional encoding for keys and values, respectively. We apply multi-head attention to capture a mixture of affinities by computing N single-head attentions in parallel on x_o. The final output is achieved by concatenating the results from each head. Finally, we reshape the tokens and upsample to the original size of the feature map.

Based on the segmentation mask, we classify each 3D volume as one of three target labels, i.e., cancer, benign tumor, or normal. To achieve a explainable classification, we use a simple, non-parametrized approach to give final patient-level decision. At first, we construct a graph on all voxels predicted as normal esophagus and esophageal abnormalities (cancer+benign tumor). We compute all connected components in such graph and only keep the largest connected comment to filter out the outliers. Then, a 3D volume is considered as normal if less than K mm^3 of voxels are predicted as abnormal. K is tuned to achieve a specificity of 99% for the model on the validation set. To further classify the abnormal case as cancer or benign, the label with more segmented voxels is predicted.

3 Experiments

Datasets and Annotation: We build a dataset of CT scans of 741 patients and among them, 481 have esophageal tumors, either cancer or benign (leiomyomas), and the other 260 are normal cases. The dataset is randomly split into a training and a testing set but according to the distribution of cancer stages and tumor types. The resulting training set includes 324 cancer, 57 benign, and 180 normal. The testing set has 80 cancer, 20 benign, and 80 normal. Both cancer and benign cases are confirmed by pathology reports. Normal cases are confirmed by radiology reports and 2 years of follow-up. An experienced radiation oncologist (10 yrs esophageal specialist) manually annotates tumors in the training set (on either radiotherapy CT or contrast-enhanced CT). Esophagus masks are automatically generated by a segmentation model which is trained on a public dataset [12] with a self-learning process [27] on our training set.

Implementation Details: Each CT volume is resampled into $0.8 \times 0.8 \times 3.0$ mm spacing and then normalized into zero mean and unit variance. In self-attention layer, (m_h, m_w, m_d) is set as (12, 12, 6) and N=4 heads is used. During training, extensive data augmentation [8] was applied on the fly to improve the generalization, including random rotation and scaling, elastic deformation, additive brightness, and gamma scaling. The objective for optimization is the sum of binary cross entropy and the dice loss. The networks were optimized with RAdam with the initial learning rate as 0.001 and we set the maximum epoch as 1000 following the nnUNet.

	Two-class			Three-class(%)		
Method	AUC	Sens. (%)	Spec. (%)	Cancer	Benign	Normal
Ours	0.990 (0.986-0.993)	93.1 (91.8-94.4)	97.5 (96.6-98.4)	91.3	60.0	97.5
nnUNet-S4C [8, 29]	0.961* (0.954-0.967)	85.1 (83.3-86.8)	100.0 (100.0-100.0)	90.0	45.0	100.0
LENS (detection) [23]	0.954** (0.947-0.961)	88.1 (86.5-89.6)	96.3 (95.3-97.3)	91.3	15.0	96.3
Mean doctors WOTC	-	75.0	83.8	74.4	28.8	83.8

Table 1. Results on two-class classification: abnormal (esophageal cancer + benign) vs. normal, and three-class classification: esophageal cancer vs. benign vs. normal. WOTC: without time constraint. Sens.: Sensitivity. Spec.: Specificity. *for $p < 0.05$, **for $p < 0.01$.

Evaluation Metrics and Reader Study. Evaluation metrics include AUC, sensitivity, and specificity in the 2-class classification task (abnormal vs. normal), and class accuracy in the 3-class classification task. To measure the performance of tumor localization, we compare the segmentation mask with the ground truth and define the localization is successful if the intersection (Dice score) over the human annotation is larger than 0.1. Four readers, including two radiation oncologists (5 and 14 yr [12 yr esophageal specialist], respectively) and one radiologist (8 yr) from the 1st affiliated hospital of Zhejiang University and one radiologist (13 yr [6 yr esophageal specialist]) from Sun Yat-sen University Cancer Center (SYSUCC), are invited for the reader study. They read the 180 noncontrast CTs in the testing set without time constraints and give a forced three-class decision for each CT: esophageal cancer, leiomyomas, or normal. Patient information and records are not provided. Readers are informed that the dataset might contain more tumor cases than the standard prevalence observed in screening, but the proportion of case types is not informed. The first three readers view and interpret these CTs in the radiation planning viewer used in their daily work environment. The reader from SYSUCC uses ITK-SNAP, with three years of user experience for research purposes.

Comparison with Other Algorithms and Readers. We compared our method with two approaches representing strong segmentation-based and detection-based methods. The segmentation-based one uses the "Segmentation for classification (S4C)" [29] paradigm but its standard U-Net is updated with the more powerful nnUNet [8]. The detection-based one has proven its performances as a competitive universal lesion detector [23]. Standard deviation of the AUCs, Sens. and Spec. values are obtained from 1000 bootstrap replicas of the test dataset. DeLong test is performed for statistical analysis between two AUCs (Ours vs. comparison method). Results can be seen in Table 1 and two illustrative examples are shown in Fig. 2. Most medical segmentation and detection models focus on local texture and structure and thus lack the ability to model the global context. From the table and figure, we could see improved results of our method in finding abnormal patients and detecting benign tumors. The performance of

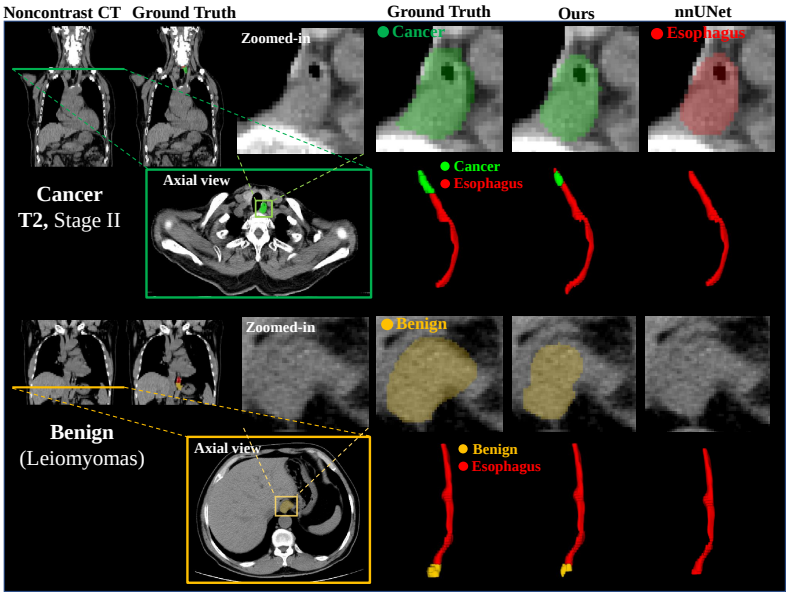

Fig. 2. Two cases (a cancer and a benign) miss-detected by all readers in the test set. Our methods can successfully locate the tumors while nnUNet fails in these cases.

all four doctors is below our model's predictions (ROC curve, Fig. 3 (A)). Our sensitivity in finding tumor (93%) outperforms the best sensitive doctor (79%) by a large margin and also performs better in predicting normal patients than the best specific doctor (Specificity 98% vs. 96%). For the 3-class task, our model perform much better than the mean doctors by absolute margins of 17%, 31%, and 14% for cancer, benign, and normal, respectively (Table 1; confusion matrix, Fig. 3 (B)).

Subgroup Analysis Stratified by Tumor Stage. More specificity, we evaluate performance of our model in benign and malignant tumor, as shown in Table 2. We first report patient-level detection rate values across benign and each T stage cancer. Then we compare predictions with ground-truth annotations to see if predicted tumors are detected correctly. Tumor-level localization evaluates how segmented masks overlap with the ground-truth cancer or leiomyoma regions (Dice > 0.1 is used for correct detection). Radiologist 2 (the best sensitive reader) in finding abnormalities is also compared. A detection is considered successful if the intersection (between the ground truth and segmentation mask) over the ground truth is > 0.1; otherwise, it is considered a misdetection. Our model performs better in detecting benign and T2 stage tumor and providing more accurate tumor location.

Comparison with Established Screening Tools. Compared with a state-of-the-art blood test screening [11], at a similar specificity level, our solution achieves much better results in detecting early-stage esophageal cancer (stage I-

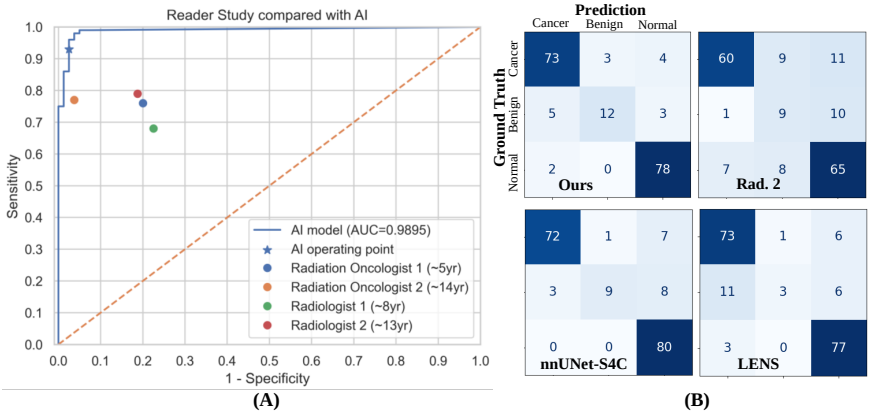

Fig. 3. (A) ROC curve for our model results versus all participated experts' referrals on the test set of $n = 180$ patients for 2-class classification (abnormal vs. normal). (B) Confusion matrices with patient numbers of predictions for our model, the two baseline methods, and Radiologist 2 (the best 3-class accuracy reader).

		Benign	Esophageal cancer			
Method	Criteria	-	T1	T2	T3	T4
Ours	Patient-level	60.0(12/20)	44.4(4/9)	94.1(16/17)	97.1(33/34)	100(20/20)
	Tumor-level	40.0(8/20)	44.4(4/9)	88.2(15/17)	97.1(33/34)	100(20/20)
nnUNet-S4C	Patient-level	45.0(9/20)	55.6(5/9)	82.4(14/17)	97.1(33/34)	100(20/20)
	Tumor-level	35.0(7/20)	44.4(4/9)	70.6(12/17)	97.1(33/34)	100(20/20)
Radiologist 2	Patient-level	45.0(9/20)	55.6(5/9)	64.7(11/17)	76.5(26/34)	90.0(18/20)

Table 2. Patient-level detection and tumor-level localization results over the two types of esophageal abnormalities (benign and cancer).

II) using the similar size of testing patients (Table 3). Moreover, our method can detect benign tumors (leiomyomas, which needs surgery) while the blood test cannot. We also compare our model with a state-of-the-art endoscopy AI system (trained on 15,000 patients data with fully supervision) [13] for upper gastrointestinal cancer detection. Note that the definition of sensitivity and specificity in the endoscopy screening scenario is slightly different from our radiological scenario. To facilitate the comparison, we report the results of cancer vs. (benign + normal) by following [13]. We could see a (slightly) lower (91.3% vs. 94.2%) sensitivity with a slightly higher specificity (92.9% vs 92.3%). In fact, for the 80 cancer cases in our test set, our method detects 76 of them, with three being misclassified as benign, which are leiomyomas, and surgery will be recommended in the noncontrast CT screening scenario. As such, our method only misses four cancer cases, which equals a sensitivity of 95% (76/80) for cancer tumor detection. In contrast, nnUNet-S4C missed seven cancer cases and classified them as normal patients which shows a sensitivity of 91% (73/80), as shown in Fig.3 (B).

	Sensitivity					Specificity
	Benign	Esophageal cancer (TNM Stage)				Normal
Method	-	I	II	III	IV	-
Ours	60.0(12/20)	37.5(3/8)	93.1(27/29)	100.0(21/21)	100.0(22/22)	97.5
Blood Test [11]	N/A	12.5(1/8)	64.7(11/17)	94.1(32/34)	100.0(40/40)	99.5

Table 3. Comparison with a state-of-the-art blood test on esophageal cancer detection.

4 Conclusion

In this paper, we investigate a relatively convenient, simple opportunistic screening solution of esophageal cancer with noncontrast CT scans. To better capture global context and detect early-stage tumors, we propose a position-sensitive self-attention to augment convolutional layers with non-local interactions in the encoder block. We achieve high sensitivity and specificity on a large-scale dataset and outperform the mean doctors by large margins. Compare with other tools like blood test and endoscopy, our work suggests the good feasibility of using noncontrast CT scans as a promising clinical tool for large-scale esophageal cancer opportunistic screening with no extra costs.

References

1. U.S. Preventive Services Task Force (USPSTF), "Recommendations,". `https://www.uspreventiveservicestaskforce.org/uspstf/topic_search_results?topic_status=P`
2. Arnal, M.J.D., Arenas, Á.F., Arbeloa, Á.L.: Esophageal cancer: Risk factors, screening and endoscopic treatment in western and eastern countries. World Journal of Gastroenterology **21**(26), 7933 (2015)
3. Cheng, N.M., Yao, J., Cai, J., Ye, X., Zhao, S., Zhao, K., Zhou, W., Nogues, I., Huo, Y., Liao, C.T., Wang, H.M., Lin, C.Y., Lee, L.Y., Xiao, J., Lu, L., Zhang, L., Yen, T.C.: Deep Learning for Fully Automated Prediction of Overall Survival in Patients with Oropharyngeal Cancer Using FDG-PET Imaging. Clinical Cancer Research **27**(14), 3948–3959 (2021)
4. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. In: MICCAI. pp. 424–432. Springer (2016)
5. Doki, Y., Ajani, J.A., Kato, K., Xu, J., Wyrwicz, L., Motoyama, S., Ogata, T., Kawakami, H., Hsu, C.H., Adenis, A., et al.: Nivolumab combination therapy in advanced esophageal squamous-cell carcinoma. New England Journal of Medicine **386**(5), 449–462 (2022)
6. Gehrung, M., Crispin-Ortuzar, M., Berman, A.G., O'Donovan, M., Fitzgerald, R.C., Markowetz, F.: Triage-driven diagnosis of barrett's esophagus for early detection of esophageal adenocarcinoma using deep learning. Nature Medicine **27**(5), 833–841 (2021)
7. Heinrich, M.P., Jenkinson, M., Brady, M., Schnabel, J.A.: MRF-based deformable registration and ventilation estimation of lung CT. IEEE transactions on medical imaging **32**(7), 1239–1248 (2013)

8. 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. Nature Methods **18**(2), 203–211 (2021)
9. Jang, S., Graffy, P.M., Ziemlewicz, T.J., Lee, S.J., Summers, R.M., Pickhardt, P.J.: Opportunistic osteoporosis screening at routine abdominal and thoracic ct: normative l1 trabecular attenuation values in more than 20 000 adults. Radiology **291**(2), 360–367 (2019)
10. Jin, D., Guo, D., Ho, T.Y., Harrison, A.P., Xiao, J., Tseng, C.K., Lu, L.: Deeptarget: Gross tumor and clinical target volume segmentation in esophageal cancer radiotherapy. Medical Image Analysis **68**, 101909 (2021)
11. Klein, E., Richards, D., Cohn, A., Tummala, M., Lapham, R., Cosgrove, D., Chung, G., Clement, J., Gao, J., Hunkapiller, N., et al.: Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. Annals of Oncology **32**(9), 1167–1177 (2021)
12. Lambert, Z., Petitjean, C., Dubray, B., Kuan, S.: Segthor: Segmentation of thoracic organs at risk in ct images. In: 2020 Tenth International Conference on Image Processing Theory, Tools and Applications (IPTA). pp. 1–6. IEEE (2020)
13. Luo, H., Xu, G., Li, C., He, L., Luo, L., Wang, Z., Jing, B., Deng, Y., Jin, Y., Li, Y., et al.: Real-time artificial intelligence for detection of upper gastrointestinal cancer by endoscopy: a multicentre, case-control, diagnostic study. The Lancet Oncology **20**(12), 1645–1654 (2019)
14. Pickhardt, P.J., Graffy, P.M., Zea, R., Lee, S.J., Liu, J., Sandfort, V., Summers, R.M.: Automated ct biomarkers for opportunistic prediction of future cardiovascular events and mortality in an asymptomatic screening population: a retrospective cohort study. The Lancet Digital Health **2**(4), e192–e200 (2020)
15. Qin, Y., Wu, C.W., Taylor, W.R., Sawas, T., Burger, K.N., Mahoney, D.W., Sun, Z., Yab, T.C., Lidgard, G.P., Allawi, H.T., et al.: Discovery, validation, and application of novel methylated dna markers for detection of esophageal cancer in plasma. Clinical Cancer Research **25**(24), 7396–7404 (2019)
16. Rice, T., Ishwaran, H., Hofstetter, W., Kelsen, D., Apperson-Hansen, C., Blackstone, E.: Recommendations for pathologic staging (ptnm) of cancer of the esophagus and esophagogastric junction for the 8th edition ajcc/uicc staging manuals. Diseases of the Esophagus **29**(8), 897–905 (2016)
17. Rustgi, A.K., El-Serag, H.B.: Esophageal carcinoma. New England Journal of Medicine **371**(26), 2499–2509 (2014)
18. Siegel, R.L., Miller, K.D., Jemal, A.: Cancer statistics, 2021. CA: A Cancer Journal for Clinicians **71**(1), 7–333 (2021)
19. Valanarasu, J.M.J., Oza, P., Hacihaliloglu, I., Patel, V.M.: Medical transformer: Gated axial-attention for medical image segmentation. In: de Bruijne, M., Cattin, P.C., Cotin, S., Padoy, N., Speidel, S., Zheng, Y., Essert, C. (eds.) MICCAI 2021. pp. 36–46. Springer International Publishing, Cham (2021)
20. Wang, H., Zhu, Y., Green, B., Adam, H., Yuille, A., Chen, L.C.: Axial-deeplab: Stand-alone axial-attention for panoptic segmentation. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M. (eds.) ECCV 2020. pp. 108–126. Springer International Publishing, Cham (2020)
21. Wei, W.Q., Chen, Z.F., He, Y.T., Feng, H., Hou, J., Lin, D.M., Li, X.Q., Guo, C.L., Li, S.S., Wang, G.Q., et al.: Long-term follow-up of a community assignment, one-time endoscopic screening study of esophageal cancer in china. Journal of Clinical Oncology **33**(17), 1951 (2015)

22. Xia, Y., Yao, J., Lu, L., Huang, L., Xie, G., Xiao, J., Yuille, A., Cao, K., Zhang, L.: Effective pancreatic cancer screening on non-contrast ct scans via anatomy-aware transformers. In: de Bruijne, M., Cattin, P.C., Cotin, S., Padoy, N., Speidel, S., Zheng, Y., Essert, C. (eds.) MICCAI 2021. pp. 259–269. Springer International Publishing, Cham (2021)
23. Yan, K., Cai, J., Zheng, Y., Harrison, A.P., Jin, D., Tang, Y.B., Tang, Y.X., Huang, L., Xiao, J., Lu, L.: Learning from Multiple Datasets with Heterogeneous and Partial Labels for Universal Lesion Detection in CT. IEEE Trans. Med. Imaging **40**, 2759–2770 (Oct 2021)
24. Yao, J., Shi, Y., Cao, K., Lu, L., Lu, J., Song, Q., Jin, G., Xiao, J., Hou, Y., Zhang, L.: Deepprognosis: Preoperative prediction of pancreatic cancer survival and surgical margin via comprehensive understanding of dynamic contrast-enhanced ct imaging and tumor-vascular contact parsing. Medical Image Analysis **73**, 102150 (2021)
25. Ye, X., Guo, D., Tseng, C.k., Ge, J., Hung, T.M., Pai, P.C., Ren, Y., Zheng, L., Zhu, X., Peng, L., et al.: Multi-institutional validation of two-streamed deep learning method for automated delineation of esophageal gross tumor volume using planning-ct and fdg-petct. arXiv preprint arXiv:2110.05280 (2021)
26. Yousefi, S., Sokooti, H., Elmahdy, M.S., Peters, F.P., Shalmani, M.T.M., Zinkstok, R.T., Staring, M.: Esophageal gross tumor volume segmentation using a 3d convolutional neural network. In: MICCAI. pp. 343–351. Springer (2018)
27. Zhang, L., Gopalakrishnan, V., Lu, L., Summers, R.M., Moss, J., Yao, J.: Self-learning to detect and segment cysts in lung CT images without manual annotation. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). pp. 1100–1103. IEEE (2018)
28. Zhou, D., Huang, G., Li, J., Zhu, S., Wang, Z., Ling, B.W.K., Pun, C.M., Cheng, L., Cai, X., Zhou, J.: Eso-net: a novel 2.5 d segmentation network with the multi-structure response filter for the cancerous esophagus. IEEE Access **8**, 155548–155562 (2020)
29. Zhu, Z., Xia, Y., Xie, L., Fishman, E.K., Yuille, A.L.: Multi-scale coarse-to-fine segmentation for screening pancreatic ductal adenocarcinoma. In: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.T., Khan, A. (eds.) MICCAI 2019. pp. 3–12. Springer International Publishing, Cham (2019)