A Multi-Phase Multi-Graph Approach for Focal Liver Lesion Classification on CT Scans

Tran Bao Sam[1], Ta Duc Huy[2], Cong Tuyen Dao[1], Thanh Tin Lam[1], Van Ha Tang[1,3], and Steven Q.H. Truong[1]

[1] Vinbrain JSC, Hanoi, Vietnam
[2] Australian Institute for Machine Learning, University of Adelaide, Australia
[3] Department of Computer Science, Le Quy Don Technical University, Vietnam

Abstract. Liver cancer remains a leading cause of global mortality, driving interest in computer-aided diagnosis for liver tumor detection. Existing methods typically focus on individual lesions and avoid the impact of neighboring tumors on diagnostic accuracy. This study introduces a novel multi-phase multi-graph (MPMG) approach to improve liver tumor classification using contrast-enhanced computed tomography (CECT) scans. The MPMG method models inter-lesion relationships, including the ratio of diameters, semantic similarity, physical distance, and neighbor influence score as graph edge embeddings, while multiphasic features extracted from a proposed deep convolutional neural network form the node representations. By analysing different edge embedding formations, we find through extensive experiments that the proposed MPMG model outperforms several state-of-the-art methods in liver tumor diagnosis.

Keywords: Liver lesion classification · Computer-aided diagnosis · Graph neural networks.

1 Introduction

The classification and characterization of focal liver lesions (FLLs) are crucial for accurately diagnosing and determining the appropriate treatment for patients. Various medical imaging modalities, including computed tomography (CT), magnetic resonance imaging (MRI), ultrasound, and X-ray, are employed for liver evaluation. Among these, CT is the most prevalent and effective technique for detecting and characterizing FLLs due to its high-resolution and detailed cross-sectional imaging capabilities.

The contrast-enhanced computed tomography (CECT), or multi-phase CT imaging, is widely used in clinical practice to detect and diagnose focal liver lesions. CECT is typically divided into four phases: Non-contrast, arterial, portal venous, and delayed. Radiologists meticulously analyze these phases, often simultaneously, to identify and evaluate the informative regions of lesions for diagnosis. This process, however, is heavily reliant on the expertise of radiologists, making it inherently subjective and variable based on individual experience. To enhance the diagnostic process and reduce the workload on radiologists,

computer-aided diagnosis systems (CADx) have been developed. These systems aim to provide systematic and precise detection and classification of FLLs, particularly in the early stages of disease. CADx offers the potential to standardize diagnostic criteria, reduce human error, and improve overall diagnostic accuracy.

The evolution of methods in FLL classification has been marked by a significant transition from traditional techniques, which relied on predefined features, to machine learning algorithms, and more recently, the advent of deep learning approaches. These advanced techniques facilitate automatic feature extraction from CT images, thereby enhancing diagnostic accuracy and efficiency.

Radiomics [1], which involves the extraction of statistical and gray-level information from regions of interest (ROIs), has proven effective when integrated with machine learning predictors, underscoring the potential of combined approaches in medical image analysis [2–5]. [6] constructs a regression model to predict the recurrence and survival after curative liver resection for hepatocellular carcinoma (HCC) using clinical data. Similarly, [7] utilized clinical measurements, such as hemoglobin and hematocrit levels, as features for support vector machines and random forest classifiers. [8–10] defined a set of intensity and texture features from the local region of the lesions on images and used classifiers to identify lesion types. These methods laid the foundation for the more complex deep learning techniques that currently dominate the field of medical image analysis.

Due to the superiority of convolutional neural networks (CNNs), deep learning-based approaches are widely applied to various image processing tasks. Deep feature representation has shown its effectiveness compared to low- or mid-level features [11–18]. CNNs have revolutionized the field of medical image analysis [19–21] by enabling automatic and hierarchical feature extraction, which is particularly beneficial for complex tasks such as liver lesion classification. [18] introduced the deep learning-based framework ResGL-BDLSTM, which combines ResNet with a global and local pathway and a bi-directional long short-term memory (BD-LSTM) for the classification of FLLs in multi-phase CT images. [13] proposed an end-to-end hierarchical deep network to distinguish liver metastases from colorectal cancer and benign cysts in abdominal CT images of the liver. [14] deployed a weakly-supervised CNN for liver lesion localization without requiring pixel-level ground truth annotation. The inferred results from this CNN were then used to train a second CNN for liver lesion classification (healthy, cyst, hypodense metastasis). This method showcases the utility of weak supervision in reducing annotation costs while maintaining high classification performance. [16] combined collaborative learning and transfer learning for the liver lesion segmentation and classification problem. [17] incorporated the self-attention mechanism to enhance liver lesion classification tasks.

Despite the success of these methods, most existing approaches focus on analyzing individual lesions in isolation, without considering the potential influence of neighboring lesions. This limitation overlooks the interdependencies between multiple lesions, which can provide crucial diagnostic information. To address this gap, we propose a novel graph-based method that leverages a multi-phase, multi-graph neural network architecture. Our approach integrates inter-lesion

relationships across different phases of CT scans, incorporating factors such as similarity, physical distance, and predictive confidence. By capturing the dynamic changes and complex interactions between lesions, our method aims to significantly enhance the accuracy and efficiency of liver tumor diagnosis, providing a robust tool for clinical decision-making. The main contributions of this paper can be highlighted as follows:

1. We propose a multi-phase classifier that integrates lesion characteristics across different phases to identify FLLs, highlighting the importance of multi-phase imaging in liver tumor classification.
2. This study introduces a novel graph-based neural network architecture that leverages inter-lesion relationships within a patient, significantly enhancing the accuracy and efficiency of liver tumor diagnosis.
3. Our method mimics medical radiologists' workflows by utilizing characteristics and predictions of larger neighboring tumors to identify smaller tumors, reflecting real-world diagnostic practices.
4. Extensive experimental validation using abdominal CT scan datasets shows that our model significantly improves medical image diagnosis. It outperforms several existing methods in terms of F1-score and area under the receiver operating characteristic curve (AUC).

The organization of the paper is as follows. Section 2 describes our methodology. Section 3 presents the experimental results, comparisons with other state-of-the-art (SOTA) methods, and analyzes the effects of different components. Section 4 concludes our paper.

2 Proposed Method

This section presents a new multi-phase multi-graph (MPMGs) approach for classifying FLLs. Figure 1 illustrates an overview of our proposed method. The novelty of MPMGs lies in two aspects: i) the multi-phase CNN module extracts important features from CT volumes, and ii) the multi-graph model combines lesion information in cases to predict lesion pathology.

2.1 Data Pre-processing

For each lesion, we utilize a lesion mask and three corresponding volume images from the arterial phase, venous phase, and delay phase. These phases capture distinct aspects of the lesion's characteristics, providing a comprehensive view that enhances accurate classification. Abdominal CT images are typically high-resolution, sized at $(512 \times 512 \times C)$, where $C \in [100, 600]$ represents the number of slices. Including the original full shape directly in the model could be computationally expensive and may dilute the focus on the region of interest (ROI). Therefore, our initial step involves reprocessing the images to extract the ROI before feeding them into the CNN model. Each lesion is isolated using a lesion mask, and a bounding box is computed to encompass the lesion with additional

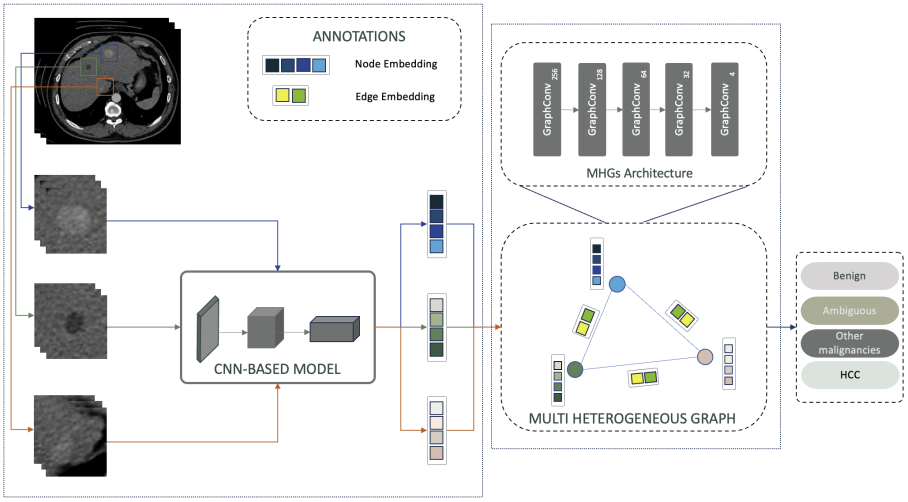

Fig. 1. An illustration of our approach for focal liver tumor classification. It consists of two main components: i) a multi-phase CNN module that extracts key features from CT volumes, and ii) Multi Heterogeneous Graphs model that combines lesion embeddings within cases to predict lesion pathology.

padding. Subsequently, we crop the three volume images based on this bounding box to extract the ROI, thereby reducing the input size and ensuring the model concentrates on the lesion and its immediate surroundings.

To further optimize the input data, we select the central slice from each of the three phases to create a 3-channel image representation. The intensity values of these cropped volumes are normalized to a standard range, and the 3-channel images are resized to fixed dimensions suitable for input into the CNN, ensuring uniformity across all samples. This multi-channel input effectively integrates complementary information from different phases, enhancing the model's ability to capture comprehensive lesion characteristics. Additionally, to boost the model's generalization capability, we employ data augmentation techniques such as rotation, translation, and flipping on the ROI volumes. These augmentations simulate various clinical scenarios and assist the model in learning robust features that are crucial for accurate liver tumor classification.

2.2 Multi-Phase Classification CNN-Based Approach (MP-CNN)

The goal of our MP-CNN approach is to create a mapping function from input medical images to diagnostic labels by leveraging the complementary information present in multi-phase data. The CNN multi-phase model processes each lesion independently, without considering the relationships between lesions within the same patient case. The architecture of the multi-phase network is designed to automatically extract features from images and then map these computed fea-

tures to predictions via two fully connected layers. We adopt the ConvNeXt architecture [22] as the feature extractor within the multi-phase model due to its superior performance in capturing complex patterns in medical images. The output of this model includes the sample's prediction (diagnostic label), the probability of the prediction (confidence score), and the extracted features. The extracted features from the multi-phase model serve as inputs for the subsequent multi-graph model, allowing for a comprehensive understanding of the underlying samples and their correlations, thereby facilitating more accurate and efficient liver tumor diagnosis. By leveraging the multi-phase data and advanced feature extraction capabilities of ConvNeXt, our approach ensures a detailed and robust classification of focal liver lesions, setting a strong foundation for further analysis in the multi-graph model.

2.3 Multi Heterogeneous Graphs Classifier (MHGs)

To enhance the robustness of the image classifier, we employ the MHGs to combine multiple features extracted from different tumors within a single patient case. Our MHGs provide two key benefits for multi-phase FLLs classification. Firstly, MHGs enable the connection of insights from all lesions, allowing for information sharing among lesions in the same case. This is particularly valuable when identifying small lesions in medical contexts. Small tumors, which might not exhibit common traits, are often identified based on their proximity to larger tumors, such as primary and satellite tumors. This ability to link lesions based on spatial relationships is a unique advantage that surpasses the limitations of previous methods, which were constrained by fixed inputs. Secondly, MHGs consider both physical distance, semantic similarity, and confidence scores among tumors to evaluate the influence of one central tumor on other tumors within the same case. Another strength of using a graph-based approach is that it allows our model to analyze the types of tumors that often co-occur. By leveraging these relationships, the multi-graph model enhances diagnostic accuracy and provides a more comprehensive understanding of the interdependencies between lesions.

Node Embeddings in the MHG-Based Model Let S denote the total number of samples in the dataset, represented by the collection of heterogeneous graphs $\{G_s\}_{s=1}^{S}$. Each sample G_s corresponds to a heterogeneous graph with its associated class label y_s. These graphs consist of a set of vertices N and a collection of edges E. The function of the multi-phase model is denoted by $f(x)$, where x represents each input lesion. Two primary types of nodes exist within a graph:

- Central node representation: $N_c(x) = f(x)$ represents the central node embedding, which corresponds to the feature representation of the central lesion extracted by the multi-phase model.
- Neighbor node representation: $N_n(x) = \{f(x_k)\}_{k=1}^{K}$ represents the feature representations of the top K nearest neighbor lesions for each central node,

with k indicating the rank of the neighbor based on certain criteria such as diameter ratio, physical distance, semantic similarity, or neighbor influence score.

Edge Embeddings in the MHG-Based Model For each lesion x, we have a lesion mask M_x and spatial resolutions (r_w, r_h, r_d) for the width, height, and depth of each voxel in the CT volume. The diameter of the lesion D_x is calculated using the following formula:

$$D_x = \max\left(\max_i \sum_{j,k} M_x(i,j,k)\cdot r_w, \max_j \sum_{i,k} M_x(i,j,k)\cdot r_d, \max_k \sum_{i,j} M_x(i,j,k)\cdot r_h\right)$$

In the proposed MHG-based model, the edge embeddings E represent four primary types of connections between the central node and its neighbor nodes. These embeddings capture various relationships, including the ratio of diameters, physical distances, semantic similarity, and neighbor influence score.

1. Diameter Ratio (E_1): The edge embedding E_1 is calculated as the ratio between the diameter of the central node D_c and the diameter of a neighboring node D_n. This ratio helps in understanding the relative size difference between lesions.

$$E_1 = \frac{D_c}{D_n}$$

2. Physical Distance (E_2): The edge embedding E_2 represents the physical distance between the central node and a neighboring node. Let $I_c = (I_c^x, I_c^y, I_c^z)$ and $I_n = (I_n^x, I_n^y, I_n^z)$ denote the center points of the central node and the neighbor node, respectively. The physical distance is computed by considering the Euclidean distance between these center points, adjusted for voxel dimensions (r_w, r_h, r_d), and subtracting the sum of the diameters of the central and neighbor nodes.

$$E_2 = \left(\sqrt{(I_c^x - I_n^x)^2 + (I_c^y - I_n^y)^2 + (I_c^z - I_n^z)^2}\right)\cdot\sqrt{r_w^2 + r_h^2 + r_d^2} - (\frac{D_c}{2} + \frac{D_n}{2})$$

3. Semantic Similarity (E_3): The edge embedding E_3 represents the semantic similarity between the central node and a neighboring node. This similarity is computed using the softmax of the negative cosine similarity between two nodes, F_c and F_n, extracted by the multi-phase model. The softmax function with scaled factor of 0.1 ensures that the similarities are normalized and converted into probabilities, which sum to 1. This helps in emphasizing the relative similarities of different neighboring nodes.

$$E_3 = \frac{\exp\left(-\frac{F_c\cdot F_n}{\|F_c\|\|F_n\|}\cdot\frac{1}{0.1}\right)}{\sum_{n'}\exp\left(-\frac{F_c\cdot F_{n'}}{\|F_c\|\|F_{n'}\|}\cdot\frac{1}{0.1}\right)}$$

4. Neighbor Influence Score (E_4): The edge embedding E_4 captures the effective level of influence of the neighbor on the central node by combining two

main metrics: the highest predicted probability of neighbor P_n and the similarity between the central and neighbor nodes E_3. E_4 reflects that when a neighbor node has both a higher confidence score (indicative of stronger prediction certainty) and a higher similarity with the central node, its influence on the central node is weighted more heavily.

$$E_4 = \frac{2 \times P_n \times E_3}{P_n + E_3}$$

These edge embeddings are then combined to form a comprehensive representation of the relationships between lesions within the same patient case. The final edge embedding E is given by:

$$E = [E_1, E_2, E_3, E_4]$$

By incorporating these diverse types of connections, the MHG-based model effectively captures the complex interdependencies between lesions, enhancing the robustness and accuracy of liver tumor diagnosis.

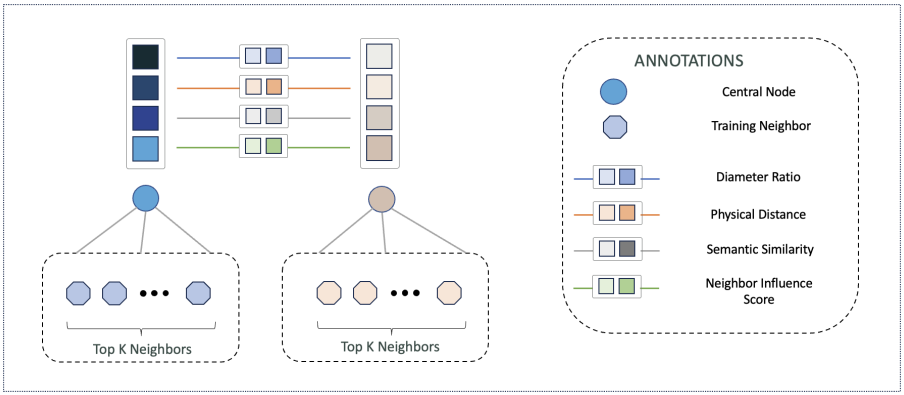

Fig. 2. Visualization of the edges between nodes in the proposed MHGs, illustrating the relationships between the central node and its top K neighbors.

MHGs Construction Strategy In this study, each patient case is represented by a heterogeneous graph (HG), constructed from the lesions within the case and their relationships. We employ two strategies to build a MHGs for FLLs classification.

First, nodes within each HG are connected to other nodes in the same case based on the four relationships outlined above. This intra-case connectivity captures the spatial and semantic relationships between lesions within a single patient.

Second, each node in the HG is also connected to nodes representing the top K nearest neighbor lesions in the training set. These training-set neighbors are selected by calculating the cosine similarity between the central node and every lesion in the training set. The K neighbors with the highest similarity, and a similarity score above a specified threshold, are chosen. If no neighbors in the training set exceed the threshold, the central node is considered to have no training-set neighbors. In this study, we set the similarity threshold to 0.5.

By combining these two strategies, the MHGs captures both intra-case and inter-case relationships, enhancing the model's ability to leverage comprehensive lesion information for accurate FLL classification. The visualization of our strategy is shown in Figure 2

MHGs Architecture In terms of architecture design, the proposed MHG-based model consists of a sequence of four GraphConv [23] layers with varying numbers of units: 256, 128, 64, and 32. The final GraphConv [23] layer has a number of units equivalent to the number of output classes, which are used to compute the prediction probabilities, thereby replacing the need for a fully connected layer. Figure 1 illustrates the detailed architecture of our model.

3 Experiments

3.1 Datasets

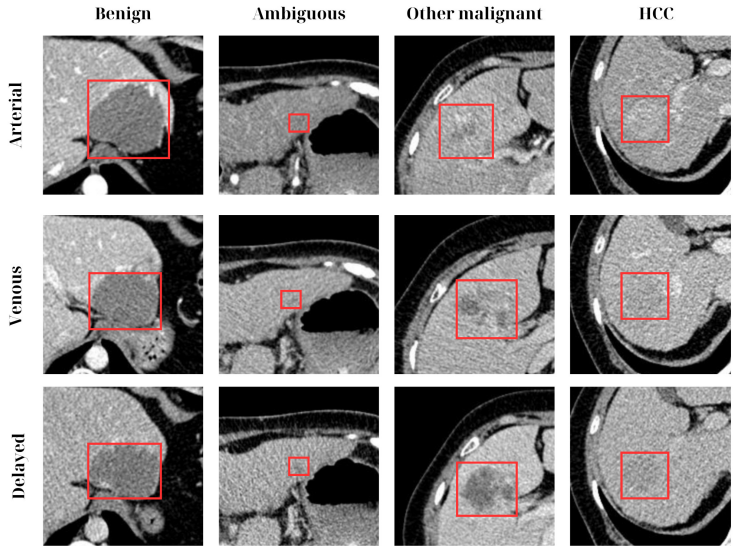

Fig. 3. Examples of four different lesion classes in FLLs classification: Benign (Cyst), Ambiguous (LI-RADS 3), Other Malignant (Metastasis), and HCC.

Group	Data		Number of Tumors				
	# Patients	# Samples	Benign	Ambiguous	Malignant	HCC	Total
Train	625	659	561	195	147	491	1394
Test	252	261	176	43	56	165	440
Total	877	920	737	238	203	656	1834

Table 1. The statistics of the liver tumor dataset.

Our method's performance is evaluated on a private medical dataset comprising 920 abdominal CT examinations conducted between 2016 and 2022. These exams originate from 877 patients, ranging in age from 5 to 96 years, and contain four distinct 3D CT images corresponding to different phases of contrast enhancement: Non-contrast, arterial, venous, and delayed phases. We stratify the dataset into two groups using label stratification: a training set (659 cases) and a test set (261 cases). Our classification categorizes lesions into four distinct classes based on their pathological characteristics: benign, ambiguous, malignant other than HCC, and HCC tumors.Examples of these four classes are shown in Figure 3. Benign lesions encompass a variety of liver tumors such as focal nodular hyperplasia (FNH), cyst, hemangioma, abscess, differential diagnosis (DD), and adenoma. The ambiguous label contains two types of liver tumors, namely LI-RADS 3 and ambiguous DD. The malignant group contains three types of tumors: intrahepatic cholangiocarcinoma (ICC), metastases, and malignant DD. Comprehensive dataset statistics are listed in Table 1.

3.2 Training and Evaluation Metrics

Our experiments are built on the PyTorch framework [24] and conducted on a single A100 GPU with 40GB RAM. We trained the proposed architecture using the following parameters: a batch size of 32, a default training epoch of 120, an SGD optimizer with a momentum of 0.9, a one-cycle learning rate scheduler, and the cross-entropy loss function.

We evaluated our models primarily in terms of F1-score, AUC, Precision, and Recall for the FLLs classification task on CT exams. The F1-score provides a balance between Precision and Recall, making it a suitable metric for our classification task where both false positives and false negatives are important. AUC is used to evaluate the model's ability to distinguish between classes, with higher values indicating better performance. Precision measures the accuracy of positive predictions, while Recall (or sensitivity) measures the ability to identify all relevant instances. These metrics together provide a comprehensive evaluation of our model's performance.

3.3 Experiment Results and Analysis

Our contribution is highlighted by two main experimental approaches. First, we compare single-phase and multi-phase models, demonstrating a significant

performance advantage for the latter. Second, we investigate the impact of incorporating MHGs within the multi-phase model, further enhancing its ability to identify FLLs.

Single and Multi Phase Approaches For the single-phase model, we trained an end-to-end model using the ConvNeXt architecture [22] as the main backbone. Each phase yielded a separate model capable of recognizing lesion labels based on that phase's volume. Three distinct models were developed to discriminate liver lesion types across the three different CT scan phases.

For the multi-phase methods, we adopted the intra-phase and inter-phase approaches introduced by Weibin Wang et al. [25]. In their work, lesion ROIs from non-contrast, arterial, and venous phases were used to predict early recurrence in HCC patients. To adapt this method to our task of FLLs classification and our specific dataset, we retained the phase attention module and adjusted the input phases and the output's model accordingly. The proposed MP-CNN models were trained and evaluated using the ConvNeXt architecture [22] as the feature extractor. Additionally, we employed the Vision Transformer (ViT) [26] as a feature extractor for comparison.

Phase	**Method**	**F1-score**	**AUC**	**Precision**	**Recall**
Single	Arterial	59.55	84.91	59.56	61.68
	Venous	58.79	82.78	62.68	61.93
	Delay	60.64	82.83	59.85	62.88
Multi	Phase Attention [25]	63.63	88.53	64.04	66.10
	MP-ViT	63.12	87.44	63.34	65.39
	MP-CNN (ConvNeXt)	**73.44**	**89.74**	**74.59**	**73.26**

Table 2. Performance comparison of single-phase and multi-phase methods for FLLs classification.

Table 2 summarizes the performance metrics achieved by the single-phase and multi-phase models. Single-phase models exhibit moderate performance, with the delay phase model achieving the highest F1-score of 60.64% and the arterial phase model achieving the highest AUC of 84.91%. This suggests that individual phases capture some discriminative features, but not enough for robust classification. In contrast, multi-phase models significantly outperform single-phase models. Among these methods, the MP-CNN model demonstrates the highest performance, achieving an F1-score of 73.44% and an AUC of 89.74%. This improvement over single-phase models is substantial, with F1-score increasing by approximately 13% and AUC by around 7%. These results highlight the importance of exploiting information across multiple phases for FLL classification, aligning with the multi-phase imaging protocols used in clinical practice.

While the phase attention model [25] and MP-ViT models also achieve good performance compared to single-phase models, the MP-CNN model surpasses them in terms of F1-score by a significant margin (approximately 10%). This

difference can be attributed to the nature of the input data. The phase attention model [25] modifies both input phases and output, potentially introducing additional complexity. Conversely, MP-ViT, despite its success in larger datasets, might not be fully optimized for this specific task size. CNN architectures often perform well on mid-sized datasets like ours, potentially explaining the superior performance of the MP-CNN model in this case.

MHGs Methods In this study, we conducted six main experiments to investigate the effectiveness of MHGs in the FLL classification task as outlined in Section 2.3. Five experiments were designed to explore intra-case lesion relationships using different criteria: diameter ratio (MHG-E_1), physical distance (MHG-E_2), semantic similarity (MHG-E_3), and influence score (MHG-E_4), along with a combined approach (MHG-E). Additionally, one experiment assessed the impact of highly similar neighbors from the training set, referred to as MHG-based intra-inter case (MHG-II). In the MHG-II setting, we selected $K = 7$, meaning each lesion was paired with its seven most similar training neighbors, based on cosine similarity, to provide more confidence in model predictions. This approach aimed to leverage both intra-lesion and inter-lesion relationships to improve diagnostic accuracy.

Incorporating the MHGs significantly enhances the performance of classification models. As demonstrated in Table 3, the F1-score improves from 73.44% with the MP-CNN (ConvNeXt) model to 85.08% when using MHG-II. Similarly, the AUC increases from 89.74% to 93.49%, showcasing the effectiveness of modeling intra-lesion and inter-lesion relationships.

Relationship	**Method**	**F1-score**	**AUC**	**Precision**	**Recall**
	MP-CNN (ConvNeXt)	73.44	89.74	74.59	73.26
Intra-case	MHG-E_1	82.36	92.64	84.01	81.06
	MHG-E_2	77.50	90.41	79.44	76.72
	MHG-E_3	79.57	91.23	81.10	78.51
	MHG-E_4	81.44	92.62	82.99	80.15
	MHG-E	83.55	93.39	86.17	81.75
Intra-Inter	MHG-II-E_1	82.37	93.04	83.56	81.50
	MHG-II-E_2	77.93	90.00	79.45	77.40
	MHG-II-E_3	82.08	90.89	83.64.45	80.81
	MHG-II-E_4	82.30	93.14	84.43	80.76
	MHG-II-E	**85.08**	**93.49**	**86.72**	**83.81**

Table 3. Performance comparison of different MHG-based construction approaches, including diameter ratio (E_1), physical distance (E_2), semantic similarity (MHG-E_3), and influence score (E_4), as well as a combined approach (E) and the MHG-based intra-inter case approach (MHG-II).

A detailed analysis of the results reveals the contributions of different types of embeddings used in the MHGs. E_1-embedding, which considers the ratio of

tumor diameters, performs the best among the four relationships, achieving an F1-score of 82.36% and an AUC of 92.64%. This suggests that the size relationship between tumors is a crucial factor in accurate FLL classification. Conversely, E_2-embedding, which measures physical distance, shows less effectiveness, with an F1-score of 77.5% and an AUC of 90.41%. This indicates that while spatial relationships are important, they may not be as critical as other factors. The semantic similarity (E_3) and neighbor influence score (E_4) also contribute positively to the model's performance, achieving F1-scores of 79.57% and 81.44%, and AUCs of 91.23% and 92.62%, respectively. These results highlight the importance of considering both the similarity in features and the influence of neighboring lesions in this task. The combined embedding ($E = [E_1, E_2, E_3, E_4]$) further enhances performance, with an F1-score of 83.55% and an AUC of 93.39%. This comprehensive representation captures the multifaceted relationships between lesions, leading to more robust and accurate predictions.

On the other hand, the combined inter-case information from training neighbors shows a significant enhancement in performance metrics, especially in F1-score and Recall. MHG-II demonstrates that leveraging both intra-case and inter-case relationships leads to superior diagnostic performance, achieving the highest F1-score of 85.08% and Recall of 83.81% among all methods tested. Addition, the improvement from 79.57% (MHG-$E3$) to 82.08% (MHG-II-$E3$) underscores how incorporating semantic similarity as an edge embedding in both intra-case and inter-case connections provides the model with greater insights.

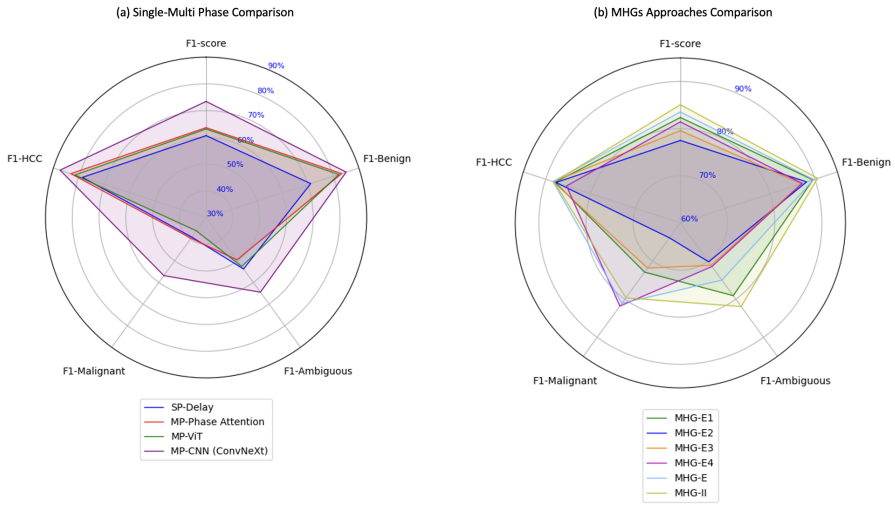

Fig. 4. F1-score comparison across four classes.

Figure 4 (a) illustrates the F1-scores for four classes in the test set, comparing single-phase and multi-phase models. The highest-performing single-phase model

is the delay phase (SP-Delay), which is compared against various MP models. Although MP models like MP-ViT and MP-Phase Attention outperform SP-Delay in identifying benign and HCC tumors, SP-Delay shows competitive results in other malignant and ambiguous classes. This can be explained by the need for multi-phase information in clinical practice to accurately identify HCC, as it provides evidence of tumor characteristic changes across phases, aiding in the distinction of HCC tumors from others. The analysis of the radar chart in Figure 4 (b) reveals significant improvements in classifying ambiguous and malignant tumors, with F1-score increases of approximately 15% and 24%, respectively. These results highlight the effectiveness of MHGs in handling tumors that frequently appear in clusters, such as metastases. Notably, MHG-E_4, despite not achieving the highest overall F1-score, excels in classifying malignant tumors, demonstrating the substantial impact of the neighbor influence score. This suggests that integrating neighbor influence can significantly enhance classification performance. The overall superior performance of MHG-II, leveraging both intra-case and inter-case relationships, confirms that a comprehensive approach yields the most robust diagnostic outcomes.

Overall, the MHG-based experiments underscore the importance of incorporating key lesion characteristics—such as tumor diameter, location, features, and confidence score—and their relationships, including diameter ratio, physical distance, semantic similarity, and influence score, in effectively identifying FLLs. The significant improvements in performance metrics demonstrate that MHGs provide a powerful tool for enhancing the accuracy and reliability of liver lesion classification models.

4 Conclusion

This paper introduces a novel MHGs-based approach for FLLs classification that leverages both intra- and inter-lesion relationships. Through extensive experimental validation on an internal dataset, we demonstrate that our proposed method significantly enhances diagnostic precision compared to existing single-phase and other multi-view techniques. The results underscore the effectiveness of modeling complex lesion relationships in improving classification accuracy.

Our study opens up a promising approach to constructing and developing efficient CADx systems that can perform well in practical clinical settings. The significant improvements observed with our MHGs-based approach suggest its potential to benefit real-world medical applications, providing more accurate and reliable diagnostic support to healthcare professionals.

Ongoing work involves exploring novel MHGs architectures for broader medical imaging analysis, focusing on other diseases such as mammary cancer, lung cancer, and brain cancer across various modalities, including mammography and MRI. Future research will particularly emphasize exploring lesion relationships and integrating expert knowledge to construct even more powerful and effective MHGs.

References

1. Lambin, P., Rios-Velazquez, E., Leijenaar, R., Carvalho, S., Van Stiphout, R.G.P.M., Granton, P., Zegers, C.M.L., Gillies, R., Boellard, R., Dekker, A., and others: Radiomics: extracting more information from medical images using advanced feature analysis. In: European Journal of Cancer 48(4), 441–446. Elsevier (2012)
2. Liu, X., Khalvati, F., Namdar, K., Fischer, S., Lewis, S., Taouli, B., Haider, M.A., Jhaveri, K.S.: Can machine learning radiomics provide pre-operative differentiation of combined hepatocellular cholangiocarcinoma from hepatocellular carcinoma and cholangiocarcinoma to inform optimal treatment planning?. In: European Radiology 31(1), 244–255. Springer (2021)
3. Wu, J., Liu, A., Cui, J., Chen, A., Song, Q., Xie, L.: Radiomics-based classification of hepatocellular carcinoma and hepatic haemangioma on precontrast magnetic resonance images. In: BMC Medical Imaging 19(1), 1–11. BioMed Central (2019)
4. Severn, C., Suresh, K., Görg, C., Choi, Y.S., Jain, R., Ghosh, D.: A Pipeline for the implementation and visualization of explainable machine learning for medical imaging using radiomics features. In: Sensors 22(14), 5205. MDPI (2022)
5. Tang, V.H., Duong, S.T.M., and others: Wavelet radiomics features from multiphase CT images for screening hepatocellular carcinoma: analysis and comparison. In: Nature Journal (2023)
6. Shim, J.H., Jun, M.J., Han, S., Lee, Y.J., Lee, S.G., Kim, K.M., Lim, Y.S., Lee, H.C.: Prognostic nomograms for prediction of recurrence and survival after curative liver resection for hepatocellular carcinoma. In: Annals of Surgery 261(5), 939–946. LWW (2015)
7. Wibowo, V.V.P., Rustam, Z., Hartini, S., Setiawan, Q.S., Aurelia, J.E.: Comparison between Support Vector Machine and Random Forest for Hepatocellular Carcinoma (HCC) Classification. In: International Conference on Decision Aid Sciences and Application (DASA) 2020, pp. 618–622. IEEE (2020)
8. Gletsos, M., Mougiakakou, S.G., Matsopoulos, G.K., Nikita, K.S., Nikita, A.S., Kelekis, D.: A computer-aided diagnostic system to characterize CT focal liver lesions: design and optimization of a neural network classifier. In: IEEE Transactions on Information Technology in Biomedicine 7(3), 153–162. IEEE (2003)
9. Huang, Y.L., Chen, J.H., Shen, W.C.: Diagnosis of hepatic tumors with texture analysis in nonenhanced computed tomography images. In: Academic Radiology 13(6), 713–720. Elsevier (2006)
10. Mougiakakou, S.G., Valavanis, I.K., Nikita, A., Nikita, K.S.: Differential diagnosis of CT focal liver lesions using texture features, feature selection and ensemble driven classifiers. In: Artificial Intelligence in Medicine 41(1), 25–37. Elsevier (2007)
11. Nguyen, H.T.X., Tran, S.B., Nguyen, D.B., Pham, H.H., Nguyen, H.Q.: A novel multi-view deep learning approach for BI-RADS and density assessment of mammograms. In: 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (2022)
12. Cao, S.E., Zhang, L.Q., Kuang, S.C., Shi, W.Q., Hu, B., Xie, S.D., Chen, Y.N., Liu, H., Chen, S.M., Jiang, T., Ye, M., Zhang, H.X., Wang, J.: Multiphase convolutional dense network for the classification of focal liver lesions on dynamic contrast-enhanced computed tomography. In: World Journal of Gastroenterology 26(25), 3660 (2020)
13. Romero, F.P., Diler, A., Bisson-Gregoire, G., Turcotte, S., Lapointe, R., Vandenbroucke-Menu, F., Tang, A., Kadoury, S.: End-to-end discriminative deep network for liver lesion classification. In: IEEE 16th International Symposium on Biomedical Imaging (ISBI) 2019, pp. 1243–1246. IEEE (2019)

14. Shapira, N., Fokuhl, J., Schulthei, M., Beck, S., Kopp, F.K., Pfeiffer, D., Dangelmaier, J., Pahn, G., Sauter, A.P., Renger, B.: Liver lesion localization and classification with convolutional neural networks: a comparison between conventional and spectral computed tomography. In: Biomedical Physics & Engineering Express 6(2), 025016. IOP Publishing (2020)
15. Alahmer, H., Ahmed, A.: Computer-aided classification of liver lesions from CT images based on multiple ROI. In: Procedia Computer Science, vol. 90, pp. 80–86. Elsevier (2016)
16. Heker, M., Greenspan, H.: Joint liver lesion segmentation and classification via transfer learning. In: arXiv preprint arXiv:2004.12352 (2020)
17. Chen, X., Lin, L., Liang, D., Hu, H., Zhang, et al.: A dual-attention dilated residual network for liver lesion classification and localization on CT images. In: International Conference on Image Processing (ICIP), pp. 235–239. IEEE (2019)
18. Liang, D., Lin, L., Hu, H., Zhang, Q., Chen, Q., Han, X., Chen, Y.-W., et al.: Combining convolutional and recurrent neural networks for classification of focal liver lesions in multi-phase CT images. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 666–675. Springer (2018)
19. Huy, T.D., Huyen, H.C., Nguyen, C.D.T., Duong, S.T.M., Bui, T., Truong, S.Q.H.: Adversarial contrastive Fourier domain adaptation for polyp segmentation. In: IEEE 19th International Symposium on Biomedical Imaging (ISBI) 2022, pp. 1–5. IEEE (2022)
20. Quan, T.M., Thanh, H.M., Huy, T.D., Chanh, N.D.T., Anh, N.T.P., Vu, P.H., Nam, N.H., Tuong, T.Q., Dien, V.M., Van Giang, B., and others: XPGAN: X-ray projected generative adversarial network for improving Covid-19 image classification. In: IEEE 18th International Symposium on Biomedical Imaging (ISBI) 2021, pp. 1509–1513. IEEE (2021)
21. Huynh, T.M., Nguyen, C.D.T., Huy, T.D., Huyen, H.C., Bui, T.H., Truong, S.Q.H., VinBrain, JSC, Park, V., and others: Diffeomorphism Matching for Fast Unsupervised Pretraining on Radiographs. In: 32nd British Machine Vision Conference (2021)
22. Liu, Z., Mao, H., Wu, C.-Y., et al.: A ConvNet for the 2020s. In: The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022)
23. Morris, C., Ritzert, M., Fey, M., et al.: Weisfeiler and Leman go neural: Higher-order graph neural networks. In: Association for the Advancement of Artificial Intelligence (AAAI) (2019)
24. Paszke, A., Gross, S., Massa, F., et al.: PyTorch: An imperative style, high-performance deep learning library. In: Advances in Neural Information Processing Systems 32, pp. 8024–8035. Curran Associates, Inc. (2019)
25. Wang, W., Wang, F., Chen, Q., Ouyang, S., Iwamoto, Y., Han, X., Lin, L., Hu, H., Tong, R., Chen, Y.-W.: Phase attention model for prediction of early recurrence of hepatocellular carcinoma with multi-phase CT images and clinical data. In: Frontiers in Radiology, vol. 2, pp. 856460. Frontiers (2022)
26. Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. In: International Conference on Learning Representations (ICLR) (2021)