Joint Prediction with Dummy Node for Abdominopelvic Lymph Node Metastasis

1st Xingyu Zou
College of Computer Science
Sichuan University
Chengdu, China
zouxingyu@stu.scu.edu.cn

2nd Yiji Mao
College of Computer Science
Sichuan University
Chengdu, China
2023223040057@stu.scu.edu.cn

3rd Yuling Zheng
College of Computer Science
Sichuan University
Chengdu, China
zyl021011@gmail.com

4th Xiao Yu
College of Computer Science
Sichuan University
Chengdu, China
2023223045159@stu.scu.edu.cn

5th Yuxuan Ji
College of Computer Science
Sichuan University
Chengdu, China
jiyuxuan0918@gmail.com

6th Mingxuan Tian
College of Computer Science
Sichuan University
Chengdu, China
1647787177@qq.com

7th Haixian Zhang*
College of Computer Science
Sichuan University
Chengdu, China
zhanghaixian@scu.edu.cn

***Abstract*—Abdominopelvic lymph node (LN) metastasis prediction using preoperative computed tomography (CT) is a critical task for personalized disease progression assessment and treatment planning for colorectal cancer (CRC) patients. However, there are notable challenges in aligning CT images with pathology data. Moreover, manual annotation of LNs has inherent limitations, such as annotation omission, leading to low accuracy in prediction models. To address these issues, we propose a joint prediction framework guided by node-level and patient-level tasks, which leverages patient-level pathological labels to predict LN metastasis in CT scans. Key innovations in the framework include: (1) Pretraining encoder using node discrimination from nine views, (2) Predicton with a dummy node, which simulates unlabeled LN to bridge the gap in annotation omission, (3) Intra-class contrastive learning based on neural memory ordinary differential equations (nmODE), which distinguishes subtle and continuous features from dynamic evolution, and (4) Prediction with inter-class contrastive learning based on gated attention for better nodes integration within patient. These techniques allow the model to make reliable joint prediction in such a difficult task and even in imperfectly labeled data. We validate the proposed framework on a private dataset of 517 CRC patients and a public chest CT dataset of 1,595 scans, showing superior performance. Additionally, we provide visualized explanations of the embedding expression.**

***Index Terms*—Colorectal cancer, Lymph node metastasis, Multi instance learning, Contrastive learning, Joint prediction**

I. Introduction

Colorectal cancer (CRC) is one of the most prevalent malignant tumors globally, ranking third in incidence and second in mortality worldwide [1]. Lymph node (LN) metastasis, a critical prognostic factor that profoundly influences CRC patients' disease-free survival and local recurrence [2], significantly exacerbates disease severity. Preoperative assessment of abdominopelvic LN metastasis status is therefore essential for personalized disease progression evaluation, precise treatment planning, and early clinical intervention [3]. Computed tomography (CT) scans are widely used in preoperative clinic imaging for CRC, leveraging this economic modality to predict node status holds great clinical value. However, current diagnosis via CT relies heavily on manual interpretation by clinicians, which is time-consuming and prone to errors such as missed or misdiagnosed nodes [4]. Thus, developing an automated CT-based method for LN metastasis prediction is imperative to assist clinical decision-making and improve diagnostic accuracy.

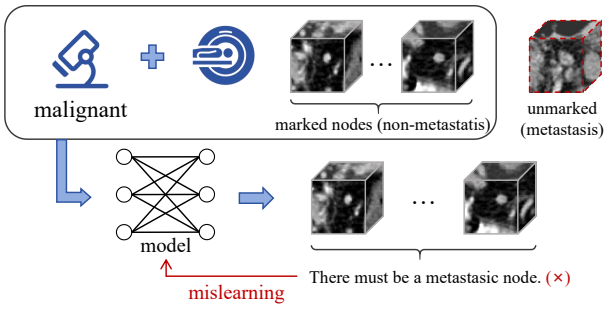

Fig. 1. Example of a mislearning due to manual annotation omission.

Pathological examination, the gold standard for diagnosis [5], faces challenges in alignment with CT imaging due to factors such as tissue displacement, small size, and tissue adhesion. These issues impede the precise mapping of LNs identified on CT scans to those sampled during surgery. While pathology labels can confirm the presence of LN metastasis, they fail to specify which individual LN is affected. Even with the adoption of a two-level annotation and cross-review workflow [6], manual annotations remain susceptible to omissions or inaccuracies, which can adversely impact the performance of deep learning models (as unmarked node leads model to mislearning in Fig. 1). Such noisy datasets hinder the model's ability to learn robust features, thereby restricting its predictive power [7].

*Corresponding author.

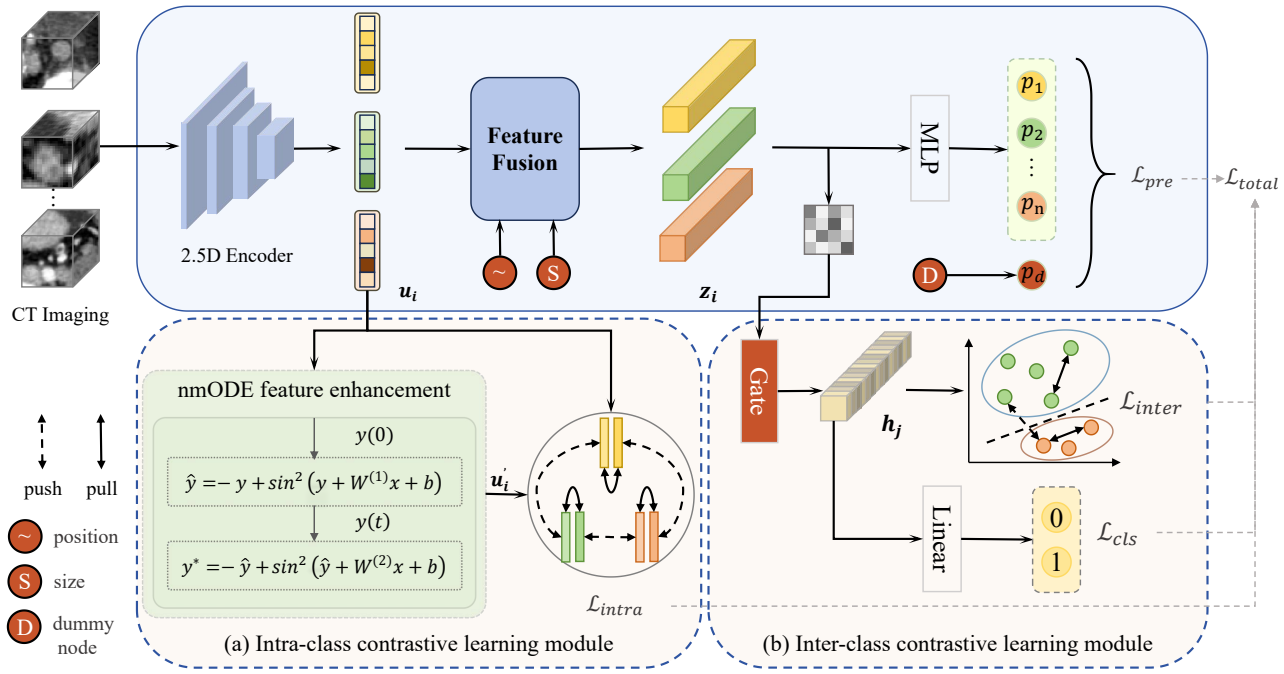

Fig. 2. The architecture of the proposed joint prediction framework.

To address these, we propose a joint prediction framework that leverages pathologic labels to predict LN metastasis in CT scans with manual annotation omission. The method consists of several key components:

- We pretrain encoder with self-supervised contrastive learning at multi-view level and add 2.5D fusion layer during subsequent task.
- We introduce predicton with a dummy node to bridge the gap between malignant pathology results and unlabeled metastatic LNs.
- We present an nmODE-based intra-class contrastive learning module to guide toward capturing subtle and continuous feature variations.
- We progress patient-level prediction with inter-class contrastive learning based on gated attention.

Additionally, a dynamic training loss function gradually shifts focus from contrastive representation learning in early stages to classification in later stages, ensuring effective feature extraction and optimal classification. The code and data example are available in https://github.com/zxy630/IICL.git.

II. Related Work

A. Intelligent Approach of Lymph Node Metastasis Prediction

Existing node-level methods are mainly developed for cancers other than CRC, such as esophageal cancer and thyroid cancer [8]. For example, Cui et al. [9] treated node features as graph nodes with clinical attributes, achieving 82.3% AUC. Li et al. [10] proposed a semi-supervised framework that exploits node-level annotations by jointly leveraging patient-level labels and teacher–student predictions. Despite their effectiveness, these approaches are based on node-level annotations, which are difficult to obtain in CRC clinical practice.

Multiple instance learning (MIL) has emerged as a weakly supervised solution. In MIL, labels are assigned solely at the bag level, which fits our task perfectly: each patient acts as a bag, and nodes are treated as instances. Recent CT-based methods predominantly adopt the MIL paradigm and focus on improving instance aggregation [12]. Xie et al. [13] proposed an attention-based MIL framework with global–local cross-attention fusion of CT slices and LN embeddings, achieving 74.7% accuracy and 76.8% AUC. Similar global–local fusion strategies were further explored in [14].

However, these methods ignore manual annotation omissions, which commonly arise from annotator fatigue, distraction, and variations in professional expertise [15], leading to incomplete instance-level supervision and low AUC. Moreover, these approaches underutilize priors such as node size and anatomical position. Recent studies indicate that node size and morphology are key indicators of metastasis [16], and spatial position within the abdominopelvic region also plays an important role [17].

B. MIL in Medical Image Classification

MIL has been widely adopted for pulmonary nodule classification, particularly on the Data Science Bowl (DSB) 2017[1] dataset, where only patient-level cancer labels are available. Liao et al. [18] systematically evaluated patient-level cancer prediction by aggregating candidate nodules, highlighting the importance of instance selection in weakly labeled CT data. Zheng et al. [19] proposes a fine-grained contrastive learning

[1]https://www.kaggle.com/competitions/data-science-bowl-2017/data

method to guide the model in understanding fine-grained similarities and differences between nodule patterns. Xu et al. [20] extracted multi-scale 3D nodule cubes as postive pairs for supervised contrastive learning. All achieve SOTA performance and demonstrate strong transfer learning capabilities.

Besides, MIL has been extensively studied in computational pathology [21], where whole slide images are treated as bags and image patches as instances. However, the extreme sparsity and class imbalance of tumor regions pose significant challenges for conventional MIL sampling strategies. So recent works enhance MIL with different strategies. For instance, Perez et al. [22] improved patch representations using self-supervised contrastive learning combined with attention-based MIL, while Zheng et al. [23] introduced a reinforcement learning–driven dynamic sampling framework. Further studies [24], [26] incorporate supervised contrastive learning and curriculum strategies to mitigate class imbalance and noisy instance supervision, offering valuable insights for robust MIL modeling under weak supervision.

III. Methods

Fig. 2 presents the architecture of our proposed joint prediction framework.

A. 2.5D Encoder with Pretraining

To enhance node representation learning, we pretrain the encoder using a self-supervised contrastive learning framework, as illustrated in Fig. 3. Each node is first cropped as a 3D cube and projected into multiple 2D views from different spatial orientations via view filters, forming a set of nine views. These views are encoded with shared weights and pulled together in the embedding space as positive pairs, while views from different nodes are pushed apart as negative pairs. To account for the subtle morphological differences between nodes, we further incorporate an informative negative sample selection strategy that dynamically identifies morphologically similar nodes within each mini-batch, forcing the encoder to focus on fine-grained and discriminative features rather than coarse inter-object differences.

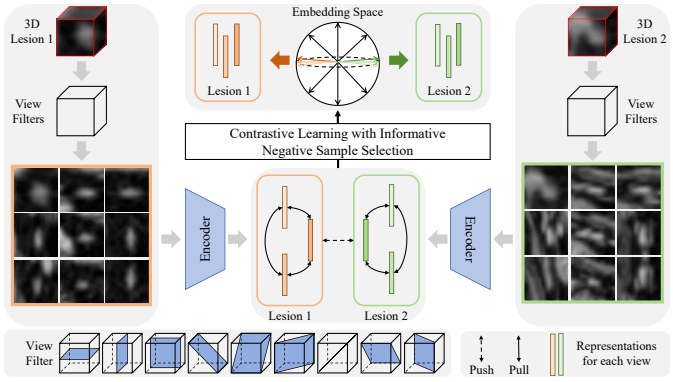

Fig. 3. The architecture of self-supervised contrastive learning pretraining for lymph node from nine views.

Given the redundancy in CT volumes, we structure the input as a $3 \times 32 \times 32$ patch centered on each node, consisting of the target slice and its adjacent slices. We augment the pretrained 2D encoder with strategically placed 2.5D convolutional fusion layers that concatenate embeddings from neighboring slices and integrate them via 1×1 convolutions (Fig. 4). Fusion is introduced only after the first two convolutional blocks, as integrating contextual information at moderate depth is more effective for capturing node–related morphological cues while avoiding unnecessary redundancy. The resulting encoder produces a high-dimensional patch embedding u_i.

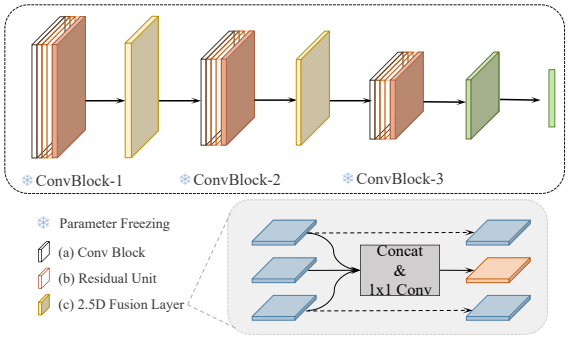

Fig. 4. Sketch of 2.5D encoder.

B. Predicton with Dummy Node

Guided by clinical prior knowledge, we incorporate anatomical position and size as auxiliary cues for diagnosis. Specifically, sinusoidal positional embeddings (PE) are introduced to encode the 3D spatial coordinates (x, y, z) of each node and are added to the patch embeddings. In addition, the long-axis diameter, a clinically important indicator for metastasis risk, is integrated by projecting the scalar size s_i into the feature space and jointly fusing it with image and positional features. The final node embdding is formulated as:

$$z_i = \sigma\left(u_i + PE(x, y, z) + (w_s s_i + b_i)\right), \tag{1}$$

where $w_s \in \mathbb{R}^{D\times 1}$, $b_i \in \mathbb{R}^D$, $\sigma(\cdot)$ is an activation function, and z_i represents the fused node embedding, which is subsequently used to predict the node-level metastasis probability p_i via a lightweight MLP.

Due to incomplete manual annotations in CT scans, patient-level positive labels may not be fully attributable to the annotated nodes, potentially leading to false positive learning. To address this issue, we introduce a dummy node that represents unannotated or missing metastatic nodes and explicitly models their contribution to patient-level prediction. Let p_d denote the learnable malignancy probability of the dummy node. The overall patient-level metastasis probability P_N is computed using leaky noisy-or formulation:

$$P_N = 1 - (1 - p_d)\prod_i (1 - p_i), \tag{2}$$

where p_d is optimized jointly with the network parameters. This design allows the model to attribute unexplained positive labels to the dummy node, preventing overestimation

of malignancy in benign nodes and improving robustness under incomplete annotations. The predictor is trained using a weighted binary cross-entropy loss $\mathcal{L}_{pre}$:

$$\mathcal{L}_{\text{pre}}(y, P_N) = -\alpha \cdot y \log(P_N) - (1-\alpha) \cdot (1-y) \log(1-P_N), \tag{3}$$

where $y \in \{0, 1\}$ represents the true label, $P_N \in (0, 1)$ represents the predicted transition probability by the model, α is the weight for the positive class.

C. nmODE-based Intra-class Contrastive Learning Module

LN morphology exhibits continuous and gradual variations across adjacent slices, which are difficult to use discrete data augmentation, such as crop and rotate. To model continuous morphological changes while preserving semantic consistency, we propose an intra-class representation learning module based on neural memory ordinary differential equations (nmODE [25]). It defines a nonlinear mapping from external input x to a global attractor y^* where $y(0)$ represents the initial value, and y(t) represents the state of the network at time t. The feature dynamics are defined as:

$$\hat{y} = -y + \sin^2(y + W^{(1)}x + b), \tag{4}$$

$$y^* = -\hat{y} + \sin^2(\hat{y} + W^{(2)}x + b), \tag{5}$$

where $y \in \mathbb{R}^D$, the initial state $y(0) = u_i$ is the patch embedding, $W^{(1)}$ and $W^{(2)}$ are learnable weight matrices $\in \mathbb{R}^{D\times D}$, while $b \in \mathbb{R}^D$. The decay and growth terms respectively capture feature degradation caused by imaging noise and feature enhancement associated with pathological progression, while the bounded $\sin^2(\cdot)$ nonlinearity ensures smooth and stable feature evolution.

To exploit the continuity between original and evolved features, we construct positive pairs from the original embedding u_i and its nmODE-enhanced counterpart u_i', while features from different nodes serve as negative pairs. Furthermore, we adopt a hard negative mining strategy by selecting the top-K most similar negative samples based on cosine similarity within each mini-batch. The resulting contrastive loss encourages compact intra-class representations while enhancing discrimination among morphologically similar nodes:

$$\mathcal{L}_{\text{intra}} = -\log \frac{\exp(\text{sim}(u_i, u_i^{'})/\tau)}{\exp(\text{sim}(u_i, u_i')/\tau) + \sum_{k=1}^{K} \exp(\text{sim}(u_i, u_k^{'})/\tau)}. \tag{6}$$

Among them, $\text{sim}(\mathbf{u}, \mathbf{v}) = \mathbf{u}^\top \mathbf{v}/(\|\mathbf{u}\|\|\mathbf{v}\|)$ represents the cosine similarity, and $\tau > 0$ is the temperature hyperparameter. $u_k^{'}$ is the feature of the difficult negative sample selected through filtering.

This feature-space continuous augmentation preserves anatomical plausibility and enables the model to focus on fine-grained morphological cues critical for LN metastasis prediction.

D. Prediction with Inter-class Contrastive learning

Clinical diagnosis of CRC focuses on identifying a few critical abdominalpelvic LNs rather than treating all nodes equally. To achieve selective aggregation, we employ a gated attention–based pooling mechanism to fuse node embeddings into a patient-level representation. Given a node set $\mathcal{I} = \{LN_0, LN_1, \ldots, LN_n\}$ and their corresponding embeddings $\mathcal{Z} = \{z_0, z_1, \ldots, z_n\}$, the gated attention module assigns adaptive importance weights to each node and aggregates them as:

$$h_i = \sum_{i=0}^{n} g_i \odot a_i \cdot z_i, \tag{7}$$

where the attention score a_i and gate value g_i are computed as:

$$a_i = \text{softmax}\big(w^\top \sigma(V z_i^\top)\big), \tag{8}$$

$$g_i = \sigma_g\big(w^\top \sigma(V z_i^\top)\big). \tag{9}$$

Here, $g_i \in (0, 1)$ controls the information flow from each node, $\odot$ denotes element-wise multiplication, σ, σ_g represents the $tanh, sigmoid$ activation respectively, and $w \in \mathbb{R}^{n\times 1}$ and $V \in \mathbb{R}^{n\times D}$ are learnable parameters. This mechanism effectively suppresses noisy or less informative nodes and adaptively highlights diagnostically relevant nodes, mimicking clinical reading behavior.

To further improve inter-class discrimination at the patient level, we introduce supervised contrastive learning on the aggregated patient representations. Given a mini-batch of patient-level embeddings $X = \{\mathcal{H}\} = \{x_1, x_2, \ldots, x_N\}$ with corresponding labels $Y = \{y_1, y_2, \ldots, y_N\}$, embeddings sharing the same label are treated as positive pairs, while those with different labels are treated as negative pairs. The loss is defined as:

$$\mathcal{L}_{\text{inter}} = -\frac{1}{N} \sum_{m=1}^{N} \log \left(\frac{\sum_{k=1}^{N} \mathbb{I}_{[y_m = y_k]} \exp\left(\text{sim}(x_m, x_k)/\tau\right)}{\sum_{k=1}^{N} \exp\left(\text{sim}(x_m, x_k)/\tau\right)} \right), \tag{10}$$

where $\mathbb{I}$ is the indicator function, 1 for $y_m = y_k$ and 0 otherwise.

Finally, the patient-level embedding is fed into a linear classifier followed by a sigmoid function to produce the overall metastasis probability P_P, and the model is optimized using the loss $\mathcal{L}_{\text{cls}}(y, P_P)$ as defined in (3).

E. Bridging Representation to Classification

Inspired by curriculum learning, we design a joint optimization strategy that prioritizes learning meaningful and fine-grained representations before focusing on prediction accuracy, thereby alleviating overfitting caused by noisy or incomplete clinical annotations. The overall loss jointly optimizes the node–level and patient-level predictors and is defined as:

$$\begin{aligned} \mathcal{L}_{total} = \lambda_t \mathcal{L}_{\text{intra}} + (1-\lambda_t)\mathcal{L}_{\text{pre}}(y, P_N) \\ + \lambda_t \mathcal{L}_{\text{inter}} + (1-\lambda_t)\mathcal{L}_{\text{cls}}(y, P_P). \end{aligned} \tag{11}$$

Here, the weighting factor $\lambda_t \in [0, 1]$ decays over training iterations according to a curriculum schedule, enabling a

smooth transition from representation learning to classifier optimization.

Specifically, a linear decay strategy is adopted:

$$\lambda_t = \lambda_0 \times \left(1 - \frac{t}{T}\right), \tag{12}$$

where λ_0 is the initial decay coefficient, t denotes the current training epoch, and T is the total number of decay epochs. At early training stages, a larger λ_t emphasizes the intra- and inter-contrastive losses to encourage consistent node embeddings and discriminative patient-level embeddings. As training progresses, λ_t gradually decreases, shifting the optimization focus toward the prediction losses $\mathcal{L}_{\text{pre}}$ and $\mathcal{L}_{\text{cls}}$ to refine classification boundaries.

IV. Experiments

A. Experimental Settings

1) Dataset: This study employs a dataset from cooperative hospital, LNMPDataset, consisting of 517 preoperative abdominopelvic CT scans. Each CT scan was labeled by postoperative pathology, with LN positions annotated manually by CRC specialists. Among scans, 344 were negative (without LN metastasis), and 173 were positive (with LN metastasis). In total, 11,871 LNs were identified and 6,610 were mesangial LNs. We use 11,871 for pretrain and 6,610 for prediction task. Mesangial LNs are the primary drainage site for CRC and the first lymphatic spread station, critically influencing local tumor progression and staging. Notably, mesangial LNs are typically small, averaging 8.16 mm (Fig. 5). Additionally, we validated the model on the DSB2017 pulmonary nodule dataset, where candidate bounding boxes were generated using LeakyNet [18] and 8,552 nodules were obtained in 1,595 chest CT scans.

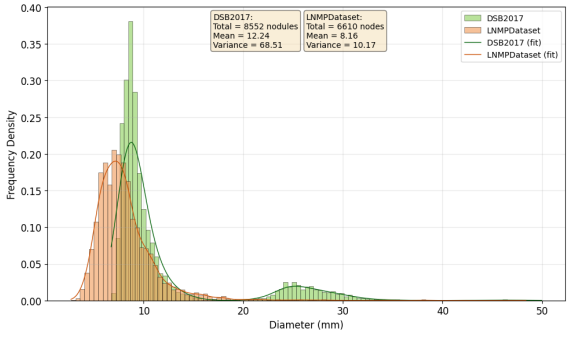

Fig. 5. Distribution of Diameter (mm) in LNMPDataset and DSB2017.

To ensure a more accurate evaluation of system performance, the LNMPDataset was stratified and divided into five subsets as shown in Table I, with cross-validation experiments conducted. We splited the DSB2017 dataset into training (1116), validation (319), and test (160) subsets.

2) Implementation Details: $LNs \times 3 \times 32 \times 32$ patches are cut with LNs as the center. We set $\tau = 0.4$, $D = 128$, $\alpha = 0.75$ (the balance factor in $\mathcal{L}_{cls}$ and $\mathcal{L}_{pre}$), $p_d = 0.99$, $K = 5$ and $\lambda_t = 1$ at the start of training. The Adam optimizer is employed with a weight decay of 0.01. The learning rate scheduler is configured to plateau with an initial learning rate of 1×10^{-5}. The training process is conducted over 200 epoches, using a mini-batch size of 16. Experiments are carried out on a Linux-based server equipped with an Intel Core i9-10900X processor (3.7 GHz), an NVIDIA RTX A4000 GPU, and 64 GB of RAM.

TABLE I
Descriptive statistics of five subsets.

Dataset	Patients (LNs)	Pos (LNs)	Neg (LNs)	Diameter
Subset 1	103 (1217)	36 (427)	67 (790)	3.40 - 41.74
Subset 2	103 (1218)	36 (488)	67 (730)	2.69 - 34.38
Subset 3	103 (1376)	36 (520)	67 (856)	3.00 - 48.37
Subset 4	103 (1420)	31 (413)	72 (1007)	3.33 - 50.38
Subset 5	105 (1379)	34 (564)	71 (815)	3.20 - 52.31
Total	517 (6610)	173 (2412)	344 (4198)	2.69 - 52.3

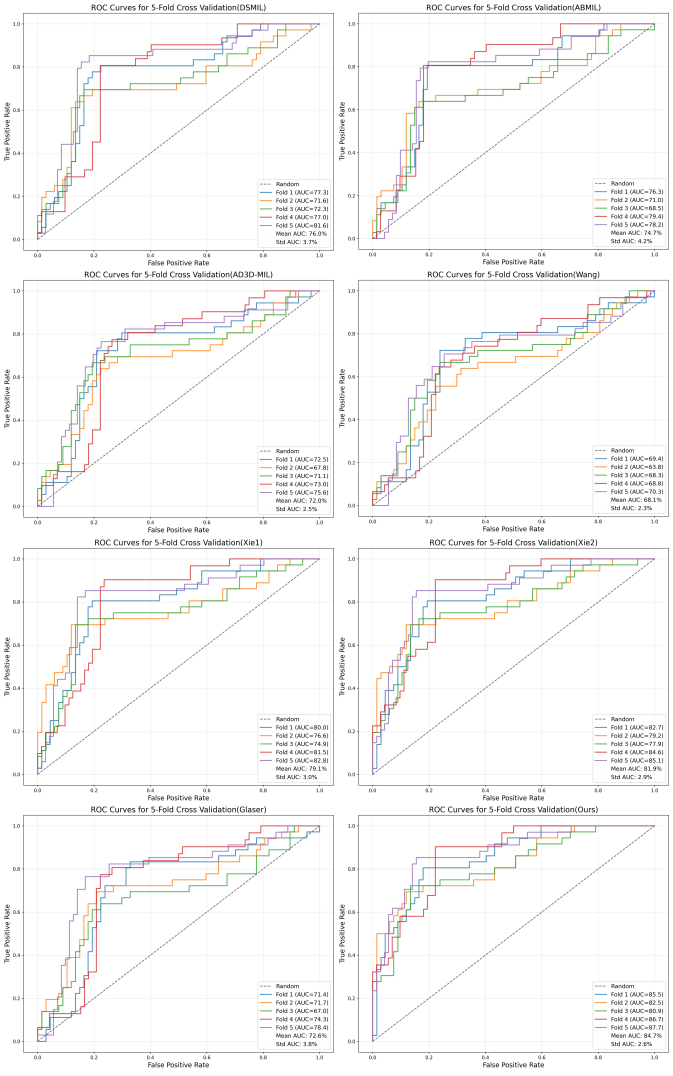

Fig. 6. ROC curves and AUC of ours and existing methods on LNMPDataset.

B. Performance Comparison with Existing Methods

We compare our method with representative weakly supervised methods. Table II shows our model consistently outperforms competitors across all metrics. Our method attained the highest accuracy (82.20%), precision (71.38%), and F1-score (74.89%). Notably, the achieved recall 79.57% and specificity 83.80% indicate a favorable balance between sensitivity and false-positive control, which is particularly critical for medical image analysis. Fig. 6 further demonstrates our superiority, achieving an average AUC of 84.68%, surpassing Xie et al. (81.95%) and DSMIL (76.06%).

TABLE II
COMPARISONS AMONG OURS AND EXISTING METHODS ON LNMPDATASET (MEAN±VARIANCE). BEST IN **BOLD**, SECOND UNDERLINED. (%)

Method	ACC	Prec.	Recall	Spec.	F1
DSMIL [21]	78.15±2.49	64.69±2.64	76.24±4.61	79.12±1.52	69.99±3.58
ABMIL [11]	77.87±3.47	64.96±1.76	73.34±5.44	80.14±3.58	68.93±2.53
AD3D-MIL [27]	72.19±2.53	56.41±2.85	73.82±4.53	71.37±2.55	63.95±3.72
Wang et al. [28]	69.25±2.32	53.09±2.46	68.45±3.55	69.67±3.45	59.78±3.24
Xie et al. [13]	80.90±3.49	69.15±2.73	77.32±4.43	82.69±3.54	73.01±2.57
Xie et al. [12]	80.94±3.47	70.23±3.67	74.55±5.44	**84.14**±3.48	72.33±3.52
Glaser et al. [29]	71.87±1.58	56.04±3.66	73.35±4.96	71.12±3.45	63.53±2.69
Ours	**82.20**±1.43	**71.38**±4.72	**79.57**±8.74	83.80±4.05	**74.89**±2.29

Table III presents the pulmonary nodule classification results on the DSB2017, where our method achieves an AUC of 88.95%. Although medical images exhibit significant intra-class variation and inter-class similarity, our framework effectively learns discriminative patient-level representations, remaining robust even with limited and noisy annotations.

TABLE III
COMPARISONS AMONG OURS AND EXISTING METHODS ON DSB2017. BEST IN **BOLD**, SECOND UNDERLINED. (%)

Method	AUC	ACC	Prec.	Recall	Spec.	F1
DSMIL [21]	84.62	78.75	60.98	60.47	86.32	60.47
ABMIL [11]	79.75	78.75	60.47	58.14	85.47	59.52
AD3D-MIL [27]	77.76	73.75	50.94	62.79	77.78	56.25
Wang et al. [28]	74.52	73.75	51.02	58.14	79.49	54.35
Xie et al. [13]	87.88	82.50	65.31	74.42	85.47	69.57
Xie et al. [12]	83.26	77.50	57.78	60.47	83.76	59.09
Glaser et al. [29]	83.52	77.50	55.56	**81.40**	76.07	66.04
Ours	**88.95**	**85.39**	**74.60**	67.30	**91.84**	**70.77**

C. Ablation Experiments

Table IV presents the results of ablation experiments, systematically validating the contribution of each module. Removing the pretraining strategy results in the most significant performance degradation (AUC drops from 84.68% to 76.42%), indicating its crucial role in learning robust representations. The architectural designs also prove essential, excluding the predicton with dummy node, feature fusion module, and 2.5D layers cause substantial AUC decreases of 6.95%, 6.87%, and 4.27%, respectively, demonstrating their effectiveness in handling annotation omission, aggregating anatomical prior information, and capturing contextual patterns. Meanwhile, both contrastive learning strategies provide complementary regularization, confirming that all components work synergistically to achieve the optimal performance. Importantly, when only node diameter is considered, the random forest classifier achieves merely 64.72% AUC, suggesting that morphological size alone is insufficient for accurate metastasis prediction and highlighting the indispensable role of imaging features.

TABLE IV
THE RESULTS OF BLATION EXPERIMENTS. THE FIRST LINE MEANS RANDOM FOREST. PRE. MEANS PRETAIN. FUSE MEANS FEATURE FUSION. PRED. MEANS PREDICTON WITH DUMMY NODE. (%)

Pre.	**2.5D**	**Fuse**	**Pred.**	**Intra**	**Inter**	AUC	ACC	Prec.	Recall	F1	Spec.
						64.72	69.23	58.82	28.57	38.46	89.86
✓	✓	✓	✓	✓	✓	84.68	82.20	71.38	79.57	74.89	83.80
×	✓	✓	✓	✓	✓	76.42	77.76	64.22	75.72	69.50	78.78
✓	×	✓	✓	✓	✓	80.41	79.69	67.35	76.30	71.54	81.40
✓	✓	×	✓	✓	✓	77.81	78.53	65.35	76.30	70.40	79.65
✓	✓	✓	×	✓	✓	77.73	76.69	67.53	75.72	71.39	81.69
✓	✓	✓	✓	×	✓	83.46	81.43	69.95	78.03	73.77	83.14
✓	✓	✓	✓	✓	×	82.40	80.85	69.07	77.46	73.02	82.56

D. Visualization Analysis

Fig. 7 shows four paired examples of patch embedding interpretability. The heatmaps consistently highlight the central regions of the nodes, with red/yellow core areas matching the brightest and most diagnostically relevant regions, while background regions appear in blue. Attention area scales with node size, and larger nodes exhibit broader attention distributions. When perinodal infiltration is present, the attention map extends toward the infiltrated region with a gradual yellow-to-cyan transition, and the heatmap boundaries closely align with the true anatomical contours of the node. These results confirm that the model focuses on pathological features rather than background or non-critical tissue, demonstrating the rationality and interpretability of our encoder.

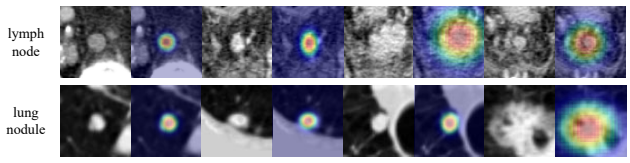

Fig. 7. Visualizing the patch embedding extraction.

V. CONCLUSION AND DISCUSSION

This paper proposes a joint prediction framework for abdominopelvic lymph node metastasis. The framework incorporates three distinct contrastive learning strategies: during pretraining, multi views of the same node are treated as positive pairs; during feature enhancement, features before and after nmODE are constructed as positive pairs; and at the patient level, samples sharing the same class label are regarded as positive pairs. The entire framework is optimized end-to-end using a joint training loss.

Extensive experiments demonstrate the superiority and robustness of the proposed framework, even though the complex abdominopelvic environment and annotation omissions. Notably, intelligent analysis of abdominopelvic lymph nodes remains limited due to the difficulty of acquiring high-quality annotations. In ongoing collaboration with clinicians, we are collecting approximately 400 cases with node-level annotations by combining intraoperative tissue marking and postoperative 3D reconstruction. Incorporating this curated dataset will further enhance the model's performance in future work.

ACKNOWLEDGMENT

We thank for the support of National Natural Science Foundation of China under Grant 62476183 and Natural Science Foundation from the Science Technology Department of Sichuan Province under Grant 2024NSFTD0051.

REFERENCES

[1] R. L. Siegel, N. S. Wagle, A. Cercek, R. A. Smith, and A. Jemal, "Colorectal cancer statistics, 2023," *CA: a cancer journal for clinicians*, vol. 73, no. 3, pp. 233–254, 2023.

[2] N. Hoshino, K. Murakami, K. Hida, T. Sakamoto, and Y. Sakai, "Diagnostic accuracy of magnetic resonance imaging and computed tomography for lateral lymph node metastasis in rectal cancer: a systematic review and meta-analysis," *International Journal of Clinical Oncology*, vol. 24, pp. 46–52, 2019.

[3] K. Ichimasa, S.-e. Kudo, H. Miyachi, Y. Kouyama, K. Mochizuki, Y. Takashina, Y. Maeda, Y. Mori, T. Kudo, Y. Miyata *et al.*, "Current problems and perspectives of pathological risk factors for lymph node metastasis in t1 colorectal cancer: systematic review," *Digestive Endoscopy*, vol. 34, no. 5, pp. 901–912, 2022.

[4] Y. Zhang, J. Li, X. Li, M. Xie, M. T. Islam, and H. Zhang, "FAOT-Net: A 1.5-stage framework for 3D pelvic lymph node detection with online candidate tuning," *IEEE Transactions on Medical Imaging*, vol. 43, no. 3, pp. 1180–1190, 2024.

[5] M. Khazaee Fadafen and K. Rezaee, "Ensemble-based multi-tissue classification approach of colorectal cancer histology images using a novel hybrid deep learning framework," *Scientific Reports*, vol. 13, no. 1, p. 8823, 2023.

[6] H. Wang, H. Huang, J. Wang, M. Wei, Y. Zhang, Z. Wang, and H. Zhang, "An intelligent system of pelvic lymph node detection," *International Journal of Intelligent Systems*, vol. 36, no. 8, pp. 4088–4116, 2021.

[7] S. Kudo, K. Ichimasa, B. Villard, Y. Mori, M. Misawa, S. Saito, K. Hotta, Y. Saito, T. Matsuda, K. Yamada, et al., "Artificial intelligence system to determine risk of T1 colorectal cancer metastasis to lymph node," *Gastroenterology*, vol. 160, no. 4, pp. 1075–1084, 2021.

[8] C. Wang, P. Yu, H. Zhang, X. Han, Z. Song, G. Zheng, G. Wang, H. Zheng, N. Mao, and X. Song, "Artificial intelligence–based prediction of cervical lymph node metastasis in papillary thyroid cancer with CT," *European Radiology*, vol. 33, no. 10, pp. 6828–6840, 2023.

[9] H. Cui, P. Xuan, Q. Jin, M. Ding, B. Li, B. Zou, Y. Xu, B. Fan, W. Li, J. Yu, et al., "Co-graph attention reasoning based imaging and clinical features integration for lymph node metastasis prediction," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2021, pp. 657–666.

[10] H. Li, Y. Wang, J. Zhu, D. Guo, Q. Yu, K. Yan, L. Lu, X. Ye, L. Zhang, Q. Wang, 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*, 2024, pp. 623–633.

[11] M. Ilse, J. Tomczak, and M. Welling, "Attention-based deep multiple instance learning," in *International Conference on Machine Learning*, 2018, pp. 2127–2136.

[12] M. Xie, Y. Zhang, X. Li, J. Li, X. Zou, Y. Mao, and H. Zhang, "An intelligent system of predicting lymph node metastasis in colorectal cancer using 3D CT scans," *International Journal of Intelligent Systems*, vol. 2024, no. 1, p. 7629441, 2024.

[13] M. Xie, Y. Zhang, X. Li, Y. Mao, X. Zou, and H. Zhang, "Predicting lymph node metastasis of colorectal cancer in CT scans using attention-based multiple instance learning," *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2023, pp. 2695–2701.

[14] Y. Mao, Y. Zhang, X. Zou, Y. Zheng, H. Huang, and H. Zhang, "Asymmetric matching in abdominal lymph nodes of follow-up CT scans," *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2025, pp. 35–44.

[15] Yishu Wei, Deng Y, Sun C, Lin M, Jiang H, Peng Y. "Deep learning with noisy labels in medical prediction problems: a scoping review," *Journal of the American Medical Informatics Association : JAMIA*, 31(7):1596–1607, 2024.

[16] L. Tan, H. Li, J. Yu, H. Zhou, Z. Wang, Z. Niu, J. Li, and Z. Li, "Colorectal cancer lymph node metastasis prediction with weakly supervised transformer-based multi-instance learning," *Medical & Biological Engineering & Computing*, vol. 61, no. 6, pp. 1565–1580, 2023.

[17] Y. Chen, Z. Shao, H. Bian, Z. Fang, Y. Wang, Y. Cai, H. Wang, G. Liu, X. Li, and Y. Zhang, "dmil-transformer: Multiple instance learning via integrating morphological and spatial information for lymph node metastasis classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 9, pp. 4433–4443, 2023.

[18] F. Liao, M. Liang, Z. Li, X. Hu, and S. Song, "Evaluate the malignancy of pulmonary nodules using the 3-D deep leaky noisy-OR network," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11, pp. 3484–3495, 2019.

[19] Yubin Zheng, Peng Tang, Tianjie Ju, Weidong Qiu, and Bo Yan. "Fine-grained contrastive learning for pulmonary nodule classification," *2024 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2024.

[20] Xiaoxian Xu, Ying Wei, Jie Zheng, Zhongxiang Ding, Yiqiang Zhan, Xiang Sean Zhou, Zhong Xue, Feng Shi, and Dinggang Shen. "Multi-scale supervised contrastive learning for benign-malignant classification of pulmonary nodules in chest ct scans," *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*, pages 1–4, 2023.

[21] B. Li, Y. Li, and K. W. Eliceiri, "Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 14318–14328.

[22] C. H. Perez, M. C. Escudero, S. Puig, J. Malvehy, and V. V. Besler, "Contrastive and attention-based multiple instance learning for the prediction of sentinel lymph node status from histopathologies of primary melanoma tumours," in *MICCAI Workshop on Cancer Prevention through Early Detection*, 2022, pp. 57–66.

[23] T. Zheng, K. Jiang, and H. Yao, "Dynamic policy-driven adaptive multi-instance learning for whole slide image classification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 8028–8037.

[24] K. Liu, W. Zhu, Y. Shen, S. Liu, N. Razavian, K. J. Geras, and C. Fernandez-Granda, "Multiple instance learning via iterative self-paced supervised contrastive learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 3355–3365.

[25] Z. Yi, "nmODE: neural memory ordinary differential equation," *Artificial Intelligence Review*, vol. 56, no. 12, pp. 14 403–14 438, Dec. 2023.

[26] D. Juyal, S. Shingi, S. A. Javed, H. Padigela, C. Shah, A. Sampat, A. Khosla, J. Abel, and A. Taylor-Weiner, "Sc-mil: Supervised contrastive multiple instance learning for imbalanced classification in pathology," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024, pp. 7946–7955.

[27] Z. Han, B. Wei, Y. Hong, T. Li, J. Cong, X. Zhu, H. Wei, and W. Zhang, "Accurate screening of covid-19 using attention-based deep 3d multiple instance learning," *IEEE transactions on medical imaging*, vol. 39, no. 8, pp. 2584–2594, 2020.

[28] H. Wang, H. Wang, L. Song, and Q. Guo, "Automatic diagnosis of rectal cancer based on ct images by deep learning method," in *2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. IEEE, 2019, pp. 1–5.

[29] S. Glaser, G. Maicas, S. Bedrikovetski, T. Sammour, and G. Carneiro, "Semi-supervised multi-domain multi-task training for metastatic colon lymph node diagnosis from abdominal ct," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2020, pp. 1478–1481.