

Medical image recognition based on VGGNet19

Yanke Liu^{1,†} Honglin Wang^{2,4,†}, Yizhuo Zhao^{3,†}

¹Computer Engineering and Science College, Shanghai University, Shanghai, 200000, China

²International College, Zhengzhou University, Zhengzhou, 450000, China

³Beijing-Dublin International College, Beijing University of Technology, Beijing, 100000, China

⁴leowp@stu.zzu.edu.cn

[†]All authors contributed equally.

Abstract. Gastrointestinal diseases are one of the common clinical diseases, which often require medical imaging for diagnosis and treatment. Recently, the development of deep learning technology has promoted the development of medical image recognition, which provides new ideas and methods for the automatic recognition and analysis of medical images. VGGNet19 is a convolutional neural network model that has attracted much attention because of its simple structure, easy training and better recognition effect. For this reason, this paper proposes an improved VGGNet19 model for medical image recognition of gastrointestinal diseases. Specifically, the project adds an additional fully connected layer and Dropout layer on top of the built VGGNet19 to achieve the recognition of medical images of stomach diseases. Extensive experiments on standard medical stomach images show that the proposed method improves the recognition performance to a certain extent.

Keywords: medical image, VGGNet19, gastrointestinal tract recognition.

1. Introduction

Gastrointestinal diseases are one of the common clinical diseases, which often require medical imaging for diagnosis and treatment. Recently, the development of deep learning technology has promoted the development of medical image recognition, which provides new ideas and methods for the automatic recognition and analysis of medical images. There are also papers in the prestigious journal Lancet that use magnetic resonance imaging diagnosis to study gastroenterology and liver diseases [1]. Among the many papers that have been read, there are more and more automatic recognition of medical images using convolutional neural networks, which has become a hot research direction. Among them are deep neural network-based dermatologist-level skin cancer classification and discussions on whether the neural network should be fully trained or fine-tuned [3]. Therefore, this paper will use the improved VGGNet19 model for medical image recognition of gastrointestinal diseases, aiming to improve the automatic recognition and analysis of medical images and provide more accurate and reliable technical support for medical diagnosis and treatment.

2. Related work

Medical image recognition is a method that uses computer vision technology to identify abnormalities or diseases in medical images. The application of medical image recognition technology involves many fields such as oncology, cardiovascular diseases, and neuroscience, and its importance and application prospects are receiving increasing attention from scholars. In the field of medical image recognition, many researchers have explored a variety of algorithms and methods. Traditional machine learning methods include Support Vector Machines (SVM), Random Forest, Decision Tree, etc. These methods have achieved good results in some medical image recognition problems, but they require manual feature selection and design, and their generalization ability and robustness are limited. With the development of deep learning, more and more researchers are trying to use convolutional neural networks (CNNs) to solve medical image recognition problems.

Among the work related to the field of medical image recognition is that of lung cancer screening using deep learning for chest tomography, and the team of this paper describes a method for lung cancer screening using a 3D deep learning approach for low-dose chest computed tomography scans [4]. Also closer to this paper is the automatic detection of caries in panoramic dental radiographs based on deep learning. The researchers used a convolutional neural network (CNN) to design a model that learns to extract features from dental radiographs and classify these features to determine the presence of caries [5]. The researchers used some classic CNN architectures, such as ResNet and DenseNet, with appropriate modifications to accommodate the characteristics of dental x-ray films.

Convolutional neural network is a deep learning model originally applied in the field of image recognition, which mimics the visual system of the human brain and is able to automatically learn features from large amounts of data, allowing it to classify and recognize images efficiently and accurately. In the field of medical image recognition, researchers have tried several different CNN structures, such as LeNet, AlexNet, VGG, ResNet, etc. These models have achieved impressive results in medical image recognition problems, such as recognition of lung cancer, breast cancer, skin cancer, and other aspects.

3. Research methodology

3.1. Selection and pre-processing of dataset

The Kvasir dataset, a dataset containing images within the gastrointestinal (GI) tract, was used in the study (Table 1). Anatomical landmarks include z-line, pylorus, and cecum, and pathological manifestations include esophagitis, polyps, and ulcerative colitis. Sorting and annotation of the dataset is performed by the physician. In this regard, Kvasir is important for both single and multi-disease computer-aided detection studies. The data were collected using endoscopic equipment from the Vestre Viken Health Trust (VV) in Norway, which consists of four hospitals providing health care services to 470,000 people. One of the hospitals (Bærum Hospital) has a large gastroenterology department from which training data have been collected and will be made available (Figure 1).

Table 1. Categories of dataset.

Class Index	Label	Data Size
0	dyed-lifted-polyps	500
1	dyed-resection-margins	500
2	esophagitis	500
3	normal-cecum	500
4	normal-pylorus	500
5	normal-z-line	500

Table 1. (continued).

6	polyps	500
7	ulcerative-colitis	500

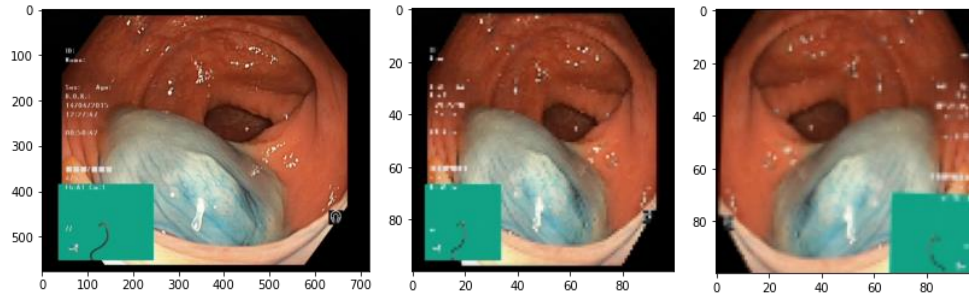


Figure 1. Pre-processing of dataset.

3.2. Principle analysis of VGGNet19 neural network

VGGNet19 is a classic model in convolutional neural network which performs well on ImageNet datasets with high accuracy and stability. Hence by using VGGNet19 model may improve the accuracy of modeling in medical image identification and classification, reduce rates of misdiagnosis and promote diagnostic efficacy. In addition, medical imaging has grand complexity and diversity, and is influenced by multiple factors. VGGNet19 has good robustness and can effectively perform recognition and classification on various types of medical images, with high generality. More specifically, utilizing VGGNet19 model in the diagnosis of medical images helps achieve automated medical work, provide with higher efficiency and diagnostic quality, reduce the workload of doctors, and analyze image data more comprehensively and objectively [7].

In order to optimize feature local information extraction and reduce the risk of overfitting, we added full connection layer and Dropout layer on the basis of VGGNet19's convolutional network, regularized the 512 features in the final layer, hence our model may filter invalid features and available information more efficiently, and improve the generalizing ability and robustness.

The structure of our model is shown in the figure 2.

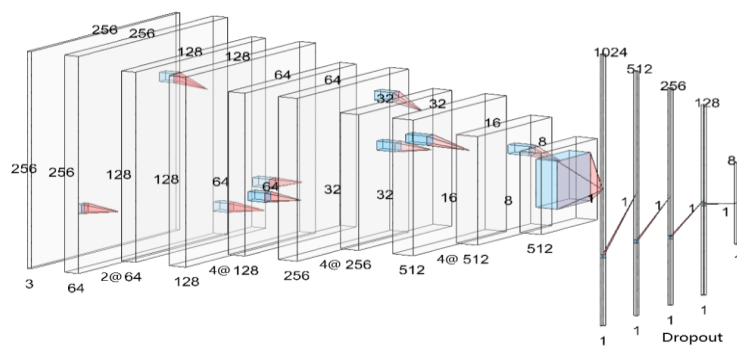


Figure 2. Structure of modified VGGNet19.

3.3. Algorithm

3.3.1. Forward propagation. Forward Propagation is a process in which the neural network integrates input data and passes the data layer by layer through connections between neurons of generally different layers, and output specific feature of the input [8]. Different forms of layers, such as convolutional layers, pooling and dropout layers, fully connected layers, etc., have distinct connection with the previous layer

and characteristic data processing patterns. Specific description and algorithms in different kinds of layers are given in figure 3.

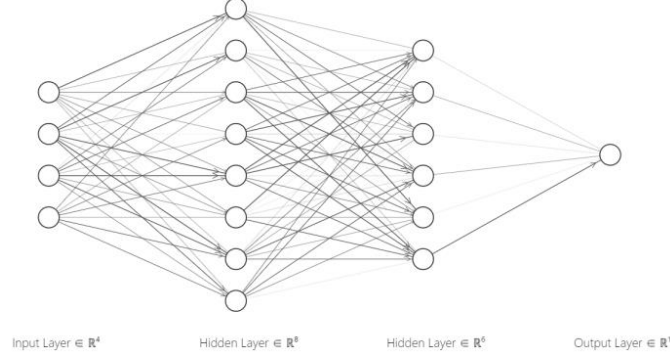


Figure 3. Forward propagation.

Fully Connected Layer: A layer formed by the all-possible combinations between any posterior layer neuron and any anterior layer neuron, its output algorithm can be expressed as the following equation.

$$y_h^{l+1} = f \left(\sum_{j=1}^k w_{h,j} y_j^l + b_h \right) \quad (1)$$

Where y_h^l represents the output value of the h-th neuron in the l-th layer, $w_{h,j}$ is the multiplicative weight with respect to the output of h-th neuron of the subsequent layer to the previous layer, b_h refers to the output threshold of the h-th neuron in the following layer, and f is the activation function.

Dropout Layer: Differed from fully connected layers, neurons in the previous layer will be randomly ignored by each input, while neurons that are not ignored will still be connected to the neurons in the subsequent layer in a fully connected pattern.

The algorithm can be expressed as follows.

$$y_h^{l+1} = f \left(\sum_{j=1}^k s_j w_{h,j} y_j^l + b_h \right) \quad (2)$$

Where s_j is a random variable following a binomial distribution: $P(s_j = 0) = p, P(s_j = 1) = 1 - p$, p is a constant set manually which represents the probability of the front layer neurons being ignored.

Activation Functions: In convolutional layers and fully connected non-output layers, this work chose ReLu function as the activation function. Its expression is shown below.

$$\text{ReLu}(x) = \max(0, x) \quad (3)$$

Making use of ReLu as the activation function in computation can improve the overall nonlinearity of the model while simplifying the propagative calculation and improving the recognition efficiency. In the output layer, we use the Softmax function as the activation function. The algorithm is shown in the following formula.

$$\text{Softmax}(y_i) = \frac{e^{y_i}}{\sum_{j=1}^k e^{y_j}} \quad (4)$$

After regularizing the outputs by Softmax, our model finally obtains the classification results.

3.3.2. Optimization. Back propagation is an algorithm to calculate and adjust the weights of neuron connections and other parameters according to the error between the current output and the real value.

The implementation of the algorithm mainly relies on the selection of optimizer and loss function. We chose Softmax as the activation function of the output layer, thus we finally selected cross entropy as the loss function to estimate the update step size of each parameter [9,10].

Specifically speaking, we adopted the Adam optimizer, which can iteratively update parameters in neural network by utilizing the method of estimating the expectation and the noncentral variance of the gradients of the input, and then calculate the update step for all parameters. In this article, we used the cross entropy as the loss function.

4. Experiment

4.1. Frameworks and parameters

The experiment was based on the configuration listed in Table 2.

Table 2. Software configuration.

Platform	Name of Extension	Version
python 3.9.12 64-bit-jupyter 1.0.0	numpy	1.23.3
	sklearn(scikit-learn)	1.1.3
	keras	2.10.0
	torchvision	1.13.0+cu117
	PIL(Pillow)	9.2.0
	tensorflow-gpu	2.10.0
	matplotlib	3.6.1
	cv2(opencv-python)	4.7.0.72
	pandas	1.5.1

The hyperparameter used in the experiment were as follows: Learning rate $\alpha = 0.001$, the exponential decay rates applied were $\beta_1 = 0.9$, $\beta_2 = 0.999$. The model was trained with input batches, where batch_size = 20. The dropout rate was $p = 0.3$.

4.2. Experimental process

This work uses OpenCV to read and resize the images, store the image data and labels in arrays, and then split the dataset into training, validation, and testing sets. Additionally, ImageDataGenerator is used for image data augmentation, including rotation, horizontal flip, and scaling [7].

Specifically, the VGGNet19 network is employed for image classification, which includes pre-trained weight parameters. After building the complete VGGNet19 network, a fully connected layer is added and the Softmax activation function is used for classification. Then we set hyperparameters such as batch size, number of training epochs, and learning rate, and use the Adam optimizer to compile the model. We then train the model using GPU acceleration and save the model's training history and weight parameters.

The last step is the visualization of the experimental results. We use matplotlib to plot the loss and accuracy curves of the model. Finally, CPU is used to evaluate the assessment of the model's performance, and visualized its performance measures including accuracy curve, confusion matrix, etc.

4.3. Experimental results and ablation experiment verification

4.3.1. Loss and accuracy function results. The loss and accuracy curves with epoch changes are plotted as shown in the figure 4. After 100 epochs, the model's training accuracy reaches about 97%.

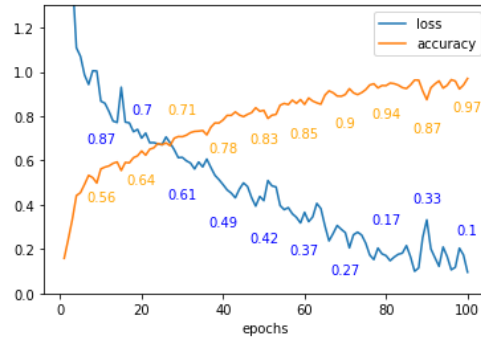


Figure 4. Loss and accuracy curve.

After training, any image can be placed in the model, which can recognize the image and correspond it to the eight class folders in the trained dataset (Figure 5).

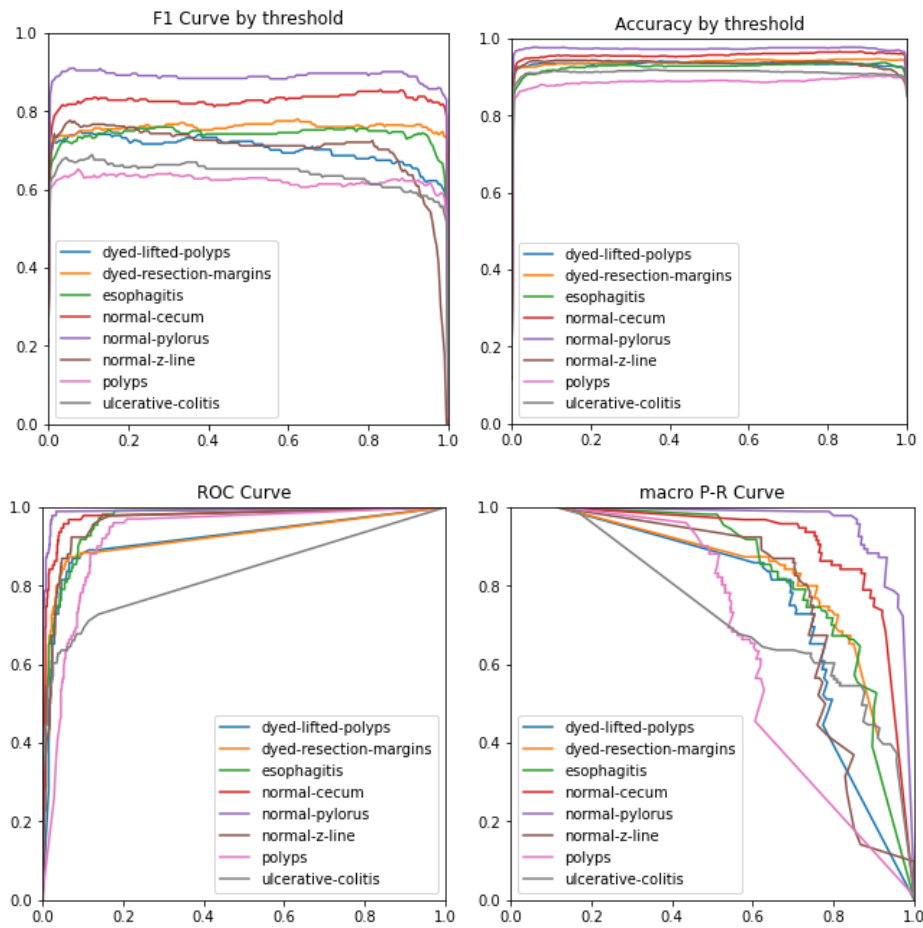


Figure 5. Performance measures.

4.3.2. Confusion matrix results. After training, any image can be placed in the model, which can recognize the image and correspond it to the eight class folders in the trained dataset. The model's accuracy on the validation set is about 74%, and the confusion matrix of the entire dataset is shown in figure 6.

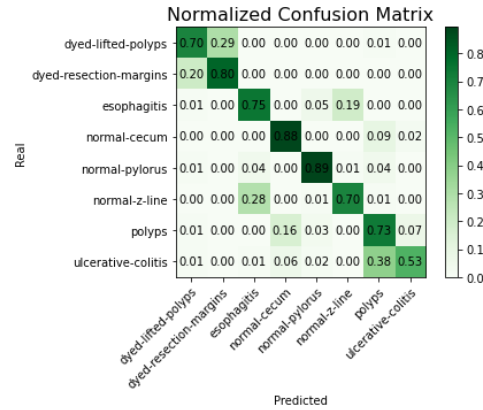
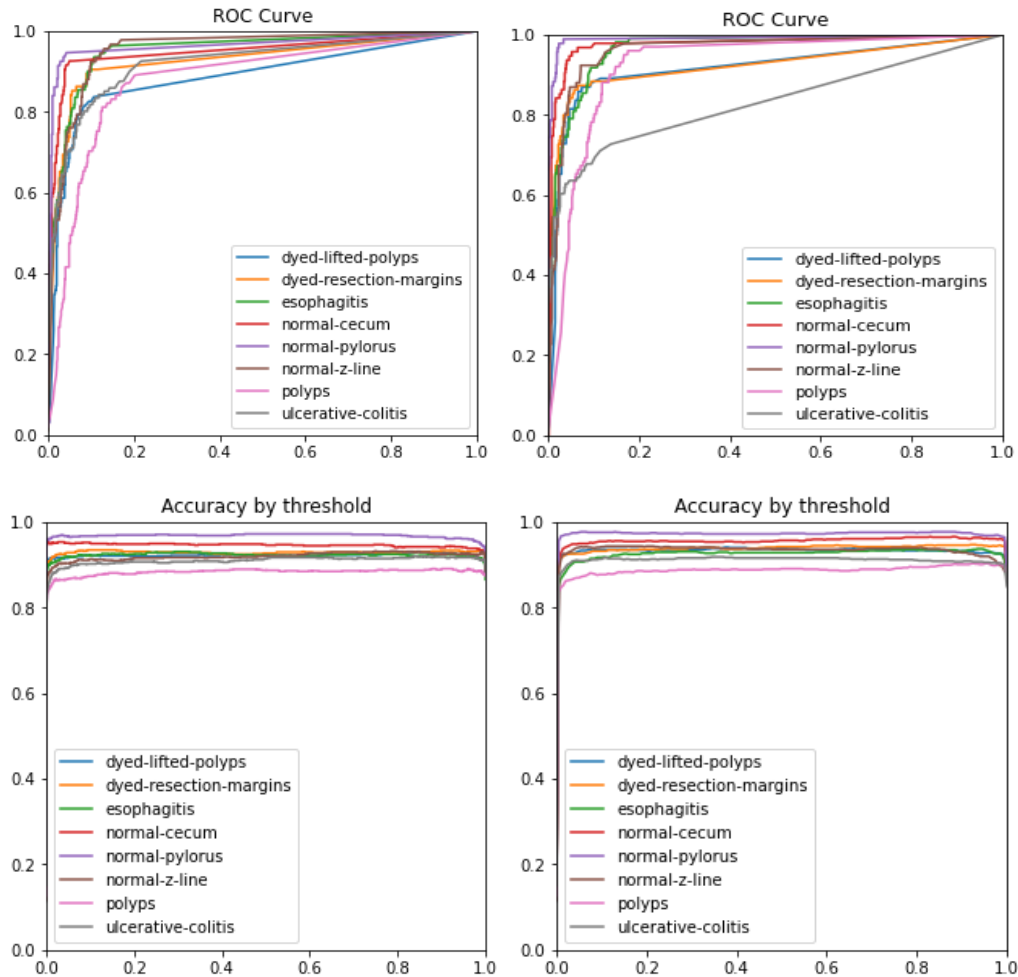


Figure 6. Normalized confusion matrix.

We expect that by adjusting the exponential decay rate, calling methods such as Reduce on Plateau to dynamically adjust the learning rate, and increasing the number of training epochs, the model's accuracy will be further improved.

4.3.3. Ablation experiment. The ROC, ACC, macro-PR, and F1 graphs for the model without the Dense and Dropout layers (left) and the model with the Dense and Dropout layers (right) on the test set with 2000 sample points are shown in figure 7.



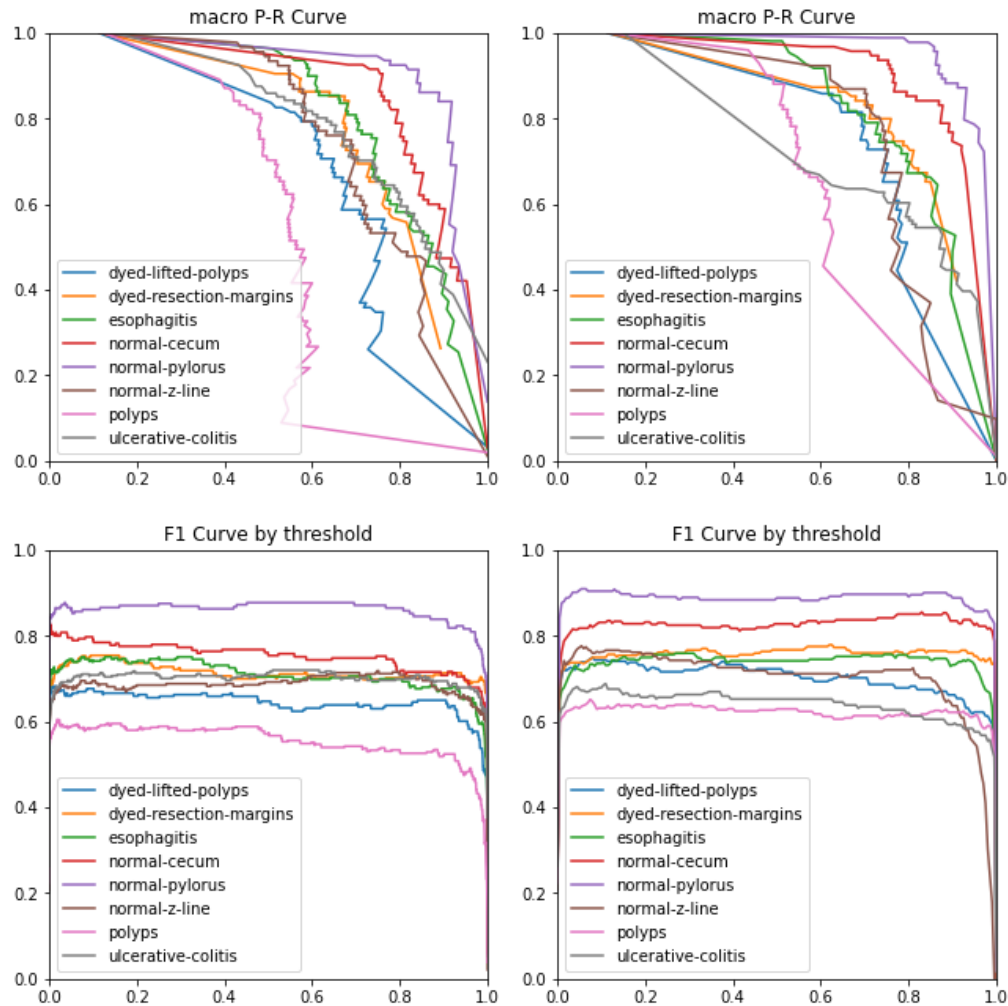


Figure 7. Performance measures in comparison.

The modified model has better F1 and macro-PR indicators and weakens the overfitting phenomenon.

5. Conclusion

This work is focused on medical image recognition of gastrointestinal diseases. Therefore, an improved VGGNet19 model is proposed for medical image recognition of gastrointestinal diseases. Specifically, the project adds an additional fully connected layer and Dropout layer on top of the built VGGNet19 to achieve the recognition of medical images of gastric diseases. Finally, the experimental results and visualize the loss function, accuracy function and confusion matrix during the training process are obtained. In order to further improve the experimental results, we conducted ablation experiments and referred to image recognition papers using ablation experiments, and a large number of experiments show that the method in this paper gets some improvement in performance.

Our preliminary results of this experiment are good, but there are still areas that can be improved. We can learn other network models of CNNs, such as the ResNet50 model, and perform classification operations on this dataset to compare which model is more accurate. In addition, this work still needs to learn in this area of data augmentation, because for medical image recognition, the amount of data is often limited and good data augmentation methods are extremely useful. For the original dataset, it would be nice to find more resources for gastrointestinal data, but specialized medical images are

extremely scarce and difficult to obtain, so this improvement can only be achieved by larger medical teams.

With the continuous accumulation of medical image data and the continuous development of medical artificial intelligence technology, medical image recognition models based on convolutional neural networks have a bright future and can play a crucial part in medical image analysis, pathological diagnosis, and treatment plan development. We expect a better combination of computer and medical fields in the future, and better results in the field of medical image recognition.

Reference

- [1] Pu, L., Liu, X., Zhang, Y., Li, J., & Li, J. Diagnostic performance of magnetic resonance imaging for detecting perianal fistula and abscess: a systematic review and meta-analysis. 2021 *PloS one*, **16(5)**, e0251838.
- [2] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J. et al. Dermatologist-level classification of skin cancer with deep neural networks. 2017, *Nature*, **542(7639)**, 115–118.
- [3] Tajbakhsh, N., Shin, J. Y., Gurudu, S. R. et al. Convolutional neural networks for medical image analysis: Full training or fine tuning? 2021, *IEEE Trans. Medi. Imag.*, **35(5)**, 1299–1312.
- [4] Wang, X., Zhang, Y., Zhang, J., Sun, X., Chen, H., He, Y., & Hu, Z. Lung cancer screening with low-dose chest computed tomography using three-dimensional deep learning. 2021 *J. X-Ray Sci. Tech.*, **29(1)**, 1-12.
- [5] Liu, Y., Wang, Y., Zhang, X., Zhang, Q., & Chen, Z. Deep learning-based automatic detection of dental caries in panoramic radiographs. 2020, *J. Dentist.*, **99**, 103407.
- [6] Simonyan, K., & Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. 2015, *Inter. Conf. Learn. Represent.* 1-14.
- [7] Wong, A., et al. Data Augmentation for Deep Learning: A Review. 2018, *ACM Computing Surveys*, **51(4)**, 1-36.
- [8] He, K., et al. Deep Residual Learning for Image Recognition. 2016, *Conf. Comput. Vis. Patt. Rec.*, 770-778.
- [9] He, K., Zhang, X., Ren, S., & Sun, J. Deep Residual Learning for Image Recognition. 2016, *Conf. Comput. Vis. Patt. Rec.* 770-778.
- [10] Buchan, S., et al. Difficulty of acquiring clinical brain MRI data for three machine learning tasks. 2019, *Scientific Data*, **6(1)**, 1-7.