

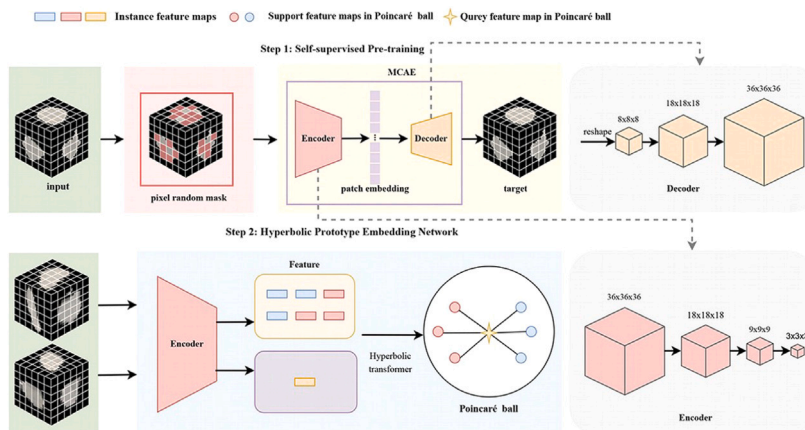
# Occult lymph node metastasis prediction in non-small cell lung cancer based self-supervised pretrained and hyperbolic theory<sup>☆</sup>

Haizhou Xu<sup>a,1</sup>, Jiaqi Wu<sup>b,1</sup>, Yujia Yu<sup>a</sup>, Wenkai Huang<sup>a,\*</sup>, Jiong Ni<sup>b,\*</sup>

<sup>a</sup> School of Mechanical and Electrical Engineering, Guangzhou University, Guangzhou 510006, China

<sup>b</sup> Department of Radiology, Tongji Hospital, School of Medicine, Tongji University, Shanghai, 200065, China

## GRAPHICAL ABSTRACT



In this research, we proposed a two-step method called OLNNet to enhance the learning and extraction of occult lymph node image features for prediction. The architecture of the OLNNet network for NSCLC occult lymph node metastasis prediction consists of two parts: MCAE (Step 1) and HPENet (Step 2). MCAE is designed to initially extract image features from the lymph nodes, while HPENet serves as a fine-tuning network to achieve the final prediction results.

## ARTICLE INFO

### Keywords:

Occult lymph nodes  
Bio-medical imaging  
Non-small cell lung cancer  
Hyperbolic space  
Few-shot learning

## ABSTRACT

Predicting occult lymph node metastasis in non-small cell lung cancer (NSCLC) patients is pivotal for tailoring appropriate surgical and therapeutic interventions. This prognostic factor remains underexplored, largely due to the intricate variability of occult lymph node characteristics and the absence of a pathologically confirmed predictive dataset. Addressing this gap, we retrospectively assembled a dataset of occult lymph node metastases (TJ-OLNM) from NSCLC patients who underwent chest Computed Tomography (CT) scans at Tongji Hospital, Tongji University from 2016 to 2021. Utilizing this dataset, we developed a novel self-supervised learning model, the Occult Lymph Node Metastasis Network (OLNM-Net), which leverages hyperbolic metric few-shot learning to enhance the prediction accuracy of occult metastases. Our comprehensive evaluations demonstrate that OLNNet significantly outperforms existing models in predicting occult lymph node metastasis, offering new insights into the preoperative assessment of NSCLC and advancing the application of machine learning in medical diagnostics.

<sup>☆</sup> This work was supported by the Guangzhou Youth Science and Technology Education Project under Grant KP2024403.

\* Corresponding authors.

E-mail addresses: [smallkat@gzhu.edu.cn](mailto:smallkat@gzhu.edu.cn) (W. Huang), [ni.jiong@163.com](mailto:ni.jiong@163.com) (J. Ni).

<sup>1</sup> Contributed equally to this work.

## 1. Introduction

Lung cancer is among the most prevalent and deadly diseases worldwide [1], with non-small cell lung cancer (NSCLC) being the most common type [2]. NSCLC is characterized by its rapid progression, poor prognosis, and tendency to develop chemotherapy resistance. Lymph node metastasis is the most common route of metastasis in NSCLC. On chest CT scans, lymph nodes with a short axis of less than 10 mm and confirmed as metastatic after surgery are defined as occult lymph node metastasis [3]. Due to the presence of occult lymph node metastasis, conventional examination methods have limited value in determining lymph node staging. Predicting the occurrence of occult lymph node metastasis before surgery is of paramount clinical significance in accurately assessing regional lymph node metastasis and assisting clinicians in selecting the most appropriate treatment plan for NSCLC patients.

Currently, preoperative assessment of lymph node metastasis in NSCLC patients includes CT, positron emission tomography/X-ray computed tomography (PET/CT), ultrasound guided fine needle aspiration biopsy, and mediastinoscopy [4–7]. CT is the most widely used detection method in clinical practice. However, conventional CT mainly evaluates the lymph node metastasis status through morphological indicators, usually using a short-axis diameter of 10 mm as the evaluation indicator, resulting in a single quantitative indicator with limited sensitivity and specificity. PET/CT can better evaluate the metastasis status of lymph nodes. However, NSCLC patients often have complications, which can readily result in false-negative results [8,9], and the image acquisition scanning is time-consuming with a high radiation dose and high cost, making it difficult for patients [10]. Ultrasound-guided fine needle aspiration biopsy and mediastinoscopy, as invasive procedures, can cause various complications during surgery that can have a negative impact on patients without lymph node metastasis [11]. However, these methods for preoperative lymph node diagnosis are limited. There are instances where lymph nodes have postoperative pathologically confirmed metastases despite showing no suspicious hints and having a short diameter (less than 10 mm) on preoperative CT or PET/CT imaging, referred to as occult lymph nodes. Therefore, it is crucial to identify a reliable imaging method to predict occult lymph node metastasis in NSCLC patients.

In recent years, the rapid development of artificial intelligence has significantly impacted medical research, particularly through the introduction of machine learning methods which are revolutionizing how medical data is analyzed and interpreted [12–15]. These methods can effectively manage and analyze high-dimensional and complex data sets from imaging and clinical records, providing a rich exploratory feature space that enhances diagnostic accuracy and predictive capabilities. Zhong et al. [16] exemplify the use of machine learning in medical diagnostics by integrating imaging features with clinical data such as tumor pathological subtype, diameter, and location. They developed a multi-parameter logistic regression model to predict mediastinal lymph node metastasis in lung adenocarcinoma. This study is particularly noteworthy for its approach to combining various types of data into a single predictive model, demonstrating the potential to increase diagnostic precision beyond the use of imaging data alone. The integration of different data types helps in capturing a more comprehensive picture of the tumor environment, potentially leading to better stratification of patients and more personalized treatment plans. Cong et al. [17] advanced this methodology by employing the least absolute shrinkage and selection operator (LASSO) [18] to constrain logistic regression, which they applied to both qualitative CT morphological features and radiomics features to predict lymph node metastasis in NSCLC patients. The use of LASSO helps in refining the model by selecting only the most relevant features, thereby reducing the risk of overfitting and improving the model's generalizability. This approach underscores the critical role of feature selection in developing robust predictive models in medical imaging, particularly in handling datasets with a high degree of multicollinearity and noise. Ferreira Junior et al. [19] took

a slightly different approach by utilizing multiple machine learning models to predict the histopathology and metastasis of lung cancer using radiomics CT features. Their research highlights the advantages of an ensemble learning approach, where multiple models are used in conjunction to provide a consensus prediction. This method can be particularly effective in medical applications, as it typically yields more stable and reliable predictions than single-model approaches. By leveraging the strengths of various algorithms, ensemble methods can enhance predictive performance, especially in complex diagnostic tasks where different models may capture different aspects of the data. These studies collectively illustrate the diverse methodologies and innovative applications of machine learning in enhancing the diagnosis and treatment of lung cancer. They highlight not only the capability of machine learning to improve the accuracy and efficiency of medical diagnostics but also emphasize the ongoing need for new techniques and models that can handle increasingly complex and voluminous data sets. The challenge now lies in integrating these advanced computational methods into clinical workflows in a way that is feasible, efficient, and maximizes patient benefit, thereby transforming theoretical innovations into practical tools for medical practitioners. However, the studies referenced earlier have exclusively focused on the research of standard lymph node metastasis, neglecting the exploration of predicting occult lymph node metastasis, which presents distinct challenges. Occult lymph node metastasis typically lacks the clear characteristics often associated with conventional lymph node metastasis, making it particularly difficult to detect these metastatic conditions through standard preoperative imaging techniques. This lack of evident features requires more sophisticated diagnostic approaches to identify subtle signs of disease progression. Additionally, the relative rarity of occult lymph node metastasis cases further complicates research efforts, as it results in a scarcity of training data. This limited data availability significantly hampers the development and application of fully supervised deep learning models that rely on large, comprehensive datasets to achieve high accuracy and reliability. The need for innovative solutions to overcome these obstacles is critical for advancing the capabilities of medical diagnostic technologies.

To tackle the issue of limited data availability, researchers have increasingly turned to few-shot learning (FSL) techniques in recent years [20–24]. The primary goal of FSL is to maximize the utility of small datasets by extracting highly informative features that enhance the model's training process, thereby improving model performance. This approach is especially crucial in contexts where data collection is challenging or costly, such as medical imaging. Traditionally, most FSL methods depend on a substantial amount of annotated data from known classes to train models that can effectively generalize to new, unseen classes. These methods are designed to create robust image representations that can predict accurately across different scenarios, bridging the gap between limited training data and the diverse range of real-world applications. However, this reliance on extensive annotated datasets presents a significant challenge in the medical field, where such data is often scarce and expensive to acquire. Annotating medical images requires expert knowledge from experienced clinicians, making the process both time-consuming and costly. Moreover, the variability in medical conditions and the subtle nuances in imaging that need to be recognized make the task even more complex. Given these challenges, there is a pressing need for innovative FSL approaches that can work effectively with minimal annotated data. The development of such methods would not only be more practical for medical applications but could also lead to breakthroughs in how machine learning models are trained to handle sparse and highly specialized datasets. By enhancing the efficiency of data usage in the training process, these advanced FSL methods could significantly improve the accuracy and reliability of diagnostic models, ultimately leading to better patient outcomes in medical fields like oncology, where early and precise diagnosis is crucial.

To address the above challenges associated with predicting occult lymph node metastasis, including the gaps in previous research in this field, the scarcity of imaging datasets, and the difficulties of accurate prediction with small samples, we have retrospectively collected and accurately annotated a dataset of pathologically confirmed occult lymph node metastases. Furthermore, we propose a novel few-shot learning model for predicting occult lymph node metastasis, named OLNNet. This model employs a self-supervised learning (SSL) strategy based on voxel mask reconstruction, combined with a Hyperbolic Prototype Embedding Network (HPENet), to predict the occurrence of occult lymph node metastasis. The SSL strategy is designed to acquire task-relevant prior knowledge through self-supervised mask reconstruction, thus reducing the dependence on extensive manually annotated data typically required for such tasks. The Hyperbolic Prototype Embedding Network is designed to enhance the model's ability to effectively learn image representations of occult lymph nodes by embedding their imaging features into a hyperbolic space. Extensive experiments have demonstrated that our predictive framework, OLNNet, achieves significant results in the prediction of occult lymph node metastasis. Overall, our main contributions are as follows:

- For the first time in the field of medical image analysis, we have developed a comprehensive CT image dataset specifically for the prediction of occult lymph node metastasis. This dataset includes accurately manually annotated occult lymph nodes, serving as a critical resource to advance research in this area. Our dataset fills a significant gap in available occult lymph nodes metastasis medical imaging data.
- We designed an innovative few-shot learning model for predicting occult lymph node metastasis, named OLNNet. In this model, we initially created a self-supervised pre-training module called MCAE to extract preliminary imaging features from lymph nodes. Subsequently, we designed the HPENet module to further refine the imaging features of occult lymph nodes. The proposed model effectively addresses the issue of data scarcity and provides accurate predictions of occult lymph node metastasis, demonstrating its efficiency and practicality in medical imaging analysis.
- We conducted extensive evaluations to benchmark the performance of OLNNet, providing a thorough analysis of its effectiveness in predicting occult lymph node metastasis. Our experiments demonstrate that OLNNet operates with robustness and delivers highly accurate predictions, confirming its potential as a transformative tool in medical diagnostics.

The rest of this paper is organized as follows. In Section 2, related works are reviewed. In Section 3, the algorithm of OLNNet is described in detail. In Section 4, the experimental data set, design, and experimental results, and a detailed analysis of the experimental results are presented. In Section 5, we summarize the full paper with future perspectives.

## 2. Related works

This section introduces the related work of the study. First, Section 2.1 delves into few-shot learning, analyzing common methods for few-shot image classification tasks and few-shot learning applications in medical imaging analysis. Section 2.2 introduces Self-Supervised Learning, outlining its methodologies. Finally, Section 2.3 focuses on Hyperbolic Embedding, discussing its research precedents in natural language processing and computer vision fields.

### 2.1. Few-shot learning

At present, existing methods for addressing few-shot image classification tasks can be divided into two categories: optimization-based and metric-based. The optimization-based approach aims to quickly adapt to new tasks by learning effective initial model parameters. Finn

et al. [20] proposed a model-agnostic meta-learning (MAML) algorithm that trains on different meta-tasks via multi-step gradient updates. Its goal is to improve the basic network model initialization state and quickly achieve good generalization performance on new tasks. Hu et al. [21] proposed a two-stage MAML algorithm compatible with noise mechanisms, alleviating overfitting during training.

The metric-based approach learns the similarity between samples by mapping images to a specific feature space and classifying them based on the relevant distance metrics in the space, with the choice of distance metric being particularly crucial. Snell et al. [22] proposed the prototypical network, which learns an embedding space by computing the average of the support set to obtain a prototype representation for each class and uses Euclidean distance to measure the similarity between the test image and the prototype. Vinyals et al. [23] proposed a matching network that compares the similarity between support set embeddings and test sample embeddings by using cosine distance. Liu et al. [24] combined the prototype network and relation network by classifying new categories based on the minimum distance between the query image and each class's prototype.

In the field of medical image analysis, FSL methods have been relatively underexplored. M. Ahmadi et al. [25] proposed using U-Net and a pretrained SAM for tumor segmentation. The pretrained SAM architecture includes a mechanism to capture spatial dependencies, thereby enhancing the generation of segmentation results. M. Ahmadi et al. [26] provided an advanced MRI segmentation technique for tumor diagnosis. Initially, the method employed a novel Quantum Matched-Filter Technique (QMFT) for preprocessing, effectively reducing noise. Subsequently, it leveraged a deep spiking neural network (DSNN) within a conditional random field framework for precise and efficient segmentation. A. K. Mondal et al. [27] proposed a few-shot 3D multimodal medical segmentation model based on generative networks that uses unannotated data to prevent overfitting through generative adversarial networks. Chen et al. [28] proposed a generative network-based one-shot Magnetic resonance imaging (MRI) segmentation method that transfers images from CT space to MRI space and uses generative adversarial learning to guide the training of the MRI-based segmentation network. Paul et al. [29] proposed a set-based learning method for chest X-ray diagnosis. Other applications of FSL have been proposed for brain imaging pattern recognition [30], voxel-wise medical image segmentation [31], and medical image classification [32].

### 2.2. Self-supervised learning

Self-supervised learning has received considerable attention in computer vision. It is used to transform unsupervised learning problems into supervised learning problems in different ways to obtain a pre-trained model for further supervised training. Currently, most self-supervised learning employs one of two approaches: pretext task-based methods and contrastive learning-based methods. Pretext task-based methods typically learn image representations through manually designed tasks such as image colorization [33], image inpainting [34], or geometric transformation recognition [35]. Contrastive learning-based methods [36–38] usually learn representations of images by measuring the similarity between two or more views.

The learning method in self-supervised masked reconstruction is gradually gaining in popularity. He et al. [39] were the first to apply the masked model to the field of vision. They designed an asymmetric encoder-decoder structure and learned image representation by randomly masking patches of input images and reconstructing missing pixels, achieving outstanding self-supervised pretraining effects. Zhou et al. [40] applied self-supervised masked reconstruction to medical image analysis tasks and achieved excellent performance in these tasks by mining the inherent features of images through data with the same target task.

Although self-supervised learning models have achieved remarkable results, almost all current self-supervised models require a large amount

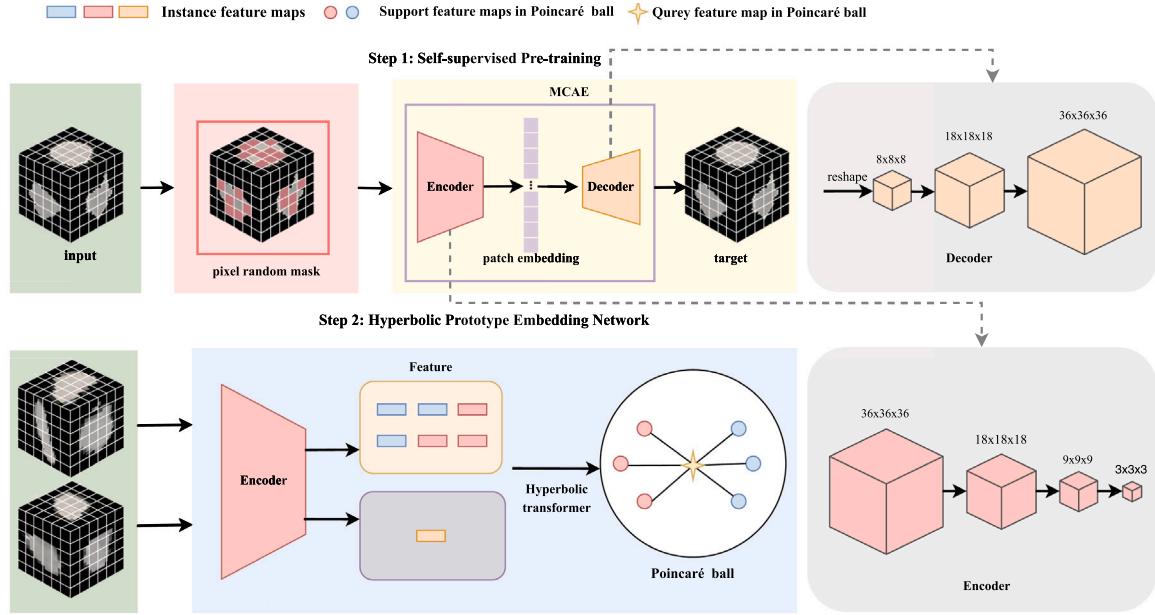


Fig. 1. Architecture of the proposed OLMN-Net network for NSCLC occult lymph node metastasis prediction. It consists of two parts, MCAE (Step 1) and HPENet (Step 2). MCAE is designed to initially extract the image features of the lymph nodes. HPENet is then introduced as a fine-tuning network to obtain the final prediction results.

of data for pretraining, and the size of the model parameters is extremely large. It is difficult to train a model for special medical cases with very little data. Therefore, reducing the size of the pretrained model and effectively mining potential features, as well as designing reasonable finite strategies to reduce the gap between the model and downstream tasks, are still major challenges.

### 2.3. Hyperbolic embeddings

The inherent embedding hierarchy in hyperbolic space, such as tree structures, endows it with low distortion capabilities [41,42]. This has made hyperbolic embeddings increasingly popular in the field of natural language processing. Nickel and Kiela [43] were the first to propose using the Poincaré model to learn embeddings, demonstrating that Poincaré embeddings can learn the latent hierarchical structure in data and that they are superior to Euclidean embeddings in terms of representational and generalization capabilities. A large body of work in natural language processing has employed hyperbolic embeddings to tackle various tasks, including text classification [44], text generation [45], word embeddings [46,47], and entity typing [48,49].

However, due to the different nature of data in natural language processing versus computer vision, it is not possible to directly extend hyperbolic space to the visual domain. Khrulkov et al. [50] observed the potential hierarchical relationships between images and utilized the second last layer of a deep neural network to learn higher-level semantic representations for image embedding. Then, the image classification task can be viewed as a certain type of geometry in the embedding space, where pairs of classes are separated by Euclidean hyperplanes. Grattarola et al. [51] constructed a hyperbolic adversarial autoencoder that outperforms autoencoders based on Euclidean and non-Euclidean geometries in semi-supervised classification, link prediction, and molecule generation tasks. Peng et al. [52] introduced Poincaré ST-GCN for skeleton-based action recognition, modeling input sequences (skeletons) using graph structure, and projecting the structural features extracted by the network into hyperbolic space for action recognition. In these works, one observes the potential of hyperbolic embedding in computer vision. In this paper, we explore for the first time the effectiveness of embedding medical image features into hyperbolic space and propose the OLMN-Net, which embeds occult lymph node image features into hyperbolic space for metastasis prediction.

### 3. Proposed method

This section will provide a detailed introduction to our proposed method. Section 3.1 introduces the self-supervised pretraining step and the MCAE model we constructed. Section 3.2 will discuss Hyperbolic Geometry. Section 3.3 expounds on Hyperbolic Arithmetic Operations. Finally, Section 3.4 presents an in-depth discussion of our newly proposed Hyperbolic Prototype Embedding Network (HPENet).

The central idea of OLMN-Net is to effectively extract 3D discriminative features of occult lymph nodes by learning the image representations from numerous unlabeled CT data in a self-supervised pretrained manner. Then, the finite strategy of few shot metric learning is used to reduce the gap with the target task. Fig. 1 illustrates the workflow of OLMN-Net. OLMN-Net primarily consists of two stages. In the first step, we utilize our specially designed lightweight MCAE module for self-supervised pretraining to initially learn the imaging characteristics of occult lymph nodes. In the second step, we employ HPENet for few-shot metric learning to predict occult lymph node metastasis. In the first step, a lightweight self-supervised mask-based pretraining model called MCAE is created. It consists of a simple asymmetric convolutional encoder-decoder. Similar to MAE [39], CT image data similar to the target task is divided into non-overlapping 3D patches. However, unlike MAE's random sampling mask strategy, there may be a large amount of irrelevant background information in medical images, which may further affect the reconstruction effect. Therefore, similar to sentence completion in natural language processing, patches outside the Region of Interest (ROI) are marked as padding tokens, and when calculating the loss, they are ignored. Then, ROI patches are randomly masked, and all patches are input into MCAE for reconstruction pretraining task. The second step is to fine-tune the target task using a few-shot learning strategy. In this step, we propose the Hyperbolic Prototype Embedding Network. First, the MCAE encoder is used to extract the occult lymph node image features, and then these features are embedded and mapped to the hyperbolic space. As shown in Fig. 2, we employ hyperbolic distance instead of Euclidean distance as the metric distance. Hyperbolic distance not only exhibits greater robustness compared to the Euclidean distance metric but also allows for learning the underlying hierarchical structure of image embeddings.



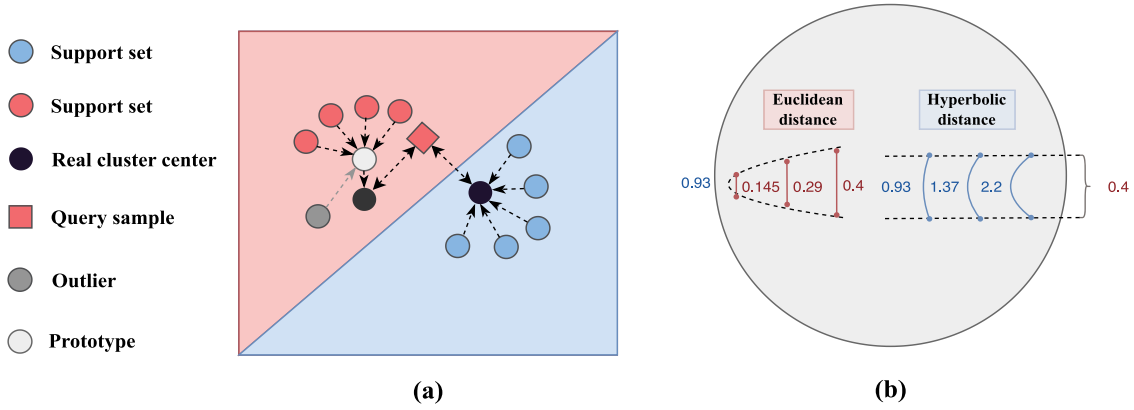


Fig. 2. (a) For the prototype embedding in Euclidean space, the query sample of the red class is misclassified as blue by the classifier because the outlier pulls the prototype away from the real clustering center in the embedding space. (b) Comparison between Euclidean distance and hyperbolic distance. In hyperbolic space (2-dimensional Poincaré disk), due to the difference in degree gauge, the farther the distance from the center of the circle, the greater the spatial curvature and the greater the geodesic or distance between points, which can better represent the potential hierarchical clustering relationship compared with Euclidean space.

### 3.1. Self-supervised pretrained

During the self-supervised pretraining step, we developed a lightweight model called MCAE. The MCAE module is specifically designed to address the challenges and requirements of medical imaging, particularly in the context of OLN prediction. This module differs from other self-supervised pretraining modules, such as the MAE, primarily because MCAE is composed of lightweight convolutional neural networks, as opposed to the large and complex transformer models. The use of convolutional neural networks in MCAE is intentional to facilitate effective and faster learning of NSCLC imaging features, especially in scenarios with limited data. Large and complex transformer models, on the other hand, can be challenging to train with limited data and are more prone to overfitting. MCAE uses a self-supervised masked reconstruction training strategy to predict the original image from randomly masked images, thereby obtaining an intuitive representation of the occult lymph node dataset and providing good prior knowledge for the next few-shot classification task. The entire process is divided into the following parts.

#### 3.1.1. Masking

The masking module is primarily responsible for converting the 3D volume image into a 3D volume mask through random masking. For a 3D image  $X \in \mathbb{R}^{C \times H \times W \times D}$ , it is first reshaped into a sequence of 3D patches  $X_p \in \mathbb{R}^{N \times (C \cdot P^3)}$ , where  $C$  is the number of channels of the input image,  $(H, W, D)$  are the height, width, and depth of the input image,  $P$  is the size of the non-overlapping patch blocks used to divide the image, and  $N = H \times W \times D / P^3$  is the length of the 3D patch sequence.

Due to the difference in information density distribution between medical images and images in the real world, the global random masking strategy of MAE may introduce unnecessary noise interference in medical images. Therefore, we use a pixel random mask strategy, where patches consisting entirely of background pixels are labeled as background patch  $P_b$  and ignored during the masking process. Patches that contain pixels in the ROI region are marked as  $P_r$ , and a random masking operation is performed during the masking process, retaining only a small subset (e.g., 25%). Finally, all patches are unpatchified into  $X_{input} \in \mathbb{R}^{C \times H \times W \times D}$ , where the length of the unmasked patch sequence is  $N_{visible} = (1 - \text{maskingratio})N_r$ , the length of the masked patch sequence is  $N_{mask} = N - N_{visible}$ , and the values of the patches marked as background and masked are set to 0.

#### 3.1.2. MCAE encoder

The MCAE encoder is a lightweight three-layer 3D convolutional neural network with a feature-embedding layer, which embeds the lymph node features extracted by the 3D convolutional neural network

into the feature space. To maintain the spatial position of each patch, we process all patches (including background patches, visible patches, and masked patches) in the encoder.

#### 3.1.3. MCAE decoder

In the MCAE decoder, we employ an asymmetric structure design, utilizing two deconvolutional layers to reconstruct the input lymph node image from the feature vector embedded in the encoder. This lightweight asymmetric design both shortens the pretraining time and encourages the encoder to explore more advantageous features at a deeper level.

#### 3.1.4. Loss function in MCAE

To train the MCAE, we construct a loss function based on the prediction of masked pixel/voxel values, using mean squared error. Specifically, for the output of the MCAE encoder,  $P_{out}$ , we first extract the masked patch using Eqs. (1) and (2):

$$P_{mask} = P \times (1 - \text{mask}). \quad (1)$$

$$\tilde{P}_{mask} = P_{out} \times (1 - \text{mask}). \quad (2)$$

where  $P$  represents the patch that was originally divided from the image, and the mask is a binary matrix. The value of the masked patch is set to 0, while the values of other patches (including background patches and visible patches) are set to 1. Then, the mean square error loss is calculated as shown in Eq. (3):

$$\mathcal{L}(P_{mask}, \tilde{P}_{mask}) = \frac{1}{N_{mask}} (P_{mask}^i - \tilde{P}_{mask}^i)^2. \quad (3)$$

where  $P_{mask}^i$  and  $\tilde{P}_{mask}^i$  denote the patch of the  $i$ th mask and the patch of the reconstructed mask, respectively.

### 3.2. Hyperbolic geometry

Before proceeding further, let us briefly introduce hyperbolic geometry, as this will be beneficial for understanding and derivation in the next subsection.

#### 3.2.1. Basics of Riemannian geometry

A differential manifold of  $n$  dimensions is locally homeomorphic to a Euclidean space  $\mathbb{R}^n$ , that is, compact Hausdorff topological space  $\mathcal{M}$  [53]. For each point,  $p \in \mathcal{M}$ , an  $n$ -dimensional vector space  $\mathcal{T}_p \mathcal{M}$  can be defined as the tangent space, which is homeomorphic to  $\mathbb{R}^n$ . The tangent space  $\mathcal{T}_p \mathcal{M}$  can be briefly defined as all curves  $c : (-\sigma, \sigma) \rightarrow \mathcal{M}$  such that  $c(0) = p$  and  $\sigma > 0$ .

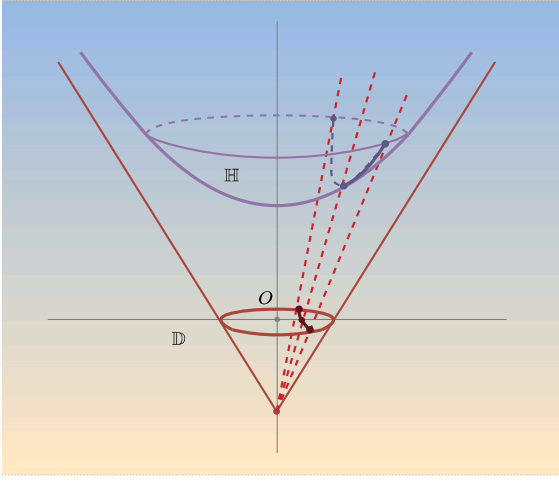


Fig. 3. A bijection between the hyperbolic space model  $\mathbb{H}^n$  and the Poincaré model  $\mathbb{D}^n$  is induced by projecting a point  $p = (0, \dots, 0, -1)$  onto the hyperbolic surface.

A Riemannian manifold [54] is a differential manifold with a metric tensor, typically represented as  $(\mathcal{M}, g)$ , where  $\mathcal{M}$  is the manifold and  $g$  is the metric tensor. The introduction of the metric tensor restores natural geometric concepts to differential manifolds. Euclidean space can be viewed as a manifold  $\mathbb{R}^n$  equipped with the Euclidean metric tensor  $g^E = I_n$  at each tangent space  $\mathcal{T}_p = \mathbb{R}^n$ .

### 3.2.2. Hyperbolic space

Hyperbolic space [55] is a type of Riemannian manifold with constant negative curvature. It has five isometric models for modeling, namely the hyperboloid (Lorentz) model, the Poincaré ball model, the Poincaré half-space model, the Klein model, and the hemisphere model. In this study, we use the Poincaré ball model.

The Poincaré model is obtained by linearly projecting every point of the hyperbolic surface model  $\mathbb{H}^n$  onto the hyperplane  $x_0 = 0$  through point  $p = (0, \dots, 0, -1)$ , as shown in Fig. 3. An  $n$ -dimensional Poincaré ball  $(\mathbb{D}^n, g^{\mathbb{D}})$  is defined as an  $n$ -dimensional Poincaré manifold  $\mathbb{D}^n = \{x \in \mathbb{R}^n : \|x\| < 1\}$  equipped with the following Riemannian metric tensor:

$$g_x^{\mathbb{D}} = \lambda_x^2 g^E. \quad (4)$$

where  $\lambda_x$  is the conformal factor:

$$\lambda_x = \frac{2}{1 - \|x\|^2}. \quad (5)$$

The geodesic distance between two points  $u, v \in \mathbb{D}^n$  on the Poincaré ball model can be obtained by the following formula:

$$d_{\mathbb{D}}(u, v) = \cosh^{-1} \left( 1 + \frac{2(\|u - v\|^2)}{(1 - \|u\|^2)(1 - \|v\|^2)} \right). \quad (6)$$

### 3.3. Hyperbolic arithmetic operations

The non-Euclidean nature of hyperbolic space makes it impossible to directly apply basic arithmetic operations such as vector addition and multiplication. One simple method for performing these computations is to use the tangent space for approximate calculations [56,57]. Another good option is to adopt the Möbius rotation vector space [58], which provides many standard operations for hyperbolic space. Similar to the previous discussion, we introduce an additional hyperparameter  $c$  to adjust the curvature of the Poincaré ball. An  $n$ -dimensional Poincaré ball is redefined as  $\mathbb{D}_c^n = \{x \in \mathbb{R}^n : c\|x\| < 1, c \geq 0\}$  equipped

with a Riemannian metric tensor  $g_x^{\mathbb{D}} = \lambda_x^{c^2} g^E$ , where the conformal factor  $\lambda_x^c$  is defined as:

$$\lambda_x^c = \frac{2}{1 - c\|x\|^2}. \quad (7)$$

Next, we introduce several fundamental arithmetic operations that are crucial to our work.

#### 3.3.1. Möbius addition

Möbius addition is a generalization of addition in Euclidean space to hyperbolic space. In the rotation vector space, the  $\oplus_c$  operation between  $u$  and  $v$  in model  $\mathbb{D}_c^n$  is defined as:

$$u \oplus_c v = \frac{(1 + 2c \langle u, v \rangle + c\|v\|^2)x + (1 - c\|u\|^2)v}{1 + 2c \langle u, v \rangle + c^2\|u\|^2\|v\|^2}. \quad (8)$$

#### 3.3.2. Geodesic distance

The geodesic distance between two points  $u$  and  $v$  in the model  $\mathbb{D}_c^n$  is defined as:

$$d_{\mathbb{D}}^c(u, v) = \frac{2}{\sqrt{c}} \tanh^{-1}(\sqrt{c}\| -u \oplus_c v \|). \quad (9)$$

#### 3.3.3. Hyperbolic average

The extension of the arithmetic mean in Euclidean space to hyperbolic space corresponds to the Einstein midpoint [59], which has the simplest form in the Klein model. Therefore, we first apply the following transformation to map the point  $p_{\mathbb{D}}$  on the Poincaré model to the point  $p_{\mathbb{K}}$  on the Klein model.

$$p_{\mathbb{K}} = \frac{2p_{\mathbb{D}}}{1 + c\|p_{\mathbb{D}}\|^2}. \quad (10)$$

Next, the hyperbolic mean can be obtained by computing the Einstein midpoint, which is defined as follows:

$$\mu_{\mathbb{K}} = \frac{\sum_{i=1}^n \gamma_i p_{\mathbb{K}}^i}{\sum_{i=1}^n \gamma_i}. \quad (11)$$

where  $\gamma_i = \frac{1}{\sqrt{1 - c\|p_{\mathbb{K}}^i\|^2}}$  is the Lorentz factor. Finally, we can easily transform the computed midpoint back to the Poincaré model to obtain the average of the Poincaré model, since the Klein model and Poincaré model are isomorphic. This transformation is defined as follows:

$$\mu_{\mathbb{D}} = \frac{\mu_{\mathbb{K}}}{1 + \sqrt{1 - c\|\mu_{\mathbb{K}}\|^2}}. \quad (12)$$

#### 3.3.4. Exponential and logarithmic maps

The above are the operational rules defined in hyperbolic space. To perform these operations, we also need to define a bijection between the Euclidean space  $\mathbb{R}^n$  and the hyperbolic space  $\mathbb{D}_c^n$ . The exponential map  $\exp_x^c(u)$  defines a transformation from the Euclidean space to the Poincaré model, defined as follows:

$$\exp_x^c(u) = x \oplus_c \left( \tanh\left(\sqrt{c} \frac{\lambda_x^c \|u\|}{2}\right) \frac{u}{\sqrt{c}\|u\|} \right). \quad (13)$$

Its inverse transformation, the logarithmic map  $\log_x^c(v)$ , is defined as follows:

$$\log_x^c(v) = \frac{2}{\sqrt{c}\lambda_x^c} \tanh^{-1}(\sqrt{c}\| -x \oplus_c v \|) \frac{-x \oplus_c v}{\| -x \oplus_c v \|.} \quad (14)$$

### 3.4. Hyperbolic prototype embedding network

#### 3.4.1. Overview

The Hyperbolic Prototype Embedding Network proposed in this paper consists of two main parts: (1) an encoder composed of convolutional neural networks, denoted as  $\mathcal{F} : \mathcal{X} \rightarrow \mathbb{R}_c^n$ , which extracts features from images and embeds them into an  $n$ -dimensional Poincaré ball, and (2) hyperbolic similarity calculation, which computes similarities in hyperbolic space for image classification.

### 3.4.2. Hyperbolic prototype embedding network

In FSL, the problem of classifying with a small number of samples is typically represented as an  $N$ -way  $K$ -shot classification problem. The training set is also known as the support set, and it contains  $N$  classes, with each class having  $K$  instances. Usually,  $K$  is a small number such as 1 or 5.

Given support set  $S = \{S_1, \dots, S_N | S_i \in \mathbb{R}^{K \times C \times H \times W \times D}\}$ , where  $C$ ,  $H$ ,  $W$ , and  $D$  denote the channel, height, width, and depth of input images, respectively, we first train a convolutional neural network to encode them into an  $n$ -dimensional feature space, i.e.,  $\mathcal{X} = \mathcal{F}_\phi(S)$ , where  $\mathcal{X} = \{\mathcal{X}_1, \dots, \mathcal{X}_N | \mathcal{X}_i \in \mathbb{D}_c^n\}$ , and  $\phi$  represents the learnable parameters of the network. Each hyperbolic prototype  $s_k$  is defined as the mean vector of Poincaré ball embeddings of the support points belonging to its category, which can be computed as follows:

$$s_k = \mu_{\mathbb{D}}(\mu_{\mathbb{K}}(p_{\mathbb{K}}(exp_u^c(\mathcal{X}_k)))). \quad (15)$$

where  $p_{\mathbb{K}}(\cdot)$ ,  $\mu_{\mathbb{D}}(\cdot)$  are operations that transform between the Poincaré model and the Klein model,  $exp_u^c(\cdot)$  is a transformation from Euclidean space to the Poincaré model, and  $\mu_{\mathbb{K}}(\cdot)$  is the Einstein midpoint operation, which calculates the mean in hyperbolic space.

For a given query set  $Q = \{Q_1, \dots, Q_N | Q_i \in \mathbb{R}^{B \times C \times H \times W \times D}\}$ , where  $B$  is the number of instances per category in the query set, we map them to  $n$ -dimensional feature vectors on the Poincaré ball:

$$q_k = exp_u^c(\mathcal{F}_\phi(Q_k)). \quad (16)$$

We can then use Eq. (6) to compute the distance between the query sample  $q$  and each support set  $s$  of each class. Finally, we can calculate the probability of the current query sample  $q_i$  belonging to each hyperbolic prototype:

$$p_\phi(y = k | q_i) = \frac{exp(-d_{\mathbb{D}}(q_i, s_k))}{\sum_{k'} exp(-d_{\mathbb{D}}(q_i, s_{k'}))}. \quad (17)$$

Assuming the true class of the sample is  $k$ , we can define the loss function as

$$\mathcal{L}(\phi) = -\log p_\phi(y = k | q_i). \quad (18)$$

We aspire to minimize the loss function, which is equivalent to maximizing the logarithm, that is, maximizing the probability of correctly classifying  $q_i$ . By minimizing the loss function, we can optimize the Hyperbolic Prototype Embedding Network.

## 4. Experiments

This section provides a comprehensive overview of the experiments conducted. Section 4.1 details the dataset's characteristics. Section 4.2 discusses the implementation details of our approach. Section 4.3 presents the results of self-supervised pretraining. Section 4.4 compares the proposed HMCAE model and other methods for predicting occult lymph node metastasis. Finally, Section 4.5 conducts ablation experiments on masking ratios, metric spaces, and self-supervised pretraining, offering insights into the effectiveness of these strategies in our methodology.

### 4.1. Dataset

This article conducts self-supervised pretraining on publicly available lymph node datasets [60–62] and comprehensively evaluates the proposed method on the TJ-OLNM Dataset, a lymph node metastasis dataset collected by our team. Below are detailed descriptions of the two datasets.

#### 4.1.1. LN dataset

The LN Dataset comprises CT images of 388 mediastinal lymph nodes from 90 patients and 595 abdominal lymph nodes from 86 patients. The lymph node locations were annotated by radiologists with the National Institute of Health Clinical Center, and the lymph node segmentation masks were manually traced and segmented based on the methodology described in [7].

#### 4.1.2. TJ-OLNM dataset

The TJ-OLNM Dataset was retrospectively collected for this study by selecting eligible cases from chest CT scans of lung cancer patients of Tongji Hospital affiliated with Tongji University. They were collected between 2016 and 2021.

Inclusion criteria for lymph node metastasis: (1) pathologically confirmed primary lung cancer; (2) CT scans of the chest showing non-suspicious target lymph nodes (short-axis diameter < 10 mm), with postoperative pathological confirmation of tumor metastasis within the lymph nodes; and (3) complete clinical data of patients.

Inclusion Criteria for non-lymph node Metastasis: (1) pathologically confirmed primary lung cancer; (2) postoperative pathological confirmation of the absence of tumor metastasis within the target lymph nodes; (3) complete clinical data of patients.

Our team imported standard chest CT images in DICOM format, which were correlated with histopathological autopsy results, into ITK-SNAP software. ITK-SNAP is a specialized medical imaging software, widely used for the analysis and visualization of medical image data, particularly adept at structure segmentation and 3D reconstruction. Subsequently, an experienced radiologist was responsible for manually delineating the lymph nodes layer-by-layer. During this process, the radiologist meticulously examined each CT image layer to ensure the precision and consistency of the annotations. Once the initial annotation was completed, another experienced radiologist conducted a thorough review to guarantee the accuracy and completeness of the markings. This dual-check mechanism significantly enhanced the reliability of the annotations and reduced the possibility of human error. A total of 40 cases met the inclusion criteria.

### 4.2. Implementation details

The method proposed in this study was implemented using the Py-Torch framework. In addition, to accurately assess the time complexity of our proposed model, we employed the Floating-point Operations (FLOPs) as a key metric. FLOPs, a crucial indicator of computational complexity, represent the number of floating-point operations required for a single forward pass of the model. This measure is vital for understanding the efficiency and computational demands of the model, especially in the context of resource-intensive deep learning applications. Our comprehensive analysis included the summation of FLOPs across all layers of the model, encompassing various operations like convolutions, pooling, and fully connected layers, along with their respective weights and operational counts. This detailed evaluation of computational complexity revealed that our model's total complexity stands at 428.50 MFLOPs (Mega Floating Point Operations). All experiments were conducted on a GeForce RTX 3090 GPU with 24 GB memory. As the data were obtained from different devices with varying spacing, we resampled them to a uniform size of 1 mm. Finally, we extracted a  $36 \times 36 \times 36$  region at the center of each lymph node based on the manually annotated masks for subsequent experimental procedures.

During the self-supervised pretraining process, we set the initial learning rate as  $l = 0.01$  for different masking rates. We then used a fixed-step decay strategy, decaying the learning rate by 0.1 every 100 epochs, until the learning rate remained  $l \geq 0.00001$ . We set the batch size as  $BS = 64$  and the maximum iteration epoch as  $E = 1000$  and utilized early stopping and l2 regularization methods. Our team used the Adam optimizer to update the network's weights.

During the fine-tuning stage of few-shot learning, in order to comprehensively evaluate the proposed model, we divided the TJ-OLNM dataset equally into five folds for five-fold cross-validation, as shown in Fig. 4. In each fold, the training set was further randomly sampled to obtain support and query sets, and the Hyperbolic Prototype Embedding Network was trained with both 1-shot and 5-shot configurations. The entire test set was used as the query set to evaluate the performance of the model. The hyperparameters of the proposed method and

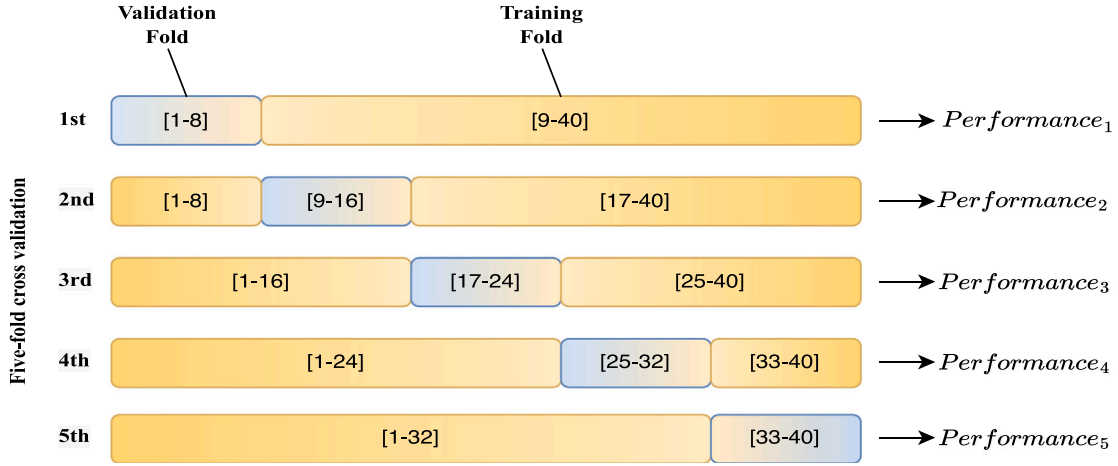


Fig. 4. Five-fold cross validation.

the comparison methods were determined by manually searching for different combinations, and the best-performing models were retained after repeated parameter training and model testing for analysis.

This article employs the mean and standard deviation of six performance metrics to evaluate the obtained prediction results, including accuracy, specificity, sensitivity, precision, F1 score, and area under the ROC curve (AUC). Accuracy represents the proportion of samples correctly classified by the model; specificity represents the ability to correctly identify negative samples; sensitivity represents the ability to correctly identify positive samples; and precision represents the ability to predict correctly among positive samples. The F1 score combines precision and recall: it is the harmonic mean of precision and recall. AUC is a metric that comprehensively considers specificity and sensitivity and thus reflects the expected generalization performance of the model. The formulas for calculating these metrics are as follows:

$$Sensitivity = \frac{TP}{TP + FN} \quad (19)$$

$$Specificity = \frac{TN}{TN + FP} \quad (20)$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (21)$$

$$Precision = \frac{TP}{TP + FP} \quad (22)$$

$$F1score = \frac{2 \times Sensitivity \times Precision}{Sensitivity + Precision} \quad (23)$$

where True Positives (TP) are instances correctly predicted as positive, True Negatives (TN) are instances correctly predicted as negative, False Positives (FP) are instances incorrectly predicted as positive, and False Negatives (FN) are instances incorrectly predicted as negative.

#### 4.3. Results of self-supervised pretraining

Fig. 5 illustrates the reconstruction capabilities of the MCAE under a masking rate of 75%. The figure is organized into three rows representing the axial, coronal, and sagittal sections of the CT images, respectively. Each row is further divided into three columns displaying the original image on the left, the masked image in the middle, and the reconstructed image on the right. The results demonstrate that MCAE is adept at restoring image patches that have been obscured by random pixel masking strategies. This ability is critical as it enables the model to cope with significant data loss, a common challenge in medical imaging scenarios where key features may be partially obscured or missing. The restoration process employed by MCAE focuses on the masked areas, leveraging learned imaging features to predict missing content accurately. It is important to note that the loss function of MCAE

specifically targets these masked patch blocks, which explains why the reconstructed images might exhibit some blurriness. This slight degradation in image quality does not detract from the utility of the process; the primary objective of MCAE is not to produce pristine reconstruction results, but rather to train the encoder module effectively. By focusing on masked areas, the encoder learns to identify and interpret critical imaging features of occult lymph nodes autonomously. This capability is pivotal for the subsequent task of predicting occult lymph nodes. The encoder's ability to discern subtle features within the lymph nodes from partially visible data enhances its effectiveness in identifying potential metastases in occult lymph nodes, which are typically challenging to detect due to their less pronounced imaging signatures. As such, the strategic use of MCAE in our methodology underlines its value not only as a tool for image reconstruction but as a foundational component in building a robust diagnostic framework for predicting occult lymph node metastases in NSCLC patients.

#### 4.4. Comparisons in OLNLM prediction

Table 1 presents the accuracy, sensitivity/recall, specificity, precision, F1 score, and AUC of the proposed OLNLM-Net model and other methods for predicting occult lymph node metastasis. Each method was evaluated with five independent repetitions, and the average values are reported. Among supervised learning methods, this study evaluated classical CNN models, such as 3D ResNet series (3D ResNet-18, 3D ResNet-34, 3D ResNet-50, 3D ResNet-101), and 3D DenseNet-121 on the TJ-OLNM Dataset. As shown in Table 1, with the increase of model depth and complexity, sensitivity, accuracy, and F1 score initially increased and then decreased, while specificity, precision, and AUC gradually increased. This suggests that the scale limitation of data will significantly affect the improvement of model performance in supervised learning methods. The model will learn to sacrifice sensitivity to increase specificity and improve classification accuracy. However, in medical classification tasks, model sensitivity is often the performance metric of primary interest. In addition, we have also included newer models such as MV-KBC [63], NASLung [64], and ResGANet(G = 2) [65] to provide a more comprehensive comparison. As shown in Table 1, these models demonstrate varying degrees of effectiveness in predicting occult lymph node metastasis. MV-KBC shows a balanced performance across all metrics with notable strengths in specificity (80.00%) and AUC (81.65%). Its overall accuracy (82.00%) and F1 score (82.35%) are commendable, suggesting its effectiveness in classification tasks. However, it falls slightly short in sensitivity when compared to OLNLM-Net, indicating a potential area for improvement in detecting true positive cases. NASLung notably excelled in specificity (83.00%) and showed strong performance in AUC (81.25%), NASLung



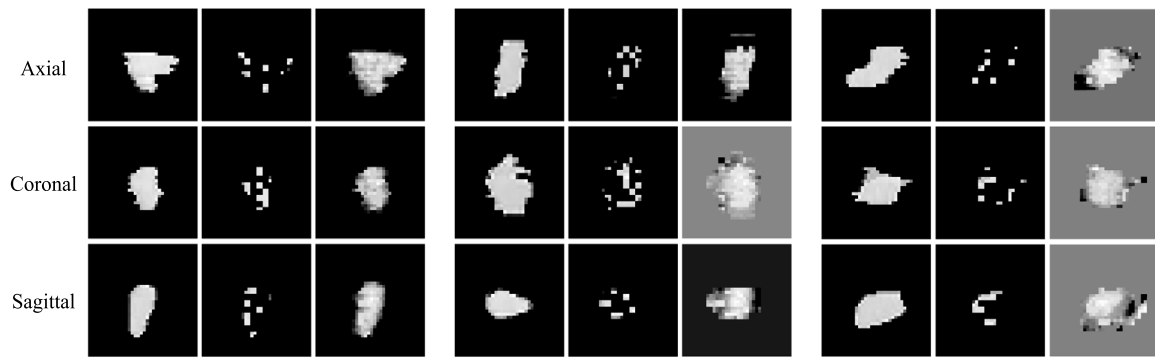


Fig. 5. Example results of self-supervised pretraining. For each triplet, we show the original image (left), the masked image (middle), and the reconstructed image (right). Each row corresponds to one of the three views of the lymph node, with a masking rate of 75%.

Table 1

Performance of different methods on TJ-OLNM dataset for occult lymph node metastasis prediction.

Inference with fully supervised learning						
Method	SEN (%)	SPE (%)	ACC (%)	AUC (%)	Precision (%)	F1-score (%)
3D ResNet-18	84.00	69.00	76.50	71.50	73.13	78.11
3D ResNet-34	<b>88.00</b>	70.00	79.00	73.30	74.42	80.59
3D ResNet-50	85.00	74.00	79.50	74.60	76.66	80.40
3D ResNet-101	76.00	80.00	78.00	76.25	79.41	77.38
3D DenseNet-121	68.00	<b>85.00</b>	76.50	76.65	82.33	74.14
MV-KBC	82.00	80.00	82.00	81.65	80.95	82.35
NASLung	79.00	83.00	81.50	81.25	80.21	82.05
ResGANet(G = 2)	80.00	<b>85.00</b>	<b>82.50</b>	<b>83.25</b>	<b>84.21</b>	<b>82.92</b>
Inference with transfer learning (MedicalNet)						
3D ResNet-18	64.00	<b>86.00</b>	75.00	72.75	<b>82.14</b>	71.89
3D ResNet-34	<b>66.00</b>	85.00	<b>75.50</b>	74.15	81.47	<b>72.91</b>
3D ResNet-50	65.00	85.00	75.00	<b>74.95</b>	81.25	72.22
Inference with K = 1 Shot						
3D ProtoNet	76.00	73.00	74.50	71.95	74.06	74.46
3D MatchingNet	79.00	82.00	80.50	78.40	<b>83.48</b>	79.74
<b>OLNM-Net</b>	<b>83.00</b>	<b>83.00</b>	<b>83.00</b>	<b>82.82</b>	83.15	<b>82.91</b>
Inference with K = 5 shot						
3D ProtoNet	76.00	77.00	76.50	75.55	77.35	76.31
3D MatchingNet	83.00	81.00	82.00	81.90	81.75	82.19
<b>OLNM-Net</b>	<b>86.00</b>	<b>88.00</b>	<b>87.00</b>	<b>88.50</b>	<b>87.82</b>	<b>86.87</b>

is a robust model. Its precision (80.21%) and accuracy (81.50%) are competitive, but it shows a slightly lower sensitivity (79.00%) compared to OLN-Net. This suggests that while NASLung is reliable in false positive reduction, it may miss some true positive cases. ResGANet(G = 2) emerges as a strong contender, with the highest scores in specificity (85.00%), precision (84.21%), and AUC (83.25%). It also boasts the highest F1 score (82.92%), indicating a balanced trade-off between precision and recall. However, its sensitivity, while impressive, does not surpass OLN-Net's performance. Comparatively, our proposed OLN-Net model demonstrates superior sensitivity, reflecting its strength in correctly identifying true positive cases, which is crucial in medical diagnostics. While it maintains competitive specificity and precision, its standout feature is the balanced sensitivity, making it particularly suitable for medical tasks where missing a positive case can have significant consequences.

Additionally, this study employed transfer learning to predict occult lymph node metastasis. Our team used a series of pretrained 3D-ResNet models from the MedicalNet project, which aggregates datasets with different modalities, target organs, and pathologies to build a relatively large dataset for extracting general medical 3D features. The results of transfer learning experiments are shown in Table 1. Due to the significant differences in features between the pretraining dataset used by MedicalNet and our TJ-OLNM dataset, the performance of the pretrained models was inferior even to that of the fully supervised learning models trained from scratch. The pretrained models had lower

sensitivity and higher specificity. However, for our occult lymph node metastasis prediction task, we need higher sensitivity to reduce the false negative rate for positive patients.

In the realm of self-supervised FSL methods, our proposed OLN-Net was compared with ProtoNet and MatchingNet, which are specifically designed for small sample classification tasks. To ensure fairness, all these methods utilized the self-supervised pretrained encoder of MCAE as the feature extraction framework. The experimental results demonstrate that our proposed OLN-Net outperforms the baseline models in various performance indicators in both 1-shot and 5-shot scenarios. In comparison to fully supervised learning and transfer learning methods, our backbone network is simpler yet exhibits higher performance efficacy. In terms of potential limitations, OLN-Net, like any model, may face challenges in handling extremely varied or noisy data, a common issue in medical imaging.

#### 4.5. Ablation study

##### 4.5.1. Masking ratio

To explore the impact of various masking rates on model performance, we conducted a series of self-supervised pretraining experiments using masking rates ranging from 50% to 90%. Fig. 6 provides a visual representation of how these different masking rates affect key performance indicators of the model. Our analysis reveals a notable trend where a higher masking rate generally correlates with improved

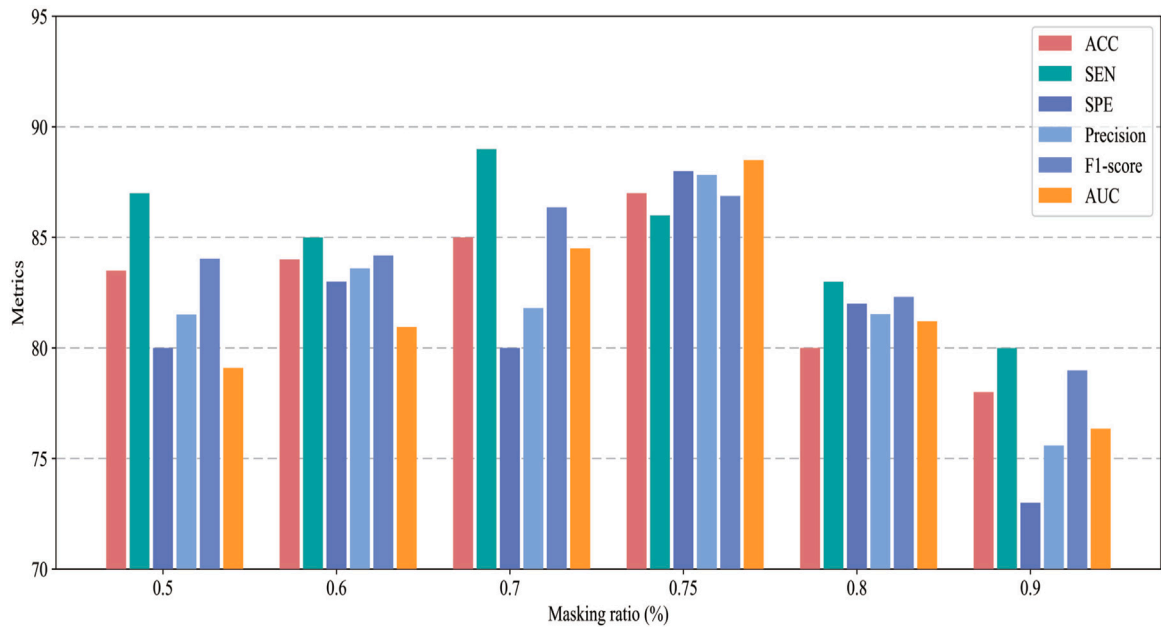


Fig. 6. The impact of different masking ratios on model performance.

model performance across several metrics. Remarkably, the optimal performance for most of these indicators was observed at a masking rate of 75%. However, it is important to highlight that sensitivity, a critical measure in medical diagnostics, reached its peak performance at a slightly lower masking rate of 70%. This nuanced observation suggests that while increasing the masking rate generally enhances the model's ability to learn robust features by forcing it to predict larger portions of missing data, there is a complex interplay between masking rate and specific performance metrics such as sensitivity. The results indicate that a balance needs to be struck to optimize all aspects of performance effectively. This balance is particularly crucial in medical applications where both high sensitivity (to minimize false negatives) and specificity (to avoid false positives) are essential for reliable diagnosis. The insights gained from these experiments are valuable for fine-tuning the self-supervised learning component of our model. By adjusting the masking rate, we can tailor the model's training process to emphasize learning features that are most relevant for the detection and classification of subtle pathological features in medical images, such as those found in occult lymph node metastases. Further, these findings underscore the importance of customizing the pretraining phase to suit specific diagnostic tasks, potentially leading to significant improvements in the efficacy and accuracy of medical imaging AI systems.

#### 4.5.2. Metric spaces

To assess the effectiveness of utilizing hyperbolic space over the traditional Euclidean space for our model, we conducted comparative experiments using both spaces. Table 2 presents the performance outcomes of these experiments, featuring the prediction results of the 3D ProtoNet model in Euclidean space against our OLN-Net model in hyperbolic space. The comparative analysis clearly demonstrates that the OLN-Net model outperforms the Euclidean-based 3D ProtoNet across all evaluation metrics. Specifically, the OLN-Net model in hyperbolic space shows substantial improvements over the Euclidean space model, with a 7% increase in sensitivity, 10% in specificity, 7.5% in accuracy, a remarkable 10.87 points increase in AUC, 9.09% in precision, and 8.45% in F1 score for the 1-shot learning scenario. Similarly, for the 5-shot learning scenario, the gains are even more pronounced: 10% in sensitivity, 11% in specificity, 10.5% in accuracy, an impressive 12.95 points increase in AUC, 10.47% in precision, and

10.56% in F1 score. These results compellingly illustrate the superior capability of the hyperbolic space representation, particularly within the Poincaré ball model, to enrich the model's embedding capacity. The Poincaré ball provides a unique geometric structure that is inherently suited to capturing complex hierarchical relationships and clustering patterns, which are often present in medical imaging data but are not effectively modeled in Euclidean space. This characteristic allows the OLN-Net to achieve higher discriminative power and more nuanced feature representation, enhancing its performance significantly. The superior performance of hyperbolic space in our experiments underlines its potential as a powerful tool for improving machine learning models in tasks involving complex and hierarchical data structures. Given these promising results, the hyperbolic space representation offers a compelling alternative to traditional Euclidean embeddings, promising advancements in how we understand and leverage geometric deep learning for medical diagnostics and beyond. This research not only contributes to the field of medical image analysis but also opens up new avenues for exploring geometric representations in other domains where understanding hierarchical and cluster-based relationships is crucial.

#### 4.5.3. Self-supervised pretraining

To evaluate the effectiveness of our self-supervised pretraining strategy, we conducted a comparative analysis between a feature extraction network that was initialized with randomly assigned weights and another that benefited from self-supervised pretraining. The results of this comparison are detailed in Table 2, where it is evident that the self-supervised pretraining approach markedly enhances various performance metrics of the model. Specifically, the model that underwent self-supervised pretraining exhibited substantial improvements across all key metrics: sensitivity, specificity, accuracy, AUC, precision, and F1 score. Notably, in the 1-shot learning scenario, the accuracy and AUC saw increases of 8% and 5.32%, respectively. The enhancements were even more significant in the 5-shot scenario, with accuracy improving by 9.5% and AUC by 7%. These results underscore the value of self-supervised pretraining in preparing the network to more effectively recognize and interpret the complex features associated with occult lymph nodes. This pretraining method leverages unlabeled data to help the model learn a rich representation of the underlying data distribution before fine-tuning on labeled data. This process not only

**Table 2**

Performance of OLNNet models with and without self-supervised pretraining on the TJ-OLNM dataset for occult lymph node metastasis prediction.

Inference with K = 1 Shot						
Method	SEN (%)	SPE (%)	ACC (%)	AUC (%)	Precision (%)	F1 score (%)
OLNM-Net w/o pretrained	80.00	70.00	75.00	77.50	72.73	76.19
<b>OLNM-Net w/ pretrained</b>	<b>83.00</b>	<b>83.00</b>	<b>83.00</b>	<b>82.82</b>	<b>83.15</b>	<b>82.91</b>
Inference with K = 5 shot						
OLNM-Net w/o pretrained	80.00	75.00	77.50	81.50	76.19	78.05
<b>OLNM-Net w/ pretrained</b>	<b>86.00</b>	<b>88.00</b>	<b>87.00</b>	<b>88.50</b>	<b>87.82</b>	<b>86.87</b>

enhances the model's ability to discern relevant features from noise but also significantly boosts its generalization capabilities. By training the network to predict parts of the data that are not directly visible, it learns to capture essential characteristics of the data, which are crucial for accurately identifying subtle signs of occult lymph node metastasis. The positive impact of self-supervised pretraining on model performance is particularly important in medical imaging, where labeled data can be scarce and expensive to obtain, and where the accuracy of predictions can directly influence clinical outcomes. These findings highlight the potential of self-supervised learning strategies to transform the landscape of medical image analysis, providing a robust foundation for developing more accurate and reliable diagnostic tools. This approach not only mitigates the challenges posed by limited labeled datasets but also enhances the efficacy of machine learning models in critical healthcare applications.

## 5. Conclusion

In this paper, we have built the first CT image dataset for the prediction of occult lymph node metastases. Furthermore we propose a novel self-supervised hyperbolic metric few-shot learning method, OLNNet, for metastasis prediction of occult lymph nodes in mediastinal CT scans of NSCLC patients. The OLNNet method attempts to overcome obstacles resulting from insufficient training samples by exploiting feature information from similar CT data. OLNNet incorporates a MCAE, a self-supervised pretraining strategy that learns a robust image representation through mask reconstruction after eliminating background noise in CT. In the fine-tuning phase, OLNNet uses a few-shot metric learning strategy to perform fine-tuning and reduce the disconnect between the pretrained model and real data. In addition, OLNNet introduces hyperbolic space to represent potential hierarchical clustering relationships, which helps better learn distinguishable features for prediction. Experimental results on our collected dataset demonstrate that the proposed OLNNet method achieves state-of-the-art results in the prediction of occult lymph node metastasis in NSCLC patients. Furthermore, although OLNNet has achieved excellent results in the prediction of occult lymph nodes, we are still exploring its utility in the early detection, prognosis, and treatment planning of various tumor diseases. OLNNet is capable of learning the radiographic features of diseases even with a small amount of data. While OLNNet has demonstrated significant promise in predicting occult lymph node metastasis in NSCLC patients, it is important to note that the model has been specifically tailored for this particular cancer type. This specialization stems from both the nature of the data used in training and the unique characteristics of NSCLC-related metastases that the model has learned to identify. Consequently, applying OLNNet to other types of cancers or diseases may require some modifications. These adaptations would involve not only retraining the model with new datasets representative of other cancers but also potentially revising the model's architecture to accommodate different patterns of disease progression and metastasis. Looking forward, our research team is committed to expanding the utility of OLNNet beyond its current application. One of the primary directions for this expansion involves adapting the model to handle multimodal data. By

integrating clinical data, such as patient history and genetic information, with imaging data, we aim to enhance the model's ability to perform in scenarios where available data are limited or incomplete. This approach is expected to improve the robustness and accuracy of the model, enabling it to provide more reliable predictions across a broader range of clinical conditions. Integrating multimodal data will also allow OLNNet to leverage the complementary strengths of various data types, enhancing its diagnostic capabilities. For example, genetic data can provide insights into the likelihood of metastasis due to inherited traits, while detailed patient histories might offer clues about environmental or lifestyle factors that could influence cancer progression. When combined with the sophisticated imaging analysis already performed by OLNNet, this comprehensive view could greatly enhance early detection, accurate staging, and effective treatment planning, ultimately leading to improved patient outcomes.

## CRedit authorship contribution statement

**Haizhou Xu:** Writing – original draft, Software, Conceptualization. **Jiaqi Wu:** Investigation, Data curation. **Yujia Yu:** Software, Resources, Investigation, Formal analysis. **Wenkai Huang:** Writing – review & editing, Resources, Project administration, Methodology, Funding acquisition. **Jiong Ni:** Visualization, Validation, Supervision.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

The research described in this article utilized a proprietary dataset from Tongji Hospital. Due to confidentiality agreements and privacy considerations, the data cannot be directly shared.

## References

- [1] K.C. Thandra, A. Barsouk, K. Saginala, J.S. Aluru, A. Barsouk, Epidemiology of lung cancer, *Contemp. Oncol.* 25 (2021) 45–52.
- [2] Y. Yun Song, CT radio genomics of non-small cell lung cancer using machine and deep learning, in: 2021 IEEE International Conference on Consumer Electronics and Computer Engineering, ICCECE, 2021, pp. 128–139.
- [3] X. qun He, T. Luo, X. Li, J. wen Huo, J. Gong, Q. Li, Clinicopathological and computed tomographic features associated with occult lymph node metastasis in patients with peripheral solid non-small cell lung cancer, *Eur. J. Radiol.* 144 (2021) 109981.
- [4] S. Bracci, M. Dolciami, C. Trobiani, A. Izzo, A. Pernazza, G. d'Amati, L. Manganaro, P. Ricci, Quantitative CT texture analysis in predicting PD-L1 expression in locally advanced or metastatic NSCLC patients, *Radiol. Medica* 126 (2021) 1425–1433.
- [5] W. Mu, L. Jiang, J. Zhang, Y. long Shi, J.E. Gray, I. Tunalı, C. Gao, Y. Sun, J. Tian, X. Zhao, X. Sun, R.J. Gillies, M.B. Schabath, Non-invasive decision support for NSCLC treatment using PET/CT radiomics, *Nature Commun.* 11 (2020).
- [6] C.M. Walker, J.H. Chung, G.F. Abbott, B.P. Little, A.H. El-Sherief, J.-A.O. Shepard, M. Lanuti, Mediastinal lymph node staging: from noninvasive to surgical, *AJR. Am. J. Roentgenol.* 199 (1) (2012) W54–64.

- [7] A. Szlubowski, M. Zielinski, J. Soja, A. Filarecka, S. Orzechowski, J. Pankowski, A. Obrochta, M. Jakubiak, J. Węgrzyn, A. Ćmiel, Accurate and safe mediastinal restaging by combined endobronchial and endoscopic ultrasound-guided needle aspiration performed by single ultrasound bronchoscope, *Eur. J. Cardio-Thorac. Surg.* : Off. J. Eur. Assoc. Cardio-Thorac. Surg. 46 (2) (2014) 262–266.
- [8] P.F. Roberts, D.M. Follette, D. von Haag, J.A. Park, P.E. Valk, T.R. Pounds, D.M. Hopkins, Factors associated with false-positive staging of lung cancer by positron emission tomography, *Ann. Thorac. Surg.* 70 (4) (2000) 1154–1159, discussion 1159–60.
- [9] R. Kanzaki, M. Higashiyama, A. Fujiwara, T. Tokunaga, J. Maeda, J. Okami, T. Kozuka, T. Hosoki, Y. Hasegawa, M. Takami, Y. Tomita, K. Kodama, Occult mediastinal lymph node metastasis in NSCLC patients diagnosed as clinical N0-1 by preoperative integrated FDG-PET/CT and CT: Risk factors, pattern, and histopathological study, *Lung Cancer* 71 (3) (2011) 333–337.
- [10] H.C. Verdusco-Aguirre, G. de Lima Lopes, E.S.-P. de Celis, Implementation of diagnostic resources for cancer in developing countries: a focus on PET/CT, *ecancermedalscience* 13 (2019).
- [11] G.A. Eapen, A.M. Shah, X. Lei, C.A. Jimenez, R.C. Morice, L. Yarmus, J.J. Filner, C. Ray, G.C. Michaud, S.R. Greenhill, M. Sarkiss, R.F. Casal, D.C. Rice, D.E. Ost, Complications, consequences, and practice patterns of endobronchial ultrasound-guided transbronchial needle aspiration: Results of the AQUIRE registry, *Chest* 143 (4) (2013) 1044–1053.
- [12] A. Ala, M. Yazdani, M. Ahmadi, A. Poorianasab, M.Y.N. Attari, An efficient healthcare chain design for resolving the patient scheduling problem: queuing theory and MILP-ASA optimization approach, *Ann. Oper. Res.* (2023) 1–31.
- [13] M. Ahmadi, A. Sharifi, M.J. Fard, N. Soleimani, Detection of brain lesion location in MRI images using convolutional neural network and robust PCA, *Int. J. Neurosci.* 133 (2021) 55–66.
- [14] A. Sharifi, M. Ahmadi, M.A. Mehni, S.J. Ghouschi, Y. Pourasad, Experimental and numerical diagnosis of fatigue foot using convolutional neural network, *Comput. Methods Biomech. Biomed. Eng.* 24 (2021) 1828–1840.
- [15] A. Davoudi, M. Ahmadi, A. Sharifi, R. Hassantabar, N. Najafi, A. Tayebi, H.A. Kargari, F. Ahmadi, M. Rabiee, Studying the effect of taking statins before infection in the severity reduction of COVID-19 with machine learning, *BioMed Res. Int.* 2021 (2021).
- [16] Y. Zhong, M. Yuan, T. Zhang, Y.-D. Zhang, H. Li, T. fu Yu, Radiomics approach to prediction of occult mediastinal lymph node metastasis of lung adenocarcinoma, *AJR. Am. J. Roentgenol.* 211 (1) (2018) 109–113.
- [17] M. Cong, H. Yao, H. Liu, L. Huang, G. Shi, Development and evaluation of a venous computed tomography radiomics model to predict lymph node metastasis from non-small cell lung cancer, *Medicine* 99 (2020).
- [18] D. Ruppert, The elements of statistical learning: Data mining, inference, and prediction, *J. Amer. Statist. Assoc.* 99 (2004) 567.
- [19] J.R. Ferreira, M. Koenigkam-Santos, F.E.G. Cipriano, A.T. Fabro, P.M.A. Marques, Radiomics-based features for pattern recognition of lung cancer histopathology and metastases, *Comput. Methods Programs Biomed.* 159 (2018) 23–30.
- [20] C. Finn, P. Abbeel, S. Levine, Model-agnostic meta-learning for fast adaptation of deep networks, in: *International Conference on Machine Learning*, 2017.
- [21] Z. Hu, Z. Gan, W. Li, J.Z. Wen, D. Zhou, X. Wang, Two-stage model-agnostic meta-learning with noise mechanism for one-shot imitation, *IEEE Access* 8 (2020) 182720–182730.
- [22] J. Snell, K. Swersky, R.S. Zemel, Prototypical networks for few-shot learning, 2017, *ArXiv abs/1703.05175*.
- [23] O. Vinyals, C. Blundell, T.P. Lillicrap, K. Kavukcuoglu, D. Wierstra, Matching networks for one shot learning, in: *NIPS*, 2016.
- [24] X. Liu, F. Zhou, J. Liu, L. Jiang, Meta-Learning based prototype-relation network for few-shot classification, *Neurocomputing* 383 (2020) 224–234.
- [25] M. Ahmadi, M.F. Nia, S. Asgarian, K. Danesh, E. Irankhah, A.G. Lonbar, A. Sharifi, Comparative analysis of segment anything model and U-Net for breast tumor detection in ultrasound and mammography images, 2023, *ArXiv abs/2306.12510*.
- [26] M. Ahmadi, A. Sharifi, S. Hassantabar, S. Enayati, QAIS-DSNN: Tumor area segmentation of MRI image with optimized quantum matched-filter technique and deep spiking neural network, *BioMed Res. Int.* 2021 (2021).
- [27] A.K. Mondal, J. Dolz, C. Desrosiers, Few-shot 3D multi-modal medical image segmentation using generative adversarial learning, 2018, *ArXiv abs/1810.12241*.
- [28] X. Chen, J.J. Xia, D. Shen, C. Lian, L. Wang, H.H. Deng, S. Fung, D. Nie, K.-H. Thung, P.-T. Yap, J. Gateno, One-shot generative adversarial learning for MRI segmentation of craniomaxillofacial bony structures, *IEEE Trans. Med. Imaging* 39 (2020) 787–796.
- [29] A. Paul, T.C. Shen, S. Lee, N. Balachandrar, Y. Peng, Z. Lu, R.M. Summers, Generalized zero-shot chest X-Ray diagnosis through trait-guided multi-view semantic embedding with self-training, *IEEE Trans. Med. Imaging* 40 (2021) 2642–2655.
- [30] S. Puch, I. Sánchez, M. Rowe, Few-shot learning with deep triplet networks for brain imaging modality recognition, in: *DART/MIL3ID@MICCAI*, 2019.
- [31] A.G. Roy, S. Siddiqui, S. Pölsterl, N. Navab, C. Wachinger, 'Squeeze & excite' guided few-shot segmentation of volumetric images, *Med. Image Anal.* 59 (2019) 101587.
- [32] R. Singh, V. Bharti, V. Purohit, A. Kumar, A.K. Singh, S.K. Singh, MetaMed: Few-shot medical image classification using gradient-based meta-learning, *Pattern Recognit.* 120 (2021) 108111.
- [33] D. Pathak, P. Krähenbühl, J. Donahue, T. Darrell, A.A. Efros, Context encoders: Feature learning by inpainting, in: *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2016, pp. 2536–2544.
- [34] C. Doersch, A.K. Gupta, A.A. Efros, Unsupervised visual representation learning by context prediction, in: *2015 IEEE International Conference on Computer Vision, ICCV*, 2015, pp. 1422–1430.
- [35] C. Doersch, A. Zisserman, Multi-task self-supervised visual learning, in: *2017 IEEE International Conference on Computer Vision, ICCV*, 2017, pp. 2070–2079.
- [36] Z. Wu, Y. Xiong, S.X. Yu, D. Lin, Unsupervised feature learning via non-parametric instance discrimination, in: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3733–3742.
- [37] T. Chen, S. Kornblith, M. Norouzi, G.E. Hinton, A simple framework for contrastive learning of visual representations, 2020, *ArXiv abs/2002.05709*.
- [38] K. He, H. Fan, Y. Wu, S. Xie, R.B. Girshick, Momentum contrast for unsupervised visual representation learning, in: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR*, 2019, pp. 9726–9735.
- [39] K. He, X. Chen, S. Xie, Y. Li, P. Doll'ar, R.B. Girshick, Masked autoencoders are scalable vision learners, in: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR*, 2021, pp. 15979–15988.
- [40] L. Zhou, H. Liu, J. Bae, J. He, D. Samaras, P. Prasanna, Self pre-training with masked autoencoders for medical image analysis, 2022, *ArXiv abs/2203.05573*.
- [41] R. Sarkar, Low distortion delaunay embedding of trees in hyperbolic plane, in: *International Symposium Graph Drawing and Network Visualization*, 2011.
- [42] F. Sala, C.D. Sa, A. Gu, C. Ré, Representation tradeoffs for hyperbolic embeddings, *Proc. Mach. Learn. Res.* 80 (2018) 4460–4469.
- [43] M. Nickel, D. Kiela, Poincaré embeddings for learning hierarchical representations, in: *NIPS*, 2017.
- [44] Y. Zhu, D. Zhou, J. Xiao, X. Jiang, X. Chen, Q. Liu, HyperText: Endowing FastText with hyperbolic geometry, in: *Findings*, 2020.
- [45] S. Dai, Z. Gan, Y. Cheng, C. Tao, L. Carin, J. Liu, Apo-VAE: Text generation in hyperbolic space, in: *North American Chapter of the Association for Computational Linguistics*, 2020.
- [46] B. Dhingra, C.J. Shallue, M. Norouzi, A.M. Dai, G.E. Dahl, Embedding text in hyperbolic spaces, in: *TextGraphs@NAACL-HLT*, 2018.
- [47] A. Tifrea, G. Bécigneul, O.-E. Ganea, Poincaré glove: Hyperbolic word embeddings, 2018, *ArXiv abs/1810.06546*.
- [48] F. López, B. Heinzerling, M. Strube, Fine-grained entity typing in hyperbolic space, 2019, *ArXiv abs/1906.02505*.
- [49] F.J. López, M. Strube, A fully hyperbolic neural model for hierarchical multi-class classification, 2020, *ArXiv abs/2010.02053*.
- [50] V. Khurlov, L. Mirvakhabova, E. Ustinova, I. Oseledets, V.S. Lempitsky, Hyperbolic image embeddings, in: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR*, 2019, pp. 6417–6427.
- [51] D. Grattarola, L.F. Livi, C. Alippi, Adversarial autoencoders with constant-curvature latent manifolds, *Appl. Soft Comput.* 81 (2018).
- [52] W. Peng, J. Shi, Z. Xia, G. Zhao, Mix dimension in poincaré geometry for 3D skeleton-based action recognition, in: *Proceedings of the 28th ACM International Conference on Multimedia*, 2020.
- [53] M.W. Hirsch, *Differential Topology*, vol. 33, Springer Science & Business Media, 2012.
- [54] S. Gallot, D. Hulin, J. Lafontaine, et al., *Riemannian Geometry*, vol. 2, Springer, 1990.
- [55] J.W. Cannon, W.J. Floyd, R. Kenyon, W.R. Parry, et al., *Hyperbolic geometry*, *Flavors Geom.* 31 (59–115) (1997) 2.
- [56] A. Lou, I. Katsman, Q. Jiang, S. Belongie, S.-N. Lim, C. De Sa, Differentiating through the fréchet mean, in: *International Conference on Machine Learning*, PMLR, 2020, pp. 6393–6403.
- [57] I. Chami, Z. Ying, C. Ré, J. Leskovec, Hyperbolic graph convolutional neural networks, *Adv. Neural Inf. Process. Syst.* 32 (2019).
- [58] A.A. Ungar, Hyperbolic trigonometry and its application in the Poincaré ball model of hyperbolic geometry, *Comput. Math. Appl.* 41 (1–2) (2001) 135–147.
- [59] Ç. Gülçehre, M. Denil, M. Malinowski, A. Razavi, R. Pascanu, K.M. Hermann, P.W. Battaglia, V. Bapst, D. Raposo, A. Santoro, N. de Freitas, Hyperbolic attention networks, 2018, *ArXiv abs/1805.09786*.
- [60] H.R. Roth, L. Lu, A. Seff, K.M. Cherry, J. Hoffman, S. Wang, J. Liu, E. Turkbey, R.M. Summers, A new 2.5 D representation for lymph node detection using random sets of deep convolutional neural network observations, in: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014: 17th International Conference, Boston, MA, USA, September 14–18, 2014, Proceedings, Part I* 17, Springer, 2014, pp. 520–527.
- [61] A. Seff, L. Lu, K.M. Cherry, H.R. Roth, J. Liu, S. Wang, J. Hoffman, E.B. Turkbey, R.M. Summers, 2D view aggregation for lymph node detection using a shallow hierarchy of linear classifiers, in: *Medical Image Computing and Computer-Assisted Intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, Vol. 17 Pt 1, 2014, pp. 544–552.



- [62] A. Seff, L. Lu, A. Barbu, H.R. Roth, H.-C. Shin, R.M. Summers, Leveraging mid-level semantic boundary cues for automated lymph node detection, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015.
- [63] Y. Xie, Y. Xia, J. Zhang, Y. Song, D. Feng, M.J. Fulham, W.T. Cai, Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest CT, *IEEE Trans. Med. Imaging* 38 (2019) 991–1004.
- [64] H. Jiang, F. Shen, F. Gao, W. Han, Learning efficient, explainable and discriminative representations for pulmonary nodules classification, *Pattern Recognit.* 113 (2021) 107825.
- [65] J. Cheng, S. Tian, L. Yu, C. Gao, X. Kang, X. Ma, W. Wu, S.-J. Liu, H. Lu, ResGANet: Residual group attention network for medical image classification and segmentation, *Med. Image Anal.* 76 (2021) 102313.