LNMER-Net：A Metabolically Enhanced Lymph Node Metastasis Recognition Model Based on Lung Lymph Nodes and Microenvironment

Lingyun Wang[1], Huiyan Jiang[1,*], Yang Zhou[1], Qiu Luan[2], Bulin Du[2], Yaming Li[2], Xuena Li[2] and Yan Pei[3]

[1] Software College, Northeastern University, No.195, Chuangxin Road, Hunnan District, Shenyang 110169, China.
[2] Department of Nuclear Medicine, The First Affiliated Hospital of China Medical University, Shenyang 110001, China.
[3] Computer Science Division, The University of Aizu, Aizu-Wakamatsu, Fukushima 965-8580, Japan.
[*] Corresponding author: Huiyan Jiang (hyjiang@mail.neu.edu.cn)

Abstract. PET/CT is the preferred device for lung cancer and lymph node metastasis diagnosis, and mining effective features from PET/CT images to identify lung lymph node metastasis has important research significance and application value. Multi-phase PET/CT has temporal properties that can better represent changes in lesions' structural and metabolic properties. Early-phase PET images can show a wide range of lesion areas. Delayed-phase PET images can show the high uptake properties of ^{18}F-FDG in malignant tumor cells. Thus, multi-phase PET represents the variability of benign/malignant lesions better in the temporal dimension. This paper first proposes a metabolic enhancement method for lung lymph nodes and their microenvironment, a lymph node metastasis recognition network (LNMER-Net). The network has three branches: multi-modal early-phase feature fusion channel, multi-modal delayed-phase feature fusion channel, and single-modal metabolic decay channel. To enhance the feature of the lymph node region, a multi-receptive field-based feature extraction and feature space optimization (MRFO) method is proposed to extract lymph node features by multi-scale convolution operations and embed them in the multi-modal fusion channel. To exploit the information on the metabolic changes of the lesion in the early-phase and delayed-phase, differential results of the multi-phase PET images are fed into the single-modal metabolic decay channel to enhance the microenvironmental features. To verify its effectiveness, a multi-phase PET/CT dataset from China Medical University is used. The proposed method achieves 84.5%/82.9% in Accuracy/Recall, which is better than SOTA methods such as Res2Net, Comformer, and NextViT.

Keywords: Lung lymph node metastasis recognition, Metabolic enhancement, Multi-phase PET/CT, Feature optimization.

1 Introduction

Lung cancer is one of the most common malignancies worldwide and has a high mortality rate among cancers [1]. Cancer can metastasize to nearby lymph nodes, tissues, organs, and other parts of the body, and metastasis is a common cause of death from cancer [2,3]. Early detection of lung lymph nodes and rapid identification of their benign and malignant nature is crucial for patient survival [4]. Currently, in clinical practice, Computed Tomography (CT) and Positron Emission Tomography (PET) are important and advanced imaging tools for the diagnosis of cancer. CT imaging extracts detailed anatomical high-resolution information, and PET imaging extracts metabolic and functional information about organs [5]. Due to the variable signs and small diameter of lung lymph nodes [6], they are more disturbed by other normal tissues. Moreover, the amount of slice data obtained from PET and CT is huge, and it is time-consuming and difficult to ensure that small nodes are not missed if physicians directly analyze and identify the lesion areas in each slice [7]. Therefore, it is worthwhile to diagnose the lung lymph nodes accurately identified by computer based on PET/CT images.

Currently, the existing classification algorithms related to lung lymph nodes and pulmonary nodules can be divided into two main categories. One is based on traditional feature descriptors to extract shallow manual features to identify malignant and benign nodules; the other is to design various deep learning methods to extract deeper abstract semantic features of images for classification. The traditional methods extract features (including texture, shape, intensity, and morphological features) from nodules manually, reflecting the heterogeneity of nodules. Then feed the features into a classifier to predict the class of nodules. Many experimental results demonstrate that traditional methods can obtain good classification results [8-14].

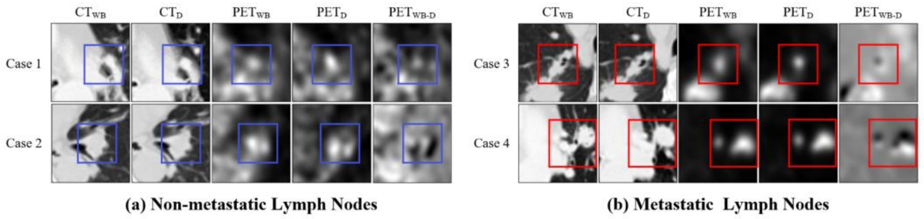
Despite the popularity of handcrafted features for classification, many limitations remain. For example, handcrafted features cannot fully characterize heterogeneous lung lymph nodes, and it can be difficult to filter key features among the numerous feature information. Researchers have recently started utilizing convolutional neural networks (CNNs) for medical image tasks. Compared with traditional methods, CNNs are automatic and adaptive models that learn highly discriminant features from various image data for classification, and omit feature design and other processes. Initially, Hua et al [15] showed that the classification of nodules in CT images using CNNs and deep confidence networks (DBNs) both outperformed traditional methods. He et al [16] proposed a deep residual network (ResNet), which introduced shortcut connections in deep learning models to alleviate the problem of excessive depth of neural networks. Huang et al [17] created dense convolutional networks (DenseNet) that can enhance feature propagation, generalize better, and prevent overfitting. Chen et al [18] proposed a dual path network (DPN) that incorporates ResNet, which focuses on feature reuse, and DenseNet, which focuses on feature generation. Gao et al [19] designed Res2net based on ResNet, which extracts global and local features of images more comprehensively and effectively through multi-receptive fields. In addition to CNNs, other neural network structures have been used in image classification studies, and Vaswani et al [20] proposed the Transformer method that relies on an attention mechanism to accomplish the classification task. Peng et al [21] fused CNN and Transformer models and proposed the first parallel hybrid network of the two, which combines the advantages of both and improves the classification effect without adding more computation. Li et al [22] designed a new hybrid model of CNN and Transformer, which model can achieve significant advantages in the classification task.

At present, the above studies are mainly based on single-phase CT images, and the existing works lack multi-modal multi-phase studies. To make lung lymph node identification more accurate and efficient, this paper combines the advantages of multi-modality and multi-phase, digs deeper into the temporal information, and introduces metabolic decay information into the network to better assist the network in diagnosis. The contributions of this paper are mainly in the following three aspects:

1. A novel model is proposed for multi-modal multi-phase data, involving three input channel branches, namely multi-modal early-phase feature fusion channel, multi-modal delayed-phase feature fusion channel, and single-modal metabolic decay channel.
2. A single-modal metabolic attenuation channel is proposed for PET images imaged using ^{18}F-FDG as a tracer, which is the first current branch of application for tumor characterization of multi-phase PET images.
3. A multi-receptive field-based feature extraction and feature space optimization method is designed, using convolutional blocks of multi-scales for feature extraction, and then filtering out the feature information that is more decisive for classification after corresponding operations.

2 Recognition of Lung Lymph Node Metastasis

CT images contain textural information about the lesion area, and PET images with ^{18}F-FDG as a tracer have metabolic information. Early-phase imaging and delayed-phase imaging have a temporal relationship, and the PET images showing metabolic information are significantly different.



(a) Non-metastatic Lymph Nodes

(b) Metastatic Lymph Nodes

Fig. 1. Schematic diagram of lung lymph nodes in different cases. (a) The diagram shows an example of non-metastatic lymph nodes in lung lymph nodes (benign). (b) The diagram shows an example of metastatic lymph nodes in lung lymph nodes (malignant). (a)(b) Each row of both figures shows the lymph nodes area of different patients (Case1-4), and each column from left to right shows the early-phase CT image, delayed-phase CT image, early-phase PET image, delayed-phase PET image, and early-phase PET - delayed-phase PET difference image, respectively, and the rectangular box indicates the lymph nodes area, blue is benign and red is malignant.

The difference between the early-phase image and the delayed-phase image is obtained as a differential image. Some of the lymph nodes are shown in Fig. 1. The different images visualize the metabolic differences in time series and show the microenvironmental information of metabolic decay. To better extract the lymph nodes and their surrounding features, this paper uses multi-modal multi-phase PET/CT images and multi-phase PET differential images.

2.1 LNMER-Net

To enhance the metabolic characteristics of lung lymph nodes and their microenvironment, this paper proposes the LNMER-Net, a lymph node metastasis recognition network model involving three channel branches. The network architecture is shown in Fig. 2 and is described as follows:

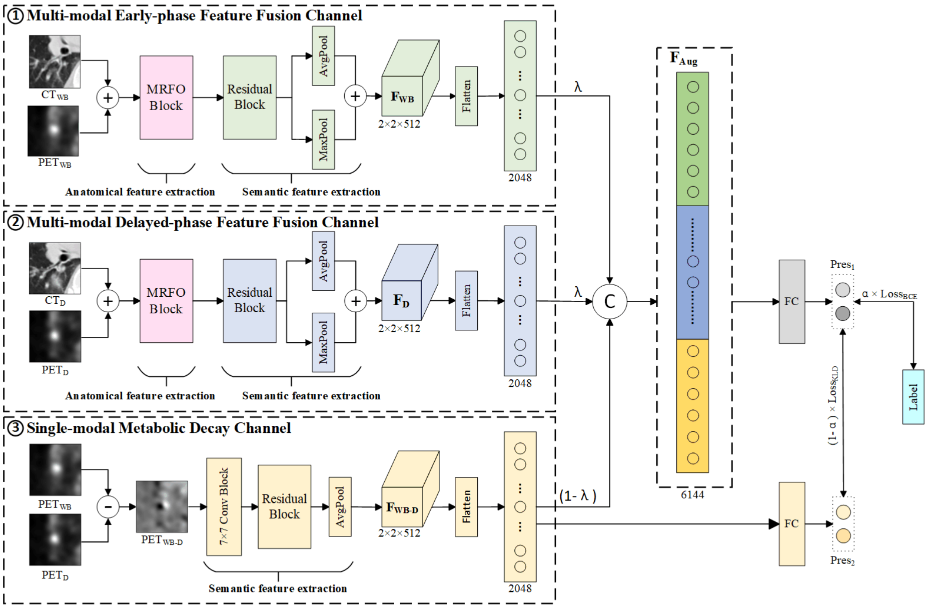


Fig. 2. Architecture of metabolically enhanced lymph node metastasis recognition network based on lung lymph nodes and their microenvironment (LNMER-Net)

① The first channel is called the multi-modal early-phase feature fusion channel, and the input of this channel is the fused image (I_{WB}) of early-phase PET (PET_{WB}) and early-phase CT (CT_{WB}). At first, this branch extracts the underlying anatomical features by the multi-receptive field-based feature extraction and feature spatial optimization method (MRFO) proposed in this paper. The shape features and texture features are extracted using multi-scale convolution blocks, and the features are spatially optimized by the corresponding operations. Then the deep semantic features are extracted after residual blocks, and flattened after compressing features in pooling layers to obtain early-phase lymph node features, denoted as F_{WB}.

② The second channel is the multi-modal delayed-phase feature fusion channel, and the input of this channel is the fused image (*I_D*) of delayed-phase PET (*PET_D*) and delayed-phase CT (*CT_D*). Similar to the multi-modal early-phase feature fusion channel, the input is flattened after the MRFO block, residual block, and pooling layer to obtain the delayed-phase lymph node feature, which is noted as *F_D*.

③ The third channel is a single-modal metabolic decay channel, where the input is the difference image (*I_{WB-D}*) between the early-phase PET (*PET_{WB}*) and the delayed-phase PET (*PET_D*), and the microenvironmental features are highlighted by metabolic decay. This metabolic decay branch focuses more on minute detail information, so MRFO containing multiple large convolutional blocks is not used. Structural features are extracted using a 7×7 convolution kernel, and deep semantic features are extracted using residual blocks, which are subsequently pooled and then flatten to obtain lymph node microenvironment features, denoted as *F_{WB-D}*. The features *F_{WB}*, *F_D*, and *F_{WB-D}* of the three channels are concatenated to obtain the region of interest enhancement features, denoted as *F_{Aug}*. The cascade is augmented by setting learnable adaptive weights in the network for the three features. The weights of *F_{WB}* and *F_D* are λ and the weight of *F_{WB-D}* is (1-λ). The *F_{Aug}* goes through the fully connected layer, resulting in the prediction *$Pred_1$*. The *F_{WB-D}* alone goes through the fully connected layer, resulting in the prediction *$Pred_2$*.

In this paper, the base-stem networks of all three paths use ResNet50 [16], which is the mainstream model in current classification networks with a wide range of applications and strong applicability. However, its receptive field is small and the long-distance dependence of spatial pixels is lost. Therefore, this method improves the lung lymph node classification task to make the network achieve better results.

2.2 Multi-receptive field-based feature extraction and feature space optimization method (MRFO)

Compared with small convolution kernels, large convolution kernels have larger receptive fields, and using large receptive fields has an irreplaceable effect compared to superimposing multiple small receptive fields. There is a higher shape bias and a stronger dependence on the target shape using large receptive fields, and a stronger texture bias and a higher dependence on the image texture for small receptive fields [23]. CT images can show clear texture information and PET images can show metabolic range. For the complex features of PET/CT fusion image information, this paper designs a feature extraction and feature space optimization module based on multi-receptive fields, denoted as MRFO. The architecture diagram is shown in Fig. 3.

The PET images (*I_{PET}*) and CT images (*I_{CT}*) are fused and inputted, and the image shape features and texture features are first extracted by a multi-receptive field-based feature extraction module, i.e., five convolutional blocks of different scales (*C_{1-5}*) and a dilation convolution (*DC*). The large convolution kernel has a large effective field of perception, can examine the feature map of a wider area, and the obtained features have global characteristics and pay more attention to the shape features of the fused image; the small convolution kernel is less computationally intensive, less likely to ignore local features, and focuses on the image texture information. Using multi-scale receptive fields, the original image is probed with multiple filters with complementary effective fields of view to obtain useful image contextual information at multiple scales, allowing feature extraction of regions of interest to ensure both shape features and texture features. Then the feature information extracted by multi-scale convolution is concatenated to obtain the feature F_C, as in *Eq.s* (1)-(3).

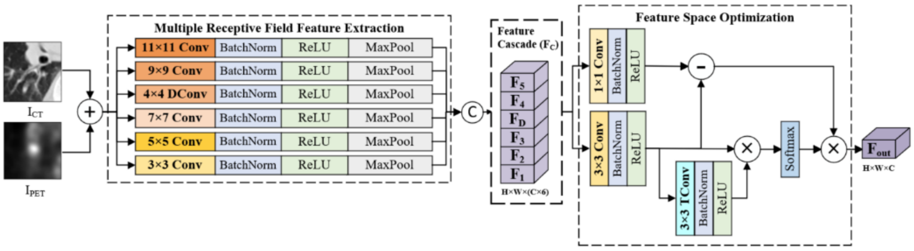


Fig. 3. Multi-receptive field-based feature extraction and feature space optimization module (MRFO) architecture diagram

The F_C enters the feature space optimization module and performs the corresponding operations in the spatial dimension to extract the spatial relationship features. The F_C is first convolved by 1×1 convolution ($C_{1\times1}$) to obtain feature F_{C1} and by 3×3 convolution ($C_{3\times3}$) to obtain feature F_{C2}, respectively. Feature F_2 is again convolved by 3×3 transposition ($TC_{3\times3}$) to obtain feature F_{C3}, as in *Eq.* (4). The F_{C2} is multiplied with the F_{C3} and then passed through Softmax and then multiplied with the difference between the F_{C1} and the F_{C2} to output feature F_{out}, as in *Eq.* (5). Among them, $C_{1\times1}$ serves to reduce the dimensionality and perform the convolution operation only in the channel direction to achieve the reduction of the number of channels and reduce the number of parameters without changing the other dimensional information. After $C_{3\times3}$, $TC_{3\times3}$ and their multiplication operations, the features are further extracted and the planar correlation of the features is found at the same time. The difference operation is performed between the F_{C1} and the F_{C2} to find the different features and filter out the features that are more decisive for subsequent image recognition. The MRFO is calculated as follows:

$$F_D = P_{max}(\sigma(BN(DC(I_{\mathrm{PET}} + I_{\mathrm{CT}})))) \tag{1}$$

$$F_n = P_{max}(\sigma(BN(C_n(I_{\mathrm{PET}} + I_{\mathrm{CT}})))) \tag{2}$$

$$F_C = Cat(F_1, F_2, F_3, F_4, F_5, F_D) \tag{3}$$

$$F_{C1} = C_{1\times1}(F_C),\ \ F_{C2} = C_{3\times3}(F_C),\ \ \ F_{C3} = TC_{3\times3}(F_{C2}) \tag{4}$$

$$F_{out} = (F_{C1} - F_{C2}) \odot Softmax(F_{C2} \odot F_{C3}) \tag{5}$$

where $C_n(\cdot)$, $C_{1\times1}(\cdot)$, $C_{3\times3}(\cdot)$ denote convolutional operations, $DC(\cdot)$ denotes dilated convolutional operation, BN denotes batch normalization operation, $\sigma(\cdot)$ denotes ReLU function, $P_{max}(\cdot)$ denotes maximum pooling, $Cat(\cdot)$ denotes concatenate operation, $TC(\cdot)$ denotes transposed convolutional operation, + denotes addition of the corresponding elements of an array, and ⊙ denotes multiplication of the corresponding elements of an array, $n \in [1,5]$.

3 Results and Discussion

3.1 Dataset and Preprocessing

The private dataset contains 99 lung lymph nodes from 51 patients from the Department of Nuclear Medicine, The First Hospital of China Medical University. All lung lymph node regions are manually outlined by experienced radiologists as the ground truth, and all lymph node metastases are determined by pathological puncture examination. The age of the patients ranged from 38 to 75 years, and the number of men and women was 28 and 23 cases, respectively. The PET image planar voxel size is 2.03642 mm × 2.03642 mm × 5.00 mm, the planar resolution is 400 × 400; the CT image planar voxel size is 0.976563 mm × 0.976563 mm × 2.00 mm, the planar resolution is 512 × 512. Each case had early-phase PET images and CT images, and delayed-phase PET images and CT images. The early-phase images were taken normally after the patient was injected with ^{18}F-FDG, and the delayed-phase images were taken about two hours after the early-phase medical images were taken.

The data is preprocessed, and the CT images are converted to Hu value truncation and then normalized. For PET images, resampling is first performed to rigidly align with CT images, which are truncated to SUV values and then standardized. PET and CT slices containing the lymph nodes are cropped to 64×64 centered on the lymph nodes according to the ground truth. A total of 959 sets of images (each set includes 4 images) are finally used for the experiments. 720 sets are used for the training set and 239 sets for the test set in the experiments. And data enhancement is performed on the data in the training set, with horizontal and vertical flips and random rotation of the images at certain angles.

3.2 Experimental Details and Evaluation Metrics

This work is conducted on an NVIDIA GeForce RTX 3060 12G server under Windows 11 operating system, based on the PyTorch framework. The Adam optimizer is set with a learning rate of 0.0001. The batch size is set to 8. The image data are preprocessed as described in Section 3.1. The network input size is 64 × 64, the network task is to determine whether the region is a lung lymph node metastasis, and the network output results in probability values for both categories. The experiments use loss functions including BCELoss, KLDivLoss, and total Loss, see *Eq.s* (6)-(8).

$$L_{BCE} = y_n \times \ln(x_n) + (1 - y_n) \times \ln(1 - x_n) \tag{6}$$

$$L_{KLD} = y_n(log y_n - x_n) \tag{7}$$

$$L_{\text{Total}} = \alpha * L_{BCE} + (1 - \alpha) * L_{KLD} \tag{8}$$

Where x_n denotes the predicted value, and y_n denotes the actual label. In the network, the enhancement feature F_{Aug} enters the fully connected layer to get the prediction result $Pred_1$ which is used with the true label to calculate L_{BCE} using *Eq.* (6). The microenvironmental feature F_{WB-D} enters the fully connected layer to get the prediction result $Pred_2$ which is used with $Pred_1$ which is treated as a weak label, to calculate L_{KLD} using *Eq.* (7). L_{BCE} and L_{KLD} set weights and sum to get L_{Total}, see *Eq.* (8). The network works best when $\alpha = 0.7$ by experiment.

3.3 Comparison with Other Methods

The lymph node metastasis recognition model based on the metabolic enhancement of lymph nodes and their microenvironment proposed in this paper is compared with other advanced networks. In ResNet50 [16], Densenet121 [17], DPN92 [18], Res2Net50 [19], Conformer [21] and Next-ViT [22] methods, multi-modal early-phase fusion images are concatenated with multimodal delayed-phase fusion images using a single branch channel input network for experiments. The results are shown in Table 1.

Table 1. Comparison with other advanced methods.

Method	Accuracy	F1-score	Precision	Recall
ResNet50 [16]	0.741	0.740	0.750	0.757
Densenet121 [17]	0.745	0.729	0.738	0.725
DPN92 [18]	0.778	0.768	0.771	0.766
Res2Net50 [19]	0.741	0.725	0.733	0.721
Conformer [21]	0.766	0.752	0.760	0.747
Next-ViT [22]	0.766	0.764	0.767	0.777
LNMER-Net (Ours)	**0.845**	**0.836**	**0.848**	**0.829**

Note: The best results are shown in bold black font.

As can be seen in Table 1, our proposed network can achieve the best results in terms of Accuracy, F1-score, Precision, and Recall, whether compared with ResNet50 [16], Densenet121 [17], DPN92 [18] base network, or Res2Net50 [19], Conformer [21], Next-ViT [22] the latest methods are improved compared with each other. The proposed method deeply explores the data features of the multi-modal single-phase and single-modal multi-phase and designs multi-modal feature extraction modules and metabolic attenuation channels that facilitate lesion identification and enhance lymph node features and microenvironmental features closely around lesion features. The current SOTA methods are poorly adapted to the data characteristics and lesion features and thus perform generally in the problem of lung lymph node metastasis identification.

3.4 Ablation Experiments

In this paper, a multi-channel branching network is designed to input multi-modal early-phase fusion images, multi-modal delayed-phase fusion images, and single-modal metabolic attenuation information into the network separately, and the extracted early-phase lymph node features, delayed-phase lymph node features, and microenvironmental features are enhanced in cascade. The early-phase and delayed-phase images extract information to enhance the lymph node features, and the multi-phase PET image difference highlights the lymph node metabolic attenuation information to enhance the microenvironmental features. For multi-modal images, PET images contain metabolic information but have low resolution, while CT images can more clearly represent the lymph nodes and the surrounding texture information, and the two are fused and input to the network to enhance the lymph node features. So, the MRFO is proposed to extract the shape and texture information from the corresponding fused images by using multi-scale receptive fields and to obtain the underlying anatomical structure features and enhance the lymph node features by filtering the feature information that is more decisive for classification through spatial optimization of the corresponding operations. In this paper, ablation experiments are performed on the network model, and the results are shown in Table 2 and Table 3.

Multi-channel Ablation Experiment. In this section, this work performs experimental comparisons by varying the number of branches of the input network, with the network branches cascaded before the fully connected layer. The experimental results for the single-branch network are shown in the first three rows of data in Table 2, the data for the two-branch network results are shown in rows 4-6 of Table 2, and the last row shows the experimental results using three channels.

Table 2. Ablation experiments of LNMER-Net under different channel inputs

Method	Channel			Accuracy	F1-score	Precision	Recall
	WB	D	WB-D				
1 Channel	√	-	-	0.766	0.759	0.757	0.760
	-	√	-	0.736	0.713	0.735	0.708
	-	-	√	0.669	0.592	0.697	0.609
2 Channels	√	√	-	0.828	0.814	0.839	0.805
	√	-	√	0.770	0.744	0.785	0.736
	-	√	√	0.791	0.772	0.800	0.764
3 Channels (Ours)	√	√	√	**0.845**	**0.836**	**0.848**	**0.829**

Note: The best results are indicated in bold black font, √ indicates that the branch is added to the network, and - indicates that the branch is not used.

From the data in the table, we can see that the network simultaneously sets up a multi-modal early-phase feature fusion channel, multi-modal delayed-phase feature fusion channel and single-modal metabolic decay channel to extract and enhance lymph node features and microenvironmental features to make the network fit better, and the best results can be obtained with Accuracy reaching 0.845 and F1-score reaching 0.836. Using the single-modal metabolic decay channel alone is the least effective, which indicates that the lymph node features extracted from multi-modal fused images are essential for image classification and that the network is not sufficient to accurately classify images by considering only learning the metabolic decay information in single-modal. The experimental results using dual channels show that enhancing lymph nodes or microenvironmental features improves the network effect.

MRFO Ablation Experiment. In this section, ResNet50 with three branch inputs is chosen as the baseline network (BL) for this work, and the results are shown in row 1 of Table 3. Each channel is selected to add or not an MRFO block before entering the residual block. Adding one MRFO block to the network results in rows 2-4 of Table 3, adding two MRFO blocks results in rows 5-7 of Table 3, and adding three MRFO blocks results in row 8 of Table 3.

Table 3. MRFO ablation experiment

Method	Channel			Accuracy	F1-score	Precision	Recall
	WB	D	WB-D				
ResNet50 (BL)	-	-	-	0.749	0.745	0.745	0.753
BL+MRFO (1)	√	-	-	0.770	0.744	0.785	0.736
	-	√	-	0.799	0.777	0.819	0.767
	-	-	√	0.766	0.752	0.760	0.747
BL+MRFO (2)	√	-	√	0.762	0.761	0.801	0.794
	-	√	√	0.824	0.820	0.818	0.826
	√	√	-	**0.845**	**0.836**	**0.848**	**0.829**
BL+MRFO (3)	√	√	√	0.736	0.724	0.727	0.723

Note: The best results are indicated in bold black font, √ indicates that the module is added to the network, and - indicates that no action is taken.

As can be seen from the data in the table, using MRFO in the two channels of multi-modal early-phase feature fusion and multi-modal delayed-phase feature fusion in the network gives the best results, with a 10.4% improvement in Accuracy and a 9.6% improvement in F1-score compared to BL. The priority of using MRFO in different branches also varies, with the most significant improvement in the multi-modal delay-phase channel. It is worth noting that using MRFO in all three channels becomes less effective instead. The reason is that for the single-modal metabolic decay channel, small and fine structural information needs to be learned from the input data, while the multi-modal early-phase channel and multi-modal delayed-phase channel have complex input fusion image information and need to extract lymph nodes and their background information. MRFO combines large receptive fields with small receptive field convolution kernels, which is more conducive to the extraction of structural information of complete large regions, and therefore the use of MRFO in multi-modal early-phase passages and multi-modal delayed-phase passages is more effective in enhancing the network.

In summary, the method proposed in this paper has good performance in all metrics, but there is still some space for improvement before it is put into clinical application. Therefore, to improve the accuracy and stability of the model performance, we need to further explore the definition of a priori features, effective feature extraction, and feature optimization, to build a model with excellent performance and realize intelligent diagnosis in the real sense.

4 Conclusion

In this paper, based on a multi-phase PET/CT dataset, the proposed network extracts multi-modal early-phase lymph node features, multi-modal delayed-phase lymph node features, and microenvironmental features. Applying the multi-modal multi-phase data, this method enhances the lymph node features while significantly enhancing the tiny fine microenvironmental features by metabolic attenuation information to assist the network in better classification. The multi-modal data combine the advantages of PET images containing metabolic information and CT images containing texture information to provide the network with more comprehensive and complementary features of lymph nodes. The proposed MRFO extracts complementary features from multi-modal fused images uses multi-receptive fields to extract more image contextual features, captures both global and local information of images, and optimizes shallow features to extract more effective deep semantic features. The experimental results show that the method in this paper improves by 6.7%, 6.8%, 7.7%, and 6.3% over the optimal method in Accuracy, F1-score, Precision, and Recall. In comparison with ResNet50 [16], with the addition of the proposed metabolic decay channel and MRFO module alone, Accuracy is improved by 7.9% and 9.6%, respectively. Therefore, the proposed network is superior and promising for research in lung lymph node classification.

Acknowledgement

This work was supported by the Natural Science Foundation of Liaoning Province (No.2021-YGJC-07).

References

1. Gridelli, C., Rossi, A., Carbone, D. P., Guarize, J., Karachaliou, N., Mok, T., Rosell, R.: Non-small-cell lung cancer. Nature reviews Disease primers. 1(1), 1-16 (2015).
2. Pham, T. D.: Classification of Benign and Metastatic Lymph Nodes in Lung Cancer with Deep Learning. In: 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE), pp. 728-733, IEEE, (2020).
3. Mehlen, P., Puisieux, A.: Metastasis: a question of life or death. Nature Reviews Cancer. 6(6), 449-458 (2006).
4. Klik, M. A. J., v Rikxoort, E. M., Peters, J. F., Gietema, H. A., Prokop, M., v Ginneken, B.: Improved classification of pulmonary nodules by automated detection of benign subpleural lymph nodes. In: 3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, pp. 494-497, IEEE, (2006).
5. Xia, X., Zhang, R.: A novel lung nodule accurate segmentation of PET-CT images based on Convolutional neural network and Graph Model. IEEE Access. 11, 34015-34031 (2023).
6. Rami-Porta, R., Asamura, H., Travis, W. D., Rusch, V. W.: Lung cancer—major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. CA: a cancer journal for clinicians. 67(2), 138-155 (2017).
7. Moltz, J. H., Bornemann, L., Kuhnigk, J. M., Dicken, V., Peitgen, E., Meier, S., Bolte, H., Fabel, M., Bauknecht, H., Hittinger, M., Kießling, A., Püsken, M., Peitgen, H. O.: Advanced

segmentation techniques for lung nodules, liver metastases, and enlarged lymph nodes in CT scans. IEEE Journal of selected topics in signal processing. 3(1), 122-134 (2009).

8. Madero Orozco, H., Vergara Villegas, O. O., Cruz Sánchez, V. G., Ochoa Domínguez, H. D. J., Nandayapa Alfaro, M. D. J.: Automated system for lung nodules classification based on wavelet feature descriptor and support vector machine. Biomedical engineering online. 14(1), 1-20 (2015).
9. Akram, S., Javed, M. Y., Hussain, A., Riaz, F., Usman Akram, M.: Intensity-based statistical features for classification of lungs CT scan nodules using artificial intelligence techniques. Journal of Experimental & Theoretical Artificial Intelligence. 27(6), 737-751 (2015).
10. Kaya, A., Can, A. B.: A weighted rule based method for predicting malignancy of pulmonary nodules by nodule characteristics. Journal of biomedical informatics. 56, 69-79 (2015).
11. De Carvalho Filho, A. O., Silva, A. C., Cardoso de Paiva, A., Nunes, R. A., Gattass, M.: Computer-aided diagnosis of lung nodules in computed tomography by using phylogenetic diversity, genetic algorithm, and SVM. Journal of digital imaging. 30, 812-822 (2017).
12. Li, X. X., Li, B., Tian, L. F., Zhang, L.: Automatic benign and malignant classification of pulmonary nodules in thoracic computed tomography based on RF algorithm. IET Image Processing. 12(7), 1253-1264 (2018).
13. Gong, J., Liu, J. Y., Sun, X. W., Zheng, B., Nie, S. D.: Computer-aided diagnosis of lung cancer: the effect of training data sets on classification accuracy of lung nodules. Physics in Medicine & Biology. 63(3), 035036 (2018).
14. Wu, W., Hu, H., Gong, J., Li, X., Huang, G., Nie, S.: Malignant-benign classification of pulmonary nodules based on random forest aided by clustering analysis. Physics in Medicine & Biology. 64(3), 035017 (2019).
15. Hua, K. L., Hsu, C. H., Hidayati, S. C., Cheng, W. H., Chen, Y. J.: Computer-aided classification of lung nodules on computed tomography images via deep learning technique. OncoTargets and therapy. 2015-2022 (2015).
16. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp. 770-778, IEEE, (2016)
17. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K. Q.: Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp. 4700-4708, IEEE, (2017).
18. Chen, Y., Li, J., Xiao, H., Jin, X., Yan, S., Feng, J.: Dual path networks. Advances in neural information processing systems. 30, (2017).
19. Gao, S. H., Cheng, M. M., Zhao, K., Zhang, X. Y., Yang, M. H., Torr, P.: Res2net: A new multi-scale backbone architecture. IEEE transactions on pattern analysis and machine intelligence. 43(2), 652-662 (2019).
20. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems. 30, (2017).
21. Peng, Z., Huang, W., Gu, S., Xie, L., Wang, Y., Jiao, J., Ye, Q.: Conformer: Local features coupling global representations for visual recognition. In: Proceedings of the IEEE/CVF international conference on computer vision, pp. 367-376, (2021).
22. Li, J., Xia, X., Li, W., Li, H., Wang, X., Xiao, X., Wang, R., Zheng, M., Pan, X.: Next-vit: Next generation vision transformer for efficient deployment in realistic industrial scenarios. arXiv preprint arXiv:2207.05501, (2022).
23. Ding, X., Zhang, X., Han, J., Ding, G.: Scaling up your kernels to 31×31: Revisiting large kernel design in cnns. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11963-11975, IEEE, (2022).