DeCOVID-CXR: Deep Learning Model for Detection of COVID-19 Infection using Chest X-Rays

Jin Tian Ci Ngiu Lai Kuan Wong

Multimedia University, Cyberjaya, Malaysia

Abstract

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is commonly called COVID-19, its pandemic outbreak occurred in December 2019 and it has brought critical impacts to the world. While the Reverse-Transcription Polymerase Chain Reaction (RT-PCR) is the standard method to detect COVID-19 infections, Chest X-Ray (CXR) serves as an alternative way that helps in the detection and allows the clinicians to diagnose the patients with the symptoms in the lungs. In this project, we have implemented two frameworks to detect COVID-19 infections based on the CXR images. The models - ResNet50 and Vision Transformer (ViT) are proposed as the based models in the frameworks and have adopted some pre-processing methods to improve the performance. The models can predict three classes which are normal, pneumonia and COVID-19. Visualization methods are applied to introduce interpretability on the important features of an image. To our knowledge, we are the first that performs an experiment on the RSNA International COVID-19 Open Radiology Database (RICORD) dataset which increases the practicality of our frameworks. We used our approaches to perform clinical evaluations using the data that comes from the University of Malaya Medical Centre clinician site.

Keywords—COVID-19, deep learning, CXR images, interpretability, RICORD dataset, clinical evaluation

I. INTRODUCTION

Coronaviruses have been around the world before the outbreak of Novel Coronavirus (COVID-19). They are known as a virus group that compromises RNA and will cause respiratory tract infections in humans and animals. As coronaviruses have a different range in level in severity, some only result in common symptoms like cold, while some cases are fatal such as Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS).

Novel Coronavirus (COVID-19) as suggested by the name, is one of the members of the coronavirus's family. One opinion about its origin is saying that the virus came from the bat as the genetic similarity with bat coronavirus is close, but there is no evidence in proving those hypotheses. COVID-19 first appeared in late December 2019 in Wuhan, Hubei Province, China with a small cluster of patients diagnosed with severe acute respiratory syndrome and it has then become unstoppably by spreading all over the world at a devastating speed. According to the statistics of the World Health Organization (WHO), COVID-19 has brought a tremendous impact to the world and resulted in over 113 million confirmed cases including more than 2 million deaths as of 2nd March 2021.

With the information given by WHO, patients who are infected by COVID-19 will experience some common symptoms which include fever, cough, fatigue, headache, diarrhea and dyspnea. However, the virus is considered lethal since it may lead to death that caused by the complications witnessed like acute lung injury, ARDS, shock and acute kidney injury. Besides, it has been found that COVID-19 can spread through human to human transmission easily. Small droplets containing the virus can come from the infected patients when they speak, sneeze or cough, the virus is then having the chance of spreading to those who contact. Thus, early detection of COVID-19 plays a key role so that the relevant authorities could take action immediately to flatten the epidemic curve.

II. PROBLEM STATEMENT

Various methods have been applied for COVID-19 diagnosis and the current gold standard is known to be Reverse-Transcription Polymerase Chain Reaction (RT-PCR). Among other detection techniques, radiography examination seems to be an alternative way in diagnosing COVID-19. Common characteristics of COVID-19 could be found on radiographic images such as Chest X-Ray (CXR) and Computed Tomography Scan (CT Scan) [1, 2]. Therefore, these radiography images could be utilized and analyzed by radiologists to screen out COVID-19 cases. Fig. 1 has shown some examples of the findings on CXR images of COVID-19 patients.

Notably, CXR images can be obtained using portable/mobile X-ray devices, which can reduce cross-infection and disruption of radiology service. Despite some detection methods are having higher sensitivity, CXR is worth to be considered as an adjunct since it is cheaper and costs a shorter time than some of the methods in obtaining the result. Besides, CXR images also allow the clinicians to investigate the patients' lung for further diagnosis. From the review in [3] and [4], AI models have shown promising results in detecting COVID-19 infection using CXR images but the majority of the models are trained on a combination of different data sources or a public dataset called COVIDx. However, most of the datasets are not from an official organization public source, hence limiting its generalization ability and practicality.

III. RELATED WORKS

Various studies have been done to come out with an approach in detecting COVID-19 infection where some utilize existing deep learning models as a base while the others implemented their network architectures from the scratch. Khobahi et. al. (2020) [5] proposed CoroNet which applied Auto Encoders on feature extractions, followed by a ResNet-18 based CNN to perform classification on COVID-19 infections. Lv D et.

al. (2020) [6] have implemented Cascade-SEMEnet which consisted of SEME-ResNet50 and SEME-DenseNet169 to

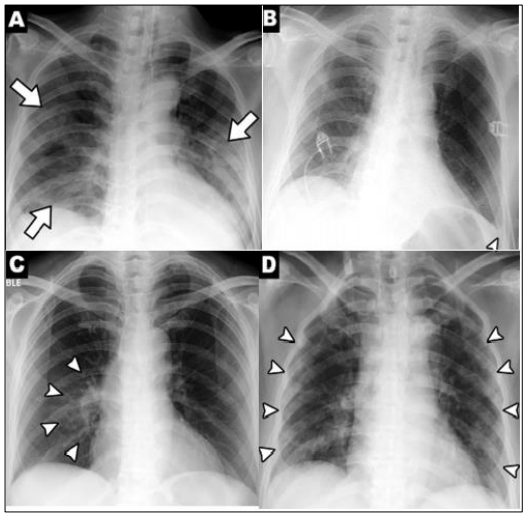

Fig. 1. Findings on chest radiographs in patients with COVID-19, (A): Patchy consolidations, (B): Pleural effusion, (C): Perihilar distribution, (D): Peripheral distribution

classify binary class and multi-class COVID-19 infections. The authors adopted Squeeze-Excitation (SE) structure and MoEx to enhance the performance of their model. A deep learning model, called COVID-NET proposed by Wang L. & Wong A. (2020) [7] made use of lightweight PEPX design patterns, selective long-range connectivity and high architectural diversity to achieve better performance while maintaining lower computational complexity. Abbas et. al. (2020) [8] investigated the medical data irregularity problem by building a deep CNN model called DeTrac which adopted class decomposition technique with transfer learning on the existing model, as class decomposition technique facilitating more flexibility to classifier's decision boundaries to deal with data irregularity.

Furthermore, CovXNet uses a multi-dilation CNN for detecting COVID-19 infection. The authors employed a stacking algorithm to optimize the prediction outputs and multi-dilation allowed the model to focus on the generalized features by broadening receptive area (Mahmud et. al., 2020) [9]. Siddhartha et. al. (2020) [10] proposed COVIDLite that was based on depth-wise separable deep neural networks. The proposed model was significantly lighter compared to other methods as it contained much lower parameters due to the application of depth-wise convolution. Hasan et. al. (2020) [11] developed CVR-Net as a COVID-19 screening method. The model leveraged multi-encoders to extract the image features from different scales and the encoders can compensate each other while one fails. An approach towards the detection of COVID-19 with CXR images was proposed as CovMUNET by Sayyed et. al. (2020) [12]. Their model was based on a modified version of U-Net and combining multiple loss functions helped to optimize the performance.

IV. Proposed Framework

The proposed frameworks are illustrated in Fig. 2 and we will discuss our approaches step by step in this section. The second framework is different from the first framework by applying the class decomposition & composition technique. The following subsections include Contras Limited Adaptive Histogram Equalization (CLAHE), deep features extraction, class decomposition & composition, transfer learning model and visualization methods.

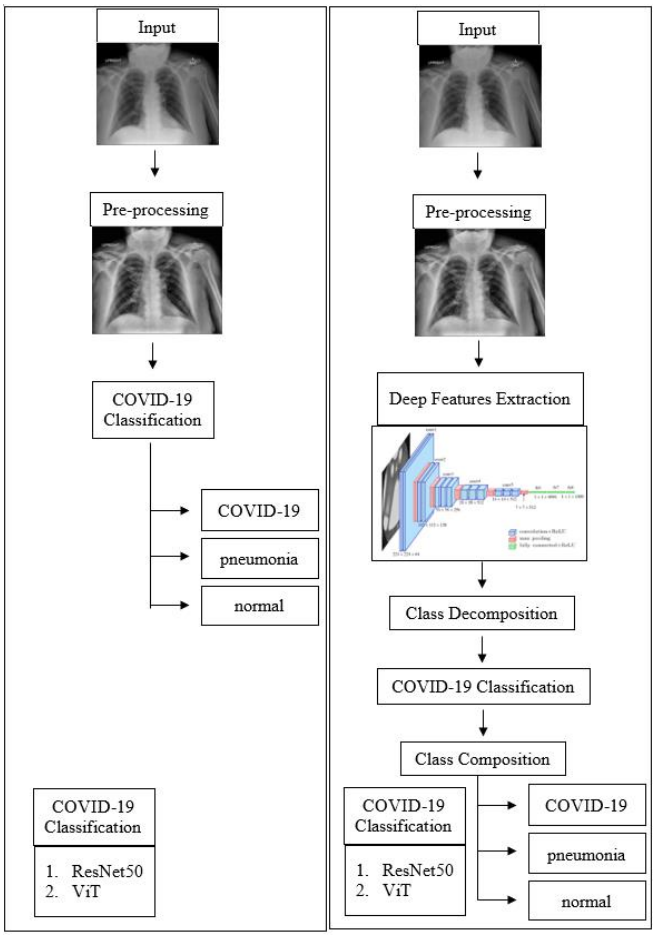

Fig. 2. Proposed Frameworks

A. Contrast Limited Adaptive Histogram Equalization (CLAHE)

CXR images may have different intensity levels and could cause a model to result in poorer performance if the images are under-exposed or over-exposed. To solve the problem, histogram equalization (HE) is one common method to enhance the images by stretching the intensity value. However, using HE has some disadvantages as it could lead to loss of details and also introduces undesirable noises. CLAHE is similarly to HE, but it can overcome the issues mentioned above by limiting the contrast amplitude within a specific range [6, 10]. The formula of CLAHE can be written as:

$$p = (p_{max} - p_{min}) * P(f) + p_{min} \quad (1)$$

where the $\boldsymbol{p}$ is the pixel value after applying CLAHE, $P(f)$ represents cumulative probability distribution function, p_{max} and p_{min} represent the pixel value of an image concerning maximum and minimum value respectively. An example is illustrated in Fig.3.

B. Deep Features Extracton

The authors of [8] have utilized pre-trained AlexNet to extract the discriminative features of three classes by adopting the last fully connected layer to initialize the weights for the three classes classification. In our case, we are

using the VGG16 model with pre-trained weights to extract the deep features from the images and the model architecture is showed as Fig. 4. We have modified the last layer to perform transfer learning on our dataset. After carrying out the training, we removed the output layer to extract features only from our CXR images. The features extracted will be used to perform class decomposition later.

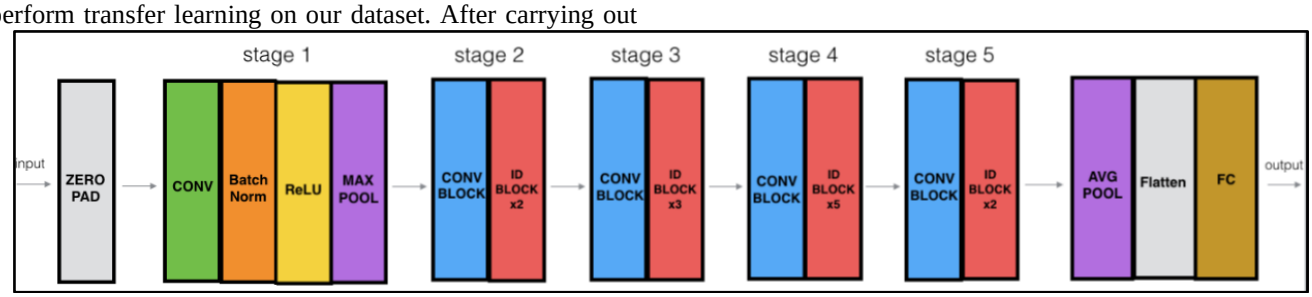

Fig 5. ResNet Architecture

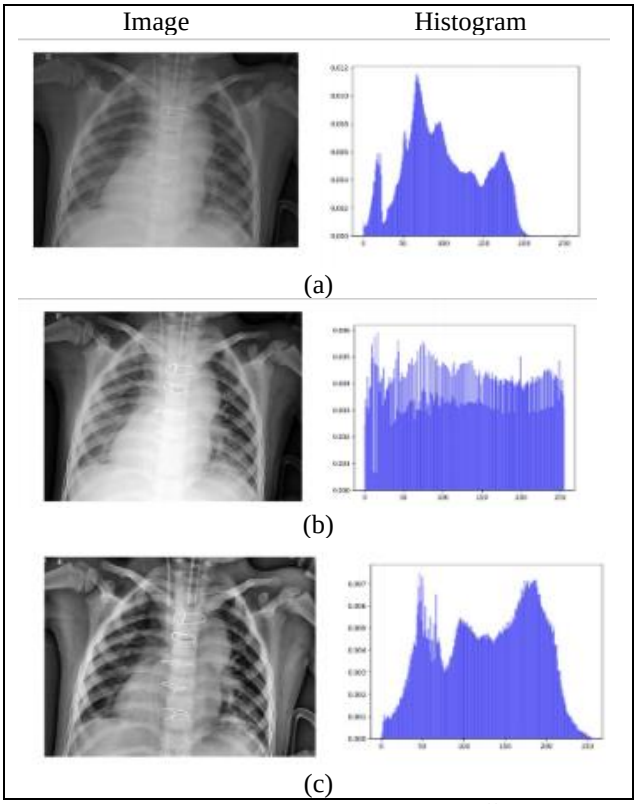

Fig. 3. Histogram equalization enhancement effect.
(a) Original, (b) HE effects, (c) CLAHE effects

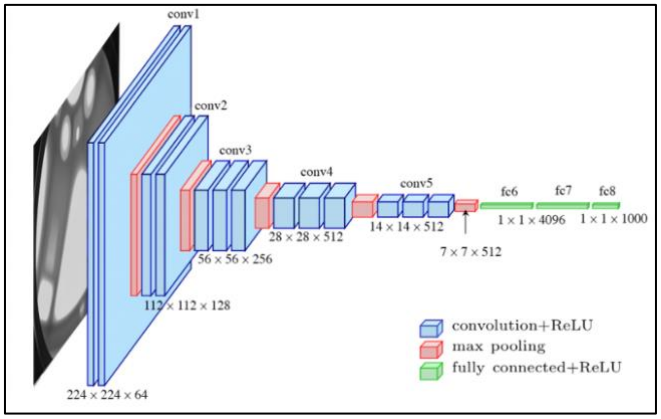

Fig. 4. Architecture of VGG16 model

C. *Class Decomposition/Composition*

The performance of a deep learning model will be affected if the dataset contains irregularities issue that confuses a model to learn precisely. When it comes to medical data, the irregularities issue is even more critical because of the inherent complexity of the decision boundaries in the medical imaging domain. To counter the issue, decomposition was

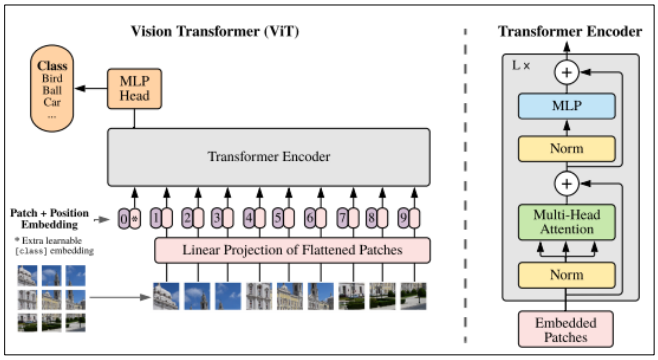

Fig. 6. Architecture of Vision Transformer (ViT)

conducted to split the original classes into several homogeneous subclasses to introduce clearer boundaries between classes [8].

K-means clustering is applied in the process to get the k number of subclasses for each original class. By using a binary class as an example with cluster k = 3, the feature space of class decomposition can be represented as:

$$A = \begin{bmatrix} a_{11} & \dots & a_{1m} & l_1 \\ a_{21} & \dots & a_{2m} & l_1 \\ a_{31} & \dots & a_{3m} & l_1 \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \dots & a_{nm} & l_2 \end{bmatrix}$$

$$B = \begin{bmatrix} a_{11} & \dots & a_{1m} & l_{11} \\ a_{21} & \dots & a_{2m} & l_{12} \\ a_{31} & \dots & a_{3m} & l_{1c} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \dots & a_{nm} & l_{2c} \end{bmatrix}$$

where A is the original dataset and B is the dataset after decomposition, a is denoted as a feature, l represents the classes and c is the number of the subclass. A composition method is adapted after the model prediction to ensemble the subclasses back to the original class.

D. COVID-19 Classification

In our approach, we have proposed two models to serve as the base model in our framework. The models are ResNet50 and Vision Transformer (ViT).

` Fig. 5 has illustrated the architecture of ResNet 50. The model contains 23 million trainable parameters and its architecture is stacked up by 50 layers of convolution layer. Each block consists of 3 convolutional layers. The stacking of convolution layers allows the model to extract deeper features to learn but it introduces the gradient vanishing issue. To mitigate the problem, the skip connections are added inside the model to allow the gradient flows through, they are located between each of the block.

ViT is a deep learning model that applies pure transformer in its model architecture to perform classification task and the architecture is illustrated in Fig. 6 [13]. This model takes an original image and split it into N number of fixed-size patches as:

$$N = HW / P^2 \quad (2)$$

N represents the number of patches, (H, W) is the resolution of the original image and (P, P) denotes the resolution of a patch image. Patch embedding is referring to the outputs of image patches that were flattened through a linear projection matrix since the standard transformer only receives input as a 1D sequence of token embedding. A 1D position embedding will be attached to every image patch to provide the transformer with the positional information of the image patch in the original image. This will help the model to evaluate the attention weights correctly as the transformer treats the inputs regardless of the order. An extra learnable embedding is added before the other image embedding to serve as the ground truth in the classification task.

The embeddings will act as the inputs to the transformer encoder to learn. A transformer encoder consists of multiple blocks of Multi-head Self Attention (MSA) and MLP that contains two layers with a GELU non-linearity. MSA is used to capture the different pattern of connectivity by computing the attention weights that flow through every block. Layer normalization (LN) is applied before every block, and residual connections are added after every block.

A fully connected MLP head is added at the end of the architecture to perform the prediction task.

E. Visualization

To increase the confidence at the model prediction, Gradient-based class activation mapping (Grad-CAM) is adapted to the ResNet50 model and the Gradient Rollout method is applied on ViT for model interpretability. Both methods will have similar output as shown in Fig. 7.

Grad-CAM works by using the gradient information flowing through the last layer of the convolutional neural network model and this helps to obtain the weights of the feature maps. A linear combination with ReLU follows after the weighted combination of activation maps to combine the features map that has a positive influence on the respective class of interest. The features map with the negative influence will be discarded to achieve higher localization performance in visualization.

Gradient Rollout visualize the features by using the attentions. In the attention map, a L layers Transformer will give a series of edges that link node ***u*** and node ***v*** as a path. A path is formed by node ***u*** in a position of L_i to node ***v*** in a position of L_j, where i > j. To compute the attentions in the particular path, it can be simply multiplying the weights of every edge in that path. As there might be more than a path between two nodes, the summation is utilized to sum up the attention of all paths. Multiplication on attention weights of every layer will be performed to acquire the total attention

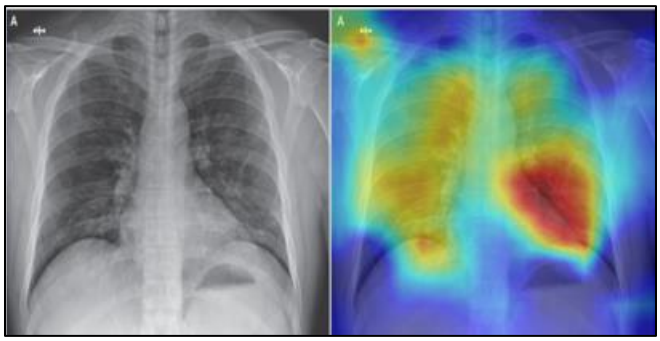

Fig. 7. Example of visualization result

weights for the transformer and then visualize the attention map.

V. Experiment

A. Datasets

To prove the efficiency of our proposed framework, we have conducted experiments on two datasets, where the details are listed in Table 1. COVIDx dataset is the first public dataset that contains CXR images of COVID-19, pneumonia and normal case. It is proposed by Wang L. & Wong A. [7] in 2020 along with their model in detecting COVID-19. The dataset will be updated in a time series manner which the size of dataset will become larger from time to time.

RICORD stands for RSNA International COVID-19 Open Radiology Database. It is funded through the National Institute of Biomedical Imaging and Bioengineering (NIBIB) under the collaboration among several organizations. The dataset is considered highly reliable as the data collection will be monitored by the organizations and there are some protocols to follow to ensure the reliability if anyone wants to contribute their data.

Table 1. Dataset details

Source	Class	Balanced	Distribution (Images)
COVIDx	COVID-19, Pneumonia, Normal	Yes	Train: 987 Validation: 246
RICORD + COVIDx	COVID-19, Pneumonia, Normal	Yes	Train: 2412 Validation:603 Test: 528

B. Evaluation Metrics

In the COVID-19 detections task, we are utilizing the following metrics to evaluate the model performance which are accuracy, precision, sensitivity, specificity and f1-score. The formulas are as follow:

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (5)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \quad (6)$$

$$F1\text{-}Score = \frac{2\times TP}{2\times TP+FN+FP} \quad (7)$$

where *TP*, *TN*, *FP* and *FN* denote true positive, true negative, false positive and false negative respectively.

C. Evaluation of Findings

Table 2 has shown the validation results of different approaches on COVIDx dataset. From the results, we found that CLAHE has improved the performance of ViT from 0.67 to 0.78 at COVID-19 sensitivity, while Decomposition & Composition managed to raise the sensitivity of ResNet50 from 0.78 to 0.89. There was not much difference in results after applying CLAHE together with Decomposition & Composition technique.

As pre-processing methods showed improvement on the results, we have decided to apply the approaches on new dataset which is the combination of RICORD dataset and COVIDx dataset. The results are illustrated in Table 3. It can be noticed that ResNet50 with CLAHE achieved the best result among the other approaches, where the overall accuracy, COVID-19 sensitivity and COVID-19 specificity reached 0.92, 1.00 and 0.97 respectively. The Decomposition & Composition method did not improve ResNet50, instead it degraded the performance slightly. However, by combining CLAHE with the Decomposition & Composition, it was able to raise the COVID-19 sensitivity to 0.92 and COVID-19 specificity to 1.00. One interesting finding is that ResNet50 is more sensitive to COVID-19 class, while ViT achieved higher value in COVID-19 specificity.

Table 4 compared the best two models with 8 CXR images from a clinical site. The table has indicated that ResNet50 outperformed the ViT model in overall by achieving 75% accuracy while ViT only reached 38% accuracy.

Table 2. Validation results on the COVIDx dataset

Model	**Class Decomposition & Composition**	**Pre-processing Methods**	**Results**		
			Overall Accuracy	**COVID-19 Sensitivity**	**COVID-19 Specificity**
ResNet50	No	-	0.83	0.78	0.97
		CLAHE	0.84	0.78	0.95
	Yes	-	**0.85**	**0.89**	0.95
		CLAHE	**0.85**	**0.89**	0.92
ViT	No	-	0.77	0.67	0.97
		CLAHE	0.83	0.78	**0.98**
	Yes	-	0.81	0.68	**0.99**
		CLAHE	0.80	0.78	0.94

Table 3. Test results on combination of RICORD dataset and COVIDx dataset

Model	**Class Decomposition & Composition**	**Pre-processing Methods**	**Results**		
			Overall Accuracy	**COVID-19 Sensitivity**	**COVID-19 Specificity**
ResNet50	No	CLAHE	**0.92**	**1.00**	0.97
	Yes	-	0.89	**1.00**	0.94
	Yes	CLAHE	0.89	**1.00**	0.92
ViT	No	CLAHE	0.89	0.90	0.99
	Yes	-	0.85	0.87	0.99
	Yes	CLAHE	0.86	0.92	**1.00**

Table 4. Test results on clinical images

Image	**Ground Truth**	**ResNet50 Results**	**ViT Results**
		CLAHE	CLAHE + Class Decomposition & Composition
Image 1	Covid-19	Covid-19	COVID-19
Image 2	Normal	Covid-19	Normal
Image 3	Covid-19	Covid-19	Normal
Image 4	Covid-19	Covid-19	Normal
Image 5	Covid-19	Covid-19	Normal
Image 6	Covid-19	Covid-19	Covid-19
Image 7	Covid-19	Covid-19	Pneumonia
Image 8	Normal	Covid-19	Covid-19
Overall Accuracy	_	**75%**	38%

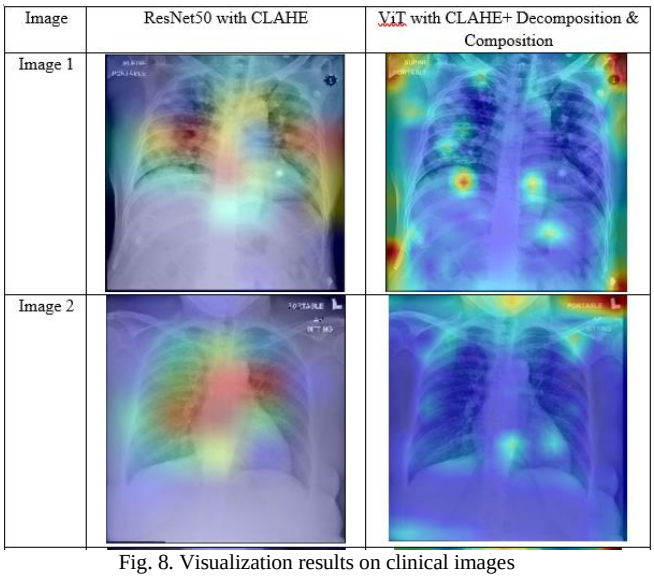

Fig. 8. Visualization results on clinical images

Since ResNet50 model with CLAHE and ViT with CLAHE and Decomposition & Composition performed the best among the proposed approaches, hence visualization methods are adopted on these 2 models. A visualization outputs is showed in Fig. 8. The figure is indicating that ResNet50 made the predictions by using the features that lie within the lungs and located at the infected area, while ViT has weighted its attention around the edges of image or the unrelevant body parts. We believe the reason that ViT did not perform well on clinical images was because the model did not put attention at the correct positions.

In short, we think that the ResNet50 is more reliable than the ViT. Furthermore, the CLAHE method has shown its efficiency in the experiments, while Decomposition & Composition is not consistent in improving the model performance.

VI. Conclusion

This paper has proposed two frameworks to compile with two different deep learning models in detecting COVID-19 infections. Moreover, visualization methods are adopted to interpret the important features of the images. The contributions of this paper are summarized as follows:

1. Implemented deep learning models to detect COVID-19 infections, which ResNet50 worked the best.
2. Proved the efficiency of CLAHE in improving the performance of a model.
3. Our work serves as the first that performed training and evaluation with CXR images of the RICORD dataset. This has increased the practicality to our proposed frameworks than the other existing approaches.
4. Performed clinical evaluation with our proposed approaches. This allowed us to validate the performance and applicability of the framework at clinician sites.
5. Investigated the ability of ViT in COVID-19 detection task.

In future, we wish to further increase the practicality of the framework by limiting the source to one that is more reliable as we were using COVIDx dataset for Non-Covid classes. To increase the performance, lungs segmentation can be done to exclude the unnecessary regions as we found the models were making predictions with the feature parts that do not lie within the lungs area. We also wish to improvise the model to gain the ability in detecting the severity of COVID-19 infections. Moreover, we would like to find out the reason that Vision Transformer has acquired good results in testing but performed bad with the clinical data.

VII. References

[1] Ng, M. Y., Lee, E. Y., Yang, J., Yang, F., Li, X., Wang, H., ... & Hui, C. K. M. (2020). Imaging profile of the COVID-19 infection: radiologic findings and literature review. *Radiology: Cardiothoracic Imaging*, *2*(1), e200034., in press.

[2] Wong, H. Y. F., Lam, H. Y. S., Fong, A. H. T., Leung, S. T., Chin, T. W. Y., Lo, C. S. Y., ... & Lee, E. Y. P. (2020). Frequency and distribution of chest radiographic findings in COVID-19 positive patients. *Radiology*, 201160, in press.

[3] Shoeibi, A., Khodatars, M., Alizadehsani, R., Ghassemi, N., Jafari, M., Moridian, P., ... & Srinivasan, D. (2020). Automated detection and forecasting of covid-19 using deep learning techniques: A review. *arXiv preprint arXiv:2007.10785*., unpublished.

[4] Shi, F., Wang, J., Shi, J., Wu, Z., Wang, Q., Tang, Z., ... & Shen, D. (2020). Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19. *IEEE reviews in biomedical engineering*., unpublished.

[5] Khobahi, S., Agarwal, C., & Soltanalian, M. (2020). CoroNet: A Deep Network Architecture for Semi-Supervised Task-Based Identification of COVID-19 from Chest X-ray Images. *medRxiv*., upublished.

[6] Lv, D., Qi, W., Li, Y., Sun, L., & Wang, Y. (2020). A cascade network for Detecting COVID-19 using chest x-rays. *arXiv preprint arXiv:2005.01468*., unpublished.

[7] 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), 1-12., in press.

[8] 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., in press.

[9] Mahmud, T., Rahman, M. A., & Fattah, S. A. (2020). CovXNet: A multi-dilation convolutional neural network for automatic COVID-19 and other pneumonia detection from chest X-ray images with transferable multi-receptive feature optimization. *Computers in biology and medicine*, *122*, 103869., in press.

[10] Siddhartha, M., & Santra, A. (2020). COVIDLite: A depth-wise separable deep neural network with white balance and CLAHE for detection of COVID-19. *arXiv preprint arXiv:2006.13873*., unpublished.

[11] Hasan, M., Alam, M., Elahi, M., Toufick, E., Roy, S., & Wahid, S. R. (2020). CVR-Net: A deep convolutional neural network for coronavirus recognition from chest radiography images. *arXiv preprint arXiv:2007.11993, unpublihsed.*

[12] Sayyed, A. Q. M., Saha, D., & Hossain, A. R. (2020). CovMUNET: A Multiple Loss Approach towards Detection of COVID-19 from Chest X-ray. *arXiv preprint arXiv:2007.14318*., unpublished.

[13] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929, unpublished.*