

Article

A Deep Learning Model with Self-Supervised Learning and Attention Mechanism for COVID-19 Diagnosis Using Chest X-ray Images

Junghoon Park , Il-Youp Kwak  and Changwon Lim * 

Department of Applied Statistics, Chung-Ang University, Seoul 06974, Korea; pjh940820@naver.com (J.P.); ikwak2@cau.ac.kr (I.-Y.K.)

* Correspondence: clim@cau.ac.kr; Tel.: +82-2-820-5547

Abstract: The SARS-CoV-2 virus has spread worldwide, and the World Health Organization has declared COVID-19 pandemic, proclaiming that the entire world must overcome it together. The chest X-ray and computed tomography datasets of individuals with COVID-19 remain limited, which can cause lower performance of deep learning model. In this study, we developed a model for the diagnosis of COVID-19 by solving the classification problem using a self-supervised learning technique with a convolution attention module. Self-supervised learning using a U-shaped convolutional neural network model combined with a convolution block attention module (CBAM) using over 100,000 chest X-Ray images with structure similarity (SSIM) index captures image representations extremely well. The system we proposed consists of fine-tuning the weights of the encoder after a self-supervised learning pretext task, interpreting the chest X-ray representation in the encoder using convolutional layers, and diagnosing the chest X-ray image as the classification model. Additionally, considering the CBAM further improves the averaged accuracy of 98.6%, thereby outperforming the baseline model (97.8%) by 0.8%. The proposed model classifies the three classes of normal, pneumonia, and COVID-19 extremely accurately, along with other metrics such as specificity and sensitivity that are similar to accuracy. The average area under the curve (AUC) is 0.994 in the COVID-19 class, indicating that our proposed model exhibits outstanding classification performance.

Keywords: COVID-19; self-supervised learning; deep learning; convolution attention; chest X-ray; Score-CAM



check for updates

Citation: Park, J.; Kwak, I.-Y.; Lim, C. A Deep Learning Model with Self-Supervised Learning and Attention Mechanism for COVID-19 Diagnosis Using Chest X-ray Images. *Electronics* **2021**, *10*, 1996. <https://doi.org/10.3390/electronics10161996>

Academic Editors: Anna Rita Fasolino and Flavio Canavero

Received: 6 June 2021

Accepted: 11 August 2021

Published: 18 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The SARS-CoV-2 virus, which causes COVID-19, has appeared across the globe and has considerably damaged all forms of human activities worldwide. As a species of the corona virus, such as the severe acute respiratory syndrome coronavirus that appeared in 2002 and the Middle East respiratory syndrome coronavirus that appeared in 2012, the SARS-CoV-2 virus spreads in the form of a droplet infection. Its spread has had adverse effects on the economies and culture of humankind. The SARS-CoV-2 virus infected 6.18 million people in just six months from December 2019 across the globe, and it continues to spread. Since discovery, approximately eighty million cases have been recorded world wide, including 35 million cases in the Americas, 26 million in Europe, and 1.8 million confirmed deaths across the globe as of 1 January 2021. Because of the rate of spread, the World Health Organization (WHO) declared the COVID-19 disease a pandemic on 11 March 2020, and the whole world has since been actively working to overcome this disease.

As the virus spreads around the world, experts are urging you to keep a distance from your acquaintances, and working hard to intensively manage confirmed cases and conduct epidemiological investigations of contacts. As the virus is highly contagious, keeping distance is the key to preventing the spread of the virus. Moreover, infected

patients require immediate treatment to prevent the spread and survive. Therefore, a quick diagnosis system is important to enable the patient to be treated and to contain the contagion. However, there is a significant problem with the detection of COVID-19.

Reverse transcriptase-polymerase chain reaction (RT-PCR) is considered to be a standard for detecting the SARS-CoV-2 virus [1], but it has some flaws, namely that is time-consuming, lacks sensitivity, and can yield a highly variable positive rate depending on the sample collection method [2,3]. Meanwhile, artificial intelligence (AI) is widely used as a solution to several problems in areas such as engineering, and economics around the world [4,5]. Particularly, convolution neural networks (CNNs) perform extremely well in image classification tasks. As patients infected by the SARS-CoV-2 or other pneumonial viruses have noticeable characteristics in their lung tissues [6], deep learning models can aid radiologists in pre-determining diseases with chest X-rays (CXRs). Compared to RT-PCR, AI using CNN has advantages such as time conservation and increased accuracy [7]. To overcome the weaknesses of RT-PCR, AI-based methods are being actively studied as tools for detecting the SARS-CoV-2 virus using CXRs [8–11]. These studies have adopted CNN-based models to classify CXR images to determine if an individual has COVID-19 or not. Furthermore, data sets of chest radiography images, such as CXRs and computed tomography (CT), of patients with COVID-19 are collected; these can be used by radiologists to analyze whether the patient is infected or not [11].

As deep learning models have millions of parameters to be trained, they cannot yield good performance if the reliable training data are not enough, leading to misclassification [12]. However, as the published CXR dataset for COVID-19 is limited compared to other pneumonial diseases, generalization performance of deep learning models might not be good enough in real application. For this reason, in this paper we considered various approaches to improve the classification performance by introducing structure-wise approach. We used a self-supervised learning method, Models Genesis [13], to ensure more robust representation learning. Then, lower-dimensional representations of CXR images can be captured; with the embedding layer that converts image features to them, we developed a classifier with an accuracy of 98.6%, thereby surpassing the accuracy of the baseline model (97.8%) by 0.8%. Our proposed system reaches improved performance by using self-supervised learning and convolutional attention module with 112,120 unlabeled CXR images and fine-tuning the weights of the encoder with COVID-19 data. This implies that our method can yield an accurately predicted class, even in unseen data. Furthermore, the results of the self-supervised learning pretext task and score-weighted class activation mapping (Score-CAM) [14] reveal that our study provides qualitative explanations, supported by high accuracy, sensitivity, specificity, and the area under the curve (AUC) of the receiver operating characteristic (ROC) curve as the quantitative metrics.

In this paper we focused on the fact that COVID-19 is a disease affecting the lungs, the same as pneumonia. Therefore, we considered Models Genesis, self-supervised learning to extract the features of lung images from a large CXR dataset. Because classification was performed by fine-tuning the weights of the encoder from the trained model by self-supervised learning, the performance of the proposed model exceeds other existing convolutional neural networks for predicting COVID-19. Quantitative and qualitative evaluations are also conducted for the self-supervised learning and downstream task of our proposed system, by examining the reasons for well-classified and misclassified cases visually. Source codes for this study is provided in <https://github.com/ngbsLab/covid19> (accessed on 12 August 2021). Overall, the main contributions in this paper are:

- Firstly, self-supervised learning is introduced to prevent the overfitting problem caused by the limited number of training images in deep learning. Models Genesis [13] was modified by adding convolutional attention module and trained using 112,120 unlabeled CXR dataset. And it was fine-tuned on a COVID-19 data set containing 1821 X-ray images. The accuracy of our model is 98.6%.
- We improved the performance of Models Genesis by adding a convolutional attention module after every convolutional layer. We conducted extensive experiments in

which we compared the performance of the modified Models Genesis containing the attention modules with that of the original for the COVID-19 classification.

- For qualitative evaluation of model results, we considered a visually explainable AI approach, Score-CAM [14]. By using it, we investigated how the proposed model makes correct/incorrect classifications to identify critical factors related to COVID-19 cases. The Score-CAM used in this paper is an improved method which resolve issues of the Grad-CAM [15].

2. Materials and Methods

This section describes the datasets and methods for the classification task based on CNN using self-supervised learning with a pretext task by a modified Models Genesis [13], which consists of UNet [16] and Convolutional Block Attention Module (CBAM) [17]. Combining these methods increases the classification accuracy, along with other metrics such as the sensitivity and AUC.

2.1. Datasets

The datasets used in this study were collected from various sources to develop classification models that could accurately identify images as belonging to the classes of normal, pneumonia, and COVID-19. Considering the balancing problem in classification tasks, the size of dataset for each of the three classes was set to 607. The NIH CXR dataset [18], which is used for self-supervised learning, consists of 122,120 images in PNG format, including images with two classes, normal and pneumonia. COVID-19 datasets were collected from various sources, as presented in Table 1.

Table 1. Number of data for each class used in this study and its sources.

Class	Sources	Number of Data
Normal	NIH CXRs dataset	607
Pneumonia	NIH CXRs dataset	607
COVID-19	COVID-19 image data collection [19]	468
	Figure 1 COVID-19 CXRs [20]	35
	Actualmed COVID-19 CXRs [21]	58
	COVID-19 Radiography Database [22]	46

These datasets are collected from same sources based on the baseline model by Lee et al. [11], to compare the model performances. All the images are resized to 512×512 , because the collected images are in different shapes. For self-supervised learning, we used 112,120 images as training dataset and 1000 images as validation dataset. For our classification task, we split the data randomly into three parts: 20% of the total data set into test set, 20% of the remaining as validation set, and the remaining as the training set; this yields 1160 training data, 290 validation data, and 363 test data. Then the pixel values of the collected images were scaled by a factor in the range of zero to one.

2.2. Existing Models

As mentioned above, since the collected datasets are from the same sources as in Lee et al. [11], we set their model as the baseline. Lee et al. [11] trained deep CNN-based models by visual geometric group (VGG) team in Oxford University, VGG-16 and VGG-19 [23]. VGG-16 and VGG-19 have depth of 16 and 19 depth convolutional layers, using an architecture with very small (3×3) convolution filters. While slightly increasing the depth of convolutional layers, image classification performance reached the highest in ImageNet Challenge 2014 submission. They constructed 12 experimental models by using different degrees of fine-tuning for VGG-16 and -19 models. Then, Lee et al. [11] conducted an experiment to compare the performance of the models for the COVID-19 classification. Fine-tuning with Conv block 4 and Conv block 5 of VGG-16 exhibited the highest AUC value of the ROC curve.

Furthermore, in addition to the baseline model by Lee et al. [11], we compare the performance of our proposed systems through two recently released models by Das et al. [24] and Rahimzadeh et al. [25]. Das et al. [24] proposed a two-stage machine learning model for classifying COVID-19 using CXR images. The proposed model extracts CXR features with the VGG19 model pretrained with ImageNet dataset, and classifies CXR images into normal and abnormal by logistic regression. Then, CXR images classified as abnormal in the first stage are again classified into pneumonia and COVID-19 using XGBoost [26]. Rahimzadeh et al. [25] introduced some training techniques that help the network learn better when there is an unbalanced dataset (with more cases from other classes and fewer COVID-19s). Also, the authors proposed a neural network connecting Xception [27] and ResNet50V2 [28] networks.

Accuracy, sensitivity, specificity, AUC and F1 score of the VGG-16 with fine-tuning by Lee et al. [11], bi-level classifier by Das et al. [24] and Xception and ResNet50V2 concated model by Rahimzadeh et al. [25] were set for comparison with our proposed method. Besides, we compared our proposed method with other well-known models, such as ResNet-50, ResNet-101 [29], MobileNet [30], and MobileNetV2 [31] which were utilized in other COVID-19 research [32–34].

2.3. Self-Supervised Learning

Self-supervised learning is used to learn the feature representations using unlabeled data. As a form of semi-supervised learning, self-supervised learning is a state-of-the-art technique that uses unlabeled data with a pretext task to capture semantic features for use in other vision tasks such as classification and to improve the performance or to prevent overfitting when the data are limited [35,36]. Image features learned from a self-supervised learning method can be good substitutes when used in a transferring approach in other vision tasks, as proven by various former studies such as Larsson et al. [37] and Zhang et al. [38]. Furthermore, introducing self-supervised learning to obtain representations of data improves the accuracy of the model and makes the model more robust to the result and even to adversarial examples [39]. Due to these benefits, self-supervised learning has been widely used in various areas including medical image management. The pretext task of Models Genesis begins by transforming the input medical images using three steps: non-linear transformation, local-shuffling, and out-painting or in-painting, as shown in Figure 1. See Zhou et al. [13] for details of the three types of input image transformation of Models Genesis.

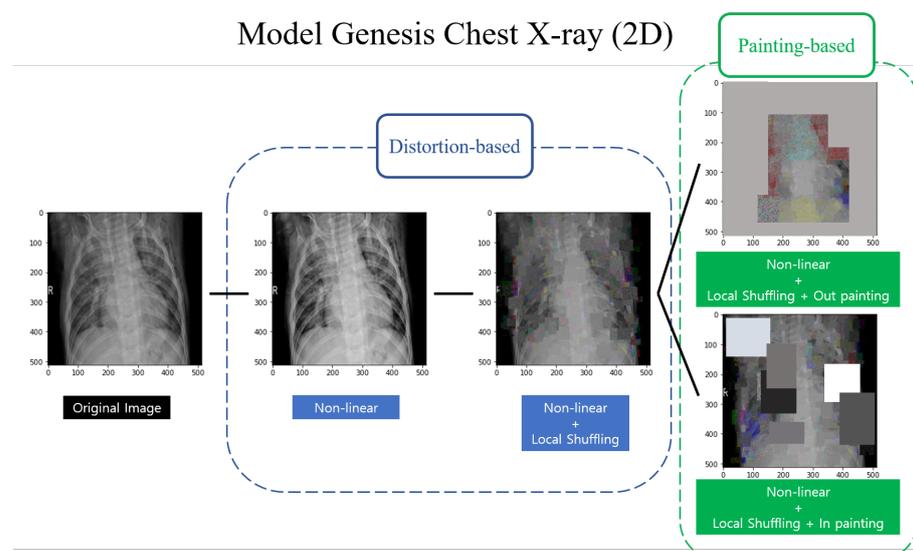


Figure 1. Example of transformed CXR (2D) by Models Genesis. CXR images are transformed using distortion-based and painting-based methods [13].

In the pretext task of Models Genesis, the model learns the CXR representation by restoring the transformed images into their original ones. UNet encoder-decoder architecture is used for the restoration task. UNet [16] is the model that first introduced a fully convolutional network in segmentation tasks, especially in the medical image field. Although it does not contain fully connected layers, it has fully convolution layers, thereby enabling the model to yield more accurate results in segmentation tasks. It is a U-shaped model, as shown in Figure 2, that consists of convolution and upsampling layers to make input and output shapes equal. UNet is divided into two parts: the encoder and the decoder. The encoder component uses a convolution layer to train input image features and the decoder component for segmentation tasks upsamples the output of the encoder gradually to yield the same shape as the input, corresponding to the same depth of the encoding component using upsampling layers. Furthermore, because the decoding component begins with a significantly large number of feature channels, the model can propagate context information to higher resolution layers [16].

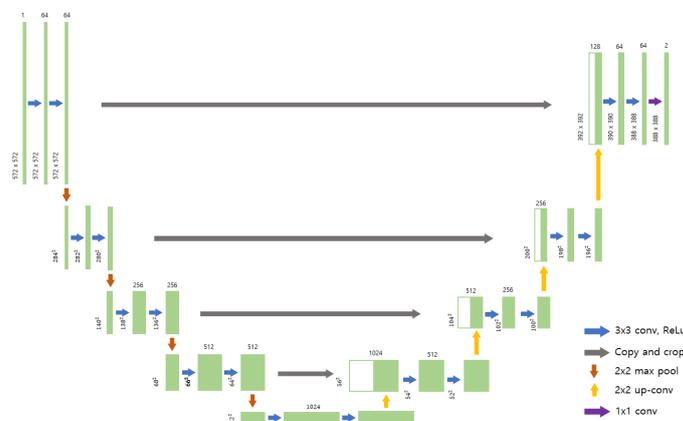


Figure 2. UNet architecture. As shown in this figure, the U-shaped model which is used especially for segmenting medical images [16].

2.4. Convolutional Attention Module

The attention method used in the deep learning field has several advantages. Attention enables a deep learning model to focus on an important point of the input and yield better interpretations of the output. Additionally, attention allows researchers to interpret the deep learning model through human perception [40]. Therefore, various approaches for applying attention to CNN-based models have been proposed. One approach is called the Convolutional Block Attention Module (CBAM), which was developed by Woo et al. [17]. Figure 3 shows the architecture of the CBAM. It consists of channel-wise and spatial-wise attentions. When a feature map F is given as input, the CBAM computes two feature maps using the channel-wise and spatial-wise attentions. Then, the overall CBAM is expressed as follows:

$$\begin{aligned}
 F &\subset R^{H \times W \times C}; \\
 M_c &\subset R^{1 \times 1 \times C}; \\
 M_s &\subset R^{H \times W \times 1}; \\
 F' &= M_c(F) \otimes F; \\
 F'' &= M_s(F') \otimes F',
 \end{aligned}$$

where H , W , and C are the height, width, and channel of the input feature map, respectively, \otimes denotes element-wise multiplication, and M_c and M_s represent the channel-wise and spatial-wise attentions, respectively. The channel attention map represents the ‘inter-channel relationship’ of features by considering the channel of an input feature map as

a feature detector [41]. Channel attention uses multi-layer perceptron, with one hidden layer applied to two different features of average-pooling and max-pooling operations. Therefore, it can be interpreted as focusing on the ‘what’ is meaningful in the input feature map [17]. Spatial attention, unlike channel attention, produces a spatial attention map focusing on ‘inter-spatial relationship’ of the input feature map. It uses convolution layers to generate spatial attention maps with a filter size of 7×7 and sigmoid function. Hence, spatial attention can be interpreted as focusing on ‘where’ to be counted as an important area of input feature map [17]. Using the channel and spatial attention, CBAM can be used on any dimensional feature map when it follows the convolution layer.

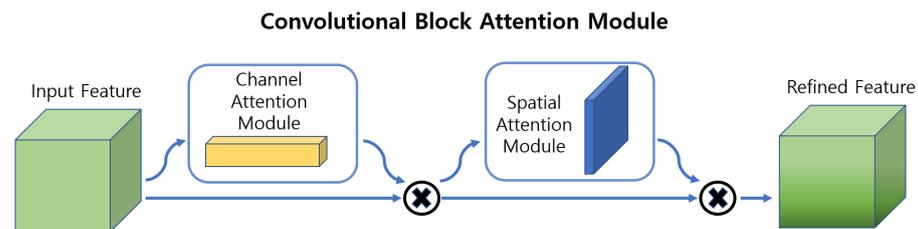


Figure 3. Architecture of the CBAM [17].

2.5. Our Proposed System

2.5.1. Self-Supervised Learning with Convolutional Attention Module

In this paper, we used the self-supervised learning method, Models Genesis [13] and modified it with CBAM [17] after every convolutional layer in Models Genesis networks to learn the representation of CXR images. For the self-supervised learning we conducted an experiment with NIH large CXR dataset, which contains 112,120 CXR images from pneumonia patients and normal cases. Following the transformation of the input images based on Models Genesis, the proposed self-supervised learning model should capture image features better than the model without self-supervised learning because the use of large amounts of CXR data helps capturing important CXR image features.

Figure 4 illustrates how our self-supervised learning task captures the image representations as pictionary. The original CXR images are expressed as X and restored images as X' . Transformed images are input into the U-shaped architecture of Models Genesis to restore their original pixel values. During this task, the U-shaped architecture and CBAM weights are updated to restore the transformed images. We conducted the self-supervised learning from scratch using the training dataset, and trainable weights are initialized randomly from zero to one. The loss function, $L(X, X')$, consists of structure similarity (SSIM) index [42] which is used to measure similarity between original images and their corresponding restored ones. SSIM index is calculated by following equations:

$$SSIM(x, y) = [l(x, y)]^\alpha [c(x, y)]^\beta [s(x, y)]^\gamma.$$

The term $l(x, y)$ equals $(2\mu_x\mu_y + C_1) / (\mu_x^2 + \mu_y^2 + C_1)$, where μ represents sample mean and C_1 is a very small constant to stabilize the metric when μ s are close to zero. $l(x, y)$ indicates the difference of luminance of the two images. $c(x, y)$ equals $2(\sigma_x\sigma_y + C_2) / (\sigma_x^2 + \sigma_y^2 + C_2)$, where σ represents sample standard deviation and C_2 is a constant close to zero. $c(x, y)$ indicates the difference of the contrast of the two images. And $s(x, y)$ is calculated by $(\sigma_{xy} + C_3) / (\sigma_x\sigma_y + C_3)$ which is a measure for the structural correlation between two images, and C_3 also plays a role in stabilizing the fraction calculations as C_1 and C_2 . The parameters, α , β and γ which are positive values, are adjustment numbers for relative importance of the three components. For our experiments, C_1 , C_2 , and C_3 are set to very small value, 0.000001, and all three parameters are equal to 1, to simplify the calculations.

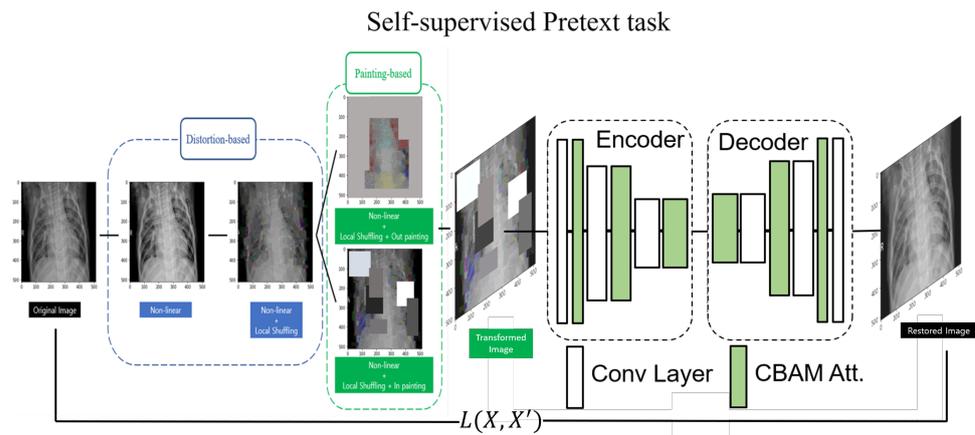


Figure 4. The modified Models Genesis with CBAM.

Therefore, as SSIM index has its value between 0 to 1 in our experiments, the $1 - \text{SSIM}$ index is used as the loss function to maximize the SSIM index while reducing the loss value.

2.5.2. Fine-Tuning the Encoder

Figure 5 shows our overall proposed system for the COVID-19 classification. As the self-supervised learning is conducted using 112,120 images, we enable our encoder-decoder architecture to capture representations of the CXR images. Our proposed classification model fine-tunes the weights of the encoder network from the pretrained self-supervised learning model.

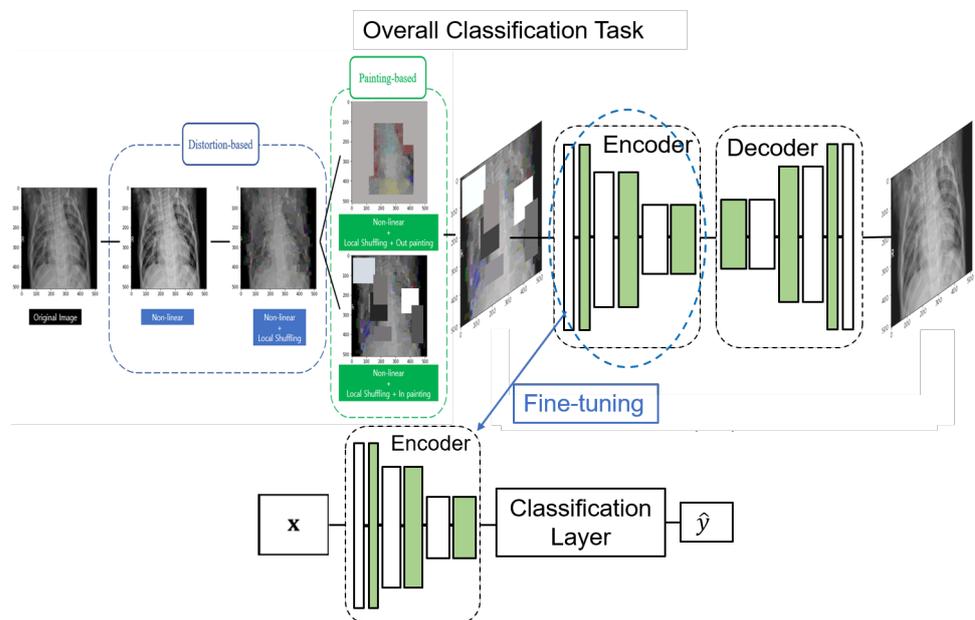


Figure 5. Our overall proposed system.

Behind the encoder network from the self-supervised learning model, we added classification layers which consist of four convolution layers, a max pooling layer, a global average pooling layer, and some fully connected layers. The four 2D convolution layers are to interpret the information of feature maps resulting from the encoder. After the convolution layers, the max and global average pooling layers follow to maintain graphical features and to produce one-dimensional nodes which connect to fully connected layers. A max pooling layer is used to down-sample the output of the two convolution layers. This procedure reduces the input dimension while maintaining the graphical information.

Furthermore, a global average pooling layer calculates the average values of each feature map, containing more implicative information than a flatten layer and prevents overfitting by reducing trainable parameters [43].

3. Experimental Study

3.1. Experimental Details

We trained the self-supervised learning model using the training dataset, and the validation set is used for hyperparameter tuning. The evaluation metric is the SSIM index and mean squared error (MSE) which considers the self-supervised learning task as a regression task to predict the pixel values of original CXR images.

For self-supervised learning, the initial learning rate was 0.0001 with an Adam optimizer [44]. In the training procedure, the key to develop classification models is to prevent overfitting. Thus, we used data augmentation techniques to make the training data more variety [45]. In this experiment, we used the three image data augmentation methods, which are flipping, zooming and width shifting to reduce the bias due to the characteristics of the CXR images. Flipping refers to a method of flipping an image left and right, and zooming is a method of performing augmentation with zooming at a certain ratio. Width shifting is an augmentation method that can reduce the bias on the position of an object in the image by moving the image up, down, left, and right by a certain distance. For the self-supervised learning task, the selected hyperparameter is the Adam optimizer with initial learning rate 0.0001. The learning rate of Adam optimizer decreases exponentially by 0.8 every 10 epochs after 40th epoch and the batch size is 16. Also, the L1 and L2 regularization value was 0.01 and the drop out ratio was 0.2. For classification task, the selected hyperparameter is the Adam optimizer with initial learning rate 0.00001. The learning rate decreases exponentially by 0.8 after 30th epoch by 0.8 and the batch size is 32.

Along with the structure-wise methods such as self-supervised learning and data augmentation, regularization, batch normalization, and dropout methods are used to manage the model. Regularization can improve the model performance and prevent overfitting by controlling trainable weights via model complexity [46]. Loss function using L1 or L2 regularization reduces the model size in training. Furthermore, batch normalization is used to reduce the chance for weights of layers to be high or low [47]. Additionally, dropout can also improve model performance by removing several connections of nodes randomly in hidden layers [48]. L1 and L2 regularization coefficients and the ratio of the total nodes to dropout were carefully tuned according to the validation accuracy in our classification task.

In this experiment, we used NVIDIA Quadro RTX 8000 in the Ubuntu 20 operating system. Entire neural networks were implemented using Keras API [49]. In the overall classification task, we used a learning rate with exponentially decaying to prevent the training procedure from falling in the local minimum of the loss function, beginning with 0.00001. An Adam optimizer was used in all training procedures.

3.2. Experimental Results

Figure 6 shows how well the modified Models Genesis with CBAM restores the transformed images. As the figures show, the output images of self-supervised learning pretext task are almost exactly the same as the original images. The training MSE of the NIH CXR dataset is 0.0228 and 0.0234 for the validation set. For the SSIM index metric of the NIH CXR dataset, the train dataset SSIM index is 0.9132, 0.9083 for the validation dataset. Organs such as the bones, lungs, and heart are restored very clearly, which implies that self-supervised learning using the modified Models Genesis pretext task can learn medical image features very well.

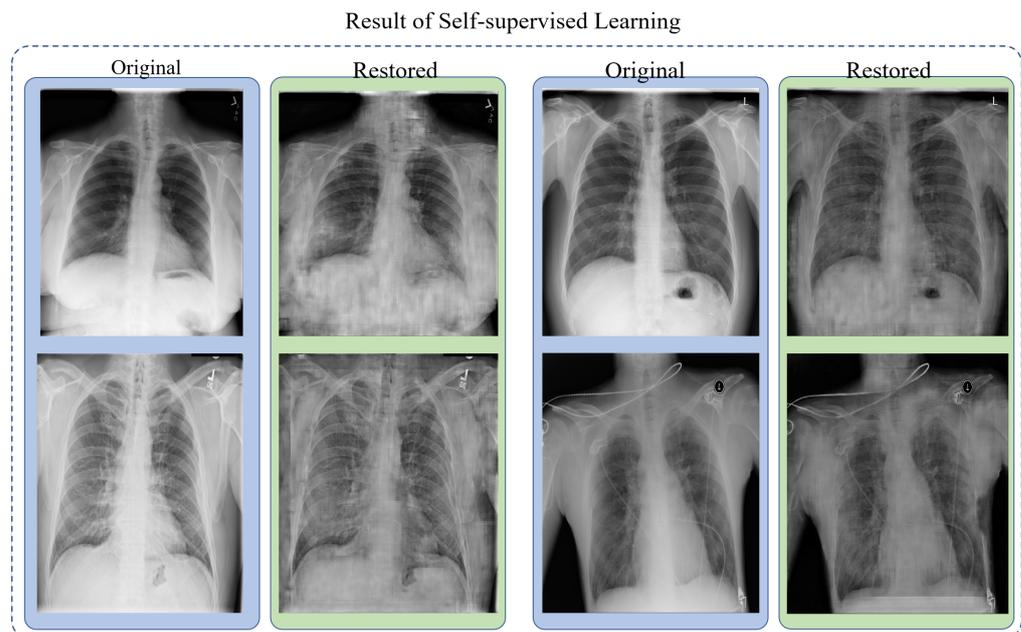


Figure 6. Result of self-supervised learning.

To evaluate the performance of our approach, the accuracy, specificity, sensitivity, AUC and F1 score of the test dataset are used. Sensitivity denotes the ability of the model to correctly detect infected patients given infected predicted cases, and specificity means to correctly detect normal people given none predicted ill cases. These metrics are calculated using the following equations:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$F1 \text{ score} = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$$

where TP and TN are the number of correctly predicted images for positive and negative cases, respectively, and FP and FN are the number of incorrectly predicted images for positive and negative cases, respectively. Using TP, TN, FP and FN, we made a plot of ROC curves and calculated the AUC to show the performance of the model at every threshold. The classification performances of our approach as well as other models are presented in Table 2.

From the table we can see that overall our model outperforms the baseline model and other CNN based models. In particular, our model scored 98.6% averaged accuracy, 0.996 specificity, 0.992 sensitivity, and 0.994 AUC in the COVID-19 case. Although it did not record the highest values in accuracy, sensitivity and specificity in all classes, the values of sensitivity and specificity vary when calculated according to the level of threshold, it is desirable to compare the overall accuracy, AUC and F1 scores to compare the classification performances. Because our proposed system has the highest AUC, F1 score and overall accuracy, it might be most powerful in diagnosing COVID-19 than other CNN-based models presented here. Also, training loss using categorical cross entropy was 0.0256 while 0.0287 in the validation set, which indicates that there is no evidence of overfitting. The confusion matrix in Figure 7 presents the result of the test dataset on the total samples that are accurately classified by our proposed methods. Furthermore Figure 8 shows the

ROC curves to depict overall classification performance of our proposed models in the test dataset, which is very excellent.

Table 2. Performance of our proposed approach and other models. N, P and C denote normal, pneumonia, and COVID-19, respectively, and Gen denotes Models Genesis.

Model Name	Class	Accuracy	Specificity	Sensitivity	AUC	F1 Score	Averaged Accuracy
Lee et al. [11]	N	0.980	0.992	0.982	0.981	0.968	0.978
	P	0.981	0.992	0.982	0.970	0.964	
	C	0.975	0.996	0.992	0.934	0.982	
Das et al. [24]	N	0.964	0.959	0.962	0.964	0.956	0.954
	P	0.954	0.949	0.952	0.954	0.955	
	C	0.954	0.959	0.957	0.954	0.953	
Rahimzadeh et al. [25]	N	0.982	0.985	0.987	0.979	0.980	0.983
	P	0.983	0.987	0.976	0.977	0.971	
	C	0.995	0.992	0.981	0.982	0.980	
ResNet50 [29]	N	0.983	0.984	0.964	0.917	0.964	0.975
	P	0.978	0.975	0.953	0.950	0.968	
	C	0.994	1.0	1.0	0.976	0.982	
ResNet101 [29]	N	0.972	0.984	0.964	0.970	0.955	0.975
	P	0.972	0.974	0.952	0.967	0.960	
	C	0.994	0.995	0.992	0.974	0.980	
MobileNet [30]	N	0.975	0.988	0.973	0.979	0.960	0.972
	P	0.974	0.987	0.973	0.976	0.971	
	C	0.994	1.0	1.0	0.956	0.982	
MobileNetV2 [31]	N	0.945	0.936	0.847	0.917	0.904	0.961
	P	0.945	0.982	0.969	0.950	0.924	
	C	0.994	1.0	1.0	0.986	0.982	
Gen + CBAM (ours)	N	0.984	0.996	0.991	0.980	0.981	0.986
	P	0.978	0.975	0.953	0.988	0.984	
	C	0.995	0.996	0.992	0.994	0.992	

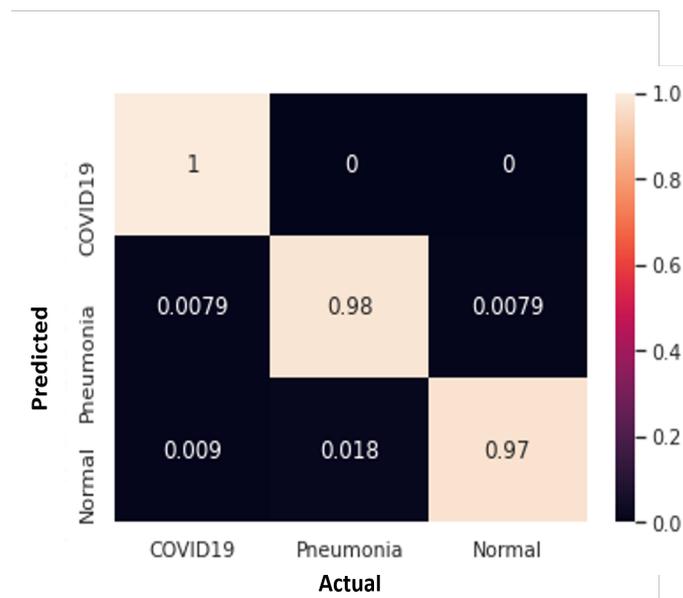


Figure 7. Confusion matrix of our proposed model for the test dataset.

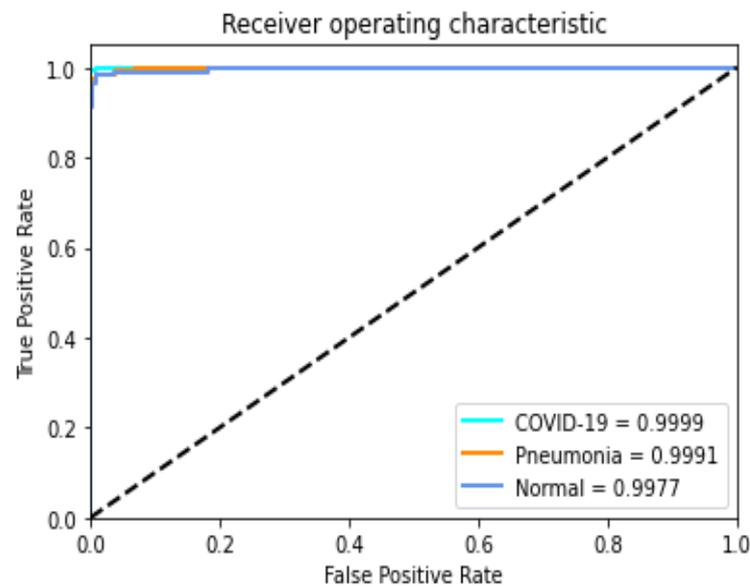


Figure 8. ROC curve of our proposed model for the test dataset.

4. Discussion

4.1. AI over RT-PCR Using CXR

Published studies have shown that a diagnosis system that uses images such as CXR and CT for COVID-19 has remarkable benefits over that of the RT-PCR [50,51]. However, while CT uses longer durations to produce the images and requires more expensive equipment, CXR is significantly cheaper and faster in producing the information from the tests. Because of these benefits, expanding the diagnosis system to CXR using AI will enhance the detection of COVID-19. Furthermore, developing systems that use AI to automatically analyze the images will reduce the time and cost of diagnosis of COVID-19 and all kinds of pneumonia diseases.

4.2. Interpretation of Classification Results Using Score-Cam

Figures 9 and 10 present examples of the Score-CAM [14]. Score-CAM can interpret which parts of images are important according to the predicted result by using gradient information flowing through the model from the last convolution layer to the input layer. Compared to other CAM-based approaches, Score-CAM eliminates the dependence on gradients by obtaining weights for activation maps through forward direction score on each class. Therefore, Score-CAM produces a linear combination of weights and achieves better performance for visual expression and fair interpretations for decision making process. As our proposed model outperforms other models considered in this paper, we employed Score-CAM to justify the model performance by confirming the activated regions of images that are important to make such decision.

Figure 9 shows the result of the Score-CAM for well-classified examples, and Figure 10 shows the misclassified samples of the test set with Score-CAM. In Figure 9, activated gradient areas are mostly in the lung tissues, with high degree compared to other regions. It is known that pneumonia is a disease that affects lung tissues [52], and COVID-19 is one of those types [53]. Therefore, it can be verified that because our model focuses mainly on lung tissues, the model could accurately diagnose whether the patients have COVID-19, pneumonia, or neither. Furthermore, considering the misclassified samples in the test dataset as shown in Figure 10, there is a significant difference that leads to misclassification. Compared to the samples in Figure 9, the activated gradient area of the both images in Figure 10 focus on the lung tissue area, but there is an alphabet 'R' in the left side of the images. Those activated gradient maps in the misclassified examples show that the proposed model focused on the character. This indicates that the activated gradient area

focuses on the wrong place because of the foreign matter in the images, thereby resulting in misclassification from normal to COVID-19 and pneumonia each. These substances distract and hinder the model that is trained to target the torso, as shown in the activated gradient region focusing on the external materials.

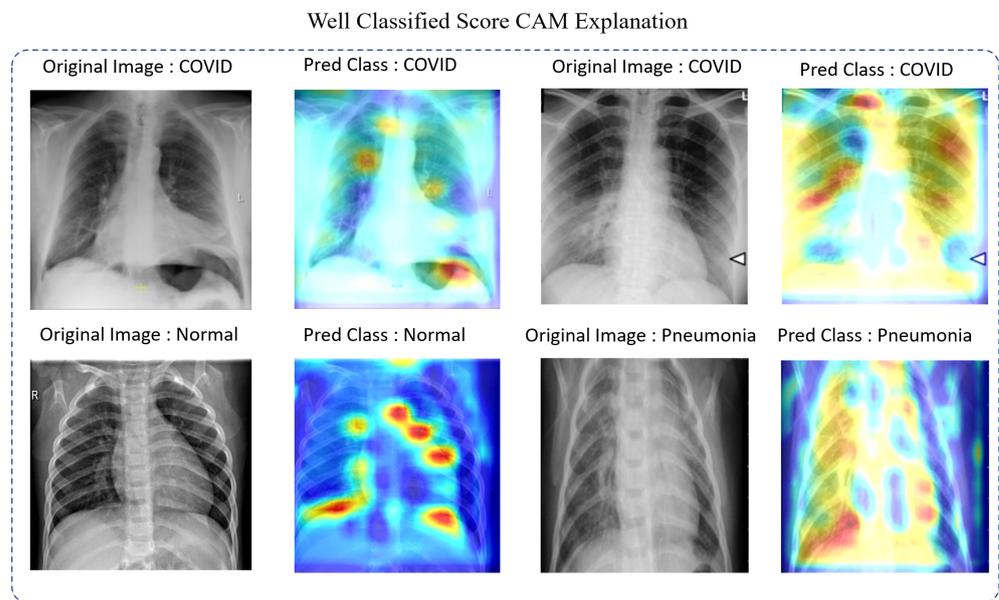


Figure 9. Examples of well-classified samples with Score-CAM.

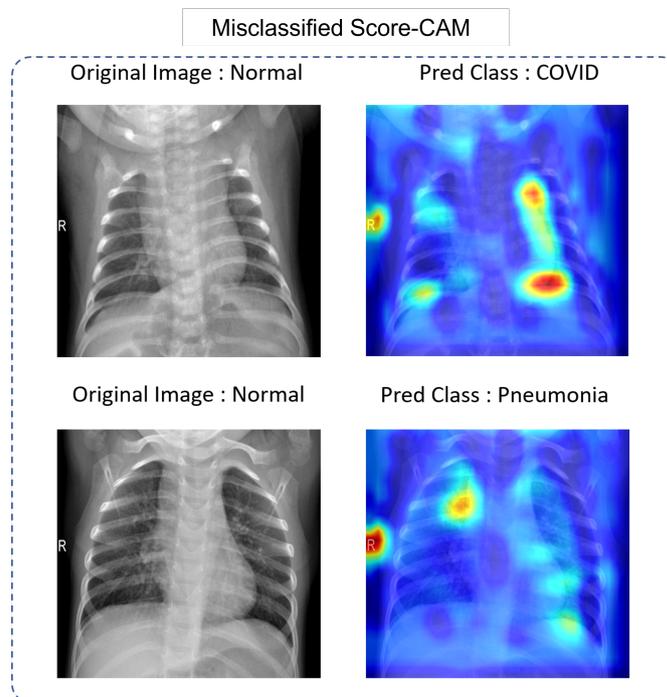


Figure 10. Examples of misclassified samples with Score-CAM.

4.3. Comparison with Other Methods for Covid-19 Classification

Since the strike of the COVID-19 globally, several new and modified deep learning models such as CVDNet [54] and transfer learning of various pre-trained deep learning models [55] have been proposed for screening the COVID-19 using AI. Among the recently published studies to diagnosis COVID-19, research topics can be divided into three main categories:

(1) Using well-known CNN based models. Hassantabar et al. [56] performed detection and diagnosis using MLP on fractal features and CNN on CXR. To extract fractal features from CXR, images were reshaped into 1-dimensional vectors at first, covariance matrix was calculated and eigenvalue and eigenvector were used for fractal features extraction. CNN architecture reached the higher accuracy, 93.2%. Khan et al. [57] proposed CoroNet using Xception architecture pretrained on ImageNet dataset. The CoroNet classifies CXR images into four classes, COVID-19, Pneumonia bacterial, Pneumonia viral and normal. This model achieved 89.6% of overall accuracy, and 95% overall accuracy in 3-class case for COVID-19, pneumonia, and normal. Ozturk et al. [58] aimed for detecting COVID-19 in early stages. The authors implemented Darknet architecture used in you only look once (YOLO) [59] to propose a real-time diagnosis method. The dataset used in this study is CXR images taken on the first day of patients infected with COVID-19. The accuracy for binary classification of COVID-19 and normal was 98.08%, and 87.02% for multi labels of COVID-19, pneumonia and normal cases. Afshar et al. [60] used Capsule Network based model to make classifications while handling small dataset. Proposed framework which consists of several Capsule and convolutional layers were pretrained by ImageNet. After transfer learning with COVID-19 dataset, classification results reached 95.7%, 90%, 95.8%, 0.97 for accuracy, sensitivity, specificity, and AUC respectively.

(2) Using self-supervised learning for feature extraction. Sriram et al. [61] trained a model using self-supervised learning based on momentum contrast method in pretraining to learn more general representations of CXR images. They used the pretrained model for downstream tasks of single image prediction, oxygen requirements predictions for greater than 6 L, and mortal prediction using multiple images sequence. The proposed model achieved AUCs of 0.742, 0.765, and 0.848 for three downstream tasks, respectively.

(3) Using an optimization method. Goel et al. [62] aimed to classify COVID-19, normal and pneumonia using CXR images. The authors proposed Optimized Convolutional Neural Network (OptCoNet) for automatic diagnosis of COVID-19. Grey Wolf Optimizer (GWO) algorithm was used to optimize the hyperparameters for training CNN. GWO algorithm selects the hyperparameter iteratively and evaluates the candidate solutions until the condition set by research is met. Using GWO, CNN based models achieved 97.78%, 97.75%, and 96.25% for accuracy, sensitivity, and specificity, respectively. The summary for comparison is represented in Table 3.

Unlike other studies, our proposed system has three important differences: (1) we proposed a system for diagnosing COVID-19 that does not overfit through the self-supervised learning pretext task of Models Genesis with a large CXR dataset to extract features of CXR images well. Pretext task of Models Genesis enables the encoder and decoder to be trained better with large CXR dataset, confirmed by the restored images in Figure 6. (2) We used convolutional attention modules to enhance the diagnosis performance of COVID-19. As our proposed system has CBAM layers after every convolutional layers, pretext task and classification task both reached improved results in SSIM index and evaluation metric. (3) We investigated the cause of misclassification visually through Score-CAM, in which the gradient is activated. Therefore, our study is considered in both aspects of quantitative and qualitative because we reached high classification accuracy as a quantitative aspect while confirming that the misclassified samples are not because of the problems in the training process of our proposed systems, but having foreign matters in Chest X-ray images as qualitative aspects.

Table 3. Details of previous research works for detecting COVID-19.

Authors [Reference Number]	Used Method	Base Architecture	Classes	Metrics	%
Das et al. [24]	Deep transfer learning with machine learning	VGG19	Normal, Pneumonia, COVID-19	Accuracy	99.26%
Rahimzadeh et al. [25]	Deep learning	Xception & ResNet50V2	Normal, Pneumonia, COVID-19	Accuracy	91.4%
Hassantabar et al. [56]	Deep learning	MLP, CNN	Normal, Pneumonia, COVID-19	Accuracy	93.2%
Khan et al. [57]	Deep transfer learning	Xception	Normal, Pneumonia bacterial & viral, COVID-19	Accuracy	95%
Ozturk et al. [58]	Deep learning	Darknet	Normal, COVID-19 Normal, Pneumonia, COVID-19	Accuracy Accuracy	98.08% 87.2%
Afshar et al. [60]	Deep transfer learning	Capsul network	Normal, Pneumonia bacterial & viral, COVID-19	Accuracy	95.7%
Sriram et al. [61]	Self-supervised learning	DenseNet	ICU transfer, intubation, mortality	AUC	74.2%
Goel et al. [62]	GWO	CNN	Normal, Pneumonia, COVID-19	Accuracy	97.78%

5. Conclusions

In this paper we proposed a novel deep learning system that contributes for screening COVID-19 efficiently. The advantage of our proposed system is in three main points. Our proposed system uses self-supervised learning, so the system can learn the features of CXR well using large amounts of CXR images data. Also, by applying the convolutional attention module in self-supervised learning task, we can focus more on important features of the CXR images. And lastly, through a qualitative evaluation using Score-CAM, we have identified the reasons for misclassified cases. The pretrained encoder from the modified Models Genesis was combined with classification layers and fine-tuned using the COVID-19 dataset. Through extensive experiments, we showed that our proposed system performs more powerful than other CNN-based classification models. The trainable parameters of the encoder and convolutional attention module in the pretext task aim to capture image representation precisely, not in just expanding numerical calculations. These lead to over 98% accuracy, with a similar or higher AUC and F1 score of our proposed system than those for other models. Furthermore, visualizing activated gradient areas using Score-CAM also verifies that our proposed model can enable diagnosing the CXR images by proper reason, as focusing on the lung tissues in CXR images.

There are some limitations of the proposed solution. Although we constructed balanced data to compare the proposed model with the baseline model proposed by Lee et al. [11], the CXR image data with three categories, normal, pneumonia and COVID-19, would be imbalanced with few COVID-19 samples in real life. Thus, the proposed solution may not perform as well as presented in this paper in that case. Furthermore, the COVID-19 data used in this paper is limited. Recent papers such as Das et al. [24] show that there are more COVID-19 data publicly available. Hence, if we could collect more data, the experimental results may be different.

For future research, one could investigate some implications of our proposed methods. As the extent of future work, we are considering the reason for misclassification appearing in our study. In our proposed method, it was confirmed through the activated gradient map from Score-CAM that foreign substances such as various medical devices on the chest and the letter 'R' on the CXR image interfered with the diagnosis and led to misclassification.

Therefore, using the point that the U-shaped model is often used in the segmentation task, it is possible to study the methodology for diagnosis that the region on images which are segmented into medical devices or external materials does not affect the classification task.

Also, our models was considered only in CXR in this paper. There will be more applications where the models are applied to chest CT data. Moreover, as COVID-19 radiology data have been collected, sufficient CXR images can enable the self-supervised learning model to capture the image representation better, thereby yielding higher classification model accuracy. Furthermore, our proposed approach manages only 3 categories; it would be more helpful in the real world to aid the radiologist when more categories and data are applied to our proposed models.

Author Contributions: Conceptualization, J.P., I.-Y.K. and C.L.; methodology, J.P., I.-Y.K. and C.L.; software, J.P.; validation, J.P.; formal analysis, J.P., I.-Y.K. and C.L.; data curation, J.P.; writing—original draft preparation, J.P.; writing—review and editing, I.-Y.K. and C.L.; visualization, J.P.; supervision, C.L.; project administration, C.L.; funding acquisition, I.-Y.K. and C.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT (NRF-2021R1F1A1056516) and the National Research Foundation of Korea (NRF) grant funded by the Ministry of Science and ICT (2020R1C1C1A01013020).

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: <https://github.com/ieee8023/covid-chestxray-dataset>, <https://github.com/agchung/Figure1-COVID-chestxray-dataset>, <https://github.com/agchung/Actualmed-COVID-chestxray-dataset>, <https://www.kaggle.com/tawsifurrahman/covid19-radiography-database> (accessed on 12 August 2021).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Wang, W.; Xu, Y.; Gao, R.; Lu, R.; Han, K.; Wu, G.; Tan, W. Detection of SARS-CoV-2 in different types of clinical specimens. *JAMA* **2020**, *323*, 1843–1844.
2. Fang, Y.; Zhang, H.; Xie, J.; Lin, M.; Ying, L.; Pang, P.; Ji, W. Sensitivity of chest CT for COVID-19: comparison to RT-PCR. *Radiology* **2020**, *296*, 115–117.
3. Wikramaratna, P.; Paton, R.S.; Ghafari, M.; Lourenco, J. Estimating false-negative detection rate of SARS-CoV-2 by RT-PCR. *Euro Surveill.* **2020**, *25*, 2000568, doi:10.2807/1560-7917.ES.2020.25.50.2000568.
4. Pham, D.T.; Pham, P.T.N. Artificial intelligence in engineering. *Int. J. Mach. Tools Manuf.* **1999**, *39*, 937–949.
5. Dirican, C. The impacts of robotics, artificial intelligence on business and economics. *Procedia-Soc. Behav. Sci.* **2015**, *195*, 564–573.
6. Parveen, N.; Sathik, M.M. Detection of pneumonia in chest X-ray images. *J. Ray Sci. Technol.* **2011**, *19*, 423–428.
7. Farooq, M.; Hafeez, A. Covid-resnet: A deep learning framework for screening of covid19 from radiographs. *arXiv* **2020**, arXiv:2003.14395.
8. Narin, A.; Kaya, C.; Pamuk, Z. Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. *arXiv* **2020**, arXiv:2003.10849.
9. Oh, Y.; Park, S.; Ye, J.C. Deep learning covid-19 features on cxr using limited training data sets. *IEEE Trans. Med. Imaging* **2020**, *39*, 2688–2700.
10. Minaee, S.; Kafieh, R.; Sonka, M.; Yazdani, S.; Soufi, G.J. Deep-covid: Predicting covid-19 from chest x-ray images using deep transfer learning. *Med. Image Anal.* **2020**, *65*, 101794.
11. Lee, K.S.; Kim, J.Y.; Jeon, E.T.; Choi, W.S.; Kim, N.H.; Lee, K.Y. Evaluation of Scalability and Degree of Fine-Tuning of Deep Convolutional Neural Networks for COVID-19 Screening on Chest X-ray Images Using Explainable Deep-Learning Algorithm. *J. Pers. Med.* **2020**, *10*, 213.
12. Mikołajczyk, A.; Grochowski, M. Data augmentation for improving deep learning in image classification problem. In Proceedings of the 2018 International Interdisciplinary PhD Workshop (IIPhDW), Swinoujscie, Poland, 9–12 May 2018; pp. 117–122.
13. Zhou, Z.; Sodha, V.; Siddiquee, M.M.R.; Feng, R.; Tajbakhsh, N.; Gotway, M.B.; Liang, J. Models genesis: Generic autodidactic models for 3d medical image analysis. In International Conference on Medical Image Computing and Computer-Assisted Intervention, Shenzhen, China, 13–17 October 2019; pp. 384–393.
14. Wang, H.; Wang, Z.; Du, M.; Yang, F.; Zhang, Z.; Ding, S.; Mardziel, P.; Hu, X. Score-CAM: Score-weighted visual explanations for convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 24–25.

15. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 618–626.
16. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical Image Computing and Computer-assisted intervention; Springer, Berlin/Heidelberg, Germany, 5–9 October 2015; pp. 234–241.
17. Woo S.; Park J.; Lee J.Y.; So K.I. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
18. Jaeger S.; Candemir S.; Antani S.; Wang Y.X.J.; Lu P.X.; Thoma G. Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. *Quant. Imaging Med. Surg.* **2014**, *4*, 475.
19. Cohen, J.P.; Morrison, P.; Dao, L.; Roth, K.; Duong, T.Q.; Ghassemi, M. Covid-19 image data collection: Prospective predictions are the future. *arXiv* **2020**, arXiv:2006.11988.
20. Chung, A. Figure 1 COVID-19 Chest X-ray Data Initiative. 2020. Available online: <https://github.com/agchung/Figure1-COVID-chestxray-dataset> (accessed on 4 May 2020).
21. Chung, A. Actualmed COVID-19 Chest X-ray Data Initiative. 2020. Available online: <https://github.com/agchung/Actualmed-COVID-chestxray-dataset> (accessed on 6 May 2020).
22. Rahman, T.; Chowdhury, M.; Khandakar, A. *COVID-19 Radiography Database*; Kaggle: San Francisco, CA, USA, 2020.
23. Simonyan, K.; and Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
24. Das, S.; Roy, S.D.; Malakar, S.; Velásquez, J.D.; and Sarkar, R. Bi-Level Prediction Model for Screening COVID-19 Patients Using Chest X-Ray Images. *Big Data Res.* **2021**, *25*, 100233.
25. Rahimzadeh, M.; Attar, A. A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2. *Inform. Med. Unlocked* **2020**, *19*, 100360.
26. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
27. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
28. He, K.; Zhang, X.; Ren, S.; Sun, J. Identity mappings in deep residual networks. In European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; pp. 630–645.
29. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 Jun 2016; pp. 770–778.
30. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
31. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 Jun 2018; pp. 4510–4520.
32. Rahaman, M.M.; Li, C.; Yao, Y.; Kulwa, F.; Rahman, M.A.; Wang, Q.; Qi, S.; Kong, F.; Zhu, X.; Zhao, X. Identification of COVID-19 samples from chest X-ray images using deep learning: A comparison of transfer learning approaches. *J. Ray Sci. Technol.* **2020**, *1–19*, in preprint.
33. Rehman, A.; Naz, S.; Khan, A.; Zaib, A.; Razzak, I. Improving coronavirus (COVID-19) diagnosis using deep transfer learning. *medRxiv* **2020**. Available online: <https://www.medrxiv.org/content/early/2020/04/17/2020.04.11.20054643.full.pdf> (accessed on 17 August 2021).
34. Wong, H.Y.F.; Lam, H.Y.S.; Fong, A.H.T.; Leung, S.T.; Chin, T.W.Y.; Lo, C.S.Y.; Lui, M.M.S.; Lee, J.C.Y.; Chiu, K.W.H.; Chung, T.W.H.; Lee, E.Y.P. Frequency and distribution of chest radiographic findings in patients positive for COVID-19. *Radiology* **2020**, *296*, E72–E78.
35. Gidaris, S.; Singh, P.; Komodakis, N. Unsupervised representation learning by predicting image rotations. *arXiv* **2018**, arXiv:1803.07728.
36. Zhai, X.; Oliver, A.; Kolesnikov, A.; Beyer, L. S4l: Self-supervised semi-supervised learning. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 27 October–3 November 2019; pp. 1476–1485.
37. Larsson, G.; Maire, M.; Shakhnarovich, G. Learning representations for automatic colorization. In European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; pp. 577–593.
38. Zhang, R.; Isola, P.; Efros, A.A. Colorful image colorization. In European conference on computer vision, Amsterdam, The Netherlands, 8–16 October 2016; pp. 649–666.
39. Hendrycks, D.; Mazeika, M.; Kadavath, S.; Song, D. Using self-supervised learning can improve model robustness and uncertainty. In International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 2712–2721.
40. Itti, L.; Koch, C.; Niebur, E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 1254–1259.
41. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning deep features for discriminative localization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2921–2929.

42. Channappayya, S.S.; Bovik, A.C.; Heath, R.W. Rate bounds on SSIM index of quantized images. *IEEE Trans. Image Process.* **2008**, *17*, 1624–1639.
43. Azulay, A.; Weiss, Y. Why do deep convolutional networks generalize so poorly to small image transformations? *J. Mach. Learn. Res.* **2018**, *20*, 1–25.
44. Zhang, Z. Improved adam optimizer for deep neural networks. In Proceedings of the IEEE/ACM 26th International Symposium on Quality of Service (IWQoS), Banff, AB, Canada, 4–6 Jun 2018; pp. 1–2.
45. Perez, L.; Wang, J. The effectiveness of data augmentation in image classification using deep learning. *Convolutional Neural Netw. Vis. Recognit.* **2017**, *11*, 1–8.
46. Girosi, F.; Jones, M.; Poggio, T. Regularization theory and neural networks architectures. *Neural Comput.* **1995**, *7*, 219–269.
47. Han, X.; Dai, Q. Batch-normalized Mlpconv-wise supervised pre-training network in network. *Appl. Intell.* **2018**, *48*, 142–155.
48. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
49. Gulli, A.; Pal, S. *Deep Learning with Keras*; Packt Publishing Ltd.: Birmingham, UK, 2017.
50. Ng, M.Y.; Lee, E.Y.; Yang, J.; Yang, F.; Li, X.; Wang, H.; Lui, M.M.; Lo, C.S.; Leung, B.; Khong, P.L.; Hui, C.K. Imaging profile of the COVID-19 infection: radiologic findings and literature review. *Radiol. Cardiothorac. Imaging* **2020**, *2*, e200034.
51. Liu, H.; Liu, F.; Li, J.; Zhang, T.; Wang, D.; Lan, W. Clinical and CT imaging features of the COVID-19 pneumonia: Focus on pregnant women and children. *J. Infect.* **2020**, *80*, 7–13.
52. Fiszman, M.; Chapman, W.W.; Aronsky, D.; Evans, R.S.; Haug, P.J. Automatic detection of acute bacterial pneumonia from chest X-ray reports. *J. Am. Med. Inform. Assoc.* **2000**, *7*, 593–604.
53. Zhao, D.; Yao, F.; Wang, L.; Zheng, L.; Gao, Y.; Ye, J.; Guo, F.; Zhao, H.; Gao, R. A comparative study on the clinical features of COVID-19 pneumonia to other pneumonias. *Clin. Infect. Dis.* **2020**, *71*, 756–761.
54. Ouchicha, C.; Ammor, O.; Meknassi, M. . CVDNet: A novel deep learning architecture for detection of coronavirus (Covid-19) from chest X-ray images. *Chaos Solitons Fractals* **2020**, *140*, 110245.
55. Marques, G.; Agarwal, D.; de la Torre Díez, I. Automated medical diagnosis of COVID-19 through EfficientNet convolutional neural network. *Appl. Soft Comput.* **2020**, *96*, 106691.
56. Hassantabar S.; Ahmadi M.; Sharifi A. Diagnosis and detection of infected tissue of COVID-19 patients based on lung X-ray image using convolutional neural network approaches. *Chaos Solitons Fractals* **2020**, *140*, 110170.
57. Khan, A.I.; Shah, J.L.; Bhat, M.M. CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest X-ray images. *Comput. Methods Prog. Biomed.* **2020**, *196*, 105581.
58. Ozturk, T.; Talo, M.; Yildirim, E.A.; Baloglu, U.B.; Yildirim, O.; Acharya, U.R. Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Comput. Biol. Med.* **2020**, *121*, 103792.
59. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
60. Afshar, P.; Heidarian, S.; Naderkhani, F.; Oikonomou, A.; Plataniotis, K.N.; Mohammadi, A. Covid-caps: A capsule network-based framework for identification of covid-19 cases from X-ray images. *Pattern Recognit. Lett.* **2020**, *138*, 638–643.
61. Sriram, A.; Muckley, M.; Sinha, K.; Shamout, F.; Pineau, J.; Geras, K.J.; Azou, L.; Aphinyanaphongs, Y.; Yakubova, N.; Moore, W. COVID-19 Prognosis via Self-Supervised Representation Learning and Multi-Image Prediction. *arXiv* **2021**, arXiv:2101.04909.
62. Goel, T.; Murugan, R.; Mirjalili, S.; Chakrabartty, D.K. OptCoNet: an optimized convolutional neural network for an automatic diagnosis of COVID-19. *Appl. Intell.* **2021**, *51*, 1351–1366.