A COVID-19 Detection Method based on Deep Learning Model Trained by Chest X-Ray Images

Guanhua Li[1] and Zhenzhu Yin[2,*]

[1]School of Computer Science and Engineering, Tianjin University of Technology, Tianjin, 300384, China
[2]School of Computer Science and Engineering, Tianjin University of Technology, Tianjin, 300384, China
*Corresponding author: vivi@stud.tjut.edu.cn

Abstract. COVID-19 is a contagious disease, with tens of millions of people worldwide infected. Taking Chest X-Ray (CXR) images is an important step during clinical diagnosis, because doctors could monitor the lung status directly by this way. In this paper, we deploy three models: pretrained ResNet152, Linear discriminant analysis (LDA), and Support Vector Machine (SVM). We train these models with different sample sizes: 100, 1500, and 3307, to observe their performances, and ResNet152 all outperforms the other two method. A 96.06% accuracy, a 96.12% precision, a 96.06% recall, and a 96.06% F1 Score are attained (the indicators are averaged weighted), demonstrating that ResNet152 has great strength and potential in CXR recognition field. Besides, we discuss the reason for the underperform of SVM and LDA. Furthermore, 10 independent repeated tests verified the prominent stability of ResNet152. The four indicators' extreme deviations obtained are all within 1.23%. This work indicate that ResNet152 is a very effective model, and can greatly assist the healthcare industry in the future.

Keywords: Image recognition, CXR Image, CNN, ResNet152, COVID-19

1 Introduction

COVID-19, also known as the Novel Coronavirus Disease, had been proved as a highly contagious epidemic which was firstly emerged in late 2019 in Wuhan, China. It had soon been recognized as a global health problem since its contact-transferred behavior ability was confirmed by scientists, epidemiologists, and biologists [1]. Many countries have tried various methods to prevent and control the epidemic [2]. According to a report by World Health Organization (WHO), COVID-19 is caused by a kind of virus called Severe Acute Respiratory Syndrome Coronavirus 2 (also known as SARS-Cov-2) [3]. The original and mutated virus strains are causing symptoms like sore throat, fever, and sneezing. In severe cases, it may cause devastating symptoms such as "White Lung". Doctors usually require patients to take Chest X-Ray (CXR) images during diagnosis and treatment in order to monitor the infection status of the patient's lungs.

B. H. Ahmad (ed.), *Proceedings of the 2023 International Conference on Data Science, Advanced Algorithm and Intelligent Computing (DAI 2023)*, Advances in Intelligent Systems Research 180,
https://doi.org/10.2991/978-94-6463-370-2_36

In recent years, Deep Learning has drawn close attention from the healthcare system. Certain studies indicates that some well-trained models have surpassed human performance in certain visual and auditory recognition tasks [4]. Kassania compared many kinds of deep convolutional neural networks and found that the DenseNet121 feature extractor with Bagging tree classifier have an outstanding performance [5]. Burlacu validated the potential of AI use in COVID-19 diagnosis on Chest X-Ray (CXR) images [6]. Khalifa trained a Deep Transfer Learning Model for COVID-19 diagnosis and measurement and achieved good results [7]. These researches are of great significance in tackling the classification problem of X-ray images and provide guidance for designing experiments.

Since researchers are exploring the performance of deeper Neural Network, and the introduction of Residual Neural Network has mitigated the "degradation" problem, we propose to train a model based on pretrained ResNet152 with CXR images [8]. The purpose is to explore whether ResNet152 is feasible on COVID-19 recognition through CXR images, and to what extent the accuracy could be. The overall structure and framework of this work is shown in Figure 1.

In summary, our main contributions are briefly outlined as follow: 1) We download CXR images from COVID-19 X-Ray Dataset, implement preliminary filtering, and then use Logical AND Operation to combine each image with corresponding Mask images [9]. 2) We use the pretrained ResNet152 model and train with the combined CXR images. Then we select to use average weighted indicators to evaluate the result. 3) We design independent repeated tests to verify the stability. Besides, in order to verify the superiority and comparability, we designed a comparative experiment with SVM and LDA, which are trained with same Dataset but under different sample sizes (100, 1500, and 3307).

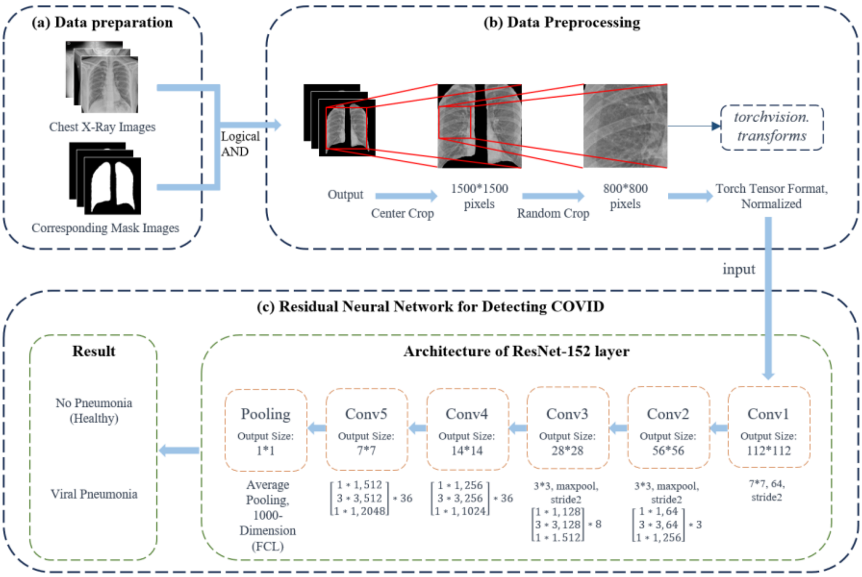

Fig. 1. Overall structure and framework of this work (Picture credited: Original)

2 Materials and Methods

2.1 SVM, LDA, CNN, and ResNet

Support Vector Machine (SVM, also Support Vector Network) is a kind of supervised learning model with associated learning algorithm. It is a robust method which could help to do text or hypertext categorization, image classification, and hand-writing recognition [10].

Linear discriminant analysis (LDA) is a generalization and realization of Fisher's linear discriminant [11]. It is often used in Data analysis and statistics. The purpose of this method is to find a linear combination among objects with different feature or characteristic. LDA could project the data in low dimension, the distance of the points after projection in each class are as close to each other, but the distance between the center of each class is as large as possible [12].

Convolutional Neural Network (CNN) is a particular class of Artificial Neural Network. CNN is widely used to analyze visual imagery [13]. The CNN use a mathematical operation called convolution to replace at least one of the network layers' matrix multiplications. The network architect is carefully designed to process pixels and process the inner data to recognize or analyze images.

ResNet (Residual Neural Network) is a kind of CNN that includes skip connections, which perform identity mappings and merge with layer outputs by addition [8]. The ResNet solved the "degradation" problem, which refers to the training accuracy reduction during stacking too much layers in a neural network [8]. ResNet152 has 152 layers, making it one of the deepest neural network architectures available. It has also been used as a pre-trained model for transfer learning in a variety of computer vision applications. The pre-trained ResNet models are available in *PyTorch* Websites (Homepage: pytorch.org).

2.2 The Experiment Procedure

Dataset Preparation and Preprocessing. Among the public CXR Datasets on the Internet, we preferred the COVID-19 X-Ray Dataset which is first published in 2020 by V7Labs (Homepage: www.v7labs.com) [9]. This dataset contains totally 6504 X-ray images so far, and has been mainly divided into 3 classes: Bacterial Pneumonia, Viral Pneumonia and No Pneumonia (healthy). During our work, we consider to use "Viral Pneumonia" and "No Pneumonia (healthy)" images to complete our work.

Considering the standardization and uniformity of the training dataset, we chose the posteroanterior (PA) and anteroposterior (AP) views (or projections) of the CXR images for training. Other views of the images are manually filtered and abandoned. Images with low quality or resolution are also abandoned. After the filtration of the CXR images, there are totally 3307 left. Among them are 1597 healthy CXR images and 1710 Viral Pneumonia CXR images. Then we computed the Mean and Standard Deviations based on the images in our dataset. These two parameters are used to normalize and standardize the input data, and to enhance the model's generalization ability and training performance [14]. They will be utilized in the subsequent ResNet training phase. In the next step, we download the Mask Image files for each CXR image from the Online Dataset Server. Each CXR image has its own unique Mask Image, which depicted the contours of the lungs. We use the Logical AND Operation to combine the CXR images with corresponding Mask Images, as the example shown in Figure 2, in order to focus on the pixels of the lung. Our intention of this step is to exclude unnecessary parts in CXR images that may affect the accuracy in the training procedure, for example, the neck or the clavicles.

Fig. 2. the Logical AND between CXR and corresponding Mask Image (Picture credited: Original)

Training Detail. The hardware equipment used throughout this work are 1x NVIDIA GeForce RTX 4090 GPU @ 24GB; 12x Intel(R) Xeon(R) Platinum 8352V CPU @ 2.10GHz; RAM size 90GB; and Hard Disk size 50GB. The System is ubuntu 20.04. The software environments are Python 3.8, PyTorch 2.0.0 and Cuda 11.8.

The training phase follows these procedures: 1) Partition of the dataset. All the CXR images are randomly shuffled, in order to disrupt the original sequence. On this basis, 80% of the CXR were selected as the training set, and the remained as the testing set. Thus, the count of CXR in the training and testing set is 2647 (1278 healthy and 1369 Viral Pneumonia) and 660 (319 healthy and 341 Viral Pneumonia), respectively. 2) Initialization of the model and parameters. Pretrained ResNet152 is used as the model in our work. Adam optimizer is selected in our work, and the learning rate is set at 1e-4. Batch size is set at 4. The epoch size is adjusted to 670. The Mean and Standard Deviations are set to 0.491 and 0.2325, respectively, after computing. 3) Transformation of CXR images. The transformation function is provided in the *torchvision* package. For each image, we firstly cropped a 1500*1500-pixels square from its center (if the image resolution is lower than 1500*1500, the shortage part will be filled with black color). Then, in that square, we randomly select an 800*800-pixels square for training. The smaller square is transformed into Torch Tensor format. 4) Training and evaluation. Cross Entropy Loss is used as the Loss-function [15]. The result and evaluation will be shown in the following chapters.

Formula of the Indicators. To evaluate the performance of this experiment, we decided to introduce the most common performance measure indicators in Deep Learning field, that are Accuracy, Precision, Recall, and F1 Score [16]. Considering that the image number in the two categories of our dataset is not equal, we choose to introduce "weighted averaged method" in the process of computing [17]. The intention of this step is taking proper processing to reduce the influence of imbalanced data during evaluation [18]. The function of computing "weighted averaged" indicators are provided in the *sklearn* package (*sklearn.metrics.precision_recall_fscore_support(...,average='weighted')*).

The formulae to compute the weight (w) is represented as equation (1), and Accuracy, Precision, Recall, F1 Score are represented as equation (2) to equation (5) [17].

$$w_i = \frac{C_i}{\sum_{i=1}^{m} C_i} \tag{1}$$

Where m means the total number of classes in the dataset, and i values from 1 to m. C_i represent the total of samples in the i^{th} Category.

$$Accuracy = \frac{\sum_{i=1}^{m} \frac{TN_i + TP_i}{TN_i + TP_i + FN_i + FP_i}}{m} \tag{2}$$

$$Precision_{Weighted} = \frac{\sum_{i=1}^{m} w_i * \frac{TP_i}{TP_i + FP_i}}{\sum_{i=1}^{m} w_i} \tag{3}$$

$$Recall_{Weighted} = \frac{\sum_{i=1}^{m} w_i * \frac{TP_i}{TP_i + FN_i}}{\sum_{i=1}^{m} w_i} \tag{4}$$

$$F1\ Score_{Weighted} = \frac{\sum_{i=1}^{m} w_i * \frac{2 * TP_i}{2 * TP_i + FP_i + FN_i}}{\sum_{i=1}^{m} w_i} \tag{5}$$

Where TP_i (True Positive Samples of the i^{th} class), TN_i (True Negative Samples of the i^{th} class), FP_i (False Positive Samples of the i^{th} class), FN_i (False Negative Samples of the i^{th} class) can be formed as a confusion matrix, which present the classification performance of the model in each category, like Table 1 showed below [16].

Table 1. TP, TN, FP, FN in a Confusion Matrix

		Prediction	
		+	-
Actual Condition	+	**TP**	**FN**
	-	**FP**	**TN**

3 Result

In this section, we detailly describe the result of our experiments. We also evaluate the model on the comparability with SVM and LDA methods. In this section, the Precision, Recall, and F1 Score are all "weighted averaged".

3.1 Result of the ResNet152

We evaluate this model by using the testing set which contains 660 samples. As shown in Figure 3, this confusion matrix presents the performance on classification of the model in each category.

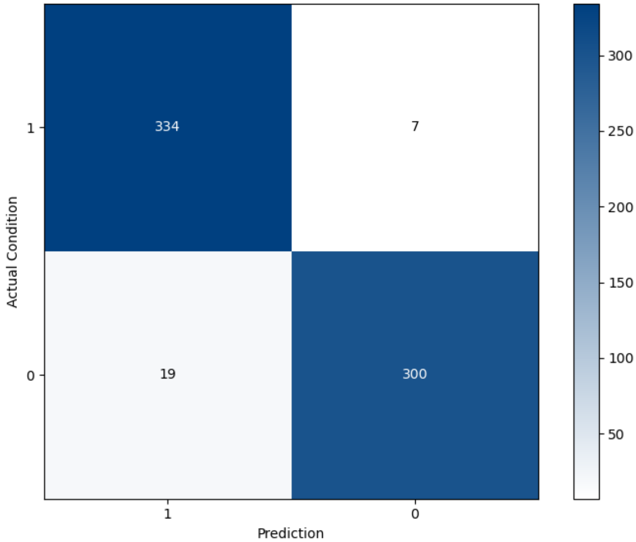

Fig. 3. The confusion matrix of the experiment result

Table 2. Result of the 10-time Independent Repeated Test

Serial Number	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
1	96.06	96.12	96.06	96.06
2	95.76	95.78	95.76	95.76
3	96.06	96.07	96.06	96.06
4	95.30	95.32	95.30	95.30
5	96.25	96.26	96.25	96.24
6	95.48	95.48	95.48	95.48
7	95.91	95.91	95.90	95.91
8	96.52	96.55	96.52	96.51
9	95.30	95.33	95.30	95.30
10	95.76	95.77	95.76	95.76
Extreme Deviation (%)	1.22	1.23	1.22	1.21

According to Figure 3, the accuracy of classifying Healthy and Viral Pneumonia CXR images is 94.04% and 97.95% respectively. Therefore, the model is more sensitive to Viral Pneumonia samples, because its accuracy is 3.91% higher. Calculated by this confusion matrix, the overall accuracy of this model is 96.06%, the precision is 96.12%, the recall is 96.06%, the F1 Score is 96.06%. These data indicate that ResNet152 has good performance in detecting viral pneumonia by CXR images.

3.2 Independent Repeated Experiments

We ran a 10-time independent, repeated experiment on the overall dataset to test the model's stability. All CXR images are randomly partitioned. We can see the results shown in Table 2. displayed that the indicators were distributed in a certain range: the accuracy from 95.30% to 96.52%, the precision from 95.32% to 96.55%, the recall from 95.30% to 96.52%, the F1 Score from 95.30% to 96.51%. Besides, the table indicates that the extreme deviations of the four indicators are 1.22%, 1.23%, 1.22%, and 1.21% respectively. These data shows that ResNet152 has stability on CXR images recognition to some certain extent.

3.3 Comparation with SVM, LDA

In order to verify the comparability with SVM and LDA, we select different sample sizes (100, 1500, and 3307) for comparative experiments. The Dataset preprocessing steps of SVM and LDA are basically same as those of ResNet152. The only difference is that we calculate the color (Grayscale) histograms, normalize, and then flatten them into a one-dimensional vector, as the image feature vectors of SVM and LDA.

The result shown in Table 3 prove that ResNet152 all performs best on recognizing CXR images under different number of samples.

Table 3. Comparative experiment result, under different sample sizes

Method	100 (Samples)				1500 (Samples)				3307 (Samples)			
	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
SVM	50.00	25.00	50.00	33.33	50.67	54.34	50.67	37.43	50.15	58.22	50.15	38.05
LDA	65.00	66.48	65.00	58.33	63.33	63.37	63.33	63.31	64.09	64.35	64.09	64.06
ResNet152	**90.00**	**91.67**	**90.00**	**90.00**	**94.33**	**94.49**	**94.33**	**94.33**	**96.06**	**96.12**	**96.06**	**96.06**

3.4 Discussions

From Table 3, a phenomenon could be found that, the result of SVM and LDA seems a bit too low. We consider that the reasons are below: 1) As we combined the original CXR images with their Masks, some parts of the sample are filled with black color. This may have an impact to the color histograms, causing a decrease in model accuracy. 2) Since we only focus on color distribution, we neglect the importance of structural and textural features, meaning that we missed many valid information [19]. 3) "Plain" SVM or LDA may have limitations to some extent, so a combination with CNN may lead to an improvement [20, 21].

Furthermore, we can see the indicators of ResNet152 increased as the number of samples went up. A reasonable conjecture is that the model will have even better and more stable performance if we could expand the Dataset, by collecting more CXR images from other Datasets, or generate additional samples by Generative Adversarial Networks (GAN) [22].

4 Conclusion

In this paper, we trained the ResNet152 model with Healthy or Viral Pneumonia Chest X-Ray image. We assess the performance of the model by a 10-time independent repeated test, and we compare the performance with SVM and LDA under different number of samples. Under the 10 repeated experiments, the accuracy obtained remains stable, with an Extreme Deviation of 1.22%. The comparation result demonstrates that ResNet152 can provide more valuable recognition judgments than SVM or LDA in this work. It also indicates that only use Grayscale histograms to calculate CXR image features is not particularly suitable for training SVM or LDA. The results verify the effectiveness, stability, and accuracy of ResNet152. It also indicates our dataset partition and preprocessing method is practicable and meaningful. According to our analysis, the model could gain a better performance with bigger CXR sample size. We believe that ResNet152 has great potential in real-life usage scenarios, and is beneficial to improve the accuracy of clinical COVID-19 diagnosis.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

1. Umakanthan S, Sahu P, Ranade AV, et al.: Origin, transmission, diagnosis and management of coronavirus disease 2019 (COVID-19). Postgrad Med J. 96(1142):753-758 (2020).
2. Bernard Stoecklin S, Rolland P, Silue Y, et al.: First cases of coronavirus disease 2019 (COVID-19) in France: surveillance, investigations and control measures. Euro Surveill. 25(6):2000094 (2020).
3. World Health Organization (WHO). Coronavirus disease (COVID-19). Retrieved from: https://www.who.int/health-topics/coronavirus
4. Lee JG, Jun S, Cho YW, et al. Deep Learning in Medical Imaging: General Overview. Korean J Radiol. 18(4):570-584 (2017).
5. Kassania SH, Kassanib PH, Wesolowskic MJ, et al. Automatic Detection of Coronavirus Disease (COVID-19) in X-ray and CT Images: A Machine Learning Based Approach. Biocybern Biomed Eng. 41(3):867-879 (2021).
6. Burlacu A, Crisan-Dabija R, Popa IV, et al. Curbing the AI-induced enthusiasm in diagnosing COVID-19 on chest X-Rays: the present and the near-future (2020).
7. Khalifa NM, Taha MH, Hassanien AE, et al. Detection of Coronavirus (COVID-19) Associated Pneumonia based on Generative Adversarial Networks and a Fine-Tuned Deep Transfer Learning Model using Chest X-ray Dataset (2020).
8. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016).
9. COVID-19 Chest X-Ray Dataset (Published by V7Labs). Retrieved from: https://darwin.v7labs.com/v7-labs/covid-19-chest-x-ray-dataset

10. Corinna C, Vladimir V. Support-vector networks. Machine Learning. 20(3):273–297. (1995).
11. Fisher RA. THE USE OF MULTIPLE MEASUREMENTS IN TAXONOMIC PROBLEMS. Annals of Eugenics, 7(2), 179–188 (1936).
12. Wei X, Li S, Zhu S, et al. Terahertz spectroscopy combined with data dimensionality reduction algorithms for quantitative analysis of protein content in soybeans. Spectrochim Acta A Mol Biomol Spectrosc. 253:119571 (2021).
13. Valueva M, Nagornov N, Lyakhov PA, et al. Application of the residue number system to reduce hardware costs of the convolutional neural network implementation. Mathematics and Computers in Simulation. 177 (2020).
14. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84–90 (2017).
15. Menache I, Mannor S, Shimkin N. Basis Function Adaptation in Temporal Difference Reinforcement Learning. Ann Oper Res 134, 215–238 (2005).
16. Goutte C, Gaussier E. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In Proceedings of the 27th European conference on Advances in Information Retrieval Research (ECIR'05). Springer-Verlag, Berlin, Heidelberg, 345–359 (2005).
17. Behera B, Kumaravelan G, Kumar.B P. Performance Evaluation of Deep Learning Algorithms in Biomedical Document Classification. 2019 11th International Conference on Advanced Computing (ICoAC) (2019).
18. He H, Garcia EA. Learning from Imbalanced Data. IEEE Transactions on Knowledge and Data Engineering, 21(9), 1263–1284 (2009).
19. Smeulders AWM, Worring M, Santini S, et al. Content-based image retrieval at the end of the early years. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(12), 1349–1380 (2000).
20. Alahmadi A, Hussain M, Aboalsamh H. LDA-CNN: Linear Discriminant Analysis Convolution Neural Network for Periocular Recognition in the Wild. Mathematics. 10(23):4604 (2022).
21. Jia D, Li Z, Zhang C. Detection of cervical cancer cells based on strong feature CNN-SVM network. Neurocomputing (2020).
22. Goodfellow I, Pouget-Abadie J, Mirza M, er al. Generative adversarial networks. Communications of the ACM, 63(11), 139–144 (2020).