2024 International Conference on Intelligent and Innovative Computing Applications (ICONIC)
7th-9th November 2024, Pearle Beach Resort & Spa, Flic en Flac, Mauritius
https://doi.org/10.59200/ICONIC.2024.004

Interpretable Protocol: A Novel Learning Strategy for COVID-19 Diagnosis on Chest-X-Ray Images

Yi Liao*, Weichuan Zhang*, Edwin Kwadwo Tenagyei*, Ugochukwu Ejike Akpudo*, and Yongsheng Gao*

*Integrated and Intelligent Systems, Griffith University, Australia

{*yi.liao2, edwinkwadwo.tenagyei, ugochukwu.akpudo*}*@griffithuni.edu.au*
{*weichuan.zhang, yongsheng.gao*}*@griffith.edu.au*

Abstract
Deep learning-based models (e.g., ResNet18 and ResNet50) have been employed for detecting COVID-19 on chest-X-ray (CXR) images which have been reported to achieve good accuracy. Although these models have the ability to correctly classify the CXR images from the dataset COVID-Xray-5k into COVID-19 and non-COVID-19 classes, our investigation shows for the first time that the good results obtained by existing methods may come from regions out of the lungs area on the images from this dataset. It is an unsolved problem that the regions used to make such decisions are automatically located in the lung area where the evidence of COVID-19 is to be found. To this end, this paper proposes an interpretable protocol for COVID-19 detection on CXR images. The proposed protocol not only improves the prediction performance but more importantly increases interpretability and trust of such prediction without the need of any region annotation. The proposed protocol is a learning strategy that can be applied to any convolutional neural networks (CNN) models. The experimental results demonstrate the superiority of the proposed strategy over the related state-of-the-arts.

Keywords: COVID-19, Deep Learning, Chest X-Ray, Trustable Learning Strategy

1. Introduction

The rare pandemic caused by novel corona-virus has resulted in millions of deaths and countless infectious patients. The critical step to control the spread of the virus is to detect all positive cases as early as possible. Reverse transaction polymerase chain reaction (RT-PCR) was considered as the main method for COVID-19 diagnosis (Wang et al., 2020). However, the accuracy of RT-PCR positive detection of throat swab samples is less than 60% (Yang et al., 2020). Chest radiographs from the cases of COVID-19 infections show airspace opacities, which refer to ground-glass (57%) and mixed attenuation (29%) (Kong & Agarwal, 2020). These abnormal airspace opacities are highly correlated with COVID-19 (Kanne et al., 2020). Therefore, the previous studies (Ahsan, Based, Haider, & Kowalski, 2021; Chen, Yao, Zhou, Dong, & Zhang, 2021; Hammoudi et al., 2021; Ke, Ellsworth, Banerjee, Ng, & Rajpurkar, 2021; Minaee et al., 2020; Ozturk et al., 2020; Rajpurkar, Joshi, Pareek, Ng, & Lungren, 2021; Shorfuzzaman & Hossain, 2020) tried to employ deep learning techniques for automatic recognition of COVID-19 from chest-X-ray (CXR) images and many works (Minaee et al., 2020) obtain the prominent performance (e.g., a recall of 98% for ResNet50).

Table 1. The Number and Proportion of Training Images with Interference and without Interference

No Interference	Interference			
	Arrows	Electrodes	Pace makers	Letters
13	10	34	12	58
15.48%	11.90%	40.48%	14.29%	69.05%

It is worth noting that CXR images are from are from real patients in hospitals. Medical equipment, such as pacemakers, electrodes, and cardiogram monitors, are essential to healing patients. The letters can indicate the photographing time and patients' gesture so they are necessary for the clinical diagnosis. Hence, it is reasonable that pacemakers, electrodes and letters exist on CXR images. For example, the CXR images obtained from the public CheXpert dataset (Irvin et al., 2019) contain letters, electrodes and pacemakers. In the published dataset COVID-Xray-5k (Minaee et al., 2020), electrodes, pacemakers and letters are unevenly distributed on 84 COVID-19 training samples. However, letters, electrodes and pacemakers are not the intrinsic features of any disease. They should be viewed as interferences when deep learning-based models are trained using data containing this interference information. For deep learning-based models in Deep-

Table 2. The Proportions of Test Images About Features to Support COVID-19 Using Deep-COVID

Features	ResNet18	ResNet50
Liver	18.9%	5.20%
Heart	12.76%	26.04%
Stomach	5.31%	11.45%
Neck and Low Jawbone	4.25%	3.12%
Acromioclavicular Joint and Axilla	20.21%	30.20%
Mediastinum	9.57%	5.20%
Arrows and Electrodes	5.31%	2.08%
Features irrelevant to the lung	75.53%	83.33%

COVID (Minaee et al., 2020), there are only 13 images out of 84 training images (15.48%) without any interference (see Table 1.). 71 images out of 84 training images (84.52%) with interferences are used for training the models. Because the models are trained on a large number of images with interferences, interferences are likely to be used for predicting COVID-19. For instance, the electrodes are likely to be determined as unique features on COVID-19 images. In this way, images containing electrodes may be diagnosed with COVID-19 positive cases during testing. Furthermore, there are many organs on CXR images such as the liver, the heart, and the stomach. The liver or stomach is possible to be viewed as the support for COVID-19 detection.

To verify the above possibilities, we follow the methods introduced in (Minaee et al., 2020) and then get a recall of 94% for ResNet18 and a recall of 96% for Resne50. The recall of 94% means 94 images out of 100 COVID-19 test samples are predicted as positive cases by using ResNet18; The recall of 96% means 96 images out of 100 COVID-19 test samples are predicted as positive cases by using ResNet50. We utilises CAM (Zhou et al., 2015) to visualize which regions to support COVID-19 diagnosis on 94 positive cases and 96 positive cases respectively. As shown in Table 2, 75.53% of 94 positive cases in ResNet18 indicate that the regions supporting COVID-19 are not relevant to lungs and 83.33% of 96 positive cases in ResNet50 indicate that the regions supporting COVID-19 are out of the range of lungs. As pneumonia caused by novel corona-virus is a lungs disease, the nidus (e.g., ground-glass and mixed attenuation) should appear in the area of lungs. Hence, both deep learning-based models (i.e., ResNet18 and ResNet50) do utilize irrelevant medical information for predicting COVID-19 which may lead to the wrong prediction.

The models should utilize the medical information from the lungs for predicting COVID-19. By now, it is unsolved that regions that are highly related to the lungs are identified automatically without region annotation from bounding boxes or segmentation masks. To this end, this paper proposes an interpretable protocol. The proposed protocol is a two-stage learning strategy. Firstly, the area highly associated with the lungs are automatically exposed to the maximum extent. Secondly, a CNN-based model is designed to learn the representative information from the area exposed in the first stage. Our main contributions can be summarized as follows:

- The proposed protocol can be widely adopted by any CNN based models for lungs disease diagnosis from CXR images.
- The proposed strategy induces models to discover the area of lungs on CXR images without extra help from bounding boxes and segmentation masks.
- The experiment results illustrate that our method can obtain not only a better performance than (Minaee et al., 2020) but a more trustworthy result.

2. Related Work

Many studies use AI models to detect COVID-19 by using CXR images. Wang and Wong (Wang et al., 2020) propose COVID-Net to diagnose COVID-19 from CXR images. Apostolopoulos and Mpesiana (Apostolopoulos and Mpesiana, 2020) use transfer learning algorithm to detect coronavirus infection from CXR images. Hossain (Hossain et al. 2020) proposes a healthcare framework that can interpret the diagnosis of COVID19 by extracting discriminative features relevant to COVID19. Abbas et al. (Abbas et al. 2021) use the pretrained CNN models such as ResNet (He et al., 2016) for diagnosis of COVID-19 from CXR images. Deep-COVID (Minaee et al., 2020) predicts COVID-19 from CXR images using transfer learning algorithm and generate heatmaps of lung regions potentially infected by coronavirus by using the perturbation method.

3. Our Method

The proposed strategy contains two main stages. The first stage is regions selection. The second stage is regions guided training. The two stages will be introduced in detail as follows.

3.1 Regions Selection

The ground truth labels have been divided into two categories (COVID-19 and non-COVID-19) in COVID-Xray-5k (Minaee et al., 2020). It is worth noting that both categories of CXR images use posteroanterior (PA) chest view. In this way, an image obtained from the non-COVID-19 group can be compared with an image from the COVID-19 group under the same local regions. Take the liver as an example, the liver is located at the left bottom corner of a PA chest image. The comparison of the left bottom

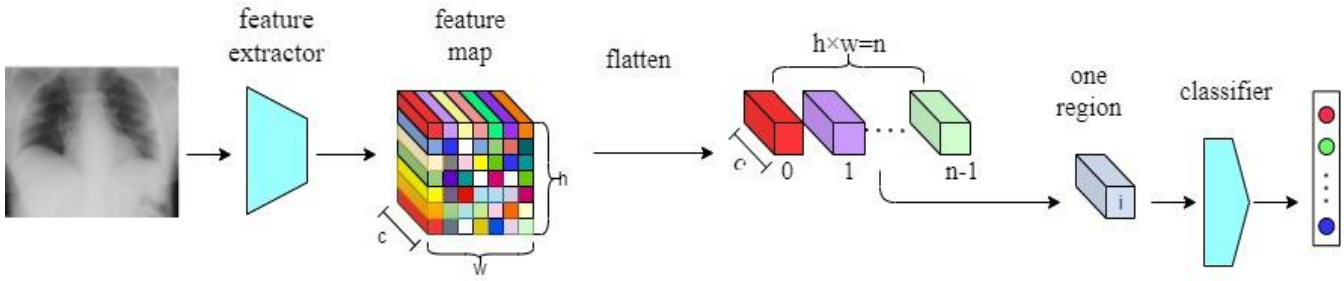

Figure 1. Overview of Regions Classification

corner regions between COVID-19 samples and no-finding subclass samples can reveal the changes in the liver.

Since the abnormalities (i.e., the ground-glass and mixed attenuation) only occur in the lungs, there should be no significant difference in the comparison of the regions representing the liver, the stomach, and the shoulders. In other words, if a comparison at the local regional level shows a significant change, it means that the regions participating in the comparison are highly likely to be located at the lungs. As all the chest radiographs are obtained from real patients in hospitals, the interferences (i.e., pacemakers, electrodes and letters) appear at relatively fixed locations on PA chest images from both categories because of the requirements of diagnosis and treatment. It is worth noting that the comparison can decrease the probability that the regional changes caused by the above interference happen.

In this paper, I represents an input image, f denotes a feature extractor, $f(I)$ represents the corresponding feature map, R is an arbitrary real number. We flatten a feature map $f(I) \in R^{c\times h\times w}$, c is the number of channels, h and w are height and width respectively. We can get n $(n = h \times w)$ vectors $V \in R^c$. A feature map $f(I)$ is flattened by the fixed orders. The horizontal order is from the left side to the right side in a row while the vertical order is from the top row to the bottom row. So, each vector has a natural label i that is used to represent the corresponding region $i \in [0, n-1]$. V_i is used to represent the vector from the region i. V_i^{covid} denotes the vector from COVID-19 images; V_i^{no} represents the vector from non-COVID-19 samples. region_i^{no} and region_i^{covid} are used to represent region i from non-COVID-19 images and COVID-19 images respectively.

A full-connected layer is employed as a classifier S, and then the classifier S , as shown in Figure.1, is trained with the help of loss function as follows,

$$\text{loss}_{\text{region } i} = \text{crossentropy}\left((V_i^{covid} \odot S), i\right) \quad , \qquad (1)$$

where classifier S is a matrix, V_i^{covid} is a vector, $\odot$ denotes matrix multiplication.

$$\text{loss} = \sum_{i=0}^{n-1} \text{loss}_{\text{region i}} \quad . \qquad (2)$$

In this way, a trained classifier S' is obtained. The trained classifier S' is actually a matrix with n rows and c columns. $S'[i]$ represents the vector of row i $(S'[i] \in R^c)$. The aim of training is to make the loss defined in (2) as small as possible. When $loss_{\text{region i}}$ is close to zero, it means inner product $< S'[i], V_i^{covid} >$ is much larger than inner product $< S'[i], V_j^{covid} > (j \neq i, j = 0,1..n-1)$. In other words, it means $S'[i]$ is close to V_i^{covid} and far away from other region vectors $V_j^{covid} (j \neq i, j = 1,2,...n-1)$. The classification score vector V_{score} can be aculated by $V_{score} = V_i^{no} \odot S'$, where $\odot$ denotes matrix multiplication operation. The score vector V_{score} consists of n scalars. Each scalar is denoted with $h_i (i = 1,2,...n-1)$. So $V_{score} = [h_0, h_2, \dots, h_{n-1}]$. We define function $h_i = H(i), i = 0,1,...n-1$.The classification mechanism of CNN based Neural Network is to look for the largest scalar from a score vector and the location index of the score vector is seen as the predicted label. Therefore, the predicted label i' for region i form non-COVID19 images is calculated as

$$i' = \text{argmax}\big(H(i)\big), H(i) \in \{h_i | i = 0,1,2...n-1\} \quad (3)$$

where i is ground truth label, i' represents predicted label. <,> represents inner product. If $i' = i$, it means $< V_i^{no}, S'[i] >$ is much larger than $< V_i^{no}, S'[j] > (j \neq i, j = 1,2,...n-1)$. In other words, V_i^{no} is closer to $S'[i]$. Because $S'[i]$ is close to V_i^{covid} accoring to the above analysis, V_i^{no} is close to V_i^{covid} and far away from $V_j^{covid} (j \neq i, j = 1,2,...n-1)$. V_i^{covid} and V_i^{no} have the similar digital distribution in c dimensional vector space. Because V_i^{covid} and V_i^{no} are both local regions feature descriptor, they are from the same region i.Because both categories of CXR images use PA chest view, the same regions from different images reflect the anatomy information from the same part of human boy. Therefore, the nidus should not be exist in region i. If $i' \neq i$, $<V_i^{no}, S'[i]>$ is not larger than $<V_i^{no}, S'[j]> (\neq i, j = 1,2,...n-1)$. V_i^{no} is far way from $S'[i]$. So, V_i^{no} is also far way from V_i^{covid} because $S'[i]$ is close to V_i^{covid}. It shows that although V_i^{covid} and V_i^{no} are from the same region i, they do not have the similar digital

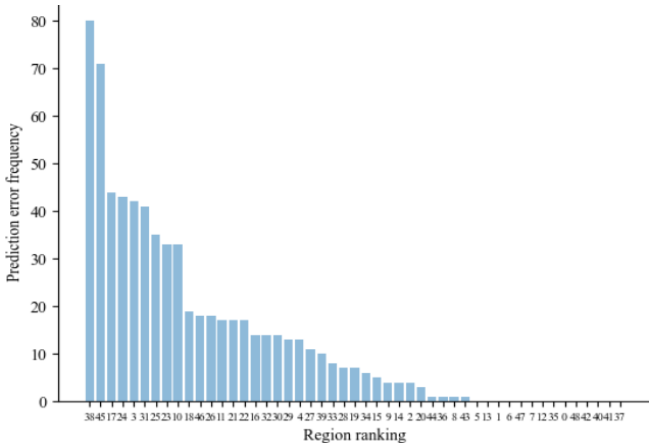

Figure 2. Frequency distribution for the prediction error across all regions.

distribution. The dissimilarity between V_i^{covid} and V_i^{no} is likely to be attributed to the nidus in the region i from COVID19 images. It also means that the region i is associated with lungs probably. Our method in the first stage aims at discovering the regions irrelevant to nidus for all CXR images. Testing single one image cannot make sure that it is impossible that the nidus appears at region i for all CXR images in dataset. It is possible that region i from one image contains the nidus while region i from another image doesn't contain the nidus. On the other hand, scale difference or translational difference existing in both categories of images may make prediction error ($i' \neq i$) happen. Therefore, it is necessary to test all non-COVID19 images from training data. K denotes the number of all non-COVID19 images from training data. For region i ($i = 0,1, \dots n-1$), $V_i^{no^k}(k = 1,2..K)$ represents local descriptor of region i from k-th non-COVID19 image. The following Equation is used to compute FREQUENCY[i]:

$$\text{FREQUENCY}\,[i] = \frac{\sum_{k=1}^{\text{K}} t_k^i}{\text{K}},\ t_k^i = \begin{cases} 1, i' \neq i \\ 0, i' = i \end{cases}$$

From the aspect of statistics, nidus occurrence can be seen as a stochastic event with a high probability. The frequencies of prediction error across all regions are calculated and sorted in descending order of the frequency of prediction error in every region. Thus, a chat of frequency distribution is shown in Figure 2. The confidence interval is set to $1-\alpha$ (a is the significance level). Finally, m regions with the highest frequency values are calculated as the following Algorithm 2.

Algorithm2: Compute m

function(frequency, α, n)
  for m=1→n
    if $1-\alpha \leq \frac{\sum_{j=0}^{m-1} \text{frequency}_\text{j}}{\sum_{j=0}^{n-1} \text{frequency}_\text{j}}$ then return m
  return -1

note: $\text{frequency}_\text{j} > \text{frequency}_{\text{j}+1} (j = 0,1,2, n-2)$. j represents a ranking number rather than a region number.

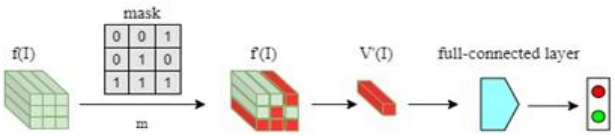

Figure 3. Overview of regions guided training.

3.2 Regions Guided Training

In the first stage, m regions are obtained by comparison between two categories of images only from training data. In the second stage, both two categories of images are used for training a new model, m regions will be utilized to make a one-hot mask. The mask will be applied to the feature map extracted from an image from the training data. It will play a role of filters in removing all regions irrelevant to nidus on the feature map. Actually it will be embedded into the model. In testing phase, it will impact the feature map extracted by model from a testing image. It is worth noticing that the mask is derived from training data rather than testing data. It will work on each image from both training data and testing data. On the other hand, once m regions have been selected in the first stage, the regions associated with nidus do not vary for each image from training set. In other words, the mask will be formed by using m regions. It will be applied to each image equally. The one -hot mask is constructed by using m regions. Firstly, m regions are denoted by $\beta = \{\text{frequency}_j | j = 0,1,2, \dots, m-1\}$, the one-hot mask is a matrix, which is denoted by $\text{M} \in N^{h \times w}$, where N denote a natural number, h *w=n and h=w. Naturally, h=w=$\sqrt{n}$. Given FREQUENCY and index is region number, M can be constructed by the following Algorithm3.

Algorithm3: Construct M

for x=0→w-1
  for y=0→ y-1
    if FREQUENCY$[x * w + y] \in \beta$ then
      $\text{M}(x, y) = 1$
    else $\text{M}(x, y) = 0$

A new backbone network is employed as the feature extractor f, $f_I(i,j)$ represents the corresponding local feature vector at i row and j column in feature map $f(I)$ from image I. The new feature map $f'(I)$ can be obtained by calculating the elementwise product (Hadamard product) between $f(I)$ and matrix M as follows,

$$f_I'(i,j) = f_I(i,j) \times \text{M}(i,j). \quad (4)$$

The Hadamard Product between $f(I)$ and matrix M can be defined as

$$f'(I) = f(I) \circ \mathrm{M} \quad (5)$$

Thus, $f'(I)$ is the region guided feature map that only utilizes the feature information relevant to pneumonia.

Only m regions participate in the process of final classification, so the final feature vector $V'(I)$ should be formed as

$$V'(I) = \frac{\sum_{i=0}^{h-1}\sum_{j=0}^{w-1} f_I'(i,j)}{m}. \quad (6)$$

Finally, $V'(I)$ is fed into a full-connected layer to generate the prediction score $\mathrm{pred} \in R^2$. A cross-entropy lost function is employed for training under the supervision of ground-truth label truth. The loss is defined as

$$\mathrm{loss} = \mathrm{crossentropy}(\mathrm{pred}, \mathrm{truth}), \quad (7)$$

which is minimized during training by using a stochastic gradient descent algorithm. Fig.3 shows a sample that illustrates the whole process of regions guided training. The matrix M plays a role of filter that removes the irrelevant regions from the feature map extracted from the original image I in the training process. The irrelevant regions should be excluded from engaging in prediction. Therefore, the mask should be employed as a filter in the testing process so that the predicted results are trustworthy.

4. Experiment

4.1 Dataset and Data Augmentation

The published dataset COVID-Xray-5k (Minaee et al., 2020) consists of the COVID-19 group and the non-COVID-19 group. There are 5,184 PA chest X-ray images in this dataset. The COVID-19 group contains 184 images from COVID Chestxray Dataset (Cohen et al., 2020). 84 COVID-19 images are used for training and 100 images are used for testing. The non-COVID-19 group collects 5,000 images including 14 subclasses from the public ChexPert dataset (Irvin et al., 2019). 2,000 samples are used for training and 3,000 images are prepared for testing. The no-finding subclass is one of 14 subclasses. 2,000 images for training include 700 images from the no-finding subclass and 100 images from each remaining 13 subclasses. 3,000 images for the test are composed of 1,700 images from the no-finding subclass and 100 images from each remaining 13 subclasses.

The COVID-19 training set has much fewer images (84 images) than the non-COVID-19 training set (2,000 images), resulting in the problem of data imbalance. The model trained on an imbalance dataset will result in a lower recall or precision in the category with fewer training samples. To balance the data, data augmentations techniques are employed to make equal numbers of two categories of images. The new images are created with the help of random transformed operations (e.g., brightness changing, contrast changing, and erasing). It worth mentioning that the created images are used only in the second stage (regions guided training). In the stage of regions selection, 84 original COVID-19 samples are used to train the classifier S. For the sake of the high frequency of prediction errors, 700 training samples from the no-finding subclass are chosen as the test set for the first stage. If we choose 84 COVID-19 images as the test set, the amount of test samples is too few to exhibit statistical significance.

Table 1 The Comparison of Results of ResNet18 for Cut-off Threshold 0.2

Metrics	ResNet18 (Deep-Covid)	ResNet18(ours)	
		α=0.05	***α=0.1***
Recall (Sensitivity)	94.00%	98.00%	97.00%
Specificity	97.10%	100%	99.90%
Precision	51.93%	100%	97.00%
Accuracy	97.00%	99.93%	99.80%
Average precision	90.94%	99.93%	99.89%
AUC	99.55%	99%	98.45%

4.2 Hyperparameters and Experiment Environment

In images pre-processing, all images are resized into 224×224 and normalized regularly before being fed into the neural network. No crop tools are adopted for all images. ResNet18 and ResNet50 are employed as the feature extractors in both stages after the last average pooling layer and the last full-connected layer are removed. In the stage of regions selection, the number of all regions is 49 (n=49) because the height and width of a feature map are 7. The batch size is set to 23 and the base learning rate is 0.0002 for training in the first stage. Adam is chosen as the optimizer and the learning rate is updated by multiplying a factor of 0.1 every 10 steps. For training, the training set is chosen for validation and the validation accuracy is computed in each epoch. The criteria to stop training is the validation accuracy is 100% and the loss value defined in Equation (2) is smaller than 0.01. The criteria can guarantee the model has been trained well. Thus, the number of epochs is dynamic. For testing, we set confidence interval to 0.95 and 0.90 (a= 0.05 and 0.1 respectively). In the stage of regions guided training, the batch size is set to 20 and the base learning rate is 0.0001. SGD is the primary optimizer. The learning rate is updated every 10 steps by multiplying a factor of 0.1. The momentum is set to 0.9. The total number of epochs is 100. It is worth mentioning that non-COVID-19 images from the test set include both normal cases (no-

Table 2. The Comparison of Results of ResNet50 for Cut-off Threshold 0.2

Metrics	**ResNet50 (Deep-Covid)**	**ResNet50(ours)**	
		α=0.05	***α=0.1***
Recall (Sensitivity)	96.00%	99.00%	100%
Specificity	97.10%	100%	99.93%
Precision	52.75%	98.02%	98.04%
Accuracy	97.10%	99.90%	99.94%
Average precision	92.97%	99.96%	99.99%
AUC	96.57%	99.47%	99.97%

finding subclass) as well as other types of diseases. The non-COVID-19 images with other disease type make it much more difficult to distinguish from COVID-19 samples than normal samples, a cut-off threshold to the prediction probability score is used to deal with this difficulty (Minaee et al., 2020). It means if the prediction probability score for a test image is larger than the cut-off threshold, the test image is predicted as COVID-19; otherwise, it is considered as non-COVID-19. In this work, the cut-off threshold is used as well, and it is set to 0.2. The implementation is done in Pytorch 1.8 and Python 3.6. The Python library used for data augmentations is Augmentor 0.2.8. The experiment is executed on a server providing 4 NVIDIA RTX 2080 TI GPU with 128 G memory and an Intel i9 CPU.

Table 3. The Proportion of Test Images about Features to Support COVID-19 Using Our Method.

Features	**ResNet18 (α=0.1)**	**ResNet50 (α=0.1)**
Liver	0	0
Heart	35.42%	32.29%
Stomach	4.17%	2.08%
Neck and Low Jawbone	3.13%	3.13%
Acromioclavicular Joint and Axilla	0	3.13%
Mediastinum	3.13%	11.46%
Arrows and Electrodes	0	0
Features irrelevant to the lungs	45.85%	52.09%

4.3 Experimental Results

The proposed algorithm is compared with the Deep-COVID method introduced in (Minaee et al., 2020) by using Resnet18 and Resnet50. The comparisons about the results yielded by the following two approaches are exhibited in Table 3 and Table 4. The metrics used for comparison include recall (sensitivity), specificity, precision, accuracy, average precision, and Area Under the Curve (AUC). According to the results of the experiments, the proposed approach is much better than the Deep-COVID method. The results shows the proposed method can make the models focus on learning the feature information highly associated with nidus in the lungs and exclude as much irrelevant information as possible. Figure 4. visualize which regions are used for COVID-19 detection with CAM by using both methods. Table 5 shows the proportion of test images about features from a feature map. This issue is expected to be solved in the future works.

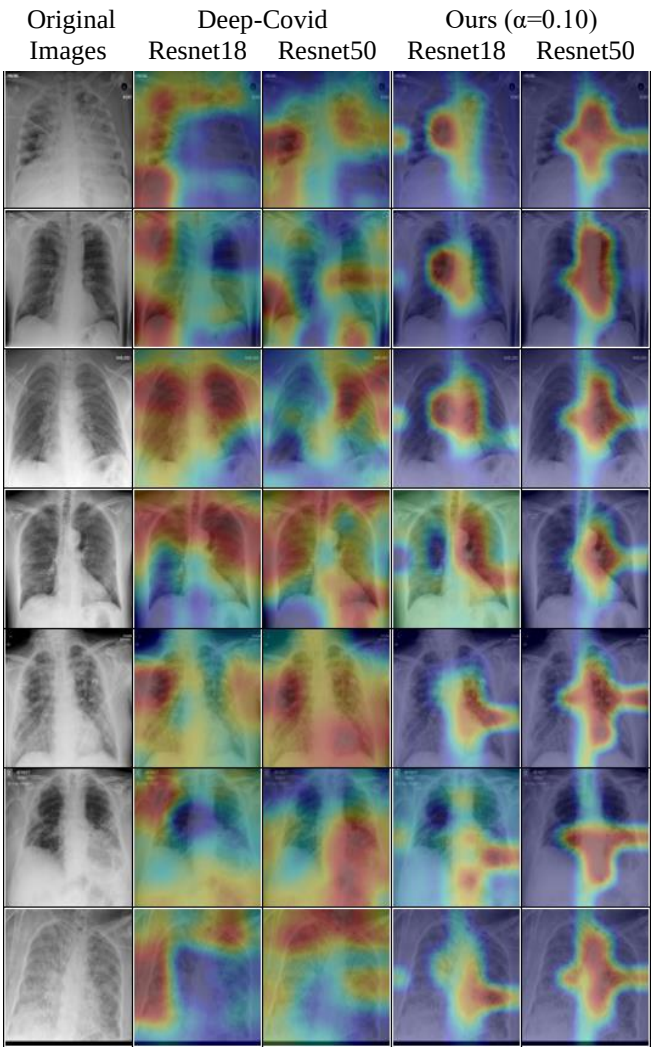

Figure 4. The comparison of heatmap between Deep-Covid method and our method.

5. Conclusion

In this work, we propose an interpretable protocol to assist CNN-based models to learn the trustable discriminative feature information for COVID-19 diagnosis from CXR images. The proposed protocol is a novel learning strategy by which ResNet models will be able to discover the regions highly associated with lungs automatically without any extra help (such as bounding boxes or segmentation masks). Our method

can help ResNet18 and ResNet50 networks to obtain a superior performance despite being trained on the CXR images with interference. Furthermore, the results are more trustworthy than the related state-of-the-arts, as the visualization with CAM illustrates.

5. References

Abbas, A., Abdelsamea, M. M., & Gaber, M. M. (2021). Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. *Applied Intelligence*, *51*(2), 854-864. https://doi.org/10.1007/s10489-020-01829-7

Apostolopoulos, I. D., & Mpesiana, T. A. (2020). Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. *Physical and Engineering Sciences in Medicine*, *43*(2), 635-640. https://doi.org/10.1007/s13246-020-00865-4

Ahsan, M., Based, M. A., Haider, J., & Kowalski, M. (2021). COVID-19 detection from chest X-ray images using feature fusion and deep learning. *Sensors*, *21*(4), 1480.

Chen, X., Yao, L., Zhou, T., Dong, J., & Zhang, Y. (2021). Momentum contrastive learning for few-shot COVID-19 diagnosis from chest CT images. *Pattern Recognition*, *113*, 107826.

Cohen, J. P., Morrison, P., Dao, L., Roth, K., Duong, T. Q., & Ghassemi, M. (2020). COVID-19 Image Data Collection: Prospective Predictions Are the Future. *ArXiv*, *abs/2006.11988*.

Irvin, J. A., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., Marklund, H., Haghgoo, B., Ball, R. L., Shpanskaya, K. S., Seekins, J., Mong, D. A., Halabi, S. S., Sandberg, J. K., Jones, R., Larson, D. B., Langlotz, C., Patel, B. N., Lungren, M. P., & Ng, A. (2019). CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison. *AAAI Conference on Artificial Intelligence*.

Hammoudi, K., Benhabiles, H., Melkemi, M., Dornaika, F., Arganda-Carreras, I., Collard, D., & Scherpereel, A. (2021). Deep learning on chest X-ray images to detect and evaluate pneumonia cases at the era of COVID-19. *Journal of Medical Systems*, *45*(7), 75.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Hossain, M. S., Muhammad, G., & Guizani, N. (2020). Explainable AI and mass surveillance system-based healthcare framework to combat COVID-I9 like pandemics. *IEEE Network*, *34*(4), 126-132. https://doi.org/10.1109/MNET.011.2000458

Kanne, J. P., Little, B. P., Chung, J. H., Elicker, B. M., & Ketai, L. H. (2020). Essentials for Radiologists on COVID-19: An Update—Radiology Scientific Expert Panel. *Radiology*.

Ke, A., Ellsworth, W., Banerjee, O., Ng, A. Y., & Rajpurkar, P. (2021). CheXtransfer: performance and parameter efficiency of ImageNet models for chest X-ray interpretation. *Conference on Health, Inference, and Learning*.

Kong, W., & Agarwal, P. P. (2020). Chest Imaging Appearance of COVID-19 Infection. *Radiology: Cardiothoracic Imaging*, *2*.

Minaee, S., Kafieh, R., Sonka, M., Yazdani, S., & Soufi, G. J. (2020). Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning. *Medical Image Analysis*, *65*, 101794–101794.

Ozturk, T., Talo, M., Yildirim, E. A., Baloglu, U. B., Yildirim, O., & Acharya, U. R. (2020). Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Computers in biology and medicine*, *121*, 103792.

Rajpurkar, P., Joshi, A., Pareek, A., Ng, A. Y., & Lungren, M. P. (2021). CheXternal: Generalization of deep learning models for chest X-ray interpretation to photos of chest X-rays and external clinical settings. *Conference on Health, Inference, and Learning*.

Shorfuzzaman, M., & Hossain, M. (2020). MetaCOVID: A Siamese neural network framework with contrastive loss for n-shot diagnosis of COVID-19 patients. *Pattern Recognition*, *113*, 107700–107700.

Wang, W., Xu, Y., Gao, R., Lu, R., Han, K., Wu, G., & Tan, W. (2020). Detection of SARS-CoV-2 in Different Types of Clinical Specimens. *JAMA*.

Wang, L., Lin, Z. Q., & Wong, A. (2020). COVID-Net: a tailored deep convolutional neural network

design for detection of COVID-19 cases from chest X-ray images. *Scientific Reports, 10*(1), 19549. https://doi.org/10.1038/s41598-020-76550-z

Yang, Y., Yang, M., Yuan, J., Wang, F., Wang, Z., Li, J., Zhang, M., Xing, L., Wei, J., Peng, L., Wong, G., Zheng, H., Wu, W., Shen, C., Liao, M., Feng, K., Li, J., Yang, Q., Zhao, J., &Liu, Y. (2020). Laboratory Diagnosis and Monitoring the Viral Shedding of SARS-CoV-2 Infection. *The Innovation, 1*.

Zhou, B., Khosla, A., Lapedriza, À., Oliva, A., & Torralba, A. (2015). Learning Deep Features for Discriminative Localization. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2921–2929.