Journal of Artificial Intelligence and Technology, 2023, 3, 25-31
https://doi.org/10.37965/jait.2023.0153

ISTP
RESEARCH ARTICLE

Lightweight Classification Network for Pulmonary Tuberculosis Based on CT Images

Junlin Tian,[1] Yi Zhang,[2] Junqiang Lei,[3,4] Chunyou Sun,[1] and Gang Hu[5]
[1]School of Mathematics and Statistics, Lanzhou University, Lanzhou 730000, People's Republic of China
[2]The First Hospital of Lanzhou University, Lanzhou 730000, People's Republic of China
[3]Radiological Clinical Medicine Research Center of Gansu Province, Lanzhou 730000, People's Republic of China
[4]Intelligent Imaging Medical Engineering Research Center of Gansu Province, Lanzhou 730000, People's Republic of China
[5]Computer Information Systems Department, State University of New York at Buffalo State, Buffalo, NY 14222, USA

(Received 19 December 2022; Revised 09 January 2023; Accepted 11 January 2023; Published online 11 January 2023)

Abstract: With the continuous development of medical informatics and digital diagnosis, the classification of tuberculosis (TB) cases from computed tomography (CT) images of the lung based on deep learning is an important guiding aid in clinical diagnosis and treatment. Due to its potential application in medical image classification, this task has received extensive research attention. Existing related neural network techniques are still challenging in terms of feature extraction of global contextual information of images and network complexity in achieving image classification. To address these issues, this paper proposes a lightweight medical image classification network based on a combination of Transformer and convolutional neural network (CNN) for the classification of TB cases from lung CT. The method mainly consists of a fusion of the CNN module and the Transformer module, exploiting the advantages of both in order to accomplish a more accurate classification task. On the one hand, the CNN branch supplements the Transformer branch with basic local feature information in the low level; on the other hand, in the middle and high levels of the model, the CNN branch can also provide the Transformer architecture with different local and global feature information to the Transformer architecture to enhance the ability of the model to obtain feature information and improve the accuracy of image classification. A shortcut is used in each module of the network to solve the problem of poor model results due to gradient divergence and to optimize the effectiveness of TB classification. The proposed lightweight model can well solve the problem of long training time in the process of TB classification of lung CT and improve the speed of classification. The proposed method was validated on a CT image data set provided by the First Hospital of Lanzhou University. The experimental results show that the proposed lightweight classification network for TB based on CT medical images of lungs can fully extract the feature information of the input images and obtain high-accuracy classification results.

Keywords: tuberculosis case classification; CNN; Transformer; lightweight network

I. INTRODUCTION

Medical image classification plays an important role in clinical image diagnosis. Among them, the classification of pulmonary tuberculosis (TB) cases aims to make more objective judgments and explanations on its occurrence, development, pathological changes, clinical manifestations, prognosis, and prognosis. It also can solve the phenomenon of coexistence of pulmonary TB and lung cancer in basic level hospitals, difficulty in clinical and X-ray identification, and easy missed diagnosis, so as to guide treatment and prevention. It is of great significance for clinical guidance.

In recent years, with the continuous development of medical informatization and digital diagnosis, the medical monitoring indicators are growing, and the amount of data is becoming larger and larger. Therefore, a strong data processing capability is urgently needed to provide strong support for the medical field. As a hot branch of AI field, deep learning has been developed rapidly in speech recognition and computer vision, and its application in medical field has become more and more mature. The application of deep learning to various medical diagnostic tasks is very effective, even surpassing human experts in some aspects. Among them, the application of deep learning technology to medical image classification is one of the hot spots of current research. The most common application scenario is to apply deep learning technology to disease auxiliary diagnosis. For example, magnetic resonance medical imaging can be used to diagnose whether intra-abdominal masses have pathological changes; classification of pulmonary TB cases by pulmonary computed tomography (CT). Gao et al. [1] proposed an automatic learning feature through depth learning technology to classify the severity of image nucleus cataract under slit lamp. Payan et al. [2] can identify and diagnose the disease status of patients with Alzheimer's disease by building an intelligent network model based on brain magnetic resonance medical images through sparse self-coding network and three-dimensional convolutional neural network (CNN). Shen et al. [3] proposed a multi-scale CNN. This multi-scale CNN can capture the heterogeneity of pulmonary nodules by alternating stacking layers and has a good classification effect on benign and malignant pulmonary nodules.

When deep learning technology is applied to the task of classification of pulmonary CT-related pulmonary TB cases,

Corresponding author: Junlin Tian (e-mail: tianjl2019@lzu.edu.cn).

the main consideration is the feature extraction of pulmonary CT information and the efficiency of classification of pulmonary TB cases. In medical image classification, deep learning, especially CNNs, has proved to have great potential in classification. CNN has the ability of representation learning and can classify the input information according to its hierarchical structure. In addition, the rapid development of modern computers provides powerful computing support for CNN training. Most of the traditional medical image classification methods need to extract the brightness, shape, gradient, gray, and texture features of the image and select the relevant features or the fusion features of the above features. These methods have complex processes, while the depth learning method can automatically train images and extract multilevel features without complex image processing steps, which improves the final classification effect. However, the CNN is limited by the receptive field of convolution, and each convolution kernel can only focus on the local area rather than the global context. That makes it difficult to capture the implicit relationship between images [4]. Global context is very important for semantic feature extraction, especially for medical image feature extraction.

In recent years, the rapidly developing visual attention mechanism helps the traditional CNN network to obtain global feature information to a certain extent. For example, the SENet proposed by Hu [5] uses the global adaptive pooling layer to capture global information on the channel domain, compress it, and then weight it on the feature channel. However, the CBAM proposed by Woo [6] and others uses the maximum pooling layer and the average pooling layer to reintegrate and reweight the information in the channel domain and spatial domain, respectively. Although pooling operation can obtain global feature representation with a small amount of parameters and computation, pooling operation inevitably ignores some important details.

However, using Transformer [7,8] as a medical image classification framework can effectively solve the problem of global context feature extraction. Transformer divides the input image into feature blocks of fixed size and obtains feature sequences after linear transformation, which makes the model to have strong modeling ability. At the same time, it uses the self-attention mechanism to perceive the global context information. It can fully capture the long-distance feature dependency relationship and establish the relationship between feature basic units (Feature tokens), thus effectively solving the problem of global feature extraction. However, the input size scale of Transformer is fixed. For image classification requiring pixel-level prediction, there is a problem of too much computational complexity [9]. At the same time, Transformer has poor ability to capture local features of a single patch and has inherent induction bias.

In addition, it is also crucial to realize the lightweight of the network in the task of medical image classification. Image classification is relatively more complex. In image training, a large number of redundant samples are collected, which leads to the problem that the training time of CNN is too long. The lightweight model has the advantages of light structure, simple calculation, and strong portability. It can improve the operation speed to the greatest extent without reducing the accuracy of the model. Therefore, it is also of great research significance and application value to realize the lightweight of deep learning network in image classification tasks.

Based on this, this paper proposes lightweight Transformer and CNNs as the network structure of classification network. The structure is composed of CNN branch and Transformer branch. In this architecture, the CNN branch not only supplements the basic local feature information to the Transformer branch at the low level but also provides different local and global feature information to the Transformer architecture at the middle and high levels of the model, enhances the ability of the model to obtain feature information, and improves the accuracy of image classification. The lightweight model can solve the problem that the training time is too long in the process of image classification and improve the speed of classification.

The main contributions of this paper are as follows:

(1) Aiming at the problem of feature extraction in existing methods, a classification network architecture combining Transformer and CNNs is proposed to make it more suitable for global and local feature extraction.

(2) Aiming at the problem that the image training time complexity, a lightweight model is proposed to solve the problem that the training time is too long.

(3) The framework proposed in this paper has been verified on the TB data set of the First Hospital of Lanzhou University. It is verified that the architecture effectively improves the effect of classification.

II. RELATED WORK

With the rapid development of computer technology and computer-aided medical diagnosis system, medical image classification has become a research hot spot in the image field. The technology related to medical image classification is mainly represented in the emerging machine learning field. CNNs, one of the current mainstream medical image classification methods, has surpassed the traditional neural network classification methods by virtue of its fast speed, high performance, and weight sharing-mechanism. In 1982, in order to simulate human visual cognitive function, Fukushima et al. [10] proposed the concept of neurocognitive machine, which is considered as the starting point of CNNs. In 1989, LeCun and others constructed the original LeNet model [11], which includes convolution layer and full connection layer. After improvement, LeCun et al. [12] proposed the classic LeNet-5 model in 1998, which has included all the basic modules of modern CNN networks. CNN started with the M-P neuron model and developed with LeNet and LeNet-5 models until Alex Krizhevsky put forward the AlexNet [13] network in 2012, using ReLU activation function and GPU grouping convolution for parallel training. The CNN model continues to develop and improve in many directions, such as architecture design, activation function, and optimization strategy improvement.

In 2017, the Google team proposed an NLP classic model – Transformer, which demonstrated strong global context modeling capabilities and overcame the limitations of CNN's local information processing. Vaswanid et al. first proposed Transformer. Because of its unique design, Transformer can handle variable length inputs and capture long-distance dependencies and sequence-to-sequence (seq2seq) characteristics. Transformer networks represented by Swin Transformer [14] explicitly interact locally by limiting attention to local windows. CeiT [15] introduced local feature learning in FFN module to model local relations. CeiT [15] and ViTc [16] added the convolution module to the Transformer to extract the underlying local information. MobileViT [17] regards Transformer as a convolutional layer and embeds it in the CNN, realizing the interaction of local information and global information. MobileFormer [18] and ConFormer [19] adopt parallel CNN and Transformer branches and realize feature fusion with shortcut.

DeiT [20] introduced the characteristics of CNN into the learning process of Transformer by introducing the distillation token with the idea of knowledge distillation.

In recent years, many lightweight network models have been proposed. Compared with other network models, lightweight models have the advantages of light structure, simple calculation, and strong portability. In terms of manually designed lightweight networks, they include SqueezeNet [21], MobileNet [22], MobileNetV2 [23], ShuffleNet [24], ShuffleNetV2 [25], improved baseline network based on Octave convolution [26], GhostNet [27], and other lightweight networks. In the automatic lightweight network design of neural network structure search, there are NasNet [28], MnasNet [29], and other lightweight networks.

Deep learning is a subfield of machine learning. Due to the growth of computing power and data, deep learning has been greatly developed and has become one of the powerful tools in the medical field. The technology of segmentation, feature extraction, and classification of medical images using the constantly deepening and mature deep learning model is increasingly mature. Current popular medical image classification and segmentation depth learning models include CNN, FCN, SegNet, U-Net, PSPNet, and MaskR CNN. Zhang et al. [30] established the "early computer diagnostic system for lung cancer" to detect pathological sections of lung cancer, which can detect several main types of lung cancer. In massive medical image data processing and classification, the lightweight network model enables typical network models, such as CNNs. It can achieve lightweight in medical image classification tasks through model compression algorithms such as quantization weight, structure thinning, and model clipping. Li et al. [31] proposed a lightweight multi-classification network MELCNet for fundus diseases. This classification network overcomes the problems of simplification of existing classification methods for eye diseases and the large number of parameters and complexity of calculation in the network model. Taking PPLCNet as the backbone network, it can focus on the key information of patients with different diseases at different scales.

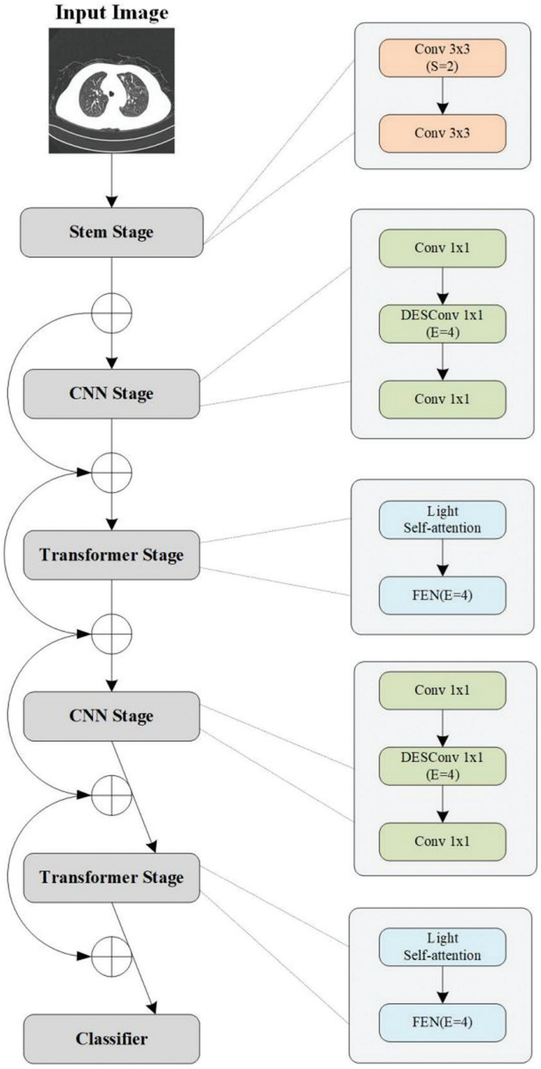

Fig. 1. Overall structure.

III. Methods

A. THE OVERALL STRUCTURE

The overall classification framework proposed in this paper for pulmonary CT-related pulmonary TB cases is shown in Figure 1.

In this work, we propose a medical image classification method based on lightweight network. CNNs pay too much attention to local information but ignore the global information of the image. Since Transformer encodes whole images, it has state-of-the-art global modeling ability. This makes Transformer an effective in extracting global context information. In order to pay attention to the acquisition of local and global features at the same time, this method mainly uses a mixture of CNN and Transformer as a classification network.

B. HYBRID NETWORKS (CNN TRANSFORMER HYBRID NETWORK)

The hybrid feature extraction network proposed in this paper aims to integrate the advantages of CNN and Transformer to complete more accurate classification tasks. CNN lead various tasks of computer vision, such as classification, detection, and segmentation. CNN extracts features through shared convolution kernel to decrease the number of network parameters and improve model efficiency. On the other hand, CNN has translation invariance, that is, no matter where features are moved to in the image, the network can detect these features. Although CNN has many advantages, its receptive field is usually small, which is not conducive to capture global features. Visual Transformer outperforms many CNN structures for many visual tasks due to its ability to capture global information about a single picture. The limited receptive field of CNN makes it difficult to capture global information, while Transformer can capture long-distance dependencies. Many works try to combine CNN and Transformer so that the network structure can inherit the advantages of CNN and Transformer and retain global and local features to the greatest extent.

The proposed framework is demonstrated in Figure 1. The framework can be divided into five phases. The first phase is Stem Stage, and the rest are four phases of alternating CNN and Transformer. At the head of stages, downsampling is performed to reduce the feature map size and increase the number of channels. Meanwhile, the network references ResNet residual connections, which are shortcuted at each stage.

The first of these, Stem, consists of two components of 3 × 3 convolution layers. The second stage is the CNNs stage. As the feature map is too large at this time, it is not suitable to use Transformer to extract global features. The CNNs phase uses Depthwise-Separable Convolution (DSConv) block to reduce the number and size of model parameters. The third stage is the Transformer stage, which follows the CNNs to extract global features. This paper adopts a lightweight Multi-head Self-Attention.

In the original self-attention module, input X is linearized into query Q, key K, and value V, where n represents the number of patches. d, dk, and dv represent dimensions of input, key, and value, respectively. Its self-attention output can be obtained by the following formula 1:

$$\mathrm{Atten}(Q,K,V) = \mathrm{Soft\ max}\left(\frac{QT^{T}}{\sqrt{d_{\mathrm{k}}}}\right)V \tag{1}$$

In order to reduce the overhead, this paper uses a k × k depth convolution with step size k to reduce the dimension of K, V, namely and. The CNNs and Transformer operations of stages 2 and 3 are repeated in subsequent stages 4 and 5. At the same time, the phase is repeated L times in each phase. Stages 1 through 5 are repeated 2, 2, 4, 2, and 8 times, respectively.

These lightweight CNNs are versatile and easy for training. These networks can easily replace heavyweight backbone networks. For example, ResNet reduces network size and improves latency in an existing task-specific model (for example, DeepLabv3). Despite these benefits, a major drawback of these approaches is that they are local in space. This work treats transformers as convolution and allows you to take advantage of convolution (for example, generic and simple training) and converters (for example, global processing) to build lightweight network structures.

IV. EXPERIMENT

A. DATA SET

In this experiment, the training and test data sets are all from CT images provided by the First Hospital of Lanzhou University. The image size of the source data set is 512 * 512, and there are 100 examples in total. Figure 2 shows some cases of the data set. The image data in each example are annotated with relevant diagnosis. In order to effectively carry out the classification task of relevant TB cases, this data set provides images of lung window and mediastinal window. The lung window image mainly includes lung texture and bilateral thorax, and mediastinal window image mainly includes trachea and esophagus in mediastinum.

The luminance value of the CT image is Hu (Hounsfield unit). The Hu value of the CT image reflects the X-ray absorption value (attenuation coefficient) of the tissue. With reference to the set value of the lung window, the data normalization between the data is performed in the range of 0 to 255. Through the feature extraction of CT images of different clinical cases, the image characteristics of pulmonary TB cases can be compared and analyzed. The training data set is used to train the network model for classification, and the test data set is used to verify the model

B. EXPERIMENTAL DETAILS

In the preprocessing phase, the source data set is in DICOM format. In this paper, all data sets are sliced, and the size of each slice is the original 512 * 512. In this paper, the DICOM file is converted to lung window png format. The original data is standardized and normalized to solve the problem of inconsistent contrast of different images. All the programs in this article are implemented under the PyTorch framework. The training process uses Nvidia GeForce 3090. The optimizer in the experiment is Adam, the momentum of the optimizer is 0.9, and other parameters are default parameters. The initial learning rate, weight attenuation, and batch size are 1e-3, 1e-5, and 16, respectively.

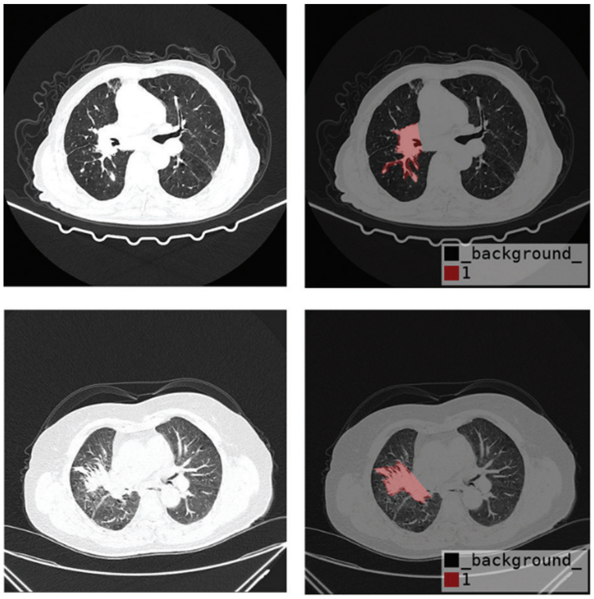

Fig. 2. Processed cases of data set.

C. TEST RESULTS

In order to verify the performance of the proposed lightweight Transformer and CNNs classification network in the task of classifying pulmonary TB cases. The experimental index results are shown in Table I. The article adopts the most common evaluation indicators in medical image classification: Precision and Recall. According to the experimental results, the method proposed in this paper has a certain accuracy and high recall.

It can be seen from Table I that the method proposed in this paper has a higher accuracy and recall rate than VGGNet and Vision Transformer. At the same time, the method proposed in this paper has much less computational complexity than the former two and has greater advantages for medical image classification Figure 3.

During the training process, the training loss (loss) and the test set loss (val_loss) of the training set and the test set are obtained, as shown in Figures 4 and 5. For the previous training, the loss drops very quickly, and for the future training, the loss decline gradually slows down and tends to be stable, and the training effect is ideal.

In addition, through the DICOM data set, the lightweight classification network proposed in this paper is tested to verify the

Table I. Comparison of different methods

	FLOPS	Accuracy rate (%)	Recall rate (%)
VGGNet	17G	93.82	94.67
Vision Transformer	49G	95.34	96.53
Ours	16M	97.23	98.41

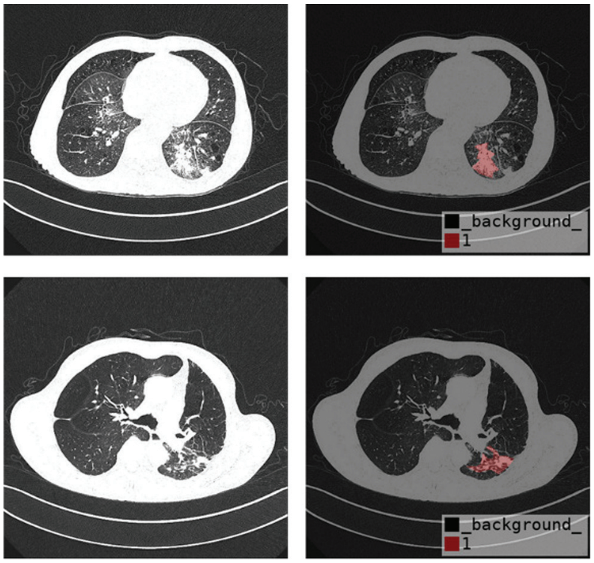

Fig. 3. Processed cases of data set.

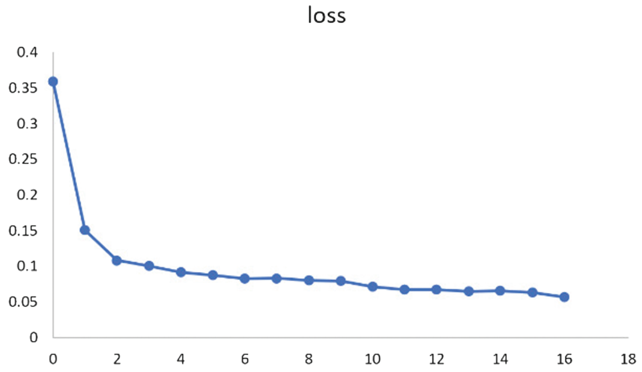

Fig. 4. Training set loss.

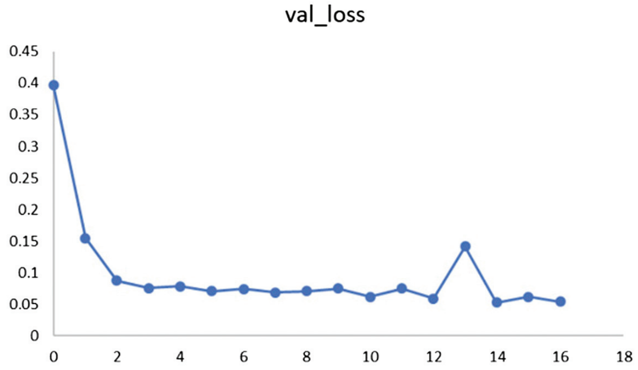

Fig. 5. Test set loss.

accuracy of the model. Compared with VGGNet and Vision Transformer, it is easy to find that the methods proposed in this paper are optimal, with an accuracy rate of 98.41%. As shown in Figures 6–8, they are the confusion matrices of VGGNet, Vision Transformer, and the method proposed in this paper. It can be found that the proposed method is very effective.

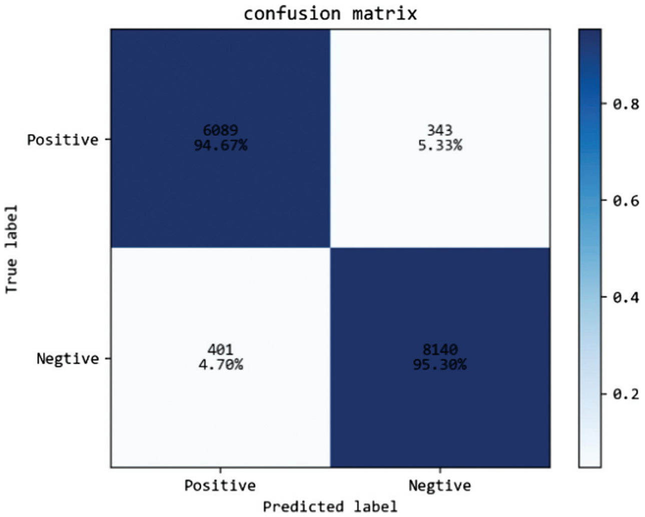

Fig. 6. VGGNet confusion matrix.

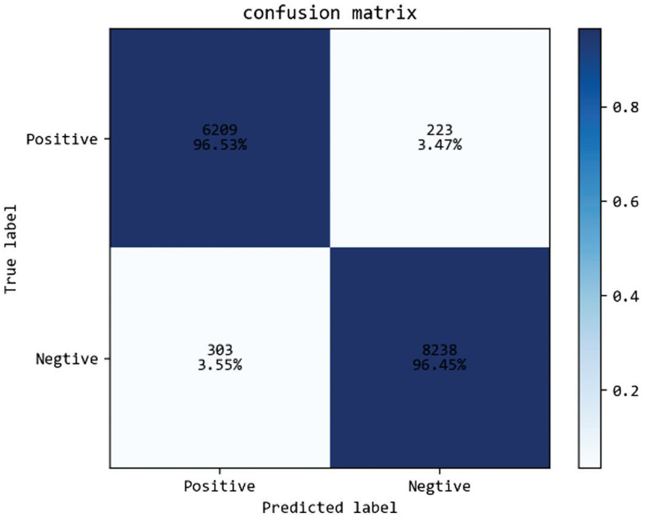

Fig. 7. Vision Transformer confusion matrix.

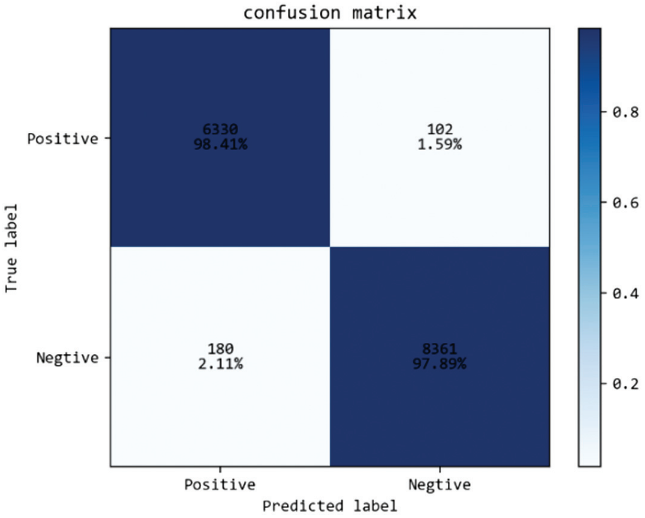

Fig. 8. Confusion matrix of the method proposed in this paper.

V. CONCLUSION

This paper proposes a lightweight medical image classification network based on the combination of Transformer and CNN, which consists of a fusion of Stem module, CNN module, and Transformer module. The Stem module is used to complete the convolutional downsampling operation and reduce the number of parameters of the network; the CNN–Transformer alternate module is used for. The alternating CNN–Transformer module is used for feature extraction and processing and finally achieves the purpose of classification. The alternating CNN–Transformer framework is shown to perform better in the classification of TB cases. The shortcut introduced in this paper can make full use of the image feature information and achieve better feature fusion than other methods. Validation against other state-of-the-art methods shows that this paper's method consistently and significantly outperforms other methods in terms of accuracy, has a certain accuracy and high sensitivity, and scores better in several classification regions. The proposed method demonstrates the superiority of lightweight networks for medical image classification tasks.

References

[1] X. Gao, S. Lin, and T. Y. Wong, "Automatic feature learning to grade nuclear cataracts based on deep learning[C]," in Springer Int. Publ., Springer International Publishing, 2014.

[2] A. Payan and G. Montana, "Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks[J]," Comput. Sci., 2015.

[3] M. Zhou et al., "Multi-scale convolutional neural networks for lung nodule classification," 2015.

[4] J. Chen et al., "TransUNet: transformers make strong encoders for medical image segmentation[J]," 2021.

[5] Y. L. Zeng et al., "Design and analysis of toilet auxiliary device based on TRIZ theory[J]," J. Nanchang Inst. Technol., vol. 39, no. 04, pp. 6l–65, 2020.

[6] R. De Ridder and C. De Blaiser, "Activity trackers are not valid for step count registration when walking with crutches[J]," Gait Posture., vol. 70, pp. 30–32, 2019.

[7] Y. Xu et al., "A medical image segmentation method based on multidimensional statistical features," Front. Neurosci., vol. 6, pp. 1–9, 2022.

[8] Z. Zhu, X. He, G. Qi, Y. Li, B. Cong, and Y. Liu, "Brain tumor segmentation based on the fusion of deep semantics and edge information in multimodal MRI," Info. Fusion, vol. 91, 376–387, 2023.

[9] Z. Liu et al., "Swin transformer: hierarchical vision transformer using shifted windows," Proc. IEEE/CVF Int. Conf. Comput. Vis., 2021.

[10] K. Fukushima and S. Miyake, "Neocognitron: a self-organizing neural network model for a mechanism of visual pattern recognition[J]," IEEE Trans. Syst. Man Cybern., vol. 13, no. 5, pp. 267–285, 1983.

[11] Y. Lecun et al., "Backpropagation applied to handwritten zip code recognition[J]," Neural Comput., vol. 1, no. 4, pp. 541–551, 1989.

[12] Y. Lecun et al., "Gradient-based learning applied to document recognition[J]," Proc. IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "lmageNet classification with deep convolutional neural networks," in Int. Conf. Neural Inf. Process. Syst., Curran Associates, Lake Tahoe, USA, 2012, pp. 1097–1105.

[14] Z. Liu et al., "Swin transformer: hierarchical vision transformer using shifted windows," in Proc. 2021 IEEE/CVF Int. Conf. Comput. Vis. (ICCV), IEEE, Montreal, Canada, 2021, pp. 9992−10002.

[15] K. Yuan, S. P. Guo, Z. W. Liu, A. J. Zhou, F. W. Yu, and W. Wu, "Incorporating convolution designs into visual transformers," in Proc. 2021 IEEE/CVF Int. Conf. Comput. Vis. (ICCV), IEEE, Montreal, Canada, 2021, pp. 559−568.

[16] T. Xiao, M. Singh, E. Mintun, T. Darrell, P. Dollár, and R. Girshick, "Early convolutions help transformers see better," in Proc. 35th Conf. Neural Inf. Process. Syst. (NeurIPS 2021), 2021.

[17] S. Mehta and M. Rastegari, "MobileViT: Light-weight, general-purpose, and mobile-friendly vision transformer," arXiv preprint arXiv: 2110.02178, 2021.

[18] Y. P. Chen et al., "Mobile-former: bridging MobileNet and transformer," *arX-iv preprint arXiv: 2108.05895*, 2021.

[19] Z. L. Peng et al., "Conformer: local features coupling global representations for visual recognition," in Proc. 2021IEEE/CVF Int. Conf. Comput. Vis. (ICCV), IEEE, Montreal, Canada, 2021, pp. 357−366.

[20] H. Touvron, M. Cord, M. Douze, M. Francisco, A. Sablayrolles, and H. Jégou, "Training data-efficient image transformers & distillation through attention," in Proc. 38th Int. Conf. Mach. Learn. PMLR, 2021, pp. 10347−10357.

[21] F. N. Iandola et al., "SqueezeNet: AlexNet-level accuracy with 50x fewer parameter sand < 0.5MB model size [EB/OL]," 2021. https://arxiv.org/abs/1602.07360.

[22] A. G. Howard et al., MobileNet: efficient convolutional neural networks for mobile vision applications [EB/OL]," 2021. https://arxiv.org/abs/1704.04861.

[23] M. Sandler et al., MobileNet V2: inverted residuals and linear bottlenecks[C]," in Proc. IEEE Conf. Comput. Vis. Pattern Recogn., Washington, DC, USA, IEEE Press, 2018, pp. 4510–4520.

[24] X. Zhang et al., "ShuffleNet: an extremely efficient convolutional neural network for mobile devices[C]," in Proc. IEEE Conf. Comput. Vis. Pattern Recogn., Washington, DC, USA, IEEE Press, 2018, pp. 6848–6856.

[25] N. Ma et al., "ShuffleNet V2: practical guidelines for efficient CNN architecture design[C]," in Proc. Eur. Conf. Comput. Vis., Berlin, Germany, Springer, 2018, pp. 116–131.

[26] Y. Chen et al., "Drop an octave: reducing spatial redundancy in convolutional neural networks with octave convolution[C]," in Proc. IEEE/CVF Int. Conf. Comput. Vis., Washington, DC, USA, IEEE Press, 2019, pp. 3435–3444.

[27] K. Han et al., "GhostNet: more features from cheap operations[C]," Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recogn., Washington, DC, USA, IEEE Press, 2020, pp. 1580–1589.

[28] B. Zoph et al., "Learning transferable architectures for scalable image recognition[C]," Proc. IEEE Conf. Comput. Vis. Pattern Recogn., Washington, DC, USA, IEEE Press, 2018, pp. 8697–8710.

[29] M. Tan et al., "MnasNet: platform-aware neural architecture search for mobile[C]," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recogn., Washington, DC, USA, IEEE Press, 2019, pp. 2820–2828.

[30] J. Chen et al., "Multiview two-task recursive attention model for left atrium and atrial scars segmentation," in 21st International Conference on Medical Image Computing and Computer Assisted Intervention—MICCAI 2018, A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-Lopez, and G. Fichtinger, Eds. vol. 11071. Granada: Springer, 2018, pp. 455–463.

[31] C. L. Li, R. F. Zhang, Y. H. Liu, "Lightweight fundus disease multi-classification network with multi-scale bilinearattention[J]," Appl. Res. Comput., vol. 39, no. 7, pp. 2183–2189+ 2195, 2022.

[32] Y. Xie et al., "Knowledge-based collaborative deep learning for benign-malignant pulmonary nodule classification on chest CT[J]," IEEE Trans. Med. Imaging, vol. 38, no. 4, pp. 991–1004, 2019.

[33] J. Q. Xu et al., “Generative adversarial networks for the classification of lung nodules malignant[J],” J. Northeastern Univ. (Nat. Sci.), vol., 39, no. 11, pp. 39–44, 2018.

[34] W. Shen et al., Multi-Scale Convolutional Neural Networks for Lung Nodule Classification: Springer International Publishing, 2015, pp. 588–599.

[35] Y. Lei et al., “Shape and margin-aware lung nodule classification in low-dose CT images via soft activation mapping[J],” Med. Image Anal., vol. 60, p. 101628, 2019.

[36] Y. Chen, J. Feng, J. Liu, B. Pang, D. Cao, and C. Li, “Detection and classification of lung cancer cells using Swin transformer[J],” J. Cancer Ther., vol. 13, no. 7, pp. 464–475, 2022.