Lymph Node Metastasis Classification with Prototype-guided Multiple Instance Aggregation and Heterogeneous Feature Fusion

Haoshen Li[1,2], Tashan Ai[3], Yirui Wang[2], Zhanghexuan Ji[2], Qinji Yu[2,4], Le Lu[2], Bin Dong[5,6], Li Zhang[1], Xianghua Ye[7], Kuaile Zhao[3], and Dakai Jin[2]

[1]Center for Data Science, Peking University, Beijing, China
[2]DAMO Academy, Alibaba Group
[3]Fudan University Shanghai Cancer Center, Shanghai, China
[4]Shanghai Jiao Tong University, Shanghai, China
[5]Beijing International Center for Mathematical Research and the New Cornerstone Science Laboratory, Peking University, Beijing, China
[6]Center for Machine Learning Research, Peking University, Beijing, China
[7]The First Affiliated Hospital, Zhejiang University, Hangzhou, China
18111230048@fudan.edu.cn, kuaile_z@sina.com, yiruiwang06@gmail.com

Abstract. Lymph node metastasis diagnosis in computed tomography (CT) scans is an essential yet very challenging task for esophageal cancer staging and treatment planning. Recent advances in deep learning have markedly improved the performance in lymph node (LN) metastasis classification. However, these methods often focus more on the averaged features of all CT slices containing a 3D LN instance, lacking effective fusion of key slice-wise features, which is important in the LN metastasis analysis by physicians. In addition, existing deep learning models are trained using CT scans in an end-to-end fashion, thus lacking the explicit incorporation of clinically relevant meta-imaging features (i.e., morphological and radiomic features). Meta-imaging features play a crucial role in LN assessment and may not be effectively captured by direct end-to-end deep learning models. To address these issues, we formulate the 3D LN metastasis classification as a multiple instance learning (MIL) problem by extracting and fusing slice-level features (instance) into a comprehensive bag representation. Building on this, we propose a two-streamed MIL framework with a prototype-guided aggregation method that effectively captures LN characteristics at both local and global scales. Furthermore, a multi-scale multi-source fusion module is introduced to integrate the heterogeneous meta-imaging features with deep learning features, enhancing the comprehensive representation of LN. Five-fold cross-validation on a cohort of 284 esophageal cancer patients with 809 pathology-confirmed LN instances demonstrate the superiority of our methods compared to the state-of-the-art approaches with +2.66% in AUROC and +4.81% in sensitivity improvements.

Keywords: Lymph node metastasis · Multiple instance learning · Heterogeneous Feature Fusion.

1 Introduction

Esophageal cancer (EC) ranks as the sixth leading cause of cancer-related deaths globally, representing 1 out of every 20 cancer deaths [19]. Lymph node (LN) metastasis serves as one of the critical prognostic indicators in EC. Accurate identification of preoperative LN metastasis is crucial for guiding the treatment decisions (whether to proceed with surgery or opt for neoadjuvant therapy) and formulating treatment strategies (surgical resection area and radiotherapy clinical target volume [CTV]) [1]. Therefore, LN metastasis assessment holds significant clinical importance in EC diagnosis and management.

Assessing LN metastatic status in CT is challenging even for experienced physicians. It depends on various factors, such as global characteristics (*e.g.*, size, shape), localized imaging characteristics (*e.g.*, intensity inhomogeneity, textures), etc. All of these characteristics contribute to the LN status, but none serve as a standalone predictive factor. For example, the existing criteria, such as RECIST [16], morphology [15] and texture characteristics [2], show limited performance in previous studies [3, 5, 10, 12, 20] (*e.g.*, sensitivity ranged from 39.7% to 67.2% with specificity around 80.0%). Thus, proposing an effective computer-aided diagnosis (CAD) solution for this task is highly desirable.

Previous work have used deep learning for LN abnormality diagnosis. Roth *et al.* develops a two-stage convolutional neural network (CNN) network with a 2.5D universal image decomposition representation and random aggregation to detect and classify enlarged LNs [13, 14]. Lee *et al.* attempts various CNNs architectures to diagnose cervical LN metastasis in CT using 202 patients with thyroid cancer [9]. Kann *et al.* train a dual 3D network, which jointly learns the global and local features of LNs, to classify metastatic LNs and extranodal extension (ENE) using CT scans of 270 patients [8]. Li *et al.* proposes a semi-supervised framework to better handle unlabeled LNs, which achieves the state-of-the-art classification performance [11]. Although these methods demonstrate their effectiveness, they still face two major limitations. First, previous methods focus more on the averaged features of all CT slices containing a 3D LN instance, lacking effective fusion of key slice-wise features. For example, the leading 2.5D method [11] simply merges multiple sets of slice features by concatenation, making it difficult to identify key slice information when handling with LNs in thinner CT scan that can occupy more than 10 to 20 CT axial slices. For metastatic LNs, its malignant characteristics are more likely to be evident in partial slices than in all slices. Therefore, it is important to integrate and capture the key slice features that most correlate to the metastatic status. Second, previous deep learning-based methods also lack explicit consideration of clinic-relevant meta-imaging features, such as long, short axis diameters, heterogeneity intensity, central necrosis, etc. These meta-imaging features may not be effectively captured by the direct end-to-end deep learning model.

To address these limitations, we formulate the 3D LN metastasis classification as a multiple instance learning (MIL) problem by extracting and combining slice-level features (instance) into a comprehensive bag representation. Building on this, we propose a two-streamed MIL framework with a prototype-guided aggre-

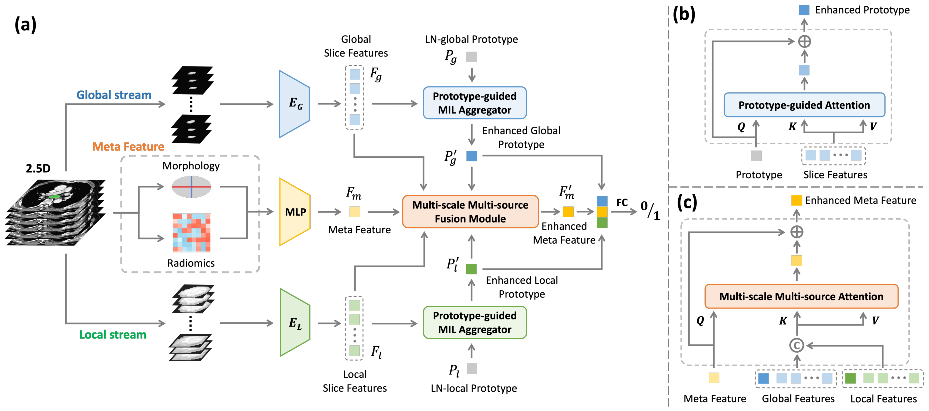

Fig. 1: (a) Illustrates the pipeline. A LN CT patch is transposed into multiple 3-slice images with original and zoom-in size, as the inputs for the global and local stream, respectively. In addition, extracted morphology and radiomics features from the LN patch are introduced as the meta feature. The prototype-guided MIL aggregator is introduced to fuse slice features into the improved LN prototype, as shown in (b). The Multi-scale Multi-source Fusion Module is introduced to enhance the meta feature by multi-scale image features, as shown in (c).

gation method that captures LN characteristics at both local and global scales. Furthermore, a multi-scale multi-source fusion module is introduced to integrate the heterogeneous meta-imaging features (such as morphological and radiomics features) with deep learning features, enhancing the comprehensive representation of LN. We collected and curated 284 EC patients with 809 pathology-confirmed LN instances. In five-fold cross-validation, our proposed method exhibits superior performance, significantly exceeding the state-of-the-art LN classification approaches and other MIL-based methods and achieving the highest AUROC score of 84.68% (+2.66% improvement) in this challenging task.

2 Method

An overview of our method is shown in Fig. 1. We develop a two-stream MIL framework with slice-wise prototype-guided MIL aggregation module to enhance the key LN features in both streams, as detailed in Sec. 2.1. To integrate meta-imaging features with deep learning features, a multi-scale multi-source fusion module is introduced that enhances meta-features by computing cross-attention with multi-scale LN imaging features, as explained in Sec. 2.2.

2.1 Two-streamed Prototype-guided Multiple Instance Aggregation

Metastatic LN imaging characteristics may only be present in a few slices of an LN 3D CT volume. Inspired by this observation, we naturally formulate the 3D

LN classification problem as an MIL task by extracting and fusing slice-level features (MIL instance) into a comprehensive bag representation for volume-level LN metastasis classification. Our two-streamed slice-wise MIL framework is shown in Fig. 1.

While global characteristics (*e.g.*, size, shape) can indicate LN metastasis, the use of localized features (*e.g.*, texture, inhomogeneity) can facilitate more accurate and robust LN metastasis modeling. Motivated by this, instead of extracting imaging features solely from the original CT input, we take a two-stream approach to simultaneously encode both the original and resized input that spatially zoomed in or out to fill the input slice, which implicitly promotes learning localized fine-scale features. This two-stream encoding method not only provides complementary features for subsequent multi-scale feature fusion but acts as a mutual regularization to avoid the network overfitting to a single-scale input. Previous methods either directly extract LN features from a pure 3D network [8] or leverage a simple 2.5D strategy [11] to treat and fuse adjacent slices equally without considering their relative importance. Here, we view LN metastasis classification as a multi-instance learning problem, group the CT slices into multiple sets of 3-channel images, and transform a 3D LN CT image with a volume-level metastasis label in multiple 2D slice groups that can be processed by pre-trained 2D network, acting as a bag of instances in the MIL paradigm. During the prototype-guided aggregation, the fusion weight of each instance is recalibrated so that the slice with more suspicious LN metastasis features can be enhanced by assigning a higher weight to the fused feature. In implementation, we adopt MobileNetv3 [6] with pre-trained weights on ImageNet [4] as the image encoder to generate the two-streamed multiple slice-group features. To effectively aggregate two-streamed slice-group features, we propose a prototype-guided MIL Aggregator. As shown in Fig. 1, we randomly initialize two learnable prototypes: LN-global Prototype and LN-local Prototype, denoted as $P_g, P_l \in \mathbb{R}^{1\times d}$, to guide the aggregation of global stream and local stream features, respectively. Taking the global stream as an example, the learnable prototype P_g and the global slice features $F_g \in \mathbb{R}^{N\times d}$ (N represents the number of slice groups) are fed into the aggregator, which is implemented as a cross-attention layer. Formally, the learnable global prototype P_g serves as the query vector Q_g to search for slice-group features that are most relevant to LN metastasis, while the global-stream slice features F_g act as both the key K_g and the value V_g. The enhanced global prototype is then updated as follows:

$$P'_g = \text{Softmax}(\frac{Q_g K_g^T}{\sqrt{d}} V_g) + P_g \tag{1}$$

The operation for the local stream is the same as the global stream. Under the guidance of P_g and P_l, the prototype-guided MIL aggregator can effectively capture slices that are highly correlated with LN metastasis status, thereby enabling better aggregation. The enhanced prototypes P'_g, P'_l are subsequently used as a compact, metastasis-aware representation for final classification.

2.2 Multi-scale Multi-source Fusion

In LN metastasis classification, clinic-relevant meta-imaging features, such as round morphology, intensity heterogeneity, are shown to be effective. However, these meta-imaging features may not be easily captured by direct end-to-end trained deep learning models. Therefore, we explicitly compute these meta-imaging features and integrate these heterogeneous features into one deep learning framework by a new multi-scale multi-source fusion module. Specifically, we measure the long and short diameters of each 3D LN instance, calculate its radiomic first-order features, and concatenate them as the meta-imaging feature. Then, a linear layer is used to map the meta-imaging feature to the same dimensionality as the deep learning feature, denoting it as $F_m \in \mathbb{R}^{1\times d}$. Next, we leverage the multi-scale deep features to further enhance the meta-imaging features. As shown in Fig. 1(c), this process is achieved by the cross-attention mechanism. Specifically, enhanced local and global deep prototypes, P'_g, P'_l, and original slice features F_g, F_l are concatenated together as key $K_{g,l}$ and value $V_{g,l}$. The meta-imaging feature F_m is taken as query Q_m. The interaction operation is as follows:

$$F'_m = \text{Softmax}(\frac{Q_m K_{g,l}^T}{\sqrt{d}} V_{g,l}) + F_m \tag{2}$$

After deriving the enhanced meta feature F'_m, along with previous P'_g and P'_l, we simply fuse them through channel-wise concatenation. The fused representation is then projected to a logit space for LN metastasis classification via a learnable linear transformation and optimized using binary cross-entropy (BCE) loss, formulated as:

$$\mathcal{L}_{\text{cls}} = \mathcal{L}_{\text{BCE}} \left(\mathbf{W}_p \cdot \left[P'_g \, \| P'_l \| \, F'_m \right]^T, Y \right) \tag{3}$$

where $\|$ denotes feature concatenation along the channel dimension, $\mathbf{W}_p \in \mathbb{R}^{1\times 3d}$ represents the trainable projection weights. $Y \in \{0, 1\}$ is the metastasis status label (0: benign, 1: malignant).

3 Experiments and Results

3.1 Experimental Settings

Dataset: We collected a dataset of 284 esophageal cancer patients who underwent esophagectomy treatment. Each patient has a preoperative contrast-enhanced CT scan and a detailed post-operative pathological report indicating whether there are metastatic LNs in the surgical dissection area. The median CT scan size is $512\times512\times358$ voxels with a median resolution of $0.785\times0.785\times$ 1.0mm. We first detected all visible LN instances using an ensemble of recent automatic LN detection models [21, 22]. Out of these detected LNs, 809 LN instances have metastasis status labels (208 positive and 601 negative), which is confirmed by the consensus of two radiologists according to the pathology report. The LN mask, along with the CT scan, is then cropped using a 64×64×32

ROI centered on each 3D LN. Experiments are conducted using five-fold cross-validation with a 70%/10%/20% training, validation, and testing split (at the patient level) for each fold.

Implementation details: A lightweight model MobileNetv3 [6] is used as the image encoder for each of the two streams. We considered 18 slices (6 slice groups) as the encoder's input, which exceeds 80% LN's slice number. The feature dimension d of the slice features and mapped meta-features is 960. For the meta-imaging feature input, we use 20 features in total, including two morphological features (short diameter and long diameter), and 18 radiomics features (original first-order features (skewness, elongation, etc.) calculated by Radiomics library in Python 3.9). SGD optimizer with a learning rate of 1.6e-4 and cosine annealing decay is adopted, and the network is trained by 300 epochs with a mini-batch size of 32.

Comparison methods: We compare with other methods in the following categories. 1) *Traditional deep learning category:* four state-of-the-art LN metastasis classification work are evaluated: two 3D-based methods of MobileNetv3_3d [6] and 3D DualNet [8]; two 2.5D-based methods of MobileNetv3_2.5D [6] and dual-stream 2.5D [10]. 2) *MIL category:* We compare to six popular MIL methods, including Max-pooling, Mean-pooling, ABMIL [7], GAMIL [7], TransMIL [17], and DTMIL [23]. For fair comparison, MobileNetv3_2.5D is used as the backbone in the MIL methods. 3) *Multi-source MIL category:* We also examine a leading multi-source MIL fusion method ViLa-MIL [18], which combines both deep learning and meta features.

Evaluation metrics: To evaluate the performance comprehensively, we compute the area under the receiver operating characteristic curve (AUROC), accuracy at a specificity rate of 75% (A@S75), accuracy at a recall rate of 75% (A@R75), recall at a specificity rate of 75% (R@S75), and specificity at a recall rate of 75% (S@R75).

3.2 Experimental Results

Quantitative comparison results: Table 1 summarizes the quantitative comparison results. Several observations can be drawn. (1) Both 2.5D and MIL methods outperform 3D methods with an improvement of 2-5% in AUROC. This indicates that the 3D deep network fail to learn and extract the most essential LN characteristics in a 3D fashion. One reason for this may be the lack of pre-trained model weights in 3D. (2) For the methods in 3D and 2.5D categories, multi-scale approaches (DualNet [8], dual-stream [11]) achieve better performance, with an improvement of more than 2% in AUROC as compared to the single-stream counterpart. This confirms that the local and global features of LNs are both important for its metastatic classification. (3) Slice-based MIL approaches yield decreased performance ([79.89%, 80.68%] AUC) when compared to the specifically designed dual-stream 2.5D LN classification method [11] (81.09% AUC). This shows that applying previous MIL methods in single stream to the LN classification task would not exhibit improved accuracy. In contrast, with the proposed two-streamed prototype-guided MIL aggregation method, our method

Table 1: Quantitative performance of LN metastasis classification. Four groups correspond to 3D methods, 2.5D methods, MIL methods and MIL methods combined with meta imaging features, respectively. Meta represents the meta imaging features (morphological and radiomics features).

Method	Meta	AUROC	A@R75	A@S75	S@R75	R@S75
mobilenetv3_3D [6]	×	75.61±4.34	67.99	71.47	64.94	66.16
DualNet_3D [8]	×	77.67±4.95	67.98	72.31	64.96	67.20
mobilenetv3_2.5D [6]	×	78.94±2.96	68.89	72.40	66.63	66.95
dual-stream_2.5D [11]	×	81.09±3.88	72.66	73.21	71.91	69.50
Max-pooling	×	80.10±3.16	70.05	72.55	68.59	68.60
Mean-pooling	×	80.40±3.03	71.49	72.95	70.35	67.30
ABMIL [7]	×	80.34±3.92	71.80	72.24	70.41	69.28
GAMIL [7]	×	80.51±4.31	71.66	73.12	70.29	68.67
TransMIL [17]	×	79.89±2.32	69.73	72.20	68.09	66.99
DTMIL [23]	×	80.68±4.54	71.88	72.81	70.45	69.41
Ours	×	**82.54±2.10** (+1.45%)	**76.57** (+3.91%)	**73.97** (+0.76%)	**76.84** (+4.93%)	**72.37** (+2.87%)
ViLa-MIL [18]	✓	82.02±2.90	73.18	73.47	72.27	71.69
Ours	✓	**84.68±2.00** (+2.66%)	**78.53** (+5.35%)	**74.68** (+1.21%)	**79.48** (+7.21%)	**76.50** (+4.81%)

exceeds the dual-stream 2.5D method [11] (indicating that multiple slices aggregation is important) and other image-based MIL methods, and achieves the best performance of 82.54% AUC, 76.84% S@R75 and 72.37% R@S75. (4) When further incorporating meta-imaging features, our method further boosts the AUROC to 84.68% with +2.66% improvement compared to the multi-source fusion method ViLa-MIL [18].

Qualitative results: Fig. 2 illustrates the attention weight distributions of prototype-guided MIL aggregation module in the global and local streams (mean values of all attention weights in the test split), showing the importance of the positions of the slices in the classification contribution. It is observed that in both streams, the importance of the LN slices decreases from center to periphery, indicating that the malignant features are more likely to be near the central of a 3D LN instance. Fig. 3 presents three malignant LN examples, each showing two different slices. The upper row presents the CT slice that has a higher attention weight, while the lower row shows the CT slice with a low attention weight. The examples demonstrate that our prototype-guided MIL aggregation method is effective in identifying the most important CT slices for metastasis status.

Ablation results of incorporating meta-imaging features: We separately demonstrate the effects of incorporating shape features and texture features in Table 2. We can see that individually adding each of these two features both improve the model's performance. In contrast, shape features provide a greater

Table 2: Effect of the different meta-imaging features: shape and texture.

shape	texture	AUC	S@R75	R@S75
		82.54±2.10	76.84	72.37
✓		83.76±2.18	77.54	74.68
	✓	83.38±3.03	76.01	74.12
✓	✓	84.68±2.00	79.48	76.50

Table 3: Effect of different input LN slice numbers.

slice number	AUC	S@R75	R@S75
9	81.12±2.78	73.51	70.67
18	82.54±2.10	76.84	72.37
27	81.77±2.60	73.63	71.89

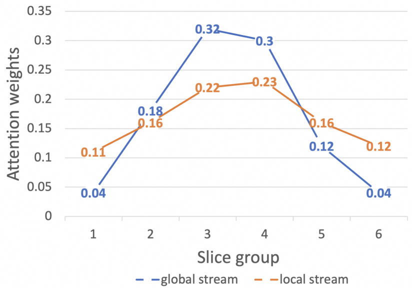

Fig. 2: Attention weight distributions of the prototype-guided MIL aggregation in two streams.

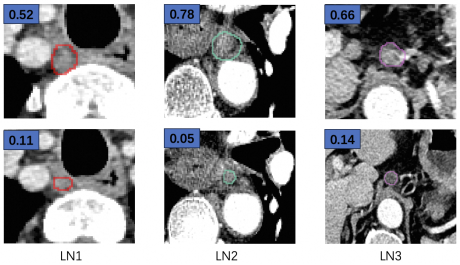

Fig. 3: Three qualitative examples of LN CT image with its attention weight assigned to the slice.

improvement, indicating that the global information (shape, size) plays a more important role in the task. Finally, combining the shape and texture features, our full model improves AUROC, S@R75, and R@S75 to 84.68%, 79.48%, and 76.50%, demonstrating its effectiveness.

Ablation results of different numbers of input slices: We select 9, 18, and 27 slices of LN, respectively, to combine into slice-wise instances as input to the prototype-guided MIL aggregation module. As shown in Table 3, we observe that selecting 18 slices achieves the best results. Since LNs are relatively small, 18 slices can fully encompass most of LNs. Selecting 9 slices may miss metastatic information, and selecting 27 slices leads to feature redundancy hindering the effective fusion of key features, both leading to suboptimal performance.

4 Conclusion

In this work, we formulate the 3D LN metastasis classification as an MIL problem by extracting and combining slice-level features (instance) into a complete bag representation. Building on this, we propose a two-streamed MIL framework with a prototype-guided aggregation method that effectively captures LN characteristics at both local and global scales. Furthermore, a multi-scale multi-source fusion module is introduced to integrate the heterogeneous meta-imaging features with deep learning features, enhancing the comprehensive representation of LN. Five-fold cross-validation on a cohort of 284 esophageal cancer patients with 809 pathology-confirmed LN instances demonstrate the superiority of our methods compared to the state-of-the-art approaches with +2.66% in AUROC and +4.81% in sensitivity improvements.

Disclosure of Interests The authors have no competing interests to declare that are relevant to the content of this article.

Acknowledgement This work was supported by Alibaba Group through Alibaba Research Intern Program, and Zhejiang Provincial Natural Science Foundation of China under Grant No. 2024-KYI-00I-I05. Li Zhang was partly supported by NSFC 81801778, NSFC 12090022 and Clinical Medicine Plus X-Young Scholars Project of Peking University PKU2023LCXQ041. Bin Dong was partly supported by NSFC 12090022, and the New Cornerstone Investigator Program.

References

1. Ajani, J.A., D'Amico, T.A., Bentrem, D.J., Chao, J., Corvera, C., Das, P., Denlinger, C.S., Enzinger, P.C., Fanta, P., Farjah, F., et al.: Esophageal and esophagogastric junction cancers, version 2.2019, nccn clinical practice guidelines in oncology. Journal of the National Comprehensive Cancer Network **17**(7), 855–883 (2019)
2. Bayanati, H., E Thornhill, R., Souza, C.A., Sethi-Virmani, V., Gupta, A., Maziak, D., Amjadi, K., Dennie, C.: Quantitative ct texture and shape analysis: can it differentiate benign and malignant mediastinal lymph nodes in patients with primary lung cancer? European radiology **25**, 480–487 (2015)
3. Boiselle, P.M., Patz Jr, E.F., Vining, D.J., Weissleder, R., Shepard, J., McLoud, T.C.: Imaging of mediastinal lymph nodes: Ct, mr, and fdg pet. Radiographics **18**(5), 1061–1069 (1998)
4. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009)
5. Foley, K., Christian, A., Fielding, P., Lewis, W., Roberts, S.: Accuracy of contemporary oesophageal cancer lymph node staging with radiological-pathological correlation. Clinical radiology **72**(8), 693–e1 (2017)
6. Howard, A., Sandler, M., Chu, G., Chen, L.C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., et al.: Searching for mobilenetv3. In: ICCV. pp. 1314–1324 (2019)
7. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International conference on machine learning. pp. 2127–2136 (2018)
8. Kann, B.H., Aneja, S., Loganadane, G.V., Kelly, J.R., Smith, S.M., Decker, R.H., Yu, J.B., Park, H.S., Yarbrough, W.G., Malhotra, A., et al.: Pretreatment identification of head and neck cancer nodal metastasis and extranodal extension using deep learning neural networks. Scientific reports **8**(1), 14036 (2018)
9. Lee, J.H., Ha, E.J., Kim, J.H.: Application of deep learning to the diagnosis of cervical lymph node metastasis from thyroid cancer with ct. European radiology **29**, 5452–5457 (2019)
10. Li, B., Li, B., Jiang, H., Yang, Y., Zhang, X., Su, Y., Hua, R., Gu, H., Guo, X., Ye, B., et al.: The value of enhanced ct scanning for predicting lymph node metastasis along the right recurrent laryngeal nerve in esophageal squamous cell carcinoma. Annals of translational medicine **8**(24) (2020)

11. Li, H., Wang, Y., Zhu, J., Guo, D., Yu, Q., Yan, K., Lu, L., Ye, X., Zhang, L., Wang, Q., et al.: Semi-supervised lymph node metastasis classification with pathology-guided label sharpening and two-streamed multi-scale fusion. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 623–633. Springer (2024)
12. Liu, J., Wang, Z., Shao, H., Qu, D., Liu, J., Yao, L.: Improving ct detection sensitivity for nodal metastases in oesophageal cancer with combination of smaller size and lymph node axial ratio. European Radiology **28**, 188–195 (2018)
13. Roth, H.R., Lu, L., Liu, J., Yao, J., Seff, A., Cherry, K., Kim, L., Summers, R.M.: Improving computer-aided detection using convolutional neural networks and random view aggregation. IEEE transactions on medical imaging **35**(5), 1170–1181 (2015)
14. Roth, H.R., Lu, L., Seff, A., Cherry, K.M., Hoffman, J., Wang, S., Liu, J., Turkbey, E., Summers, R.M.: A new 2.5 d representation for lymph node detection using random sets of deep convolutional neural network observations. In: MICCAI. pp. 520–527. Springer (2014)
15. Schreuder, A., Jacobs, C., Scholten, E.T., van Ginneken, B., Schaefer-Prokop, C.M., Prokop, M.: Typical ct features of intrapulmonary lymph nodes: a review. Radiology: Cardiothoracic Imaging **2**(4), e190159 (2020)
16. Schwartz, L., Bogaerts, J., Ford, R., Shankar, L., Therasse, P., Gwyther, S., Eisenhauer, E.: Evaluation of lymph nodes with recist 1.1. European journal of cancer **45**(2), 261–267 (2009)
17. Shao, Z., Bian, H., Chen, Y., Wang, Y., Zhang, J., Ji, X., et al.: Transmil: Transformer based correlated multiple instance learning for whole slide image classification. In: Advances in neural information processing systems. vol. 34, pp. 2136–2147 (2021)
18. Shi, J., Li, C., Gong, T., Zheng, Y., Fu, H.: Vila-mil: Dual-scale vision-language multiple instance learning for whole slide image classification. In: CVPR. pp. 11248–11258 (2024)
19. Siegel, R.L., Miller, K.D., Fuchs, H.E., Jemal, A.: Cancer statistics, 2022. CA: a cancer journal for clinicians **72**(1), 7–33 (2022)
20. Wu, L., Yang, X., Cao, W., Zhao, K., Li, W., Ye, W., Chen, X., Zhou, Z., Liu, Z., Liang, C.: Multiple level ct radiomics features preoperatively predict lymph node metastasis in esophageal cancer: a multicentre retrospective study. Frontiers in oncology **9**, 1548 (2020)
21. Yu, Q., Wang, Y., Yan, K., Li, H., Guo, D., Zhang, L., Shen, N., Wang, Q., Ding, X., Lu, L., Ye, X., Jin, D.: Effective lymph nodes detection in ct scans using location debiased query selection and contrastive query representation in transformer. In: Computer Vision – ECCV 2024. pp. 180–198. Springer Nature Switzerland (2025)
22. Yu, Q., Wang, Y., Yan, K., Lu, L., Shen, N., Ye, X., Ding, X., Jin, D.: Slice-consistent lymph nodes detection transformer in ct scans via cross-slice query contrastive learning. In: MICCAI. pp. 616–626. Springer (2024)
23. Zhang, H., Meng, Y., Zhao, Y., Qiao, Y., Yang, X., Coupland, S.E., Zheng, Y.: Dtfd-mil: Double-tier feature distillation multiple instance learning for histopathology whole slide image classification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 18802–18812 (2022)