Visual Deep Learning Model Improvement, Comparison, and Integration: Prediction of Pneumonia Based on Chest X-Ray Images

Ziang Niu *

School of Mathematical Sciences, Beijing Normal University, Beijing, 100875, China

* Corresponding Author Email: 202111130012@mail.bnu.edu.cn

Abstract. Using machine learning to process lung medical images can greatly improve hospital efficiency and save costs. With the increase in the number of patients, the demand for pneumonia pathologic recognition systems is increasing. Therefore, the organic combination of the two is of great significance to reduce the pressure on the medical and health systems. This paper presents a deep-learning method to identify and predict pneumonia. Using Convolutional Neural Networks, ResNet, and DenseNet as well as improved and integrated models, the known pneumonia images and normal lung images were used as training sets to identify lung images and determine whether pneumonia is present. The results showed that the original CNN and the improved ResNet network had the best effect, and the F1-score for pneumonia recognition reached 0.88. Therefore, this paper integrated these two neural networks. Finally, the F1-score of the integrated model reached 0.89, which was able to predict more accurately. This paper provides a new idea for selecting, integrating, and applying the model in the medical field.

Keywords: CNN, ResNet, DenseNet, Integrated model, pneumonia.

1. Introduction

The recognition of pathological images has always been one of the problems that the medical community wants to solve. As the number of people in the world increases, the need for pathological identifying systems in the clinical field is increasing. Such systems can not only greatly save medical costs, but also facilitate the early identification and determination of pathological characteristics and early intervention treatment. If the staff can use a mature system to achieve preliminary screening and prediction according to patients' pathological images, the value of the public health field is immeasurable.

The outbreak of COVID-19 in 2020 has put new demands on public health to respond the pneumonia. Based on identification, it is necessary to subdivide what kind of pneumonia is so that more targeted treatment can be carried out. The proposed requirement can not only provide convenience for medical staff but also increase a guarantee for the safety of patients. More importantly, the proposed method also puts forward a new idea for the development of the public health field, that is, the combination of neural networks and medical treatment. The performance of neural network models is improved through real medical data, and the medical field can obtain more accurate prediction results through the upgraded neural network, benefiting a wider range of patients.

At present, some studies have tried to combine neural networks with pneumonia prediction: Zhao studies the classification, retrieval, segmentation, and other tasks of the Contrastive Language-Image Pre-training model for the joint task of image text, and points out that CLIP can effectively improve the accuracy of image classification, segmentation, and retrieval, especially in the case of a small amount of annotated data [1]. Lafraxo carries out more accurate pneumonia detection by mixing and improving ACNN and LSTM models [2]. An uses a combination of deep convolutional neural network and attention integration model to detect pneumonia, which significantly improves the model's ability to focus on key features and reduces the misjudgment rate [3]. Mabrouk built an integration of multiple CNN, demonstrating the advantages of an integrated approach in improving model performance and stability [4].

Although some studies have made some preliminary explorations on neural networks and pneumonia prediction, there are still some gaps and challenges. The generalization ability of the model is limited in the face of different complex cases. The interpretability of the model and the clinical effect of practical application still need to be strengthened. The robustness of the model facing the problems of data set limitation, imbalance, and low image quality is also an urgent problem to be solved.

In this study, a variety of visual deep-learning models were used to identify, classify, and predict lung images. The model performance and output accuracy are improved through model improvement and integration. Finally, the performance of different models is compared and summarized. The method provides a new idea for image recognition in the medical field.

2. Method

2.1. Data Set

Pneumonia data from Chest X-Ray Images (Pneumonia) on Kaggle were selected. The source data is obtained from Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification, the dataset has been published by DOI, publishing house Elsevier, with high downloads and reliability [5]. Data contains three folders: train, test, and val. This method does not use val because there is too little data in the val folder. The Train folder contains 1,341 normal lung scans and 3,875 pneumonia scans; The Test folder contains 234 normal lung scans and 390 pneumonia scans. The variable is set to pneumonia/normal to distinguish.

2.2. Methods and Models

Three different neural network models, which are CNN, DenseNet, and ResNet are trained on the data set and the prediction accuracy is obtained. The prediction accuracy includes the confusion matrix and the classification report, including accuracy rate, recall rate, and F1-score. Then, three models are optimized and integrated, and similar predictions are written.

2.2.1. Model Overview

CNN is a common neural network model for image classification and prediction. Its main working principle is that the image is multi-channel encoded through the input layer, and then the feature is extracted through the convolution layer of the core step. The convolution layer is composed of a convolution kernel, which is a set of weight matrices, and generates a feature map to extract features by performing dot product operation on local region point by point. Then, nonlinear features are introduced by the nonlinear activation function, and dimensionality is reduced in the pooling layer. These layers are superimposed to form the convolution body. Finally, restore and output through the full connection layer. The advantages of CNN are automatic feature extraction, a small training set, and easy generalization.

DenseNet is an improved CNN, whose main feature is the introduction of a dense connection mechanism so that the NTH layer can receive information from the former n-1 layer (in contrast, CNN can only receive information from the former 1 layer). In this way, the problem of gradient disappearance is reduced and the parameter efficiency is high.

Unlike DenseNet, ResNet uses residual learning, in which input from one layer can skip several layers to a later layer, and focuses on solving the problem of information loss. It has the advantage of being more stable and lighter.

2.2.2. Model Optimization and Integration

The preliminary CNN model of this method consists of four convolution layers with the size of convolution kernel (3,3), and the layers have 32, 64, 128, and 128 convolution kernels respectively. After processing by the convolutional layer, a fully connected layer of 512 neurons is used for binary classification using the sigmoid activation function. Then, use the Adam optimizer with a learning rate of 0.001 and the binary_crossentropy loss function. The number of training rounds is set to 10.

The improved CNN model also contains four convolution layers, but the convolution kernel size and layer configuration are adjusted. The first two layers use the convolutional kernel of (5, 5), with 32 and 64 convolutional kernels respectively, and the second two layers use the convolutional kernel of (3, 3), with 128 convolutional kernels each. A BatchNorm layer is added after each convolution layer to improve training stability. After passing through the convolutional layer, the model passes through a fully connected layer of 1024 neurons and adds the Dropout layer (drop rate 0.5) to reduce overfitting.

The preliminary DenseNet model consists of an initial convolutional Layer, multiple Dense blocks, and a Transition Layer. The initial convolution layer uses 64 convolution kernels of size (7,7). The Dense Block contains multiple convolution layers, each with convolution kernel sizes of (1,1) and (3,3). The Transition Layer uses (1,1) convolution and pooling to reduce the number of feature maps. The Adam optimizer with a learning rate of 0.001 and the binary_crossentropy loss function are also used. The number of training rounds is set to 10.

The improved DenseNet model optimizes the convolution kernel of the initial convolution layer by changing (7,7) to (5,5) and increasing the number of convolution nuclei to 128. The convolution layer of some Dense blocks is also fixed channel adjustment. The number of neurons in the fully connected layer increases to 1024. Increase the training period from 10 to 20.

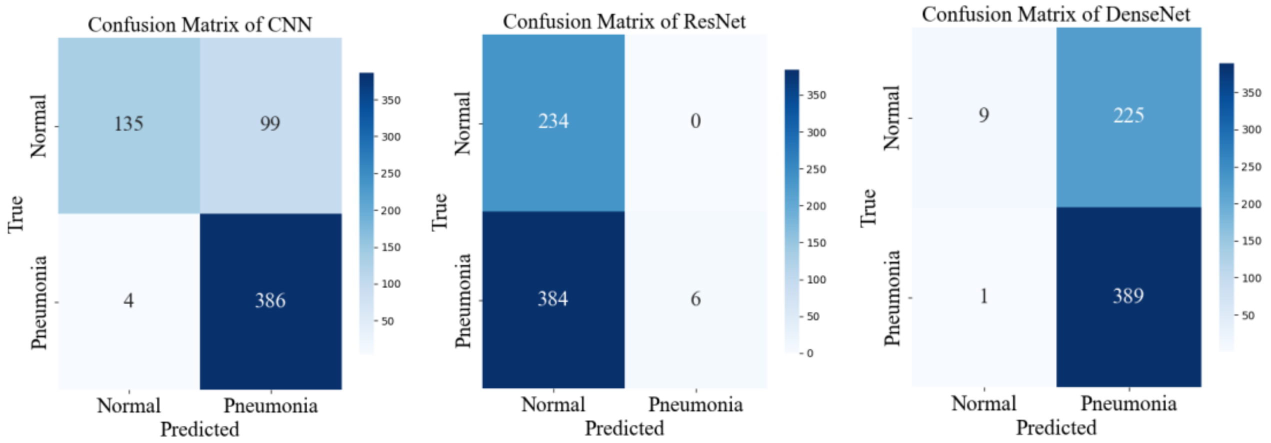
The preliminary ResNet model includes multiple residual blocks for feature extraction. The initial convolution layer uses 64 convolution kernels (7,7). The model has four groups of residual blocks: the first group has three 64-channel blocks; The second group of three 128-channel blocks; The third group has three 256-channel blocks; The fourth set of three 512-channel blocks, the first of which all use a convolution of step 2. Each residual block consists of (3, 3) and (1, 1) convolution layers, combined with Batch Normalization and ReLU activation functions. The number of training rounds is set to 10.

The improvement points include: adjusting the convolution kernel of the initial convolution layer from (7, 7) to (3, 3). Adding Dropout layer after the fully connected layer to reduce the risk of overfitting; The learning rate scheduler is introduced to adjust the learning rate dynamically. Increase the training period from 10 to 20.

Then, according to the prediction accuracy, the preliminary CNN model is integrated with the improved ResNet, and the Soft Voting method is adopted. This method calculates the average probability of each class based on the prediction probability of each model and then selects the class with the highest prediction probability as the final prediction result.

3. Results

According to the preliminary model, Figure 1 shows the three confusion matrices obtained from the model prediction and the real situation. From these confusion matrices, accuracy, recall, and F1-scores can be obtained, as shown in Table 1. The precision rate represents the proportion of samples that the model predicts to be of a certain class that belongs to that class. F1-score is the harmonic average of accuracy rate and recall rate and is an index that comprehensively considers accuracy rate and recall rate. The recall rate represents the proportion of samples that belong to a class that is correctly predicted by the model to be that class.



(a) (b) (c)

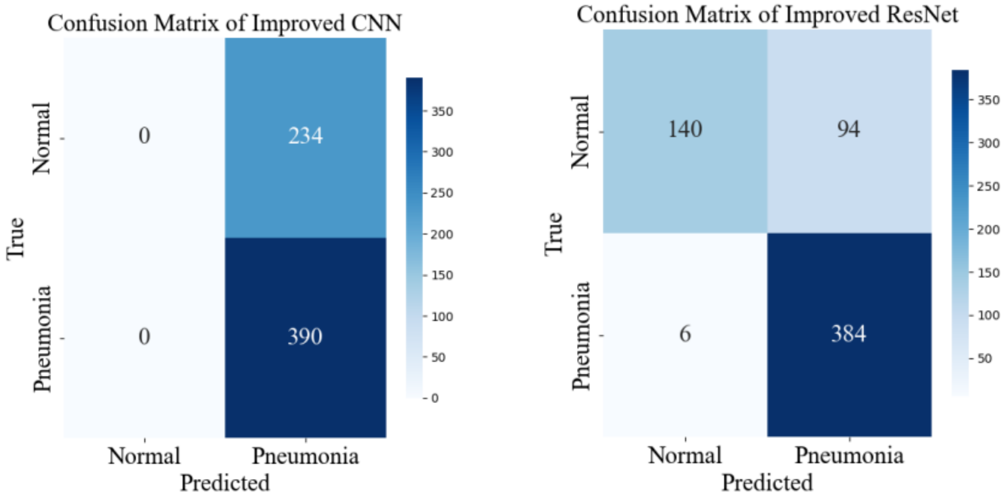
Figure 1. Confusion matrices of 3 rudimentary models, (a) CNN's confusion matrix, (b) ResNet's confusion matrix, (c) DenseNet's confusion matrix (Photo/Picture credit: Original).

Table 1. Precision rates, recall rates, and F1-score of 3 rudimentary models

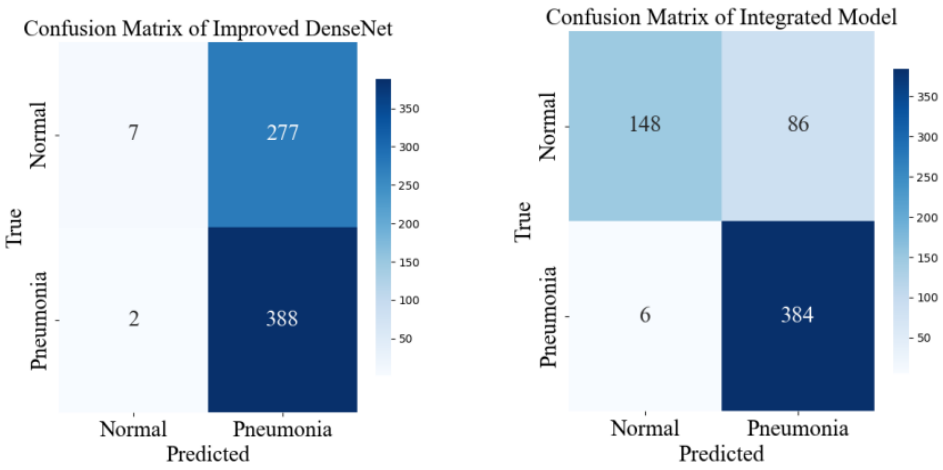
	Precision	Recall	F1-Score
CNN-Normal	0.97	0.58	0.72
CNN-Pneumonia	0.80	0.99	0.88
ResNet-Normal	0.38	1.00	0.55
ResNet-Pneumonia	1.00	0.02	0.03
DenseNet-Normal	0.90	0.04	0.07
DenseNet-Pneumonia	0.63	1.00	0.77

It is not difficult to see that CNN has the best model classification and prediction ability and the highest F1-score, which reflects its strong feature extraction ability for such images. ResNet's forecast was too heavy on Normal, and DenseNet's was too heavy on Pneumonia. This reflects that the lung image features are not well extracted.

Then, let the improved model and the final integrated model predict the output. According to the four confusion matrices drawn in Figure 2, the accuracy rate, recall rate, and F1-score are obtained, as shown in Table 2.



(a) (b)



(c) (d)

Figure 2. Confusion matrices of 3 improved models and the integrated model, (a) Improved CNN's confusion matrix, (b) Improved ResNet's confusion matrix, (c) Improved DenseNet's confusion matrix, (d) Integrated Model's confusion matrix (Photo/Picture credit: Original).

Table 2. Precision rates, recall rates, and F1-score of 3 improved models and the integrated model

	Precision	Recall	F1-Score
Improved-CNN-Normal	0.00	0.00	0.00
Improved-CNN-Pneumonia	0.62	1.00	0.77
Improved-ResNet-Normal	0.96	0.60	0.74
Improved-ResNet-Pneumonia	0.80	0.98	0.88
Improved-DenseNet-Normal	0.78	0.03	0.06
Improved-DenseNet-Pneumonia	0.63	0.99	0.77
Integrated-Model-Normal	0.96	0.63	0.76
Integrated-Model-Pneumonia	0.82	0.98	0.89

As shown in Table 2, the improved CNN model went to the extreme, all predicting pneumonia. ResNet's predictions work well. DenseNet's prediction is still too heavy on Pneumonia.

After selecting the two models with the strongest prediction ability, namely the initial CNN and the improved ResNet, it is found that the F1-score has improved, which indicates that the model performance has been upgraded.

4. Discussion

This method verifies the strong ability of CNN and ResNet in lung image feature extraction. Faced with the problems of information loss in CNN and the over-complexity of ResNet, scholars have made improvements by comparing, testing, and proposing new models. Krizhevsky achieved a significant breakthrough in the ImageNet classification task through deep convolutional neural networks, demonstrating the effectiveness of deep learning in large-scale image recognition and triggering a wide range of research and applications [6]. Shazia compared the performance of multiple neural networks in detecting COVID-19 in chest X-ray images and found that the integrated model outperformed a single network in both accuracy and sensitivity, showing the advantages of different combinations of models [7]. The integrated model can not only take advantage of residual learning in ResNet but also avoid the phenomenon of over-extraction in DenseNet.

Regarding the strong preference nature of model predictions, there are some guesses and methods to help solve these problems. Mehrabi mentioned that the sources of bias include data bias, algorithmic bias, and social bias [8]. The types of bias include data-driven bias, model-driven bias, and bias in the decision-making process. Kraidia studied compression and optimization techniques to improve the robustness and efficiency of neural networks against adversarial attacks [9]. Wang discussed the origin, development, and influence of deep learning in the field of modern artificial intelligence. These have significant implications for follow-up research.

5. Conclusion

In general, both CNN and ResNet can achieve better pneumonia prediction results on this data set. The integrated model also improved performance, resulting in a recall control of 0.98 for pneumonia predictions and an F1-score of 0.89, which is a good performance. The model performance can also be partially improved by changing the number of layers and the size of the convolution kernel. However, it also faces the problem of skewed data forecasts.

The application of visual neural networks in pathological image analysis has a broad prospect, which not only shows great potential in the detection and diagnosis of lung images but also can be extended to the image analysis of other organs, such as the stomach, intestine, liver, etc. Through deep learning technology, neural networks can identify diseased areas in pathological images of these organs, predict the malignancy of the disease, and assist doctors in making a more accurate diagnosis. For example, in stomach image analysis, neural networks can help identify gastric precancerous lesions for early intervention. In intestinal image analysis, neural networks can assist in identifying problems such as inflammatory bowel disease or colon cancer. This technique not only improves the accuracy of diagnosis but also helps to detect the disease early, thus improving the cure rate and survival rate of patients.

For the future, expect more innovative models and approaches to emerge. These new technologies should not only solve the existing computational bottleneck, but also make breakthroughs in improving the accuracy of the model, reducing the misdiagnosis rate, and enhancing the generalization ability of the model. Through these efforts, visual neural network technology can truly benefit the majority of patients and improve the quality and efficiency of medical care.

References

[1] Zhao, Z., Liu, Y., Wu, H., Li, Y., Wang, S., Teng, L., ... & Shen, D. Clip in medical imaging: A comprehensive survey. arXiv preprint arXiv:2312.07353, 2023.

[2] Lafraxo, S., El Ansari, M., & Koutti, L. A new hybrid approach for pneumonia detection using chest X-rays based on ACNN-LSTM and attention mechanism. Multimedia Tools and Applications, 2024, 1 - 23.

[3] An, Q., Chen, W., & Shao, W. A deep convolutional neural network for pneumonia detection in x-ray images with attention ensemble. Diagnostics, 2024, 14 (4), 390.

[4] Mabrouk, A., Diaz Redondo, R. P., Dahou, A., Abd Elaziz, M., & Kayed, M. Pneumonia detection on chest X-ray images using ensemble of deep convolutional neural networks. Applied Sciences, 2022, 12 (13), 6448.

[5] Kermany, D., Zhang, K., & Goldbaum, M. Labeled optical coherence tomography (oct) and chest x-ray images for classification. Mendeley data, 2018, 2 (2), 651.

[6] Krizhevsky, A., Sutskever, I., & Hinton, G. E. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 2012, 25.

[7] Shazia, A., Xuan, T. Z., Chuah, J. H., Usman, J., Qian, P., & Lai, K. W. A comparative study of multiple neural networks for detection of COVID-19 on chest X-ray. EURASIP journal on advances in signal processing, 2021, 1 - 16.

[8] Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. A survey on bias and fairness in machine learning. ACM computing surveys (CSUR), 2021, 54 (6), 1 - 35.

[9] Kraidia, I., Ghenai, A., & Belhaouari, S. B. Defense against adversarial attacks: robust and efficient compressed optimized neural networks. Scientific Reports, 2024, 14 (1), 6420.

[10] Wang, H., & Raj, B. On the origin of deep learning. arXiv preprint arXiv: 1702. 07800, 2017.