Research on Early Prediction of Lung Cancer Based on Deep Learning

Zhijun Qu
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
E-mail: quzj158@163.com

Zhongsheng Wang
State and Provincial Joint Engineering Lab. of Advanced Network, Monitoring and Control
Xi'an Technological University
Xi'an, 710021, Shaanxi, China
E-mail: wzhsh1681@163.com

***Abstract*—Cancer of the lung is a principal cause of mortality due to cancer on a global scale. Traditional imaging techniques suffer from subjectivity limitations. Meanwhile, convolutional neural networks (CNNs) within deep learning, though highly effective in image classification, still have limitations when dealing with complex and data-scarce medical images. To address this challenge, this paper proposes a data-efficient image Transformer (DeiT) model based on the Transformer architecture with a self-attention mechanism, enhanced through knowledge distillation. This model can capture global information in images and improve the classification accuracy of lung cancer images under small-sample conditions by leveraging a teacher model. Through model training and evaluation, results demonstrate that the DeiT model achieves an impressive prediction accuracy of 99.96% under small-sample medical imaging conditions. This highlights the advantages of the Transformer architecture in medical image analysis. The findings provide a new perspective for early lung cancer detection and underscore the powerful performance of the DeiT model in handling complex small-sample data conditions.**

Keywords-component; Lung Cancer Detection; Deep Learning; Knowledge Distillation; DeiT Model; Medical Image Analysis; Small-Sample Learning

I. INTRODUCTION

Over the past few years, AI's dramatic progression has spurred substantial breakthroughs in deep learning (DL), especially in computer vision. Convolutional neural networks (CNNs) are widely used in image classification and object detection. These models have found widespread applications in domains like facial recognition and autonomous driving, and security surveillance due to their powerful feature extraction capabilities. These advancements have improved image processing efficiency and accelerated the development of intelligent healthcare, making medical image analysis a key application of deep learning [1].

In the medical domain, particularly in tumor diagnosis and early detection, the analysis of medical imaging data poses significant challenges. Traditional imaging techniques like X-rays, CT, and MRI rely on physicians' expertise. However, the vast amount of image data and the complex nature of lesion morphology make these methods vulnerable to human error, increasing the chances of misdiagnosis or missed diagnosis. Lung cancer continues to be a leading cause of mortality globally, with early detection being essential for enhancing survival rates. However, early-stage lung cancer presents subtle symptoms, and its imaging data is complex, making traditional methods insufficient for efficient and accurate detection. Pathological image analysis depends on manual interpretation by pathologists, but due to the intricate details of tissue slices, this process is time-consuming and prone to errors, especially when detecting subtle cellular changes.

Deep learning has demonstrated great potential in medical image analysis. Conventional convolutional neural networks (CNNs), like VGG and ResNet, have attained remarkable outcomes in image classification and object detection through the incorporation of deeper network architectures and residual connections. However, these traditional CNNs typically focus on local features, making it difficult to effectively capture global

contextual information. This limitation is particularly evident when processing complex medical images, as local features may not fully represent the true nature of the disease. Additionally, medical image resources are often scarce, which presents another challenge. In this scenario, DeiT, a data-efficient image processing model that utilizes the Transformer architecture, has been recognized as a notable development in deep learning research in recent years.

The DeiT model not only captures global information from images through a self-attention mechanism but also enhances the ability to learn from small sample data through knowledge distillation, leveraging powerful teacher models. Compared to traditional CNN models, DeiT has an advantage when processing limited medical image data and has shown excellent performance in tasks such as early lung cancer detection.

This paper aims to explore the application of the DeiT model in early lung cancer detection. By analyzing lung pathological images, this paper compares the performance of DeiT with traditional convolutional neural networks (such as VGG and ResNet) and evaluates its accuracy and potential applications in lung cancer detection. Through experiments and data analysis, this paper aims to validate the advantages of the DeiT model under the Transformer architecture in medical image analysis, particularly under conditions of limited data samples, they provide innovative insights and solutions for early lung cancer detection [2].

II. Related Work

Detecting lung cancer at an early stage is essential for lowering its high mortality rate. Conventional imaging diagnostic techniques, including CT scans, X-rays, and pathological image analysis, depend on the expertise of radiologists, which can introduce subjectivity and potential misdiagnosis. With the progress of computer vision and deep learning, image-based lung cancer detection methods have attracted significant attention and are increasingly being applied in medical diagnosis.

A. Lung Cancer Detection Based on Deep Learning

Lately, approaches rooted in deep learning, most notably Convolutional Neural Networks (CNNs), has exhibited outstanding performance in medical image analysis and has progressively taken the place of traditional feature extraction approaches. Investigations reveal that CNNs have achieved remarkable success in identifying lung cancer images. Numerous studies have used deep learning models for CT screening, yielding high accuracy and sensitivity. Additionally, Cohen et al. developed an automated lung cancer detection system by applying deep learning to analyze lung nodules, surpassing the average performance of radiologists. In China, deep learning has also been widely applied. For example, Li Ming et al. used an improved ResNet model to classify lung cancer CT images, achieving high accuracy. Meanwhile, Zhang Hua et al. proposed a lung cancer detection method that integrates multiple deep learning networks, further enhancing the model's detection capabilities[3].

B. Lung Cancer Detection Based on Pathological Images

Unlike CT images, pathological images provide higher resolution, cell-level images, making them of significant value in cancer detection and diagnosis. Recently, there has been a rising focus on using pathological images for lung cancer detection. For example, Liu et al. developed a CNN-based technique to classify lung cancer pathological images, effectively distinguishing between malignant and normal regions. The research shows that, despite the relative scarcity of pathological image data, deep learning models, especially CNN-based architectures, can still achieve good classification results when processing this high-resolution data[4].

C. Knowledge Distillation and Small-Sample Learning in Lung Cancer Detection

With the continuous development of deep learning technology, knowledge distillation and few-shot learning have gradually become new directions in lung cancer detection. In recent years, researchers have focused on leveraging knowledge

distillation techniques to transfer insights from larger models to smaller ones, with the goal of improving the efficiency of lung cancer detection[5]. Studies have shown that the DeiT model excels in handling rare data, especially in the field of medical imaging, demonstrating strong generalization ability.

Therefore, deep learning-based lung cancer detection methods significantly improve accuracy and efficiency compared to traditional methods, but still face challenges such as data scarcity and environmental complexity. The lung cancer detection method based on the DeiT model proposed in this paper aims to improve the accuracy of lung cancer detection under few-shot conditions[6].

III. Technical Model

A. Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNNs), as the foundation of the two traditional models in the comparative experiments of this paper, utilize convolution operations to extract image features and combine mechanisms such as pooling to reduce data dimensions, thereby optimizing the processing efficiency of high-dimensional visual data. In this paper, two classic representative CNN models, VGG16 and ResNet50, were selected to train and test the early lung cancer detection task, in order to explore their performance in classification on complex medical small-sample image datasets. Figure 1 presents the fundamental structure of the CNN.

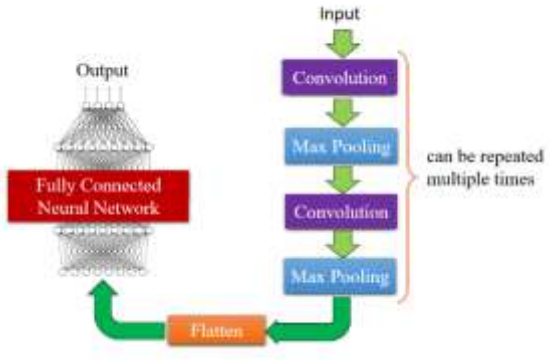

Figure 1. The basic architecture of a convolutional neural network

1) Convolution and Pooling

For Convolutional Neural Networks (CNNs), convolution acts as an essential process to extract features from the input images. This process employs a small convolutional kernel, such as a 3x3 or 5x5 matrix, to produce a feature map from the input image. The procedure entails an input image matrix I (with dimensions HxW) and a kernel matrix K (with dimensions kxk), along with a stride S. To control the size of the feature map, padding can be applied. Typical padding techniques include "valid padding" and "same padding," ensuring the output dimensions match the input image. The mathematical formulation is shown in Equation (1).

$$O(i,j)=\sum_{m=0}^{k-1}\sum_{n=0}^{k-1} I(i+m,j+n)\cdot K(m,n) \quad (1)$$

The element situated at position (i,j) within the output feature map is represented as O(i,j). I(i+m, j+n) is the corresponding element in the input image that aligns with the convolution kernel. K(m,n) represents the elements of the convolution kernel. The variable k represents the size of the convolution kernel. The dimensions of the feature map can be determined using the following equations (2) and (3).

$$\text{Output\_Height}=\frac{H-k+2P}{s}+1 \quad (2)$$

$$\text{Output\_Width}=\frac{W-k+2P}{s}+1 \quad (3)$$

Here, P represents the number of padding pixels (additional pixels at the image edges), and s is the stride. The convolution process is illustrated in Figure 2.

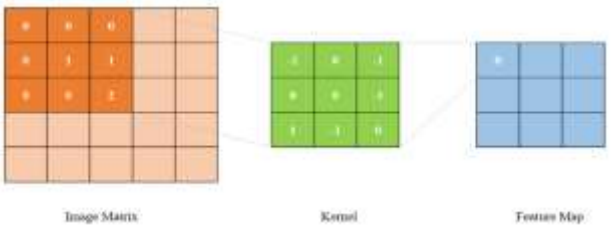

Figure 2. Convolution Process Diagram

The pooling layer in CNNs down samples feature maps, reducing dimensionality and computational complexity while enhancing translation invariance and preventing overfitting. Typically placed after the convolutional layer, it helps the model focus on essential image features. Max pooling is a widely used technique that selects the maximum value from a 2×2 or 3×3 window, typically with the stride equal to the window size. The formula is given in Equation (4).

$$O(i,j)=\frac{1}{k^2}\max_{m=0}^{k-1}\max_{n=0}^{k-1} I(i+m,j+n) \qquad (4)$$

Here, O(i,j) designates the element located at the coordinates (i,j) in the output feature map, and k represents the size of the pooling window. Figure 3 depicts the pooling operation.

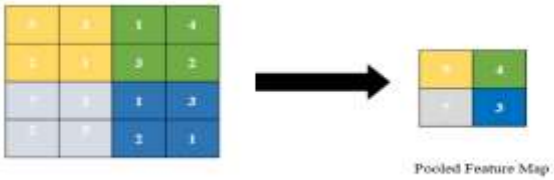

Figure 3. Pooling Process Diagram

2) VGG16 Model

VGG-16 is composed of 13 convolutional layers, each employing a 3×3 kernel, with a ReLU activation function applied after each convolution to maintain the network's non-linearity. After multiple convolutional layers, a max pooling layer is incorporated to execute down-sampling and decrease the feature dimensionality [7]. The entire model includes 5 pooling layers.

Following the convolutional and pooling layers, VGG-16 consists of three fully connected layers. The first two layers contain 4096 nodes each, while the final fully connected layer produces the classification results. The last component of the network is a Softmax output layer, which transforms the outputs of the fully connected layers into a probability distribution, indicating the probability that the image falls into each category. This architecture is depicted in Figure 4.

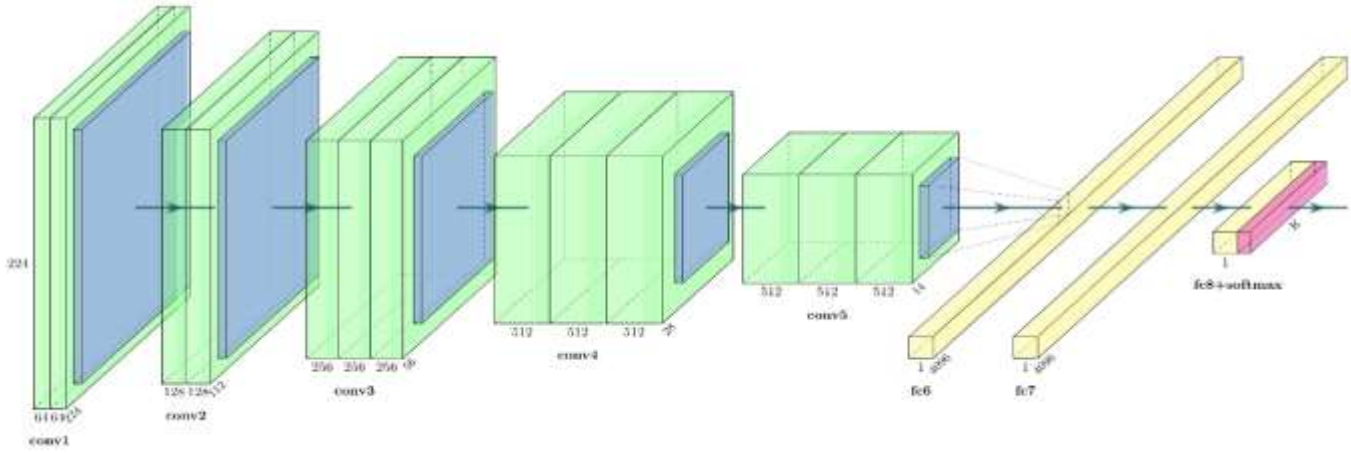

Figure 4. VGG16 Network Architecture Diagram

3) ResNet Model

Research on Convolutional Neural Networks (CNNs) has demonstrated that as deep neural networks grow deeper, they often encounter the problems of vanishing or exploding gradients during the training phase, thereby making training more challenging. In order to address this issue, the ResNet [8] model proposes a residual learning framework, which effectively overcomes the challenges associated with training deep networks.

Residual learning relies on skip connections, allowing the input to bypass layers and pass

directly to later ones. This helps the network learn a residual function instead of directly learning the mapping function. By optimizing the residual function, the network can capture complex features more effectively. The purpose of residual learning can be described by Equation (5).

$$y = F(x,\{Wi\}) + x \tag{5}$$

Assume x indicates the input, while F(x,{Wi}) illustrates the residual function, made up of nonlinear transformations (generally two or three convolutional layers) with parameters Wi. Let y denote the output. The input x is directly transmitted to the output via a skip connection, subsequently being merged with the result of F(x,{Wi}). This procedure is depicted in Figure 5.

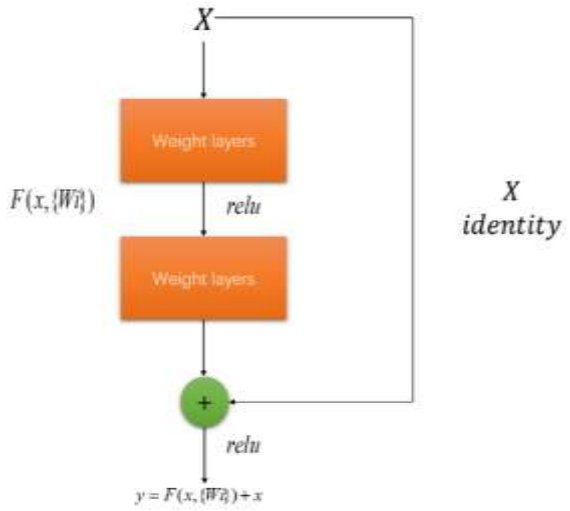

Figure 5. Residual Connection Structure Diagram

B. Transformer Model

The Transformer model employs the attention mechanism and differs from the architectures of Recurrent Neural Networks (RNN) and Long Short-Term Memory networks (LSTM) This design enables it to excel in parallel computation and effectively capture long-range dependencies. Since its introduction, the Transformer model has gained prominence, particularly in NLP and Computer Vision. Its architecture is illustrated in Figure 6.

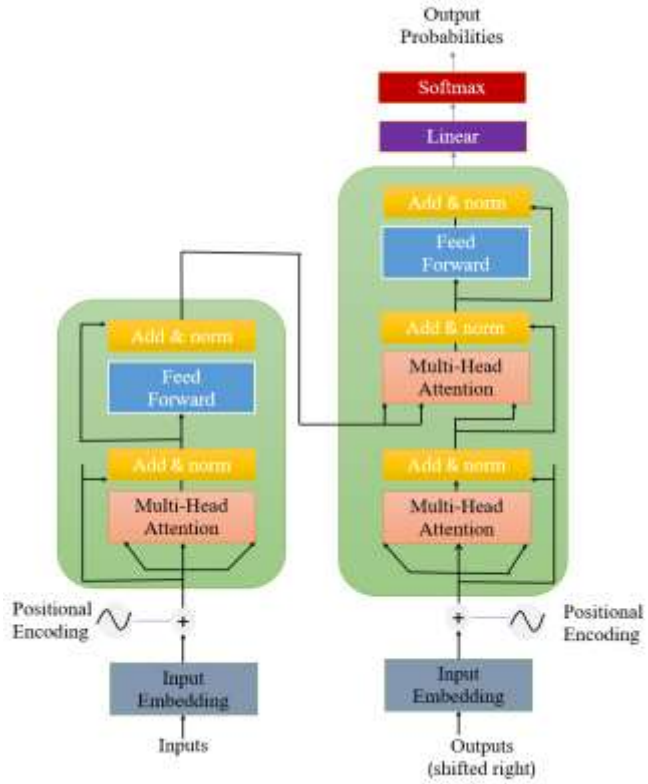

Figure 6. Transformer Architecture Diagram

1) Self-Attention Mechanism

The self-attention mechanism is an essential approach in deep learning for managing sequential data, and it is widely utilized in Natural Language Processing (NLP) and Computer Vision (CV). The core idea is to consider the influence of other parts when processing each part of the input data, allowing the model to capture relationships between different parts and helping to capture long-range dependencies, thereby improving performance on long sequences. The operation principle is as follows: Given an input sequence X = {x_1, x_2,..., x_n}, K denotes the key vectors, while V represents the value vectors. The similarity between each query vector Q and all the other key vectors K is determined by computing their dot product, then the results are normalized with a softmax operation to obtain the attention weights for each element. Equation (6) presents the computational formula.

$$Attention_{ij} = \frac{exp(Q_i \cdot K_k)}{\sum_{k=1}^{n} exp(Q_i \cdot K_k)} \quad (6)$$

Each element's representation is updated by computing a weighted sum of all value vectors Vj, where the weights are determined by the attention weights computed above. The calculation formula is provided in Equation (7). Eventually, the final result comprises a series of representations derived from this weighted sum.

$$Output_i = \sum_{j=1}^{n} Attention_{ij} \cdot V_j \quad (7)$$

Unlike traditional RNNs or LSTMs, the self-attention mechanism considers all elements of the sequence simultaneously, allowing better capture of long-range dependencies. Since it does not rely on the order of input elements, computations can be parallelized, improving efficiency. It can be applied to various data types, including text and images. Self-attention is a powerful technique for handling sequential data by modeling similarities between elements, enabling better understanding of internal relationships. It is widely used in advanced models like Transformer, ViT, and DeiT, offering significant advantages, especially for long sequences.

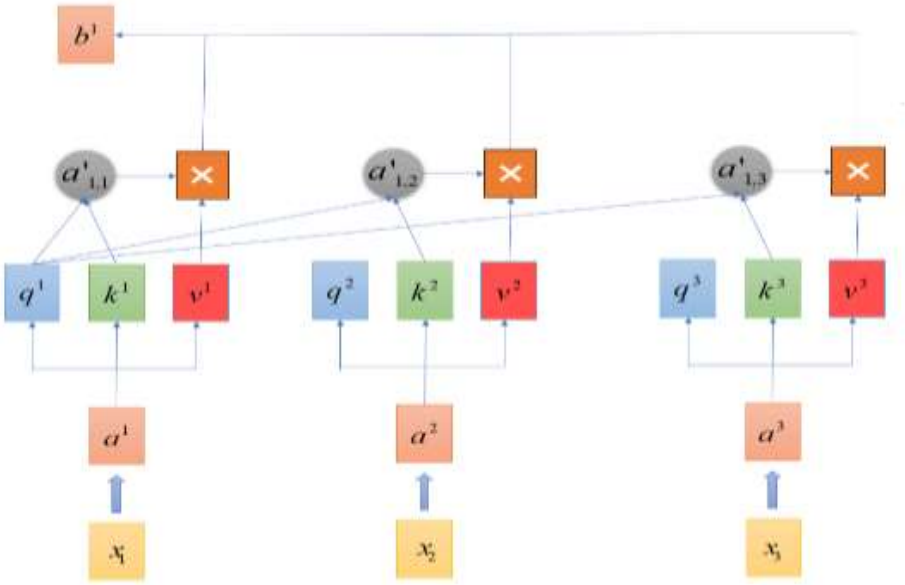

Figure 7. Self-Attention Mechanism Computation Diagram

2) ViT(Vision Transformer) Model

The Vision Transformer (ViT) is an image classification model that is founded on the Transformer architecture. It splits an image into fixed-size patches, treating each as a "token," and uses the self-attention mechanism to process the image, bypassing the convolutional layers typically found in traditional CNNs. ViT captures relationships between distant pixels in an image through self-attention, enabling global modeling[9]. Unlike CNNs, ViT performs better on large-scale datasets and can surpass traditional CNN models on massive datasets like ImageNet.

In ViT, the image is split into several patches, with each patch being transformed into a representation akin to word vectors. Positional encoding is incorporated to allow the model to detect spatial relationships among patches. The processed patches are fed into the Transformer encoder, where their representations are refined through self-attention layers. In the end, the output is categorized through a fully connected layer. Figure 8 illustrates the model architecture.

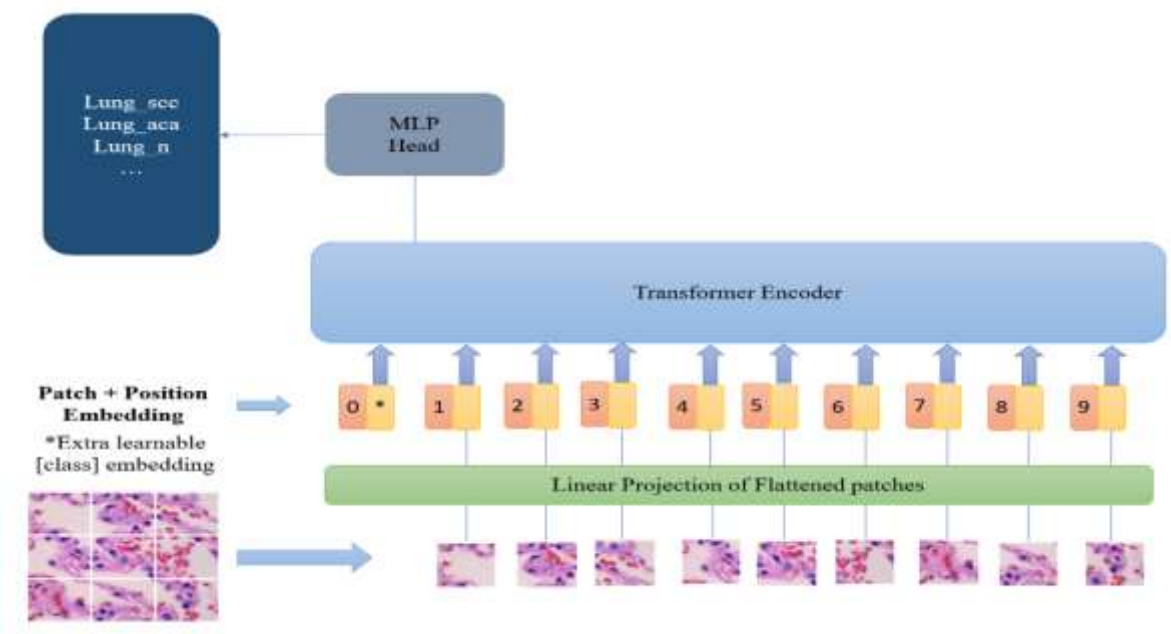

Figure 8. ViT Architecture Diagram

3) DeiT(Data-efficient Image Transformer) Model

DeiT is a variant of ViT specifically designed to improve the data efficiency of Transformer models in image classification tasks. By introducing knowledge distillation, the DeiT model successfully overcomes the inefficiency of training ViT on small datasets, this allows ViT models to perform on par with Convolutional Neural Networks (CNNs), even when there is a limited amount of data [10].

The core design principle of DeiT is based on knowledge distillation. By incorporating a teacher model, knowledge is transferred to the student model (DeiT), allowing it to learn more efficient feature representations, even with a small amount of data. This allows DeiT to achieve strong performance on small datasets while avoiding overfitting or underfitting issues commonly seen in ViT training.

The working principle of DeiT is similar to that of ViT. DeiT first divides the input image into fixed-size patches, flattens each patch, and maps them into an embedding space through a linear transformation. Each patch embedding is then enhanced with positional encoding to preserve spatial information. After positional encoding, the patches are fed into a Transformer encoder, which utilizes self-attention mechanisms to capture relationships between different patches in the image.[11]

The key innovation of DeiT lies in the introduction of the knowledge distillation mechanism to improve the training process. During training, DeiT optimizes two objectives simultaneously: Supervised loss: Computed by comparing the output of the class token with the hard labels. Distillation loss: Computed by comparing the output of the distill token with the soft labels generated by the teacher model.

Specifically, the teacher model generates a probability distribution (soft labels) for each class, capturing the similarities between categories. The distillation loss is calculated using Kullback-Leibler (KL) divergence, which estimates the divergence between the output of the student model's distill token and the soft labels generated by the teacher model [12].

To integrate these two optimization objectives, DeiT defines a total loss function (Ltotal) that combines both the supervised loss and the distillation loss. The formula is provided in Equation (8), where α and β are key weights that balance the two loss components. Here, Ltotal

denotes the total loss, and Lsupervised refers to the supervised loss, and Ldistillation is the distillation loss[13].

$$L_{total} = \alpha L_{supervised} + \beta L_{distillation} \quad (8)$$

The supervised loss Lsupervised optimizes the parameters of the class token to improve classification accuracy, while the distillation loss Ldistillation optimizes the parameters of the distill token, enabling it to learn deep feature representations from the teacher model. Additionally, both loss terms jointly optimize the shared parameters of the Transformer encoder. The distill token is a crucial innovation in DeiT, offering a pathway for the student model to receive knowledge from the teacher model, thereby significantly enhancing its performance on small datasets. Through its innovative design and distillation techniques, DeiT represents a major breakthrough in deep learning for computer vision, demonstrating its strong potential, particularly in image classification tasks. Its model architecture is shown in Figure 9.

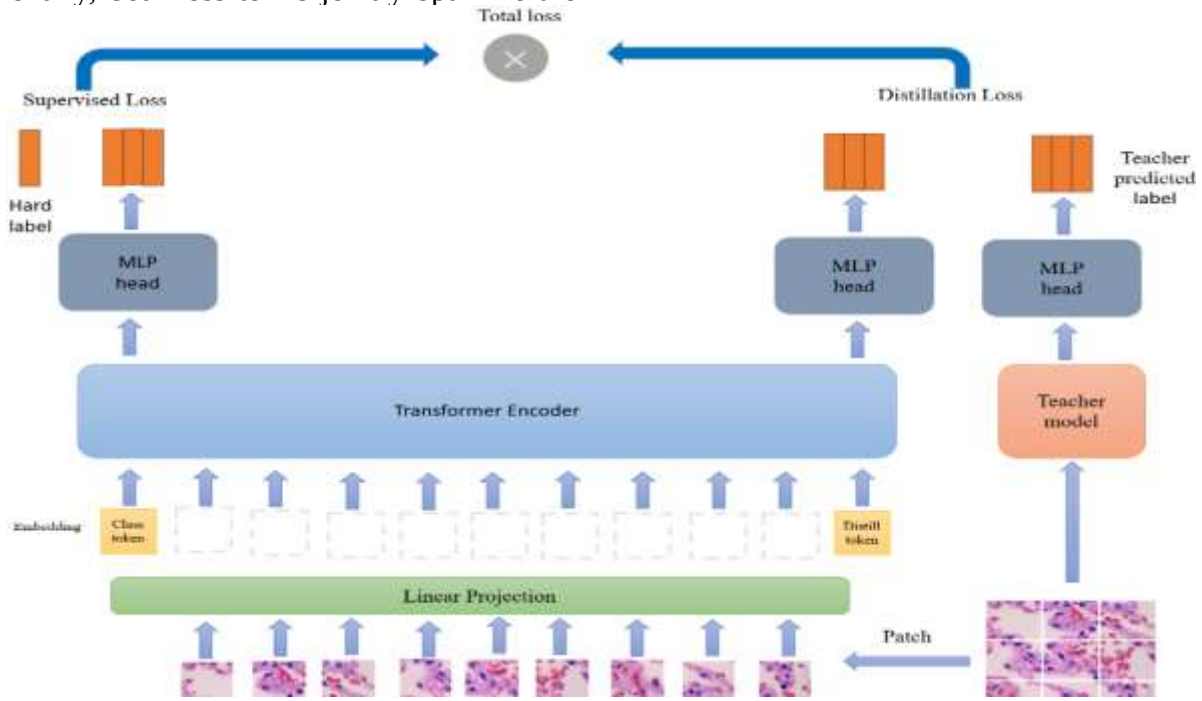

Figure 9. DeiT Architecture Diagram

IV. Experiment And Analysis

A. Experimental Environment and Model Parameters

The models utilized in this experiment were trained and fine-tuned on Kaggle with the help of the TensorFlow and PyTorch frameworks. The code was developed and run in a Jupyter Notebook environment. The hardware configuration included a GPU P100, TensorFlow version 2.16.1, Python version 3.10.14, and PyTorch version 2.4.0. The Adam optimizer, with a learning rate of 0.00001 and cross-entropy loss, was applied during training. A batch size of 64 was used, and the model was trained for 30 epochs. VGG16, ResNet50, and DeiT were evaluated on datasets containing lung adenocarcinoma, lung squamous cell carcinoma, and normal lung tissue.

B. Experimental Procedure

The lung cancer image dataset used in this study consists of 15,000 pathological images, categorized into three groups: lung_aca (lung adenocarcinoma), lung_n (healthy lung tissue), as shown in Table 1, the dataset is divided into a

training set with 10,500 images, a validation set containing 2,250 images, and a test set of 2,250 images.

TABLE I. EXPERIMENTAL DATASET TABLE

	Training Set	Validation Set	Test Set
lung_aca	3500	750	750
lung_scc	3500	750	750
lung_n	3500	750	750

In order to improve the model's capacity for generalization, a variety of data augmentation methods were utilized on the images throughout the training process. These methods included arbitrary horizontal flipping, rotation, and translation shear. The images were also resized to 224x224 pixels and normalized using the mean and standard deviation from the ImageNet dataset, with values (mean = [0.485, 0.456, 0.406] and standard deviation = [0.229, 0.224, 0.225]). Data augmentation increased the diversity of the training data, helping the model learn more transformed features, thereby improving its performance in various scenarios. The images underwent random horizontal flipping with a probability of 10%, while the rotation range was set between -10 and +10 degrees. Random cropping was applied to adjust the images to the target size (224x224), with a cropping ratio ranging from 90% to 110%. Additionally, random transformations through translation and shear were applied, with a maximum transformation of 10%, helping the model adapt to different perspectives and spatial positions. The illustration in Figure 10 displays nine arbitrarily chosen enhanced images derived from the lung cancer pathological training dataset. The images are equally distributed across three different types of lung cancer: lung adenocarcinoma, normal lung tissue, and lung squamous cell carcinoma, as mentioned in [14].

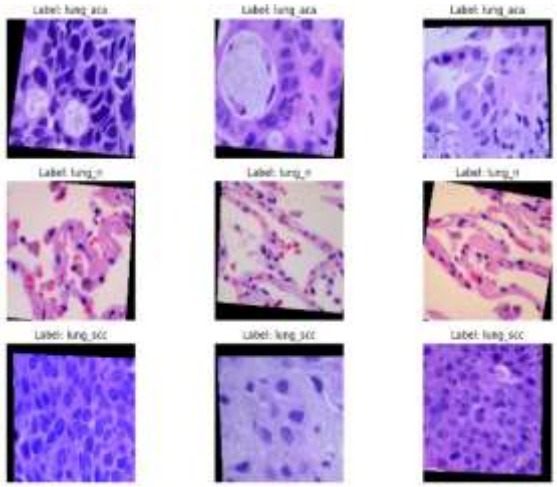

Figure 10. Randomly selected training sample images

In this study, the DeiT model (deit_base_patch16_224) based on Vision Transformer (ViT) was used and compared with traditional convolutional neural network (CNN) models, specifically VGG16 and ResNet50. During training, the Cross Entropy Loss function was used to calculate the loss, and the Adam optimizer was applied to adjust the model's parameters. A learning rate of 1e-5 was chosen, and the training spanned 30 epochs. At each epoch, the loss and accuracy for both the training and validation datasets were calculated to monitor the model's progress.

C. Experimental Results and Analysis

1) Training Results Analysis

The model's stability and ongoing performance improvement were demonstrated by plotting the loss and accuracy curves for both training and validation. As shown in Figures 11 and 12, the accuracy and loss curves for the VGG16 and ResNet50 models exhibited distinct patterns after 30 epochs of training. During the early epochs, the models had lower accuracy and higher loss values. However, as training progressed, the accuracy gradually improved, and the loss decreased. By the end of 30 epochs, both the accuracy and loss curves had stabilized, indicating convergence. Ultimately, the accuracy of VGG16 and ResNet50 reached 98.49% and 97.51%, respectively[15].

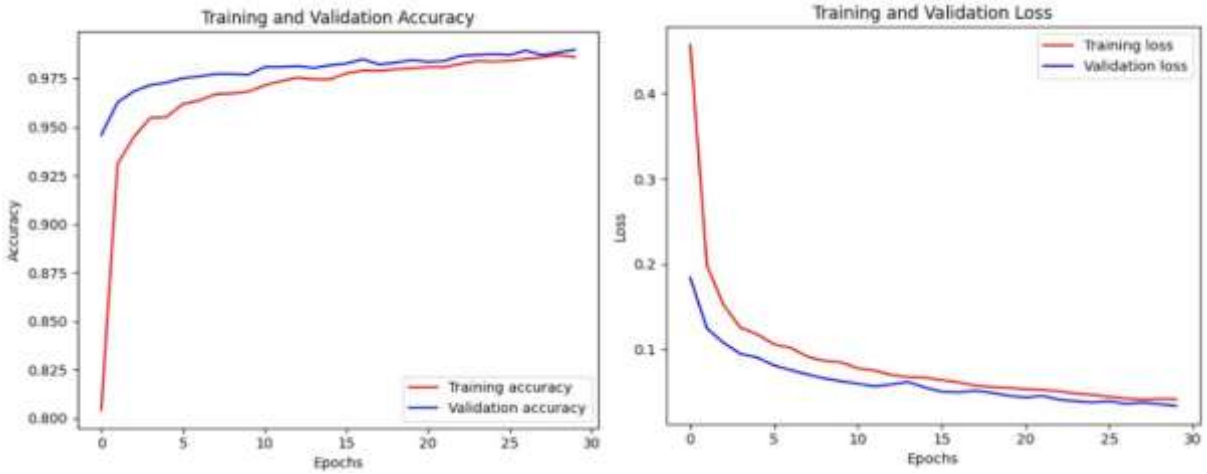

Figure 11. Accuracy and Loss Curves of theVgg16

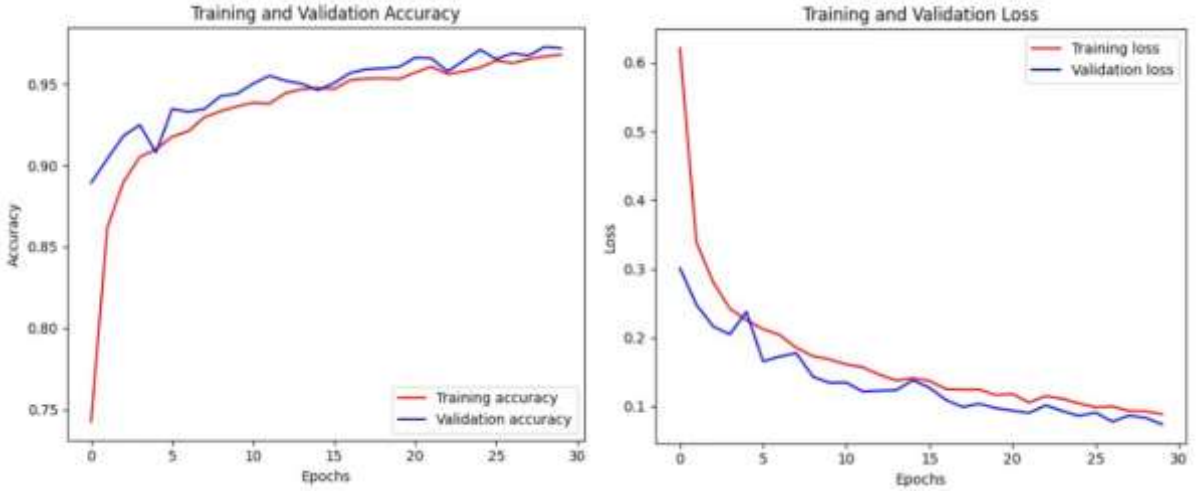

Figure 12. Accuracy and Loss Curves of the ResNet50

The DeiT model achieved the best training results and the highest accuracy, as shown in Figure 13. At the beginning of training, the DeiT model already had a high initial accuracy. This is due to the introduction of knowledge distillation and the self-attention mechanism, which enhance its ability to capture contextual information and extract complex medical image features more effectively. Additionally, the student model benefits from the guidance of the teacher model, allowing it to develop strong feature learning capabilities from the very start. After 30 epochs, the DeiT model successfully converged, achieving an accuracy of 99.96%, demonstrating its outstanding performance in medical image classification tasks. Compared to traditional CNN models, the DeiT model provides higher accuracy on small sample datasets, fully showcasing the powerful advantages of self-attention mechanisms and knowledge distillation in image classification.

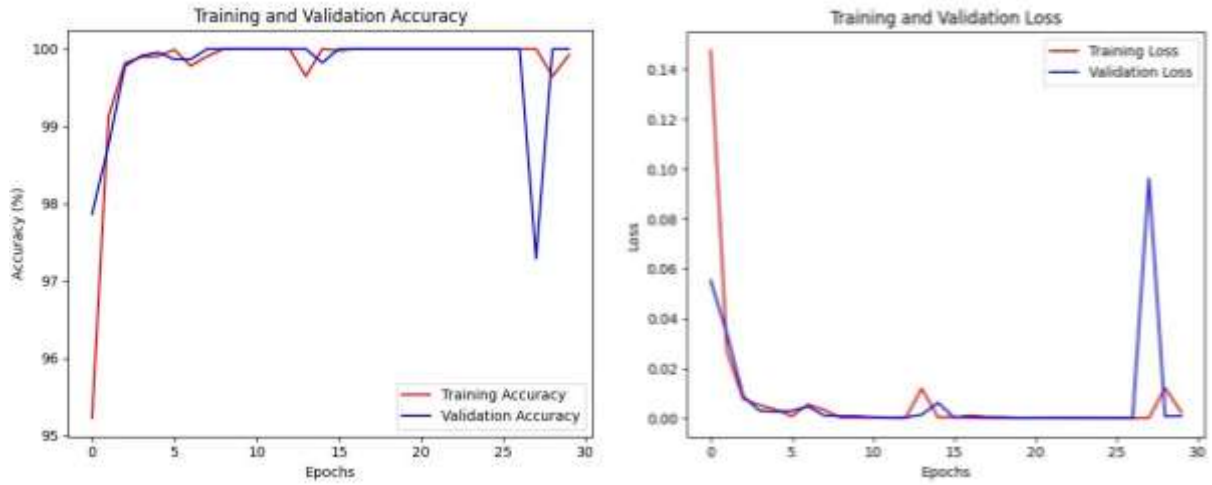

Figure 13. Accuracy and Loss Curves of the DeiT

2) Testing Results Analysis

The categorization performance of the DeiT model for various classes is graphically depicted using a confusion matrix, as illustrated in Figure 14. The results of the confusion matrix show that out of 2,250 test images, 2,249 were correctly predicted, with only one misclassification, demonstrating the exceptional performance of the DeiT model on complex small-sample medical images.

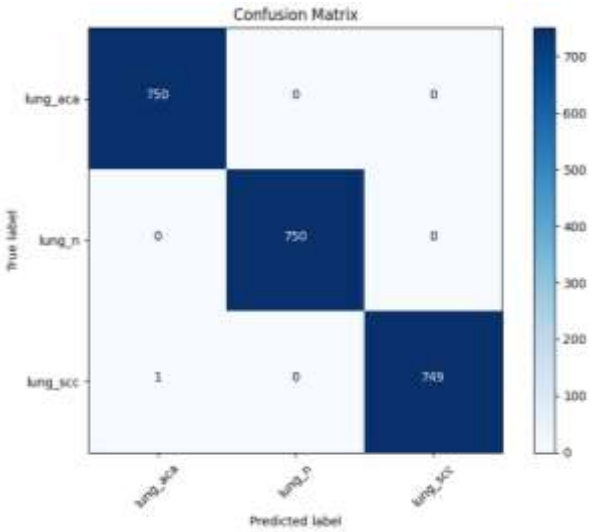

Figure 14. Confusion Matrix

Assess the model with the aid of the test set. The model's performance is assessed by calculating the loss and accuracy on the test set, along with generating a detailed classification report. This report presents the precision, recall, and F1-score for each category. Using these metrics, the overall average precision, recall, F1-score, and macro average are computed. The formulas for each test metric are presented below in Equations (9-13).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

$$Precision = \frac{TP}{TP+FP} \tag{10}$$

$$\mathrm{Re}call = \frac{TP}{TP + FN} \tag{11}$$

$$F1-Score = 2 \times \frac{\mathrm{Pr}ecision \times Recall}{Precision + Recall} \tag{12}$$

$$Macro\ Average = \frac{1}{N}\sum_{i=1}^{N} Metric_i \tag{13}$$

In this context, TP represents true positives, and TN stands for true negatives. The greater the values of TP and TN, the superior the model's predictive capability. FP refers to false positives,

while FN represents false negatives. The smaller the values of FP and FN are, the fewer errors the model commits in its predictions. N denotes the total number of categories. $Precision_i$ signifies the precision for category i, whereas $Recall_i$ denotes the recall for category i, which assesses the proportion of actual positive samples that are correctly identified. A higher recall means the model identifies more positive cases, reducing the likelihood of missing them. $F1\text{-}Score_i$ denotes the F1-score for class i, which is the harmonic mean of precision and recall, effectively balancing both metrics. A higher F1-score indicates better overall predictive performance. $Metric_i$ refers to the precision, recall, or F1-score of class i, while W_i represents the number of samples in class i. Table 2 below presents the evaluation metrics calculated after the model's execution.

TABLE II. PREDICTION METRICS FOR DIFFERENT MODELS

Model	Acc(%)	Average Precision(%)	Average Recall(%)	Average F1-Score(%)
Vgg16	98.49	98.49	98.49	98.49
Resnet50	97.51	97.51	97.51	97.51
DeiT	99.96	99.96	99.96	99.96

The test results indicate that the DeiT model performs the best. Leveraging self-attention mechanisms, it effectively captures long-range dependencies in images and employs knowledge distillation techniques, making it particularly suitable for complex medical image classification tasks on small datasets. To provide a more intuitive demonstration of the model's performance, several samples were randomly selected from the test set, comparing the DeiT model's predictions with the ground truth labels[16]. As shown in Figure 15, these images illustrate the model's classification performance across different categories, further validating its exceptional performance on small-sample, complex medical data.

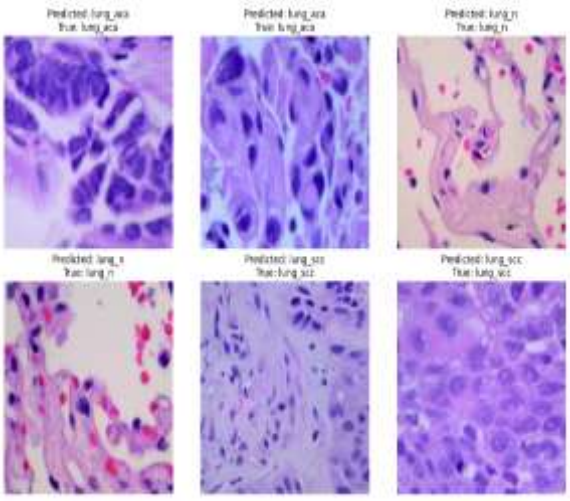

Figure 15. Random Test Plot

V. CONCLUSION AND OUTLOOK

This study explores the task of lung pathology image detection by conducting comparative experiments based on VGG16, ResNet50, and DeiT models. Given the limited dataset size, the experimental results demonstrate that the DeiT model, supported by the self-attention mechanism, can more effectively capture subtle features in pathological images. Additionally, by incorporating the knowledge distillation strategy, DeiT significantly enhances small-sample learning performance. Compared to traditional convolutional neural networks, the DeiT model achieved the highest classification accuracy on the test set, fully showcasing its potential in complex medical image analysis.

Looking ahead, the advantages of the DeiT model lay a solid foundation for novel opportunities in the domain of medical image processing. In broader medical application scenarios, such as the diagnosis of rare diseases, the combination of the DeiT model with knowledge distillation is expected to further demonstrate its capabilities on small-scale datasets, providing strong technical support for the early diagnosis of rare diseases.

REFERENCES

[1] Wu Hongjie, Tian Chuangchuang, Tao Ran, et al. Research on Building Displacement Prediction Method Based on Graph Convolution Distillation

Transformer. Journal of Suzhou University of Science and Technology (Natural Science Edition), 2024, 41(04): 128-138.

[2] Yao Yiying, Chen Junji, Ren Denghong, et al. Case Analysis of Medical Image Recognition and Diagnosis Based on Deep Learning. Application of Integrated Circuits, 2024, 41(12): 80-81.

[3] Liu Yuxin, Meng Yu, Deng Yupeng, et al. A Dual-Stream Extraction Model for High-Resolution Remote Sensing Building Images Integrating CNN and Transformer. Journal of Remote Sensing, 2024, 28(11): 2943-2953.

[4] Li Yunfei, Li Shuting, Zhang Shuai, et al. Research Progress on Deep Learning in Tumor Image Classification. Chinese Journal of Cancer Prevention and Treatment, 2024, 31(12): 719-724.

[5] Zong Haoyu, Qin Yuliang, You Ziyuan. Advances in Deep Learning Applications in Musculoskeletal Imaging. Imaging Research and Medical Applications, 2024, 8(10): 1-3.

[6] Liu Libing, Fu Liyao. Applications and Prospects of Deep Learning Technology in Medical Image Analysis. New Generation Information Technology, 2024, 7(01): 24-28.

[7] Hu Kun, Wu Guoqing, Hu Zuhui, et al. Research on Metal Surface Defect Image Classification Based on an Improved VGG16 Network. Computer Applications and Software, 2024, 41(06): 175-180.

[8] Liu Yansheng, Yu Qianru, Zhang Kun, et al. Establishment and Clinical Testing of a ResNet-Based Model for Colonoscopy Image Classification of Ulcerative Colitis. World Science and Technology - Modernization of Traditional Chinese Medicine, 2024, 26(09): 2346-2354.

[9] Lin Hailin, Chen Guoming, Tang Peiyu, et al. A Lightweight Image Classification Method Based on Convolutional Vision Transformer Fusion. Modern Computer, 2024, 30(22): 1-7.

[10] Chen Ning, Liu Fan, Dong Chenwei, et al. Few-Shot Image Classification Based on Local Contrastive Learning and New Class Feature Generation. Pattern Recognition and Artificial Intelligence, 2024, 37(10): 936-946.

[11] Wang Haibao, Liu Hongyan, Wei Zhi, et al. Research on Bone Marrow Cell Image Classification Based on Deep Learning. Genomics and Applied Biology, 2024, 43(Z2): 1872-1882.

[12] Gong Xuanjin. Long-Tailed Visual Recognition Method Based on Multi-Classifier Hierarchical Distillation. Modern Information Technology, 2024, 8(16): 49-52+59.

[13] Zhao Hongwei, Wu Hong, Mark, et al. An Image Classification Framework Based on Knowledge Distillation. Journal of Jilin University (Engineering Edition), 2024, 54(08): 2307-2312.

[14] Zhou Chengyang, Liu Wei, Wu Tianrun, et al. Rock Thin Section Image Classification Based on a Hybrid Expert Model. Journal of Jilin University (Science Edition), 2024, 62(04): 905-914.

[15] Zhang Li, Yang Minghui, Sun Jiacheng. Few-Shot Tea Leaf Disease Recognition Based on Attention Mechanism and Transfer Learning. Journal of Chinese Agricultural Mechanization, 2024, 45(10): 262-268.

[16] Zhao Tingting, Gao Huan, Chang Yuguang, et al. Fine-Grained Image Classification Method Based on Knowledge Distillation and Target Region Selection. Computer Applications Research, 2023, 40(09): 2863-2868.

.