Pneumonia Image Recognition Based on Transfer Learning

Tao Zhong[1], HuiTing Wen[1], Zhonghua Cao[2], Xinhui Zou[1], Quanhua Tang[1], and Wenle Wang[1](✉)

[1] School of Software, Jiangxi Normal University, Nanchang 330027, Jiangxi, China
wenlewang@jxnu.edu.cn

[2] School of Software and Internet of Things Engineering, Jiangxi University of Finance and Economics, Nanchang 330013, Jiangxi, China

Abstract. With the rapid development of artificial intelligence (AI), the anomalies detection in biomedical has became important in patients' health monitoring. The pneumonia, including COVID-19, is a global threat. Detecting the infected patients in time is very critical to combating this epidemics. Thus, a rapid and accurate method for detecting pneumonia is urgently needed. In this paper, a deep-learning detection model, is designed to detect pneumonia efficient. Since training a neural network needs consuming a lot of time resources and computing resources, transfer learning is used for pre-training. At the same time, in order to improve the detection efficiency, we combine various deep learning models, and then perform prediction and classification. The simulation results show that comparing with the 91.5% accuracy of the traditional CNN model, the transfer learning model consisting of vgg16VGG16, vgg19VGG19, RresNnet50 and Xxecption reached 93.27%, 93.43%, 92.31% and 90.22% respectively. Most of the models are superior to the traditional models and have excellent stability with less time consuming.

Keywords: Transfer Learning · pneumonia detection

1 Introduction

As we all know, caused by an acute respiratory infection, the pneumonia has become one of the biggest threats to human society [1]. The key to combating this disease is identify the infected people in time. Because it can be caused by lots of factors, including viruses, bacteria, or fungus. Especially from the end of 2019, the COVID-19 has brought significant damage to the world. Many people have delayed treatment due to untimely diagnosis [2]. Because the pneumonia diagnosis is involves highly skilled analysis of a chest X-ray (CXR) using focused beam of radiation and professionally confirm the diagnosis with clinical history and laboratory tests, where the whole process is time-consuming [3].

The chest X-ray examining lungs, bones, and heart, are helpful to doctors to work out the placement and extent of the pneumonia [4]. The images brought by X-ray are

Y. Cao and X. Shao (Eds.): MONAMI 2022, LNICST 474, pp. 116–126, 2023.
https://doi.org/10.1007/978-3-031-32443-7_8

of the inside of body, where the thickness of body's tissues varies because the ratio of each part of body is different. The radiology technicians need to analyse a large number of CXRs every. And because different conditions may appear as opacity, it is difficult to identify the respiratory illness in CXR.

As artificial intelligence (AI) has recently become a topic of study in different applications, including healthcare, in which timely detection of anomalies can play a vital role in patient health monitoring [5]. With the potential of AI, it can minimize the repetitive task of clinicians, and automate the initial detection of potential respiratory illness to expedite the relevant review [6]. In the medical domain, the convolutional neural networks (CNN) methods are mostly used for classification. The [7] performed a study on the large pneumonia dataset containing the train, validation & test, which encountered that the smaller the image size, the better validation score. In [8], the DenseNet and MobileNetV2 CNN were used to train models on each dataset to classify, which achieved comparable performance to DenseNet, demonstrating the efficacy of CNNs for chest X-ray abnormality detection.

For these current deep learning models, it need to consume huge time resources and computing resources training a neural network. Therefore, reusing a pretrained deep learning model as a new model for another task, namely transfer learning, is a common approach in computer vision tasks [9]. Because most of the data and tasks are related, the parameters of the pre-training model can be transferred to the new model through transfer learning, thereby speeding up and optimize the learning efficiency of the model. Among them, the VGG CNN, including VGG16 and VGG19, are the most popular CNN model due to its simplicity and practicality. It shows good results in both image classification and object detection tasks [10].

In this paper, four models VGG16, VGG19, ResNet50 and Xeption are selected as the base learners of transfer learning, because these four models are classic convolutional neural networks, which are widely used, and have high recognition accuracy and generalization ability in the field of image recognition. Some outstanding achievements have also been made in transfer learning, which is an excellent base learner.

Therefore, a deep-learning pneumonia detection model, integrated with transfer learning model containing VGG16, VGG19, RresNet50 and Xecption, is proposed to detect pneumonia efficient. The contribution of this paper is as follows:

(1) A transfer learning model is constructed with VGG16, VGG19, ResNet50 and xecption, which obtains similar accuracy and less time consuming than single CNN method.
(2) Based on the designed transfer leaning model, we propose a ensemble learning model to retrive a better classfication accuracy to the traditional models.

2 Introduction to Data and Pre-processing

2.1 Dataset Description

The data set of this project comes from the public data of "chest X-ray images (pneumonia)" of kaggle. The data set contains more than 6000 pneumonia X-ray images, including two categories: pneumonia and normal. The images are divided into training set and test set of independent patients, marked as (disease) - (random patient ID) - (image number of the patient) and divided into four directories: CNV, DME, Drasen and normal. Table 1 shows the data composition of the training set, in which the ratio of positive and negative examples is 1:3.

Table 1. Data composition of training set.

Type	Count
Normal	About 1300
Pneumonia	About 3800

It can be seen that the data proportion difference is large, so the weight of the sample needs to be adjusted. At the same time, the number of data is small, so the data needs to be enhanced.

2.2 Dataset Description

Dataset processing is as follows:

(1) Use the image data generator provided by keras for enhanced preprocessing of datasets. First, tensor image data batches are generated by real-time data enhancement, and can be iterated circularly. The principle is to flip, pan, zoom and add noise to the original image. The degree range of random rotation is 20, the random width offset, the angle of random stagger transformation and the range of random scaling are all 0.1, and random horizontal flipping is allowed. Then, use flow_From_The directory method reads the data to realize the automatic enhancement of the data. The principle is to take the folder path as the parameter, generate the data after data promotion or normalization, and generate batch data infinitely in an infinite loop.
(2) Adjust the weight of the sample. Use the class in the fit function_weight method to map the value of the class index to the weight, which is used to weight the loss function (only during the training period). The calculation method is that the weight of normal samples = the number of pneumonia samples/total, and the weight of pneumonia samples = the number of normal samples/total.

3 Model

First of all, this chapter uses several commonly used and pre trained deep learning models for transfer learning, including VGG16, VGG19, ResNet50, Xecption. By freezing some weights of these models and consuming less time for model training, the transfer learning of these models is completed and four classifiers are obtained. Then, these four weak classifiers are fused by linear regression to form a strong classifier, which realizes the integration of the model. The fusion model of integrated learning is conducive to improving the recognition accuracy and generalization ability of the model. At the same time, we also trained the traditional CNN network model to compare with the fusion model, and we can find that the performance of the fusion model is better than the traditional model in all aspects.

3.1 Transfer Learning

The pre-training models used in this paper are VGG16, VGG19, ResNet50 and Xecption. Considering that the full connection layer structure added by VGG16 and VGG19 is similar, while the full connection layer structure added by ResNet50 and xeption is similar, we take VGG16and ResNet50 as examples.

(1) VGG16 model construction. First, keep VGG16 convolution layer parameters, remove the original full connection layer of the model, and rebuild it. The model network structure before the full connection layer is called the bottleneck layer, which is used to extract the features of the image and obtain a (4, 4, 4512) feature vector; The feature vector is extracted and transformed into the classification results we need through the full connection layer. The final model structure is shown in Fig. 1.

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
vgg16 (Model)                (None, 4, 4, 512)         14714688
_________________________________________________________________
flatten_1 (Flatten)          (None, 8192)              0
_________________________________________________________________
dense_1 (Dense)              (None, 256)               2097408
_________________________________________________________________
dropout_1 (Dropout)          (None, 256)               0
_________________________________________________________________
dense_2 (Dense)              (None, 1)                 257
=================================================================
```

Fig. 1. VGG16 network structure built

As Table1 shown, we have added a flatten layer, two Dense layers and a Dropout layer for VGG16. The first Dense layer uses Relu as the activation function to accelerate training and prevent information loss. The second Dense layer uses Sigmoid as the activation function to output the training results of the model. Dropout layer can effectively alleviate the occurrence of over fitting by randomly stopping a neuron with a certain probability p, and achieve the effect of regularization to a certain extent.

Next is the model's superparameters. Through multiple feedbacks of model training results to modify the superparameters, we get the final superparameters as follows. The optimizer of the model is Adam, and the loss function is binary_crossentropy, which is a loss function commonly used in binary classification problems. Because the number of the two types of pictures is uneven, this paper also uses class_weight is used to modify the weight of loss function of different categories of data to alleviate the problem of uneven sample number. At the same time, the first several layers of the bottleneck layer are frozen to reduce the training time of the model.

The output of the final model is a number between 0 and 1. The closer it is to 1, the greater the probability that the picture is pneumonia, otherwise the opposite is true.

(2) ResNet50 model building. ResNet50 and VGG16 are generally the same in the process of transfer learning, but there are still some differences. Here are the differences. The final model structure is shown in Fig. 2.

Layer (type)	Output Shape	Param #
resnet50 (Model)	(None, 5, 5, 2048)	23587712
global_average_pooling2d_1 (	(None, 2048)	0
dense_1 (Dense)	(None, 512)	1049088
batch_normalization_1 (Batch	(None, 512)	2048
dropout_1 (Dropout)	(None, 512)	0
dense_2 (Dense)	(None, 1)	513

Fig. 2. Resnet50 network structure built

Firstly, the network structure of ResNet50 is more complex than VGG16, and there are many model parameters. Therefore, this paper uses the global pooling layer to replace the full connection layer. The advantage is that it can retain the spatial information extracted from the previous convolution layer and pooling layer, and can also effectively reduce the model parameters and reduce the training time of the model.

Because ResNet50 network structure is too deep, the network becomes difficult to train convergence and adjust parameters. This paper adds a batch processing layer, which whitens each batch, that is, the process of removing mean and variance. It can normalize the data, alleviate the gradient disappearance problem to a certain extent, accelerate the network convergence, and prevent the over fitting problem at the same time.

3.2 Integrated Learning

Because the output result of the model is between 0–1, the closer it is to 1, the greater the probability that the model considers pneumonia, and the closer it is to 0, the lower the probability that it considers pneumonia. We use linear regression to predict and build a linear regression model y $= a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + b$, where x_1 reach x_4. They are the output results of VGG16, VGG19, ResNet50 and Xecption models for a picture respectively, and y is the prediction result. a_1, a_2, a_3, a_4 and b are the parameters that need training.

Through linear regression training, we can get that the prediction result of the fusion model is a value y. We believe that when the value y is greater than 0.5, the model predicts that the picture has pneumonia. Less than 0.5 is predicted to be a normal picture.

4 Experimental Results and Analysis

The experimental data set comes from the public data of "Chest X-Ray Images (Pneumonia)" of Kaggle. The data set contains more than 6000 pneumonia X-ray images, which are classified by many experts, and the data is highly reliable. Experimental machine Intel (R) Core (TM) i5-9300 H CPU @2.40 GHz (8 CPUs), ~2.4 GHz, CPU GTX650, memory 16G Windows10 operating system. All models use PyCharm as the integrated development environment, and are implemented using the deep learning framework keras. The Loss curve and accuracy curve of the experimental results are drawn by pyplot of matplotlib to analyze the convergence of the model.

4.1 Performance Index

Error rate and accuracy are the two most commonly used performance measures in classification tasks. Error rate refers to the proportion of the number of samples with classification errors in the total number of samples, which is defined as Eq. (1):

$$error = \frac{1}{m}\sum\nolimits_{i=1}^{m} \|(f(x_i) \neq y_i) \tag{1}$$

Precision is the proportion of the number of samples with correct classification to the total number of samples, which is defined as:

$$acc = \frac{1}{m}\sum\nolimits_{i=1}^{m} \|(f(x_i) = y_i) \tag{2}$$

Precision and recall are the detection values with high adaptability to evaluate the performance of applications. For the binary classification problem, the combination of real category and algorithm prediction category can be divided into four cases: true positive (T P), false positive (F P), true negative (T N) and false negative (F N). The precision P is defined as:

$$P = \frac{TP}{TP + FP} \tag{3}$$

The recall R's formula is as follow:

$$R = \frac{TP}{TP + FN} \tag{4}$$

F1 takes into account both precision P and recall R, is measured as below:

$$F1 = \frac{2 \times P \times R}{P + R} \tag{5}$$

where ALL is the total number of samples.

In this paper, F1 value is used as the main evaluation standard, and accuracy and recall are used as auxiliary evaluation standards.

4.2 Experimental Results of Each Transfer Learning Model

The superparameters of each model are shown as table 2.

Table 2. Superparameters of each model

model	Learning rate	epoch	Frozen network layers	Lower boundary of learning rate	Learning rate decline factor
VGG16	0.00001	24	16	0.000000001	0.8
VGG19	0.00001	16	12	0.000000001	0.8
Resnet50	0.00003	24	32	0.00000000001	0.8
Xception	0.00001	24	32	0.00000000001	0.8

The following Figs. 4 and 5 show the changes of accuracy and loss of each transfer learning model with the number of epoch. Obviously, it can be found that the accuracy and loss on the training set change slightly, while the change on the verification set is larger. This is due to the small number of validation sets. At the same time, it can be found that the accuracy of each model is more than 80% at the beginning, because the pre trained weights are used, which can also significantly reduce the time cost of the training model. The number of epochs of VGG 19 is 16, and the number of epochs of the other three models is 24. This is because it is found that VGG 19 is easier to over fit during training, so the number of epochs of the model is reduced. At the same time, it is found that ResNet50 performs best in this training set, followed by VGG16. This is because the residual structure of ResNet50 can effectively prevent the gradient disappearance caused by the deepening of the network structure, so it has better performance. At the same time, VGG16 outperforms other models in the validation set, which shows that VGG16 has high generalization ability and robustness. It can be seen that VGG19 model has a significant change in accuracy and loss in the 11th epoch, which is due to the oscillation phenomenon caused by the excessive learning rate.

The changes of accuracy and loss in training dataset is shown as Fig. 3.

The changes of accuracy and loss in validation dataset is shown as Fig. 4.

The performance of each transfer learning model in the test dataset is as Table 3.

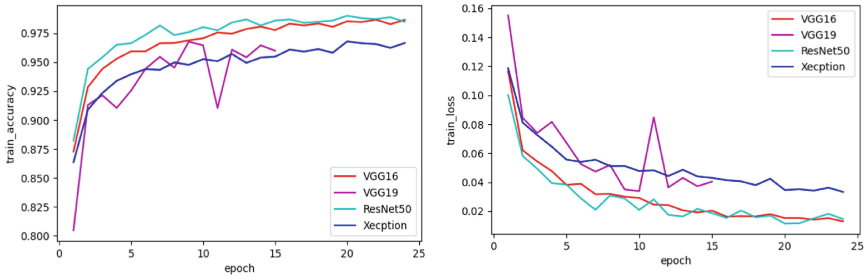

Fig. 3. The changes of accuracy and loss in training dataset.

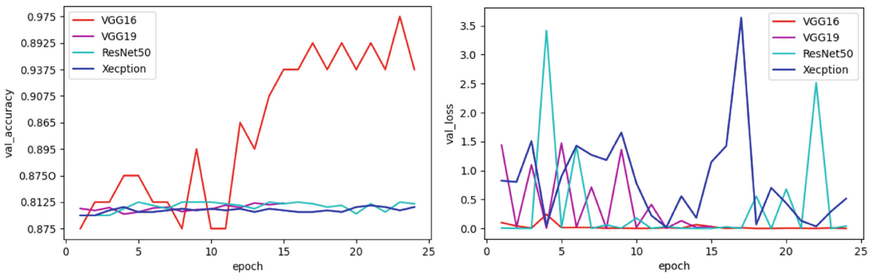

Fig. 4. The changes of accuracy and loss in validation dataset.

Table 3. Evaluation indicators of each model**1**

model	accuracy	recall	F1-score
VGG16	93.65%	91.97%	92.69%
VGG19	93.16%	92.78%	92.96%
Resnet50	92.80%	90.77%	91.61%
Xception	90.96%	88.16%	89.24%

4.3 Fusion Model

The Fig. 5 shows the weight diagram of four models fused by linear regression. It can be seen that VGG16 has the highest weight proportion, reaching 0.55, while Xecption has the lowest weight, reaching a point of almost negligible. It can be found that the higher the accuracy of the model, the greater the relative weight, which is in line with our expectations.

In order to compare the fusion model with the traditional model, this paper constructs a traditional CNN network, and the model structure is as follows Fig. 6.

The super parameter is set as: the number of iterations is 24, the learning rate is 0.00001, the learning rate reduction factor is 0.6, and the lower boundary of the learning rate is 0.000000001. It takes about seven hours to train in the local environment. The transfer learning model only needs more than three hours of training, which is twice as

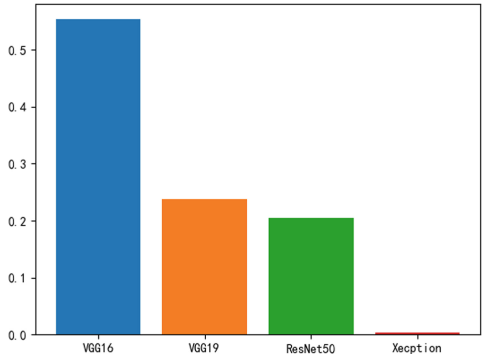

Fig. 5. Weight diagram of four models fused by linear regression

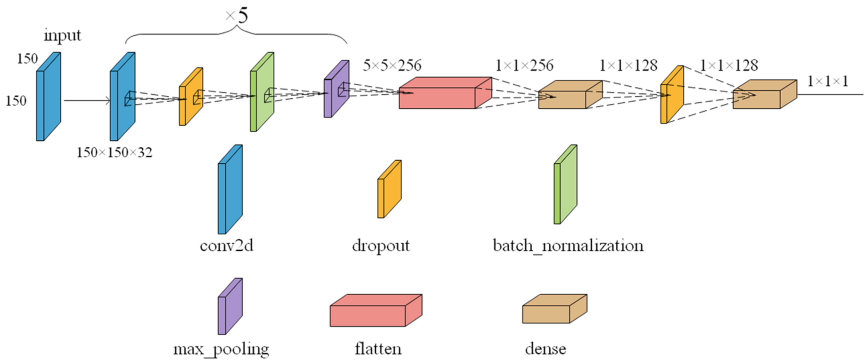

Fig. 6. The CNN network model structure constructed

fast as the ordinary CNN model. The Evaluation indicators of each model are listed in Table 4.

Table 4. Evaluation indicators of each model

model	accuracy	recall	F1-score
CNN	90.17%	94.10%	92.10%
Fusion model	94.40%	96.30%	95.34%

It can be found that the integration of multiple machine learning models can significantly improve the overall prediction ability. This is because different sub models have different expressive abilities in data. We can combine the parts they are good at to get an "accurate" model in all aspects. Through linear regression learning the weight of each model, we can get a relatively better weight of each model, which can well mine the strengths of each sub model and improve the effect of the final model.

5 Conclusion

X-ray detection of pneumonia is an important detection method in medicine. The traditional human eye detection is completely based on the judgment of doctors, which takes a long time to obtain the results and may be misjudged; CNN has a good foundation for pneumonia image detection, but the training time is long and the recognition accuracy is not high. This paper uses an integrated learning framework integrating multi transfer learning model, which greatly improves the accuracy of recognition and reduces the training time. In the model, we choose VGG16, VGG19, ResNet50, Xecption classifiers, and integrate multiple models into a strong classifier through linear regression. Compared with traditional CNN, it greatly reduces the training time of the model and improves the recognition ability of the model.

However, the model still has some limitations, including the inability to show the specific location of the lesion. This will be further studied in future work.

6 Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgment. This work has been supported in part by the Natural Science Foundation of China under grant No. 62262030, the Science and Technology Research Project of Jiangxi Provincial Department of Education (No. GJJ170234).

Data Availability. The data in the experiments, used to support the findings of this study are available from the corresponding author upon request.

References

1. Gemelli Against COVID, Post-Acute Care Study Group, et al.: Post-COVID-19 global health strategies: the need for an interdisciplinary approach. Aging Clin. Exp. Res. **32**, 1613–1620 (2020)
2. Abbas, A., Abdelsamea, M.M., Gaber, M.M.: Classification of covid-19 in chest x-ray images using detrac deep convolutional neural network. Appl. Intell. **51**(2), 854–864 (2021)
3. Lin, T.-C., Lee, H.-C.: Covid-19 chest radiography images analysis based on integration of image pre-process, guided grad-cam, machine learning and risk management. In: Proceedings of the 4th International Conference on Medical and Health Informatics, pp. 281–288 (2020)
4. Wardlaw, T., Salama, P., Johansson, E.W., Mason, E.: Pneumonia: the leading killer of children. Lancet **368**(9541), 1048–1050 (2006)
5. Lakhani, P., Sundaram, B.: Deep learning at chest radiography: automated classification of pulmonary tu-berculosis by using convolutional neural networks. Radiology **284**(2), 574–582 (2017)
6. Vu, T.H., Murakami, R., Okuyama, Y., Ben Abdallah, A.: Efficient optimization and hardware acceleration of CNNS towards the design of a scalable neuro inspired architecture in hardware. In: 2018 IEEE International Conference on Big Data and Smart Computing (Big-Comp), pp. 326–332 (2018)

7. Stephen, O., Sain, M., Maduh, U.J., Jeong, D.-U.: An efficient deep learning approach to pneumonia classification in healthcare. J. Healthc. Eng. **2019**, 4180949 (2019)
8. Pan, I., Agarwal, S., Merck, D.: Generalizable inter-institutional classification of abnormal chest radiographs using efficient convolutional neural networks. J. Digit. Imaging **32**(5), 888–896 (2019)
9. Cai, C., Wang, S., Xu, Y., et al.: Transfer learning for drug discovery. J. Med. Chem. **63**(16), 8683–8694 (2020)
10. Wang, J., Nakamura, M., Ben Abdallah, A.: Efficient AI-enabled pneumonia detection in chest x-ray images. In: 2022 IEEE 4th Global Conference on Life Sciences and Technologies (LifeTech), pp. 470–474 (2022)