

MLLCD: A Meta Learning-based Method for Lung Cancer Diagnosis Using Histopathology Images

1st Xiangjun Hu

Hainan University

Haikou, China

huxiangjun@hainanu.edu.cn

2nd Suixue Wang

Hainan University

Haikou, China

wangsuixue@hainanu.edu.cn

The co-first author

3rd Hang Li

Shenyang Normal University

Shenyang, China

lihang@synu.edu.cn

4th Qingchen Zhang

Hainan University

Haikou, China

Shenyang Normal University

Shenyang, China

zhangqingchen@hainanu.edu.cn

The corresponding author

Abstract—Lung cancer is a leading cause of death. An accurate early lung cancer diagnosis can improve a patient's survival chances. Histopathological images are essential for cancer diagnosis. With the development of deep learning in the past decade, many scholars have used deep learning to learn the features of histopathological images and achieve lung cancer classification. However, deep learning requires a large quantity of annotated data to train the model to achieve a good classification effect, and collecting many annotated pathological images is time-consuming and expensive. Faced with the scarcity of pathological data, we present a meta-learning method for lung cancer diagnosis (called MLLCD). In detail, the MLLCD works in three steps. First, we preprocess all data using the bilinear interpolation method and then design the base learner which units a convolutional neural network(CNN) and transformer to distill local features and global features of pathology images with different resolutions. Finally, we train and update the base learner with a model-agnostic meta-learning (MAML) algorithm. Clinical Proteomic Tumor Analysis Consortium (CPTAC) cancer patient data demonstrate that our proposed model achieves the receiver operating characteristic (ROC) values of 0.94 for lung cancer diagnosis.

Index Terms—MAML, CNN, transformer, lung cancer

I. INTRODUCTION

According to data from the World Health Organization (WHO) [1] in 2020, approximately one-sixth of all deaths worldwide are caused by cancer. Among all cancers, lung cancer has the second highest number of new cases, accounting for 11.4%, second only to breast cancer (11.7%). Moreover, the largest number of people die from lung cancer (18%). Cancer cells can undergo unrestricted proliferation and growth and invade and infiltrate surrounding tissues, making treatment challenging. Because lung cancer does not usually spread in the early stages, early diagnosis can increase the chance of cure, improve patient survival rate, reduce treatment time and cost, and reduce the burden on medical resources. Thus, an accurate early lung cancer diagnosis plays a crucial role.

Microscopic examination of histopathological slides is one of the methods used to diagnose cancer [2]. Digitized pathological images typically have high resolution. Doctors can observe and analyze subtle cytological features such as cell morphology, karyotype, and intercellular relationships from high-resolution pathological images that display the microstructure

of cells and tissues. However, the interpretation and analysis of pathological images require doctors to have rich experience and professional knowledge and consume considerable time and manpower.

With the deepening development of deep learning in the medical field, automated processing of pathological images can predict whether patients have cancer, assisting doctors in making faster and more accurate diagnoses. CNNs have become one of the mainstream methods to realize intelligent diagnosis of lung cancer. As an important computing intelligence technology, CNN focuses on local features of objects extracted by several convolution layers and fully connected layers [3]. Moreover, pathological images are generally cut into smaller patches for training because the size of pathological images is millions of pixels. The segmented pathological image faces the possibility of losing some cancer features. We are considering using a transformer for learning the relationships between words in the natural language processing field [4] to connect all patches. In addition, it requires a large quantity of training data to train a robust deep learning model. However, insufficient pathological images and ultrahigh resolution can easily lead to overfitting phenomenon in deep learning models, affecting cancer diagnosis accuracy.

In this work, to increase the accuracy of diagnosing carcinoma of the lungs, we propose a meta-learning method for lung cancer diagnosis (called MLLCD), which designs a new base learners and applies MAML to obtain initial parameters with strong generalization ability. Through this work, we make the following contributions:

- We resize all pathology image data to the same size for easy model training and testing and explore the base learner of MAML.
- Our method demonstrates the ability of the meta-learning algorithm MAML to accurately diagnose lung cancer in a few-shots scenario.
- This work is the first to combine meta-learning and transformers to apply tissue pathology slice data for cancer diagnosis.

The remainder of the paper is organized as follows. The related work about pathology image classification is retrospec-

tively reviewed in Section 2. Section 3 describes the proposed method. In Section 4, we describe the experiments. Finally, we conclude with a discussion in Section 5.

II. RELATED WORK

Various cancer diagnosis methods for cancer pathological images have been proposed over the years. Common methods include CNNs and transformers, etc. For example, Chen et al. [6] proposed a Dropconnect-based gastric histopathology image detection model that used position-coded transformers for high global detection performance. Chen et al. [8] introduced a hierarchical image pyramid transformer method that achieved differeft resolution image representations by two self-supervised learning methods based on student-teacher knowledge distillation. Wang et al. [9] diagnosed lung cancer subtypes using multiple-omics data and pathological images. They first constructed attention-based encoders for each modality to extract important diagnostic features, then used a generative adversarial network(GAN) to discover information of missing modal samples, and finally utilized all features for lung cancer diagnosis. Wang et al. [10] designed a deep learning ensemble to classify four subtypes of gastric cancer by pathological images.

In addition, many works study few-shot learning to improve the generalization ability of models. Koohbanani et al. [5] proposed a self-supervised CNN. When the available labelled data are insufficient, their method utilized the contextual, multi-resolution and semantic features of pathological images for training, thereby improving the classification performance of pathological images. Yang et al. [7] train a CNN to classify multinomial patterns of lung cancer using pathological images. Nair et al. [11] proposed a transfer learning model based on Inception V3 to study the impact of resolution differences on lung cancer prediction results. Chou et al. [12] compared meta-learning, matching network and prototypical network. The result showed that the preformance of meta learning is better than other methods on cancer genomics data.

The above methods have made some progress in solving the overfitting problem of model. However, only a few works apply meta-learning to histopathological images. We apply meta-learning and design the base learner to achieve the diagnosis of lung cancer.

III. METHOD

We advance an MLLCD method for lung cancer diagnosis via a transformer-based meta-learning framework (see Fig. 1). This method contains three steps. First, we preprocess all pathological images used for training and testing. Second, we design the baseline network of MAML to extract features of images. At last, we update the initial parameter by MAML.

A. Dataset Collection and Pre-processing

Our experiments focus on six CPTAC datasets: clear cell renal cell carcinoma (CCRCC) [14], cutaneous melanoma (CM) [15], lung squamous cell carcinoma (LSCC) [16], pancreatic ductal adenocarcinoma (PDA) [17], sarcomas (SAR),

and uterine corpus endometrial carcinoma(UCEC) [18]. We download a total of 1200 pathology representing six distinct types of cancer from <https://www.cancerimagingarchive.net>. Each dataset contains two classes: malignant tumors and nonmalignant tumors. Each cancer dataset is treated as a task.

Our preprocessing has four steps: pathology tiling, background removal and image resizing. (1) The pathology images are acquired in many downsampling factors, namely $1.0\times$, $4.0\times$, $8.0\times$ and so on. We crop all images into nonoverlapping patches of the same size (1024×1024) at the downsampling factor of $1.0\times$. (2) Since many patches are the background, which makes it difficult to extract useful features and increases the model parameters, we set a pixel threshold to screen out the background. The sum of the pixels of each patch is calculated. if the sum is less than the threshold, the patch is removed. Finally, we concatenate the remaining patches as a new image. (3) The number of remaining patches varies for each pathological image. The new images are resized by the bilinear interpolation method.

B. Baseline Network

A baseline network is established as the base learner to extract features of pathological images and classify malignant tumors and normal tissues. First, we crop the input images into patches of the same size ($W_p \times H_p \times C_p$, where W_p , H_p and C_p are the width, height and channels of the patches, respectively). Each input image can be cut into N patches. For extracting patch features, we design a multilayer convolution neural network (multi-CNN) including convolution layers, pooling layers, activation function and full connected layers. Second, similar to the vision transformer [13], we add a position embedding for each patch embedding to record the location information of the patch in the image. Then, we use several transformer encoders(see Fig. 2) to extract global features among patches, composed of multihead self-attention (MSA) mechanism and feed-forward networks. Finally, the multilayer perception (MLP) module is employed for classification.

C. Model Architecture

To improve the generalization ability of the baseline network, we learn a learner by MAML. Assume $T = \{T_1, T_2, \dots, T_n\}$ is a set of tasks, $X_i = [x_1, x_2, \dots, x_m]$ is the pathology image data of task T_i and $Y_i = [y_1, y_2, \dots, y_m] \in \{0 | 1\}^m$ is the label of X , where m is the number of images, and $y_j = 0$ denotes that image x_j is abnormal and diagnosed as cancer, $y_j = 1$ is normal relative to abnormal, and X_i and Y_i are divided into a support set and a query set. The MAML algorithm consists of meta-training stage and meta-testing stage. Some tasks (CCRCC, CM, PDA, SAR, UCEC) are used for meta-training, and the task on LSCC is for meta-testing. In meta-training, there are three steps. Consider a function f_W parameterized by W . Our baseline network is trained by a support set of a task T_i and the initial parameter W are updated to W_i according to stochastic gradient descent (SGD), as follows,

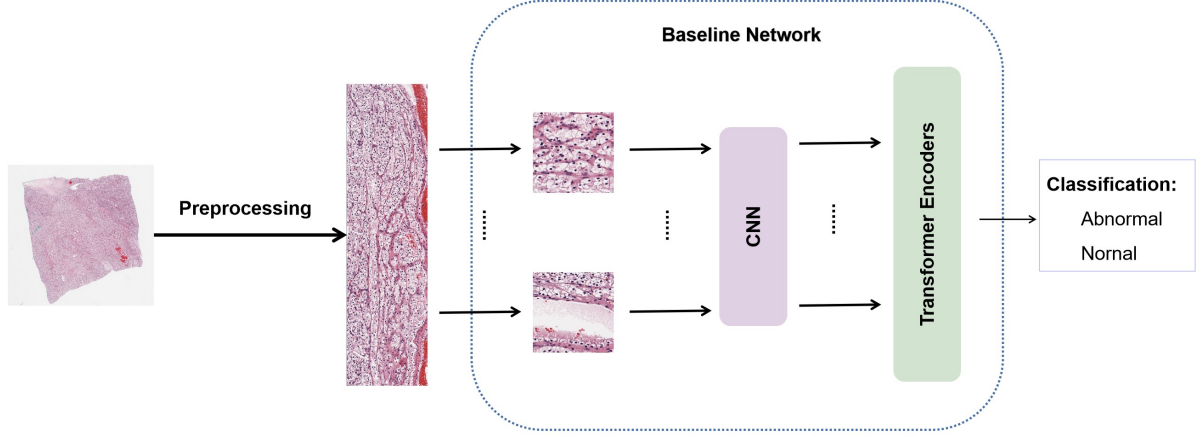


Fig. 1: The flowchart of the base learner of MAML. It contains preprocessing and a baseline network. For the baseline network, we firstly cut the input image and extract features of patches by CNN. And secondly, global features of embedded patches are extracted by transformer.

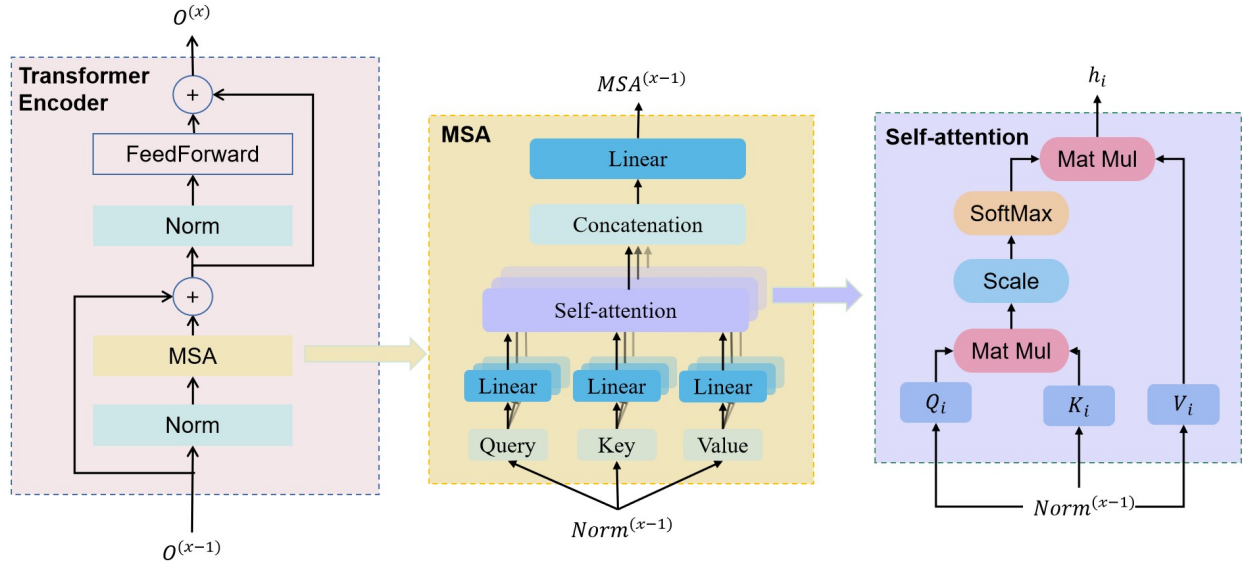


Fig. 2: The framework of a transformer encoder.

$$W_i = W - \lambda \nabla_W L_{T_i}(f_W) \quad i = 1, 2, 3, 4, 5 \quad (1) \quad W = W - \eta \nabla_W \sum_{T_i \in T} L_{T_i}(f_{W_i})$$

where λ is a learning rate for tasks, and L is a cross loss function. The loss function L is designed as:

$$L_{T_i}(f_W) = \sum_{(x^j, y^j) \in T_i} (y^j \log f_W(x^j) + (1 - y^j) \log(1 - f_W(x^j))) \quad (2)$$

Afterwards, we calculate the gradient of the model using samples in a query of T_i for meta-updating. After training all tasks, the initial parameter W is updated as follows,

$$\begin{aligned} &= W - \eta \frac{\partial(L_{T_1}(f_{W_1}) + L_{T_2}(f_{W_2}) + \dots + L_{T_5}(f_{W_5}))}{\partial W} \\ &= W - \eta \left(\frac{\partial L_{T_1}(f_{W_1})}{\partial W} + \frac{\partial L_{T_2}(f_{W_2})}{\partial W} + \dots + \frac{\partial L_{T_5}(f_{W_5})}{\partial W} \right) \\ &= W - \eta \left(\frac{\partial L_{T_1}(f_{W_1})}{\partial W_1} \cdot \frac{\partial W_1}{\partial W} + \frac{\partial L_{T_2}(f_{W_2})}{\partial W_2} \cdot \frac{\partial W_2}{\partial W} \right. \\ &\quad \left. + \dots + \frac{\partial L_{T_5}(f_{W_5})}{\partial W_5} \cdot \frac{\partial W_5}{\partial W} \right) \end{aligned} \quad (3)$$

where η is a hyperparameter for meta-updating. According to (1), $\frac{\partial W_i}{\partial W}$ is calculated as follows,

$$\frac{\partial W_i}{\partial W} = 1 - \lambda \approx 1 \quad (4)$$

So, equation (3) is as follows,

$$W = W - \eta \left(\frac{\partial L_{T_1}(f_{W_1})}{\partial W_1} + \frac{\partial L_{T_2}(f_{W_2})}{\partial W_2} + \dots + \frac{\partial L_{T_5}(f_{W_5})}{\partial W_5} \right) \quad (5)$$

Equation (5) denotes that the gradient of all training tasks can update the initial parameter W . After meta-training, our method is validated on the LSCC task. We fine-tune our model with a small number of lung cancer pathological images to update the parameters so that the baseline network can rapidly address the feature distribution of lung cancer data, thereby improving lung cancer diagnostic accuracy.

Algorithm 1 Lung Cancer Diagnosis Algorithm

Require: \mathbb{T} : A task set.

Require: λ, η : Two hyperparameters.

Require: $EPOCH_{meta}$: The number of meta updates,
 $EPOCH_{task}$: The number of iterations updates for per task.

```

1: Initialize Parameter  $W$  randomly.
2: for in  $EPOCH_{meta}$  do
3:   Sample task  $T_i \in \mathbb{T}$ .
4:   for all  $T_i$  do
5:     for in  $EPOCH_{task}$  do
6:       Sample dataset  $D_i$  to train the baseline network
       and update  $W$  by (1) and (2).
7:       Sample dataset  $D'_i$  to calculate gradient.
8:     end for
9:   end for
10:  Meta update  $W$  by (5)
11: end for
12: Return a trained model.
```

IV. EXPERIMENTS

A. Experimental Setting

We implemented our method in PyTorch, version 2.0.0, and Python 3.8 running it on a high-performance server with a Compute Unified Device Architecture (CUDA) of 11.8. The patch size was set to 256×256 . The hyperparameter λ in the meta-training stage was set to $1e-4$, and the hyperparameter η was set to $1e-2$. The batch size was 5, and $EPOCH_{meta}$ and $EPOCH_{task}$ were 20 and 30, respectively. The setting of the baseline network had 4 convolution layers, 3 fully connected layers and activation function Gaussian error linear unit (GELU) in the CNN module, 6 transformer encoders, and 4 heads of MSA. We utilized three evaluation indicators: *Accuracy*, *Precision*, *Specificity* and *Sensitivity*, calculated by true positive (TP), true negative (TN), false positive (FP), false negative (FN).

B. Experimental Results

The MLLCD model was built based on a CNN and transformer encoders. To verify the impact of different K values in few-shot learning on model performance, we conducted independent testing on the dataset LSCC. First, we prepared datasets with K values of 5, 10, and 20 for training. Then, we evaluated the impact of different training sample quantities K on model performance by drawing Receiver ROCs curves (Fig. 3(a), 3(b) and 3(c), respectively). The corresponding ROC values for diagnosing LSCC, both cancerous (also called abnormal) and normal, were 0.94, 0.94 (see Fig. 3(a)). We also observed that the performance of the 2-ways-10-shots and 2-ways-20-shots setting were slightly better than 2-ways-5-shots setting.

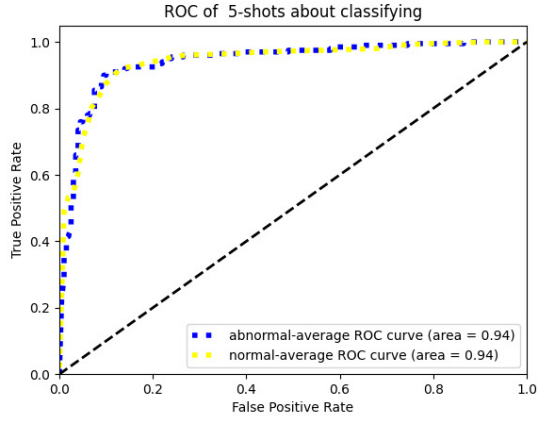
TABLE I: Performance On the Dataset

2-ways K-shots	Evaluation indicators			
	Accuracy	Precision	Specificity	Sensitivity
5-shots	85.00%	81.82%	90.00%	80.00%
10-shots	89.00%	89.00%	89.00%	89.00%
20-shots	90.50%	85.84%	97.00%	84.00%

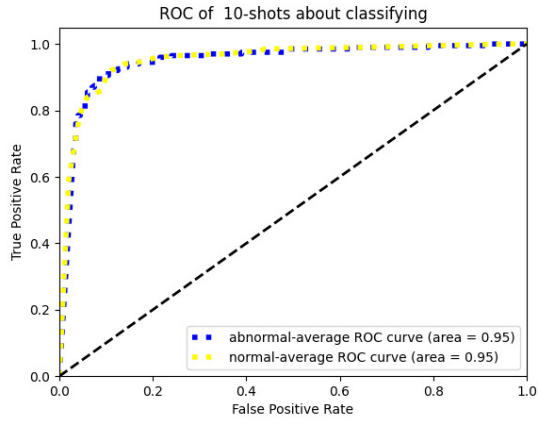
In Table I and Fig. 3, the results indicated that our model can effectively diagnose pathological images. They also showed that the more image data used for training, the better the diagnostic accuracy of the model. For lung cancer diagnosis, the lower the missed rate of the model, the better; that is, of all the samples diagnosed as normal, the more samples that were truly normal, the better. If our model diagnosed cancer patients as normal, the patients might miss the optimal treatment time. As shown in Table I, the accuracy of our model was above 85.0%, and the sensitivity was above 80.0%. Therefore, our method performed well.

In addition to analyzing the relationship between K-shots and model performance, we also explored the impact of two hyperparameters (λ, η) on model accuracy.

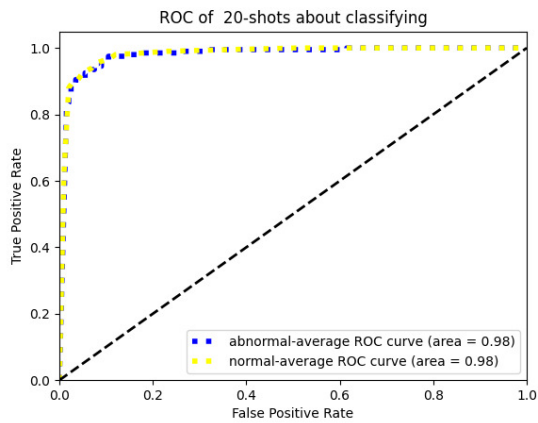
As Fig. 4 shows, the x-axis displayed the meta-update steps of model parameters from 0 to 17, and the y-axis displayed accuracy. In Fig. 4(a), for 5-shots, the model accuracy remained at basically 50.0%, indicating that the model had not learned any useful information. For 10-shots and 20-shots, the accuracy slightly improved after the seventh meta-update of the parameters. For 50-shots, the accuracy began to improve after the third parameter meta-update. We speculated that the model could not capture effective feature information during each task training due to the limited number of pathological images used for training and the hyperparameters being too small. In Fig. 4(b), for 5-shots, 10-shots and 20-shots, the accuracy significantly improved after the first meta-update of the parameters and gradually converged after the sixth meta-update. The difference was that the model learned effective meta-knowledge after the zero-parameters meta-updating. In Fig. 5, we investigated the performance differences between meta-learning and deep learning. The two models used the same dataset LSCC, with 50 positive and 50 negative samples each. For a meta-trained model, the new task required almost



(a) ROC curves of 2-ways and 5-shots scenario.

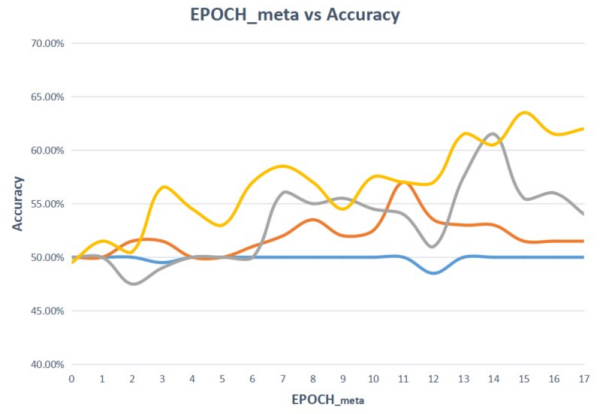


(b) ROC curves of 2-ways and 10-shots scenario.

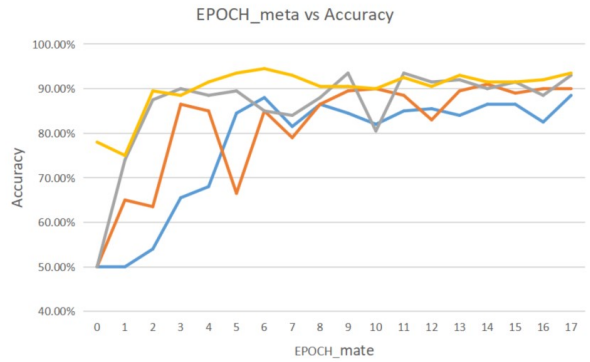


(c) ROC curves of 2-ways and 20-shots scenario.

Fig. 3: ROC curves of our model under different setting of number of training samples on dataset LSCC.



(a) $\lambda = 0.0001, \eta = 0.01$



(b) $\lambda = 0.01, \eta = 0.0001$

Fig. 4: For different hyperparameter settings, the accuracy curve changed with the meta update of model parameters.

no training to achieve good accuracy, while the deep learning model improved its accuracy after the 17th parameters update.

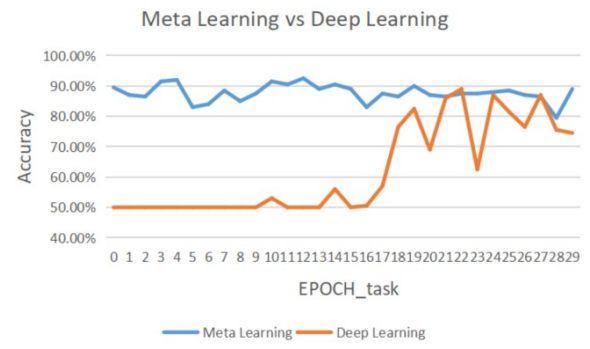


Fig. 5: The accuracy curves of the meta learning model and deep learning model for the same number of training times.

V. DISCUSSION AND CONCLUSION

In this work, we demonstrated that the proposed model can effectively diagnose lung squamous cell carcinoma with good

diagnostic performance (ROC=0.94) by using meta-learning-based ensemble models for image analysis of histopathological slides. Our method solved the problem that it is difficult for models to learn features caused by insufficient medical samples. In addition, we designed a base learner for MAML to capture local and global features of ultrahigh-resolution tissue pathology images.

To demonstrate the superiority of MAML over the baseline network, we extracted lung cancer patient data from the CPTAC database and conducted experiments under 2-ways and 50-shots settings to evaluate accuracy. We demonstrated that MAML outperforms the baseline networks and that meta trained models achieve optimal results within several fine-tuning steps.

However, our model also had some issues: we did not account for the specificity of the cancerous areas in the pathological images, increasing the training cost of the model and affecting the learning of cancerous features. Another drawback of our model is that the computational cost of model training is high, and we did not compare it to other available model architectures.

Overall, MLLCD is a promising framework based on meta-learning. The promising results indicate that MLLCD can effectively solve the problem of insufficient training data that may occur in real-world clinical practice.

ACKNOWLEDGMENT

This paper is supported by This work is supported by grants with No. KYQD(ZR)21079, No. 62162023 and No. Qhys2022-161.

REFERENCES

- [1] Colorectum fact sheet - Globocan 2020, <https://gco.iarc.fr/today/data/fact-sheets/cancers/https://gco.iarc.fr/today/data/factsheets/cancers/15-Lung-fact-sheet.pdf>. 2020
- [2] R. Rubin, D.S. Strayer, Rubin's Pathology: Clinicopathologic Foundations of Medicine. 4th edn. Philadelphia, USA: Lippincott Williams & Wilkins (2004)
- [3] X. Wang, G. Yu, Z. Yan, L. Wan, W. Wang and L. Cui, "Lung cancer subtype diagnosis by fusing image-genomics data and hybrid deep networks," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 20, no. 1, pp. 512–523, January-February 2023, doi: 10.1109/TCBB.2021.3132292.
- [4] T. Lin, Y. Wang, X. Liu and X. Qiu, "A Survey of Transformers," *arXiv*: 2106.04554, June 2021.
- [5] N. A. Koohbanani, B. Unnikrishnan, S. A. Khurram, P. Krishnaswamy and N. Rajpoot, "Self-Path: Self-supervision for classification of pathology images With limited annotations," in *IEEE Transactions on Medical Imaging*, vol. 40, no. 10, pp. 2845-2856, October. 2021, doi: 10.1109/TMI.2021.3056023.
- [6] H. Chen, C. Li, G. Wang, X. Li, M. M. Rahaman, H. Sun, et al., "GasHis-Transformer: A multi-scale visual transformer approach for gastric histopathological image detection," in *Pattern Recogn.*, vol. 130, no. C, October 2022, doi: 10.1016/j.patcog.2022.108827.
- [7] J. Yang, D. Song, H. An, and S. B. Seo, "Classification of subtypes including LCNEC in lung cancer biopsy slides using convolutional neural network from scratch," in *Sci Rep.*, vol. 12, February 2022, doi: 10.1038/s41598-022-05709-7.
- [8] R. J. Chen, C. Chen, Y. Li, T. Y. Chen, A. D. Trister, and R. G. Krishnan, "Scaling vision transformers to gigapixel images via hierarchical self-supervised learning," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, USA, pp. 16123-16134, 2022, doi: 10.1109/CVPR52688.2022.01567.
- [9] X. Wang, G. Yu, J. Wang, A. M. Zain and W. Guo, "Lung cancer subtype diagnosis using weakly-paired multi-omics data," in *Bioinformatics*, Vol. 38, pp. 5092–5099, November 2022, doi: 10.1093/bioinformatics/btac643.
- [10] Y. Wang, C. Hu, T. Kwok, C. A. Bain, X. Xue, R. B. Gasser, and et al., "DEMoS: A deep learning-based ensemble approach for predicting the molecular subtypes of gastric adenocarcinomas from histopathological images," in *Bioinformatics*, Vol. 38, pp. 4206–4213, September 2022, doi: 10.1093/bioinformatics/btac456.
- [11] T. Nair, A. F. pour and J. H. Chuang, "The effect of blurring on lung cancer subtype classification accuracy of convolutional neural networks," *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Seoul, Korea (South), pp. 2987-2989, December 2020, doi: 10.1109/BIBM49941.2020.9313192.
- [12] J. Chou, S. Bekiranov, C. Zang, M. Huai and A. Zhang, "Analysis of Meta-Learning Approaches for TCGA Pan-cancer Datasets," in *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Seoul, Korea (South), pp. 257-262, December 2020, doi: 10.1109/BIBM49941.2020.9313397.
- [13] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," *arXiv*: 2010.11929, June 2021.
- [14] National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC). The Clinical Proteomic Tumor Analysis Consortium Clear Cell Renal Cell Carcinoma Collection (CPTAC-CCRCC) (Version 10) [Data set]. The Cancer Imaging Archive, 2018, doi: 10.7937/K9/TCIA.2018.OBLAMN27.
- [15] National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC). The Clinical Proteomic Tumor Analysis Consortium Cutaneous Melanoma Collection (CPTAC-CM) (Version 10) [Data set]. The Cancer Imaging Archive, 2018, doi: 10.7937/K9/TCIA.2018.ODU24GZE.
- [16] National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC). The Clinical Proteomic Tumor Analysis Consortium Lung Squamous Cell Carcinoma Collection (CPTAC-LSCC) (Version 14) [Data set]. The Cancer Imaging Archive, 2018, doi: 10.7937/K9/TCIA.2018.6EMUB5L2.
- [17] National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC). The Clinical Proteomic Tumor Analysis Consortium Pancreatic Ductal Adenocarcinoma Collection (CPTAC-PDA) (Version 13) [Data set]. The Cancer Imaging Archive, 2018, doi: 10.7937/K9/TCIA.2018.SC20FO18.
- [18] National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC). The Clinical Proteomic Tumor Analysis Consortium Sarcomas Collection (CPTAC-SAR) (Version 9) [Data set]. The Cancer Imaging Archive, 2019, doi: 10.7937/K9/TCIA.2019.9BT23R95.
- [19] National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC). The Clinical Proteomic Tumor Analysis Consortium Uterine Corpus Endometrial Carcinoma Collection (CPTAC-UCEC) (Version 10) [Data set]. The Cancer Imaging Archive, 2019, doi: 10.7937/K9/TCIA.2018.3R3JUISW.