Using Hybrid Compact Transformer for COVID-19 Detection from Chest X-Ray

Ghadeer Almoeili[1], Abdenour Bounsiar[2]
Applied College, King Faisal University, Alhassa, Kingdom of Saudi Arabia
Ministry of Higher Education[1]
College of Computer Sciences and IT, King Faisal University, Alhassa, Kingdom of Saudi Arabia
Ministry of Higher Education[2]

***Abstract*—By the end of December 2019, the novel coronavirus 2019 (COVID-2019), became a world pandemic affecting the respiratory system. Scientists started investigating using Deep Learning and Convolutional Neural Networks (CNNs) to detect COVID-19 using Chest X-rays (CXRs). One of the main difficulties researchers reported in the detection of lung diseases is the fact that radiographic images can tell that the lungs are abnormal, but they might miss specifying the type of pneumonia exactly. Only the expert radiologist can tell the difference based on patches shapes and distribution on the affted lungs. Also CNN's require big datasets to provide good results. When new pandemics spread, The limited benchmark datasets for COVID-19 in CXR images, especially during the onset of the pandemic, is the main motivation of this research. In this research, we will introduce the use of Vision Transformers (ViTs). We consider an updated version of ViT called Compact Transformer (CT) which was proposed to reduce the expansive computations of the self-attention mechanism in ViT and to escape the big data paradigm. As a contribution of this study, We propose using a Hybrid Compact Transformer (HCT) in which a pretrained CNN is used in place of the convolutional layers in CT. Hence, with the hybrid model design, we aim to combine the localization power of CNNs, with the generalization power (attention mechanism or distanced-pixel relations) of ViTs. Based on experimental results using different performance metrics, the Hybrid Compact Transformer is shown to be superior over Compact Transformers and Transfer Learning models. Our proposed technique enjoys the benefits of both worlds; a faster training of the model due to TL with CNNs and reduced data requirements due to CT. Combining localized filters of CNNs and the attention mechanism of CT seems to provide a better discrimination between common pneumonia and Covid-19 pneumonia.**

***Keywords*—Deep convolutional neural network; CXR chest X-Ray; COVID-19 pneumonia; vision transformers; compact convolutional transformer; hybrid compact transformer**

I. Introduction

At the end of December 2019, there was a cluster of Pneumonia cases discovered in the city of Wuhan. By the end of January 2020, the World Health Organization (WHO) announced that the disease had become a world pandemic caused by Severe Acute Respiratory Syndrome Corona Virus-2 (SARS-CoV-2), commonly called COVID-19.

The test that is mostly used to detect COVID-19 infection is the Reverse-Transcription polymerase chain reaction (RT-PCR) [29]. In the early days, this test was not approved to be used for the detection of respiratory diseases because of the low sensitivity of the test and the high possibility of false negatives that might occur. Instead, chest imaging such as CXR or computed tomography (CT) was used to diagnose respiratory diseases. However, with the pandemic of Coronavirus, Using CXR is considered faster and cheaper than using RT-PCR. However, CXR interpretation requires expert radiologists. If we consider the number of radiologists in each hospital compared to the number of patients' CXR images, radiologists would not be able to study that massive amount of CXR images.

In 2016, a research paper was published demonstrating the efficacy of deep learning algorithms in the medical field [12]. This research discusses the employment of deep learning models in the task of grading for diabetic retinopathy to recapitulate the majority decision of the board-certified ophthalmologists in the US. Deep learning is a type of Machine Learning algorithm that employs neural networks (NN) to learn complex relationships among huge amounts of data.

In 2020, a research paper was published by Google team [7], introducing the new state-of-the-art image classification. Inspired by the Transformers scaling successes in NLP [45], Transformers have accomplished quick successes in computer vision. Vision Transformers (ViTs) are emerging as an architectural paradigm alternative to CNNs. However, the lack of the typical convolutional inductive bias makes these models extremely data-hungry and computationally expensive compared to common CNNs [7]. Consequently, the cost of training such models is only affordable by the lucky few at the large industrial companies.

Data Efficient Image Transformer (DeiT) [43] is one of the first papers to show that it is practical to train transformers for computer vision tasks. DeiT trained with a procedure more adapted to a data-starving regime. Subsequently, it requires far fewer data and far less computer power to produce a high-performance image classification model. Recently, numerous related work has been proposed to democratize AI research for transformers [23] [13]. To help researchers with limited resources to verify the research results and to take these results for granted. Both CNNs and Transformers have highly desirable qualities for different computer vision tasks, but each comes with their own costs [13].

Recently, many researchers have used radiology images for COVD-19 detection. Authors in [36] compared four popular neural networks for the classification of CXR images. AlexNet [22], ResNet18 [14], DenseNet201 [15], and SqueezeNet [16]. They used radiographic images from the Kaggle [33] database. Image Augmentation is applied to the

data set and three classification schemes were compared: normal vs pneumonia, bacterial vs viral pneumonia, and normal, bacterial and viral pneumonia. This paper demonstrates that the deeper the network the better the accuracy, such that DenseNet201 outperforms the other three networks in all the three classification schemes. The classification accuracy achieved was 95%.

The authors in [38] proposed an algorithm called Data Augmentation of Radiograph Images (DARI) which combines GANs architecture with generic data augmentation methods to maximize the training data. DARI algorithm is applied to the input images when the class imbalanced ratio is greater than a given threshold. The accuracy achieved training this model was 93.94%. Further, the authors of the DARI algorithm compared the performance of their proposed method with another paper called DarkCovidNet [31]. The model is designed for the diagnosis of the COVID-19 disease. In their study, the main model was inspired by the DarkNet architecture that has proven itself in deep learning. DarkCovidNet achieved an accuracy of 87.02%.

Since the early onset of COVID-19, there was a global scientific response to help in diagnosing and curing. Many of these efforts considered automated COVID-19 detection from CXRs, such as [2] or [3]. Deep Neural Networks (DNNs), and CNNs has boosted medical image analysis over the past years. This research aims to investigate the use of CNNs and ViTs for the automated detection of COVID-19 from CXR images.

Detection of COVID-19 from CXRs involves two problems: COVID-19 vs non-COVID-19 in which case a dataset like [6] can be used, and COVID-19 vs Other Pneumonia vs healthy, in which case a dataset like [4] can be used along with [6]. In this research, we will consider having a benchmark dataset that contains three classes COVID-19 Pneumonia (CP) vs Community-Acquired Pneumonia (CAP) and Normal cases for our study.

Most medical applications suffer from limited datasets, and as Deep Convolutional Neural Networks (DCNNs) are data-hungry, the medical community has almost universally adopted transfer learning to build a CNN for medical imaging. For example [31] uses initial model weights from ChestNet [37] for pneumonia disease detection. In this research, we will consider using a pretrained model on a benchmark dataset.

Moreover, as COVID-19 is a respiratory disease such that the virus directly infects the lungs area, having a model that focuses on studying the infection only by segmenting the lungs area in CXR images from other parts in the image, can add improvement in the performance of the used models as proposed by [28] and [47]. In this research, we will apply an image segmentation algorithm to segment the lungs before classification.

The considered aspects below state the contributions of this research:

1) Combining existing CXR Datasets: The main obstacle for a neural network to learn is that there are not enough data examples for training. Currently, as COVID-19 is a new disease, many websites are trying to encourage people and hospitals to contribute and share COVID-19 CXR images from patients all over the world to gather a sufficient number of CXRs and make them public for researchers. In this research, we will not gather our own dataset, but rather we will use a combination of existing ones such as [6] [33] [18]. Images are classified into three classes COVID-19 Pneumonia (CP) vs Community-Acquired Pneumonia (CAP), and Normal cases for our study.

2) Balanced Dataset: Since the available COVID-19 cases are much fewer than healthy cases or even other pneumonia types, image augmentation techniques that can enlarge the minority class, will be used to accomplish a balance between input classes to reach good accuracy of trained model [26].

3) Image Segmentation: One of the main difficulties researchers reported in the detection of lung diseases is the fact that radiographic images can tell that the lungs are abnormal, but they might miss specifying the type of pneumonia exactly. Only the expert radiologist can tell the difference [5]. Therefore, data scientists proposed using image segmentation techniques to segment the lungs so that the model will focus on studying only the lungs and the infection if available [47]. In this research, U-net model proposed by [17], will be used for the segmentation process.

4) Proposed model design: In this research, we will examine using a version of vision transformers called Compact Transformer(CT) [13]. Our proposed method, called Hybrid Compact Transformer (HCT) uses a pretrained CNN in place of the convolutional layers in the original Compact Convolutional Transformer (CCT). In addition, we will reduce the number of transformer encoding layers.

The remainder of this paper is organized as follows: Section II introduces our proposed technique and Section III explains our research methodology. Experimental results are provided in Section IV and the analysis of its results are proposed in Section V. Section VI provides conclusions and suggestions for future improvements.

II. Proposed Model Architecture

A. Segmentation Network

U-Net [17] is a deep learning architecture used for image segmentation problems and was released particularly as a solution for biomedical segmentation tasks. It is the adopted segmentation architecture for any problem domain in Kaggle [33]. Inspired by the early success of previous work [28] [47], and [30], claimed that segmentation can increase the sensitivity of the network such that the network classification result is based on the pixels related to the lung area that contains information about the disease. In this research, we will adopt U-Net to perform semantic segmentation with the benchmark dataset to separate lung and heart contour from the chest radiography images.

Instead of training a Unet model from scratch, a pretrained Unet model, that was shared by [20], is used in this research. This model was pretrained to segment CXR images. In the first phase, all images in the dataset will be segmented using the Unet model. For segmentation, images will be resized to the shape 512 X 512 X 1, as the Unet model architecture requires input images of that size [20]. All segmented images are then saved in the dataset.

TABLE I. CXR DATASETS

Data-set	Classes	Size of images	Pre-processing	Author
Italian Society of Medical and Interventional Radiology [18]	COVID-19 : 68	Image size is not fixed for all images.	Original CXR images.	Asif, Sohaib, et al [3] Ahishali, Mete, et al. [1]
GitHub repository shared by Dr. Joseph Cohen [6]	- Aute Respiratory Distress Syndrome (ARDS): 465 - COVID-19 : 319 - Middle East Respiratory Syndrome (MERS) : 481 - Pneumonia - Severe Acute Respiratory Syndrome (SARS) : 465	Image size is not fixed for all images.	Original CXR images.	Asif, Sohaib, et al [3] Sakib, Sadman, et al. [38] Waheed, Abdul, et al. [46]
Chest Imaging (Spain) at thread reader. [42]	COVID-19 : 50	Image size is not fixed for all images	Original CXR images.	Ahishali, Mete, et al. [1]
Radiopaedia [34]	- COVID-19 : 28 - Pneumonia : 3200	Image size is not fixed for all images.	Original CXR images.	Ahishali, Mete, et al. [1]
Kaggle [33]	- Normal : 1342 - Bacterial Pneumonia : 2561 - Viral Pneumonia : 1345	400 X 2000	Original CXR images.	Rahman, Tawsifur, et al. [36] Ahishali, Mete, et al. [1] Waheed, Abdul, et al. [46]
CheXpert dataset collected by Stanford ML group [19]	Contains 14 different classes with: 224,316 Chest radiographs - No COVID-19 cases	Image size is not fixed for all images.	Original CXR images.	Sakib, Sadman, et al. [38]
Japanese Society of Radiological Technology (JSRT)	247 posteroanterior chest images including normal and lung nodule cases. The images in the database were grouped according to the degree of subtlety of the lung nodule.	2048 X 2048	Original CXR images.	Oh, Yujin, Sangjoon Park, and Jong Chul Ye. [28]
Chest X-14 [48]	Contains 112,120 CXR images. 14 common thoracic diseases classes.	1024 X 1024	Original CXR images.	Wang, Kun, et al. [47]

B. Classification Network

Transformers have rapidly been increasing in popularity and have become a major focus of modern deep learning research. They are emerging as an architectural paradigm alternative to CNNs [7]. ViTs can capture global relations between image elements and they potentially have a larger representation capacity. However, the fact that ViTs are extremely data hungry and require large sets of data to be trained, leads to the exclusion of those with limited resources from research in the field [13] [43]. Moreover, based on the findings of an earlier study [32], the use of existing ViT is not optimal in the field of pneumonia detection, since feature embedding through direct patch flattening is not intended for CXR images. Fig. 1 shows the original architecture of ViT architecture.

Today, researchers persist to dispel the myth that ViTs are data-hungry and can only be applied to huge datasets. Several updates on the architecture of ViT have been proposed like [43] [23], to reduce the expansive computations of the self-attention mechanism and to escape the big data paradigm. In this study, we will use an updated version of ViT called Compact Transformer or CT proposed by [13].

1) Compact Convolutional Transformer (CCT): In [13], the author's contribution is escaping the big data paradigm, as most medical applications with AI suffer from limited data availability. The authors in [13] split their experiments into two phases: In the first phase, they proposed an updated version of the original ViT architecture (Compact Vision Transformer(CVT)), and in the second phase, they proposed using a shallow NN, composed of one convolutional layer and a Relu activation layer, to create the patches instead of extracting them directly from the original input image (Compact Convolutional Transformer (CCT)). Fig. 2 shows the architectures of both CCT and CVT.

The results show that CCT in comparison to CVT can be quickly trained from scratch on small datasets while achieving high accuracy. Based on the findings, it can be argued that transformers can perform head-to-head with state-of-the-art CNNs on small sets of data, often with better model performance, better accuracy, and fewer parameters.

In our study, we propose using a Hybrid Compact Transformer (HCT) in which a pretrained CNN is used in place of the convolutional layers in CCT. The attention mechanism allows transformer architecture to compute in parallelized manner. It can simultaneously extract all the information we need from the input and its inter-relation, compared to CNNs. CNNs are much more localized, using small filters to compress information towards a general answer. While this architecture is powerful for general classification tasks, it does not have the spatial information necessary for many tasks like instance recognition [7]. This is because convolution does not consider distanced-pixel relations (Fig. 3A), like in Vision Transformers where the attention of patches of the images and their relations to each other are calculated (Fig. 3B).

Hence, with the hybrid model design, we aim to combine the localization power of CNNs, with the generalization power (attention mechanism or distanced-pixel relations) of ViTs. Fig. 4 shows an overview of the proposed Hybrid Compact Transformer (HCT). This will allow to achieve good performance even with reduced datasets.

2) Proposed Model Design (Hybrid Compact Transformer): In the proposed HCT model design, a pretrained CNN (transfer learning) is used as a backbone feature extractor and its outputs

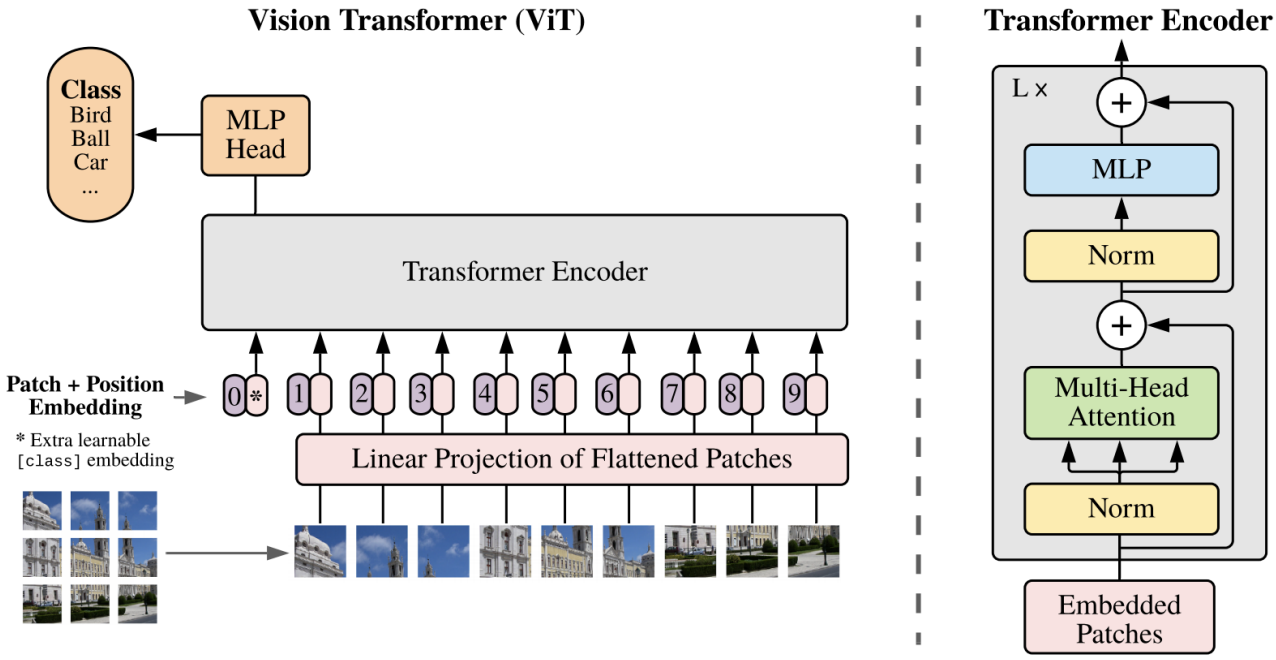


Fig. 1. Vision transformer [7].

are fed to the Compact Transformer. The input CXR image is primarily fed into a backbone CNN to create a feature map then into the compact transformer for classification. The HCT model design is illustrated in Fig. 5.

The architecture of the compact transformer differs from the original Vision Transformer architecture in the following variants, fewer transformer encoder layers (only 2 layers) and smaller dimensions of patches. A transformer encoder consists of a series of stacked encoding layers. Thus, each encoder layer consists of two sub-layers: a Multi-Layer Perceptron (MLP) head and Multi-headed Self-Attention (MSA). Each sub-layer is followed by a layer normalization (LN), and followed by a residual connection to the next sub-layer, as illustrated in Fig. 1. The compactness in the CT results in a lightweight vision transformer with as few as 200K learnable parameters only. Furthermore, in the CT, a novel sequence pooling technique was introduced to remove the needs of the class token. This sequence pooling is a linear layer placed after the transformer encoder to make the model more compact and accurate. Eventually, the output of the pooling layer is then fed into the final classification layer, as in the original ViT model, to classify the image into one of three classes: CP, CAP, or Normal.

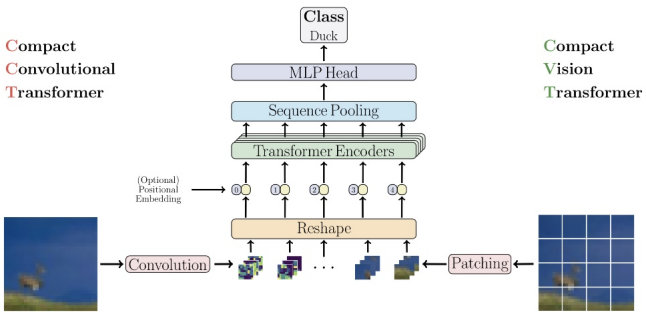


Fig. 2. Overview of CCT vs CVT [13].

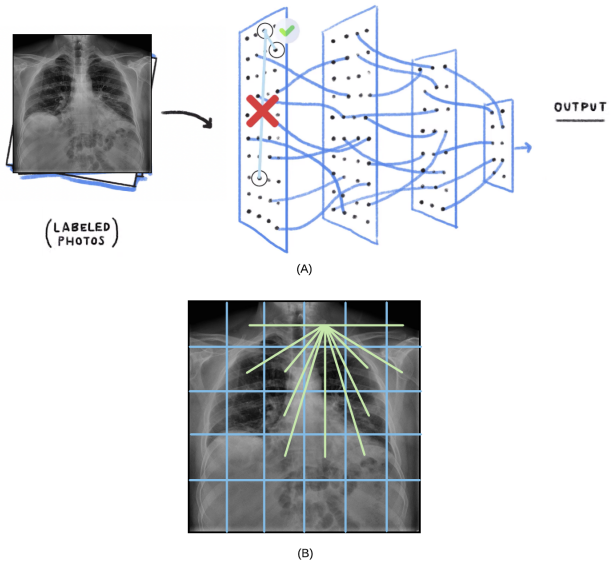


Fig. 3. Illustration of the localization power of CNNs (A), and the generalization power (attention mechanism or distanced-pixel relations) of ViTs (B).

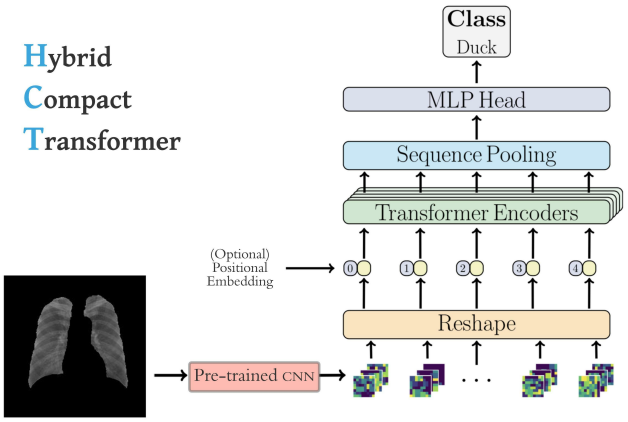


Fig. 4. Overview of Hybrid Compact Transformer (HCT).

TABLE II. TOTAL NUMBER OF CXR IMAGES BEFORE APPLYING AUGMENTATION

Class	Total Number of Images
Normal	5800
COVID-19 CXRs (CP)	6500
Non-COVID-19 CXRs (CAP)	6100

III. RESEARCH METHODOLOGY

A. Dataset

A careful analysis of the available datasets (Table I) will let us decide which data to consider in the experimental study and which possible preprocessing to be use.

We can summarize the results as follows: in our research, we will utilize many of these datasets proposed and used by previous works. Datasets, like [6], shared by Dr. Joseph Cohen, were used in most of the previous works. This dataset contains a good number of COVID-19 CXRs in comparison to all the others. The Kaggle dataset contains images of good quality and resolution for normal and other pneumonia cases. The quality of CXR images helps for better segmentation and model learning. Table II shows the total number of CXR images before to augmentation.

B. Image Augmentation

An investigation of enlarging the number of CXR images by using two different image processing methodologies was discussed in [24]. The main idea in this paper was to evaluate the test accuracy of five pretrained models (AlexNet, VGGNet16, VGGNet19, GoogleNet, and ResNet50) in four scenarios: the first scenario, when using the original dataset, second scenario, when performing data augmentation using classical data augmentation techniques [26] on input dataset to enrich the COVID-19 CXR images, the third scenario, when performing Generative Adversarial Network (GAN) [8]. And the last scenario, when combining classical data augmentation and GAN to generate more CXR images. The results show that ResNet50 is the best deep learning classifier and with augmented COVID-19 medical CXR images achieved the best performance to detect the COVID-19 from the input dataset.

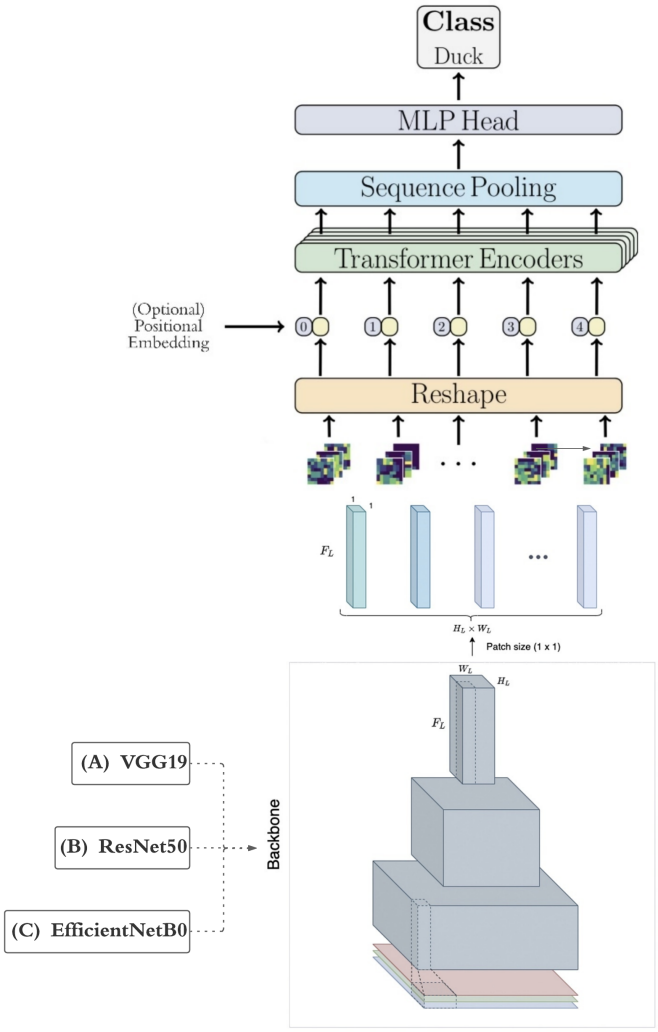


Fig. 5. Illustration of the proposed model design, HCT network. In this study, the backbone or the features extractor is going to be one of three state-of-the-art networks, namely, (A) VGG19, (B) ResNet50, and (C) EfficientNetB0.

TABLE III. TOTAL NUMBER OF CXR IMAGES AFTER APPLYING AUGMENTATION

Class	Total Number of Images
Normal	8000
COVID-19 CXRs (CP)	8000
Non-COVID-19 CXRs (CAP)	8000

In this research different methods for data augmentation techniques such as zooming, rotation, shifting, and flipping [26] were selected to be applied to the original dataset. In the end, the total number of CXR images we have in our dataset is 24000 images as shown in Table III.

C. Classification Performance Metrics

The most common performance measures in the field of deep learning are accuracy, precision,specificity, recall (or sensitivity), and Fβ score [10]. In this research we will utilize all these five performance measures in addition to the confusion matrix to measure and analyze the classification results of the CNN models.

The formulas of the measures are given below, where true positive (TP) is the correctly identified predictions for each class. True negative (TN) is the correctly rejected predictions for a particular class. False positive (FP) is the incorrectly identified predictions for a particular class, and false negative (FN) is the incorrectly rejected predictions for a certain class. After the completion of training phase, the performance of different networks for testing dataset is evaluated. The 5 metrics for performance evaluation were calculated as below:

1) Confusion Matrix: A confusion matrix is used to evaluate the results of a predictive model [49]. This matrix will be generated after making predictions on the test data. For a classifier with n output classes, the predicted values on test instances and the actual values are represented as an nxn matrix. Each row of the confusion matrix represents the samples in a predicted class, and each column represents the samples in an actual class.

2) Accuracy: It is the percentage of correct predictions among all decisions made on examples of a dataset. For a classification problem with two classes, (Eq. 1) defines Accuracy in terms of TP, TN, FP, and FN:

$$Accuracy = \frac{(TP + TN)}{((TP + FN) + (FP + TN))} \tag{1}$$

3) Precision: It represents the percentage of the classifier's correct decisions in favor of the target class (Eq. 2).

$$Precision = \frac{(TP)}{(TP) + FP)} \tag{2}$$

4) Recall (Sensitivity): It represents the percentage of the classifier's correct decisions made on the other class (Eq.3). In particular, sensitivity is the classifier's ability to identify a diseased person as diseased. Sensitivity is also known as True Positive Rate (TPR).

$$Recall(Sensitivity) = \frac{(TP)}{(TP) + (FN)} = TPR \tag{3}$$

5) Specificity: It represents the percentage of the classifier's correct decisions made on the other class (Eq. 4). That is, specificity is the ability of the classifier to correctly identify a healthy person as non-diseased while not confusing a diseased one as healthy [46]. Specificity is also known as True Negative Rate (TNR).

$$Specificity = \frac{(TN)}{(TN) + (FP)} = TNR \tag{4}$$

6) ROC curve (The Receiver Operating Characteristics curve): One of the most important evaluation metrics for checking any binary classification model's performance [50]. The ROC curve is a graphical plot, plotted with the TPR (Eq. 3), against the False Positive Rate (FPR) (Eq. 5), where TPR is on the y-axis and FPR is on the x-axis. The closer the ROC curve towards the upper left corner, the better the model's performance. The curve is created with the two variables at various threshold settings [44].

$$FPR = 1 - TNR \tag{5}$$

7) AUROC curve (Area Under The Receiver Operating Characteristics curve): Area Under the ROC curve is often used as a classification model's quality measurement. In this research, we will utilize this performance measure for each of the two binary classifiers in the two-stage CNN to visualize how well our proposed cascade classifier is performing. The AUROC for an optimal classifier is 1. In practice, the AUROC value is usually between 0.5 and 1 [44].

8) F1 Score: F1 score is the harmonic mean of precision and recall (Eq. 6) [10] . As sensitivity and precision are both important measurements in medical applications, the β value is 1 as shown in (Eq. 7).

$$F\beta = (1 + \beta^2) * \frac{(Precision * Recall)}{(\beta^2 * Precision) + Recall} \tag{6}$$

$$F1 = 2 * \frac{(Precision * Recall)}{(Precision + Recall)} \tag{7}$$

D. Pretrained Backbone CNN Models

In this research, we will investigate the detection of COVID-19 with three classifiers ResNet50 and VGG19 on the benchmark dataset. Additionally, we will study the efficiency of a new family of models invented by [40], called EfficientNets. It has been demonstrated in the paper and as the name suggests, EfficientNets are computationally efficient and achieve much better accuracy and efficiency than previous CNNs. We will study the performance of EfficientNetB0 on our benchmark dataset.

E. Experimental Setup

Python programming language was used to train the proposed deep transfer learning models. Considering having a large dataset, all experiments were performed on Google Colaboratory Pro (Google Colab Pro) on Mac Operating System using the online cloud service with Graphics Processing Unit (GPU) hardware for free. Colab Pro supports longer running notebooks, access to faster GPUs and TPUs, and provides high RAM [9]. CNN models (VGG19, ResNet50, and EfficientNetB0) were pretrained on ImageNet weights.

TABLE IV. TESTING ACCURACY OBTAINED BY THE HCT MODELS TRAINED WITH DIFFERENT BATCH SIZES

Model	Batch Size		
	32	64	128
HCT (VGG19)	92.00%	90.04%	90.62%
HCT (ResNet50)	96.82%	96.70%	97.25%
HCT (EfficientNetB0)	98.35%	98.38%	98.63%

IV. EXPERIMENTAL RESULTS

A. Segmentation Performance

Based on the segmentation results on our dataset, generally, the pretrained Unet model performance is great. Lung segmentation before classification helped in removing the irregular regions and the irrelevant objects (e.g. medical devices). An example of the segmentation results is shown in Fig. 6.

B. Classification Performance

To illustrate, all following experiments were implemented using the same dataset for multiclass classification. Moreover, the entire dataset was randomly split into training (70%), validation (20%), and testing (10%).

1) Optimal learning rate selection: To determine the optimal learning rate for all models, the models were trained using three different learning rates, .0001, .00001, and .000001. The best learning rate of a model was chosen based on the minimum loss, as such, the optimal learning rate for all three networks is 0.000001 as shown in Fig. 7.

2) Results comparison with different batch sizes: In this research, the impact of batch size on testing accuracy is studied. Table IV demonstrates the test accuracy of all networks when trained using three different batch sizes. Based on the findings, it can be argued that a stable and higher testing performance is obtained for ResNet50 and EfficientNetB0 models with a batch size of 128, and for VGG19 model with a batch size of 32.

3) Stratified K-fold Cross-Validation Results:

- Hybrid Compact Transformer (HCT) model Results:
  After training and testing, the accuracies achieved by the proposed HCT models were 92.00%, 97.25%, and 98.63% when using different features extractors: VGG19, ResNet50, and EfficientNetB0, respectively. The batch size, learning rate, and the number of epochs were experimentally set to 128, 0.000001, and 100, respectively for all models except for HCT-VGG19 the training procedure was completed in 200 epochs using a batch size of 32.
- Compact Convolutional Transformar (CCT) model Results:
  As proposed by the author of the CCT model, the batch size, learning rate, and the number of epochs were experimentally set to 128, 0.001, and 100, respectively. The model achieved an over all test accuracy of 81.79%.
- Transfer Learning (TL) model Results:
  In the fine-tuning of the individual DCNNs, we tried various combinations of batch sizes and learning rates to get the best models' performance. The batch size, learning rate, and the number of epochs, for all networks, were experimentally set to 64, 0.0001, and 100, respectively, except for VGG19 the batch size used was 32. The overall test accuracies achieved were 89.52%, 94.62%, and 96.74% for VGG19, ResNet50, and EfficientNetB0, respectively.

The detailed classification results obtained from all networks are compared in terms of various metrics and are tabulated in Table V. As can be seen from these results, the HCT model produced better accuracy scores than the fine-tuned deep CNN models. This result was considered reasonable as adding one simple ViT encoding layer (generalization power) can enhance the power of Deep CNNs. The best accuracy score overall was 98.63%, and was produced by HCT with EfficientNetB0 as backbone. It can be noticed that EfficiantNetB0 produced the best accuracies for both fine-tuned CNNs and HCT models.

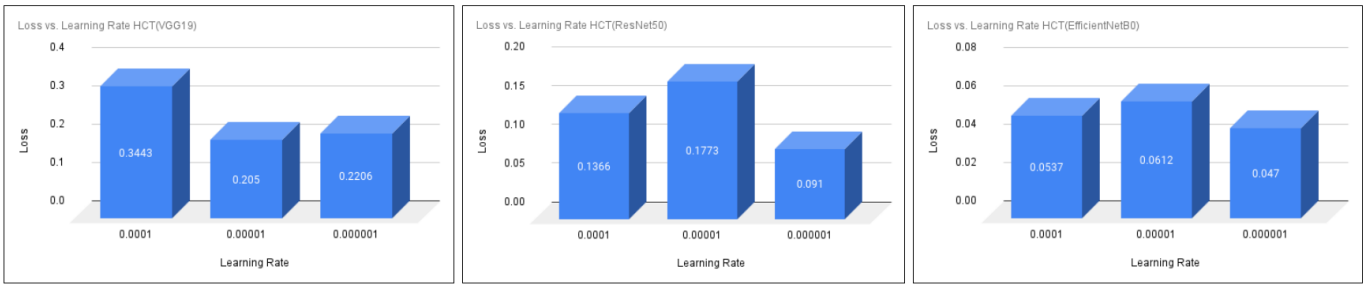
C. Analysis of Models Execution Results

1) Sensitivity-Specificity Analysis: In a diagnostic test (Neural Network), the network's sensitivity (True Positive Rate) refers to the network's ability to correctly classify COVID-19 patients who have the condition, whilst specificity (True Negative Rate) is a measure of how well a network can identify those who do not have the condition [51]. In medical applications, it is always preferable to have a network with high sensitivity such that the probability that a network produces false negatives is low [51]. FN is the proportion of positives (COVID-19) that are mislabeled as negatives (Normal) by the model. These false negatives are crucial and might threaten humans' life. Table VI and Table VII show comparisons between the fine-tuned CNN networks and the the proposed ones in terms of Sensitivity and Specificity, respectively.

It can be observed that the COVID-19 class sensitivities of the proposed models are higher than the sensitivities of the fine-tuned models. Compared to the fine-tuned models, the overall sensitivities and specificities obtained by the proposed models are higher, which confirms the efficacy of our method.

2) Confusion matrix: We presented the confusion matrices on the test data of all seven models in Fig. 8. Fig. 8(A) presents the confusion matrix obtained by our proposed HCT model, using VGG19 as a backbone feature extractor. While 757 COVID-19 samples, 739 Pneumonia samples, and 717 Normal (healthy) samples were classified correctly, 43 COVID-19 samples, 61 Pneumonia samples, and 83 Normal (healthy) samples were misclassified. Therefore, the rate of correct classification of COVID-19 samples was 95%, whilst it was 92.5% for the Pneumonia samples and was 89.62% for the Normal (healthy) cases. On the other hand, Fig. 8(B) presents the confusion matrix obtained by the HCT model, using ResNet50 as a backbone features extractor, the rate of correct classification of COVID-19 samples was 97.5%, whilst it was 98.87% for the Pneumonia samples and was 94.37% for the Normal (healthy) cases. Further, Fig. 8(C) presents the confusion matrix obtained by the HCT model, using EfficientNetB0 as a

Fig. 6. One example of the segmentation result.



(A) (B) (C)

Fig. 7. Histogram represents the loss obtained by HCT Models for different learning rates, using: (A) VGG19, (B) ResNet50, (C) EfficientNetB0 as a backbone feature extractor.

backbone features extractor, the rate of correct classification of COVID-19 samples was 98.37%, whilst it was 98.75% for the Pneumonia samples and was 98.75% for the Normal (healthy) cases. It can be observed that our proposed method with EfficientNetB0 was able to classify 98.37% of COVID-19 infection cases accurately.

In addition, we also wanted to show the confusion matrices obtained by the CCT model and the individual fine-tuned networks, to compare their classification performance to the performance of the HCT proposed models. It can be argued that the False Positives(FP) and False Negatives(FN) has been reduced for each CNN using HCT. Similarly, if we compare the percentage of the correctly classified samples between (HCT-VGG19 and pretrained VGG19), (HCT-ResNet50 and pretrained ResNet50), and (HCT-EfficinetNeB0 and pretrained EfficinetNeB0), it can be observed that HCT models yielded a higher classification accuracy for all the three classes.

Whereas the FPs and FNs obtained by the CCT model are the highest among all the seven models. Moreover, it can be observed from the confusion matrix reported in Fig. 8, that the HCT-VGG19 proposed model has detected 757 out of 800 with COVID-19 as having COVID-19, achieving a COVID-19 class sensitivity of .946 compared to the fine-tuned VGG19 that has achieved a COVID-19 class sensitivity of .924 (see Table VI). The HCT-ResNet50 proposed model has detected 787 out of 800 with COVID-19 as having COVID-19, achieving a COVID-19 class sensitivity of .975 compared to the fine-tuned ResNet50 that has achieved a COVID-19 class sensitivity of .965. Our best-proposed HCT-EfficientNetB0 model has detected 787 out of 800 with COVID-19 as having COVID-19, achieving a COVID-19 class sensitivity of .984 compared to the fine-tuned EfficientNetB0 that has achieved a COVID-19 class sensitivity of .97.

3) The AUROC: ROC curves are typically used in binary classification to study the output of a classifier. To extend the ROC curve and ROC area to multi-class classification, there are two averaging strategies: one-vs-rest (OvR) and one-vs-one (OvO) [39]. In this study, we deployed the OvR algorithm, which computes the average of the ROC scores for each class against all other classes. Fig. 9 shows the AUROC curves of all seven models implemented in this research. The achieved AUC values for the fine-tuned models and the HCT models are above 0.97, which confirms that these models are reliable when performing the classification. Whereas the CCT model achieved an AUC value of 0.93.

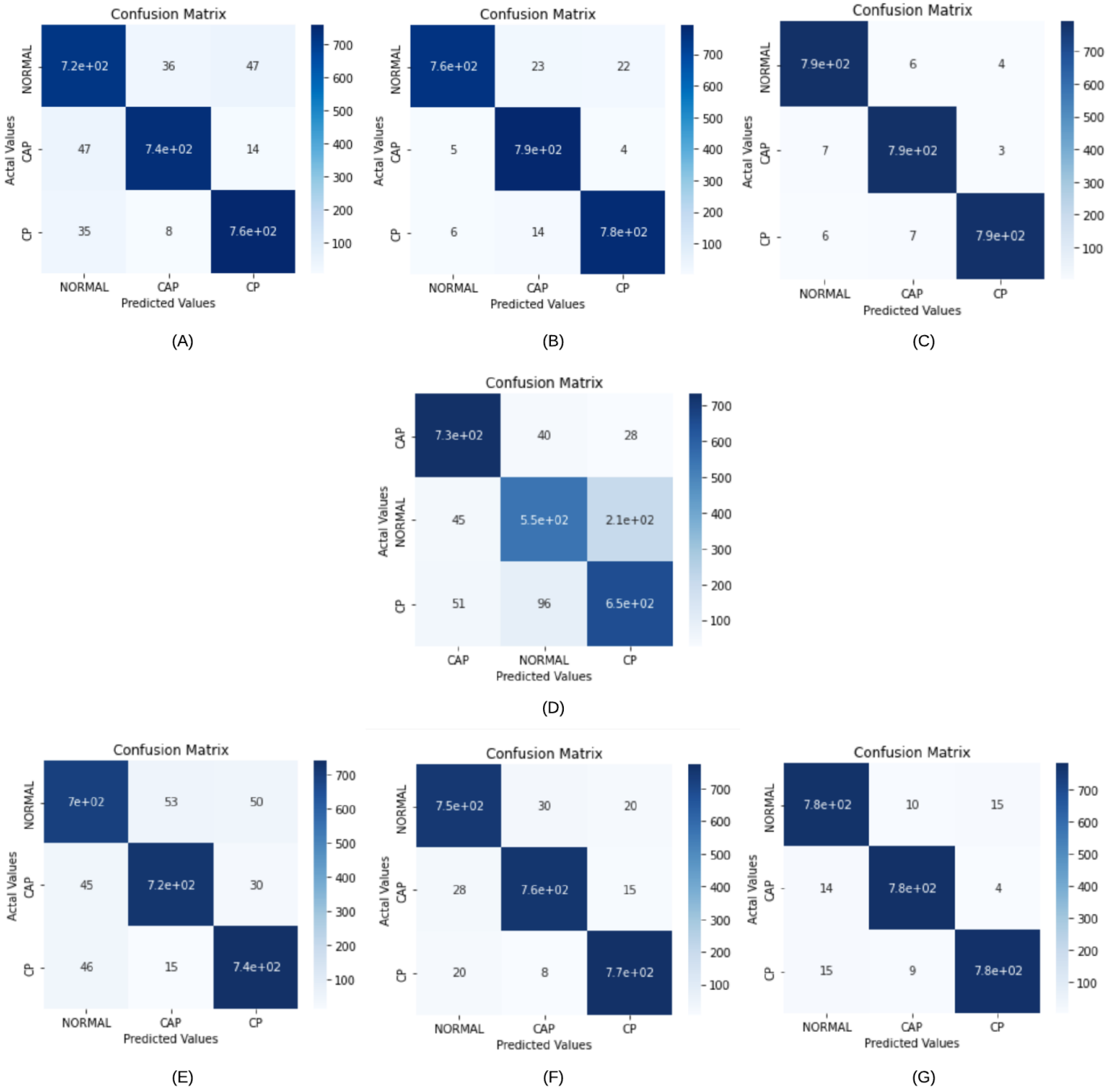


Fig. 8. Confusion matrix obtained by all the seven models implemented in this research: (A) HCT-VGG19, (B) HCT-ResNet50, (C) HCT-EfficientNetB0, (D) CCT, (E) pretrained-VGG19, (F) pretrained-ResNet50, (G) pretrained-EfficientNetB0.

TABLE V. SUMMARY OF ALL EXPERIMENTS FOR MULTI-CLASS CLASSIFICATION

Model	pretrained CNN	Accuracy	Precision	Recall	Specificity	F1 Score
TL	VGG19	89.52%	0.8811	0.8597	0.9420	0.8702
	ResNet50	94.62%	0.9737	0.9176	0.9737	0.9315
	EfficientNetB0	96.74%	0.9591	0.967	0.9793	0.9630
CCT		81.79	0.7856	0.7455	0.8972	0.7641
HCT	VGG19	92.00%	0.9217	0.9479	0.9597	0.9346
	ResNet50	97.25%	0.9702	0.9672	0.9851	0.9687
	EfficientNetB0	98.63%	0.9857	0.9862	0.9928	0.9860

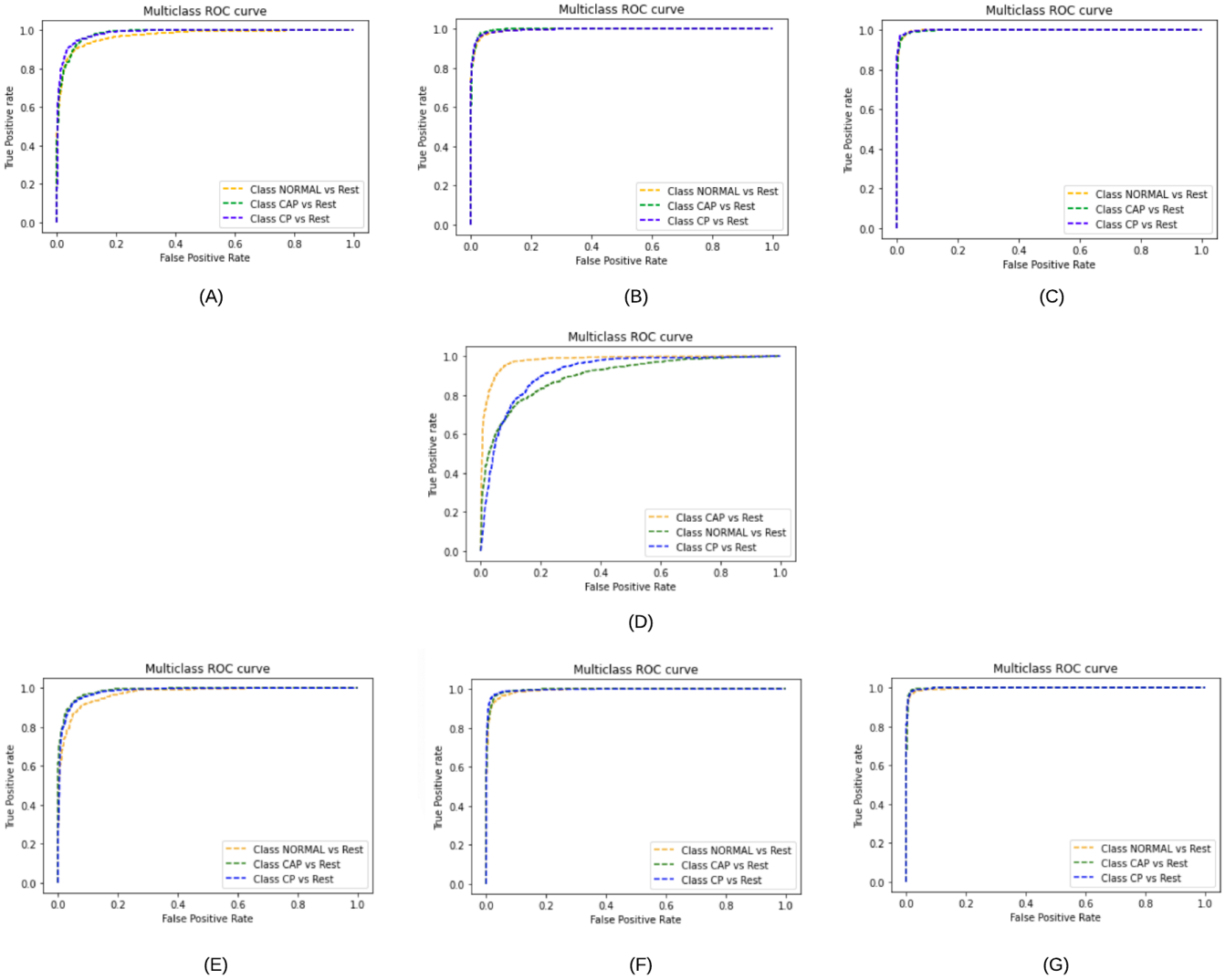


Fig. 9. ROC curve obtained by all the seven models implemented in this study: (A) HCT-VGG19, (B) HCT-ResNet50, (C) HCT-EfficientNetB0,(D) CCT, (E) pretrained-VGG19, (F) pretrained-ResNet50, (G) pretrained-EfficientNetB0.

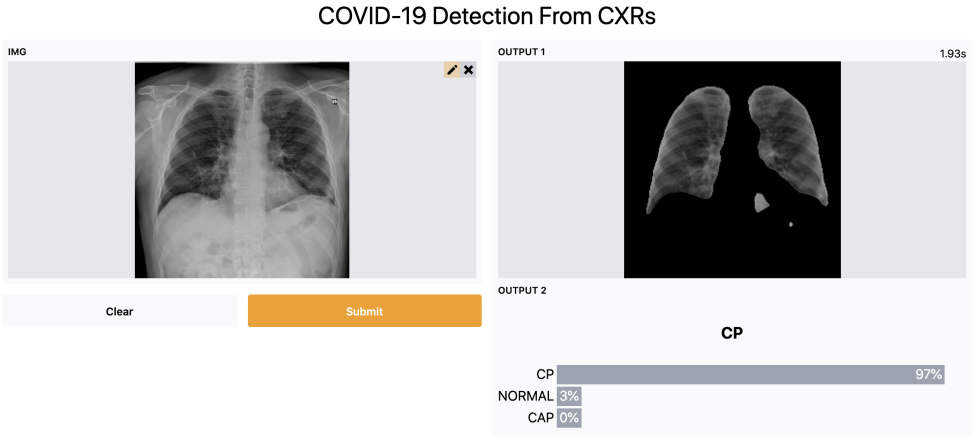


Fig. 10. Illustration of a classification result of a COVID-19 CXR image, obtained by the HCT model.

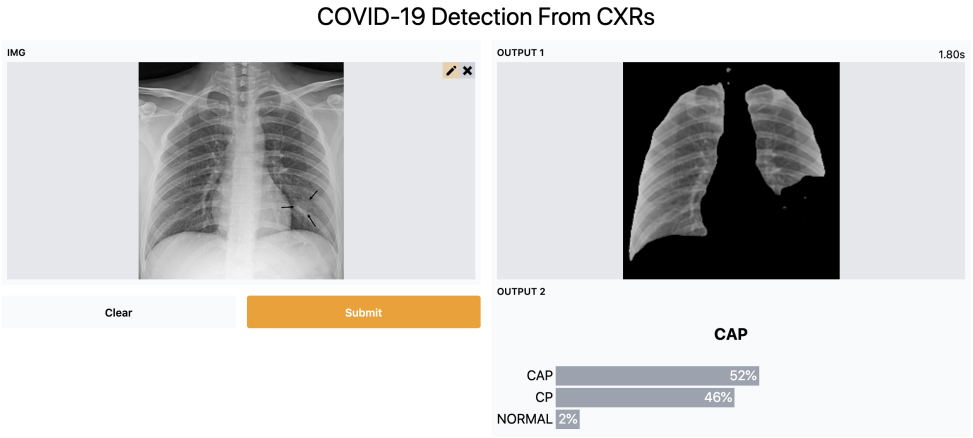


Fig. 11. Illustration of a misclassification result of a COVID-19 CXR image, obtained by the HCT model.

TABLE VI. A COMPARISON BETWEEN THE FINE-TUNED AND THE PROPOSED MODELS IN TERMS OF THEIR SENSITIVITIES

CNN \Backbone CNN Model	TL Models	HCT Models
VGG19	0.8597	0.9479
ResNet50	0.9176	0.9672
EfficientNetB0	0.967	0.9862

V. ANALYSIS AND DISCUSSIONS

A. Analysis of Classification Results

To demonstrate the performance of HCT, we randomly selected a COVID-19 CXR image, gave it input to the network, and acquired output on the image as shown in Fig. 10. The classification result was obtained by the best HCT model using EfficientNetB0 as a features extractor. Nevertheless, there is some confusion from the model sometimes between COVID-19, and Pneumonia samples, Fig. 11. This misclassification was possibly occurred because of our dataset. The Pneumonia samples in our dataset include viral and bacterial pneumonia. Considering that COVID-19 itself is a viral pneumonia disease. Furthermore, the GUI is created with Gradio, which is an open-source python library. Gradio permits researchers to quickly create easy-to-use and customizable UI components for their ML model [11].

TABLE VII. A COMPARISON BETWEEN THE FINE-TUNED AND THE PROPOSED MODELS IN TERMS OF THEIR SPECIFICITIES

CNN \Backbone CNN Model	TL Models	HCT Models
VGG19	0.9420	0.9597
ResNet50	0.9737	0.9851
EfficientNetB0	0.9793	0.9928

B. Comparison with State-of-the-Art Methods

In order to evaluate the proposed HCT model, the general performance comparison of our study with the state-of-art methods is given in this section. All networks were trained using our benchmark dataset. It can be observed that the proposed HCT-EfficientNetB0 model achieved higher performance than the other existing schemes. The results that are reported in this section are summarized in Table VIII.

Khan et al. [21] propose CoroNet, a Deep Convolutional Neural Network based on Xception architecture pretrained on

TABLE VIII. COMPARISON WITH STATE-OF-THE-ART STUDIES

Study	Model Used	Accuracy
Asif et al. [3]	Inception-v3	94.58%
Ozturk et al. [31]	DarkCovidNet	96.89%
Khan et al. [21]	CoroNet	89.30%
Mahmud et al. [25]	CovXNet	90.75%
Narin et al. [27]	Deep CNN ResNet-50	94.62%
Apostolopoulos & Mpesiana [2]	VGG19	89.52%
Proposed Method	HCT-VGG19	92%
	HCT-ResNet50	97.25%
	HCT-EfficientNetB0	98.63%

ImageNet dataset. The CoroNet model obtained an overall accuracy of 89.30%. Ozturk et al. [31] present DarkCovidNet design to the diagnosis of COVID-19. In their study, the main model was inspired by the DarkNet architecture that has proven itself in deep learning. Their model produced a 96.89% overall accuracy. Mahmud et al. [25] propose a deep neural network named CovXNet. The network architecture utilizes depth-wise convolution with varying dilation rates for efficiently extracting diversified features from CXRs. CovXNet achieved an overall accuracy of 90.75%. Oh et al. Further, Asif et al. [3], Narin et al. [27], and Apostolopoulos and Mpesiana [2], use transfer learning and obtained an overall accuracy of 94.58%, 94.62%, 89.52%, respectively. Indeed, DarkCovidNet achieved the best performance among the other proposed methods with a test accuracy of 96.89%. However, our proposed HCT-EfficientNetB0 surpasses DarkCovidNet and the other proposed methods with a test accuracy of 98.63%.

VI. CONCLUSIONS

In this research, we studied the deployment of deep learning models for COVID-19 detection using CXRs. As has been shown in research such as [24] [35] [2] [27], Transfer Learning (TL) gives the best classification results for the problem of pneumonia detection from CXR images. In our study, we investigated the use of our proposed Hybrid Compact Transformer (HCT), in which we integrate TL with Vision Transformers (ViTs) in one model. We aim to combine the localization power of CNNs, with the generalization power of ViTs. Based on the experimental results, HCT has shown satisfactory improvements over the respective accuracies of compact transformers (CTs) and models based on TL.This result could be useful for dealing with future respiratory pandemics where at their beginnings, only a few CXRs would be available for researchers.

Choosing the pretrained model can largely affect the accuracy of the final classifier. Performance results show that EfficientNetB0 pretrained model yielded the best performance among the three models. The proposed classification model with EfficientNetB0 as a feature extractor achieved more than 98% accuracy.

As a future research direction, we intend to increase the number of classes in our dataset to include: COVID-19, Viral-Pneumonia, Bacterial-Pneumonia, and Normal cases in order to have a more applicable model. We also want our model to be more interpretable. Thus, we will try to use interpretable saliency maps to correlate with the radiological findings. Moreover, the segmentation results could be further improved by training the Unet model with CLAHE enhanced CXR images, as mentioned in [41]. Authors in [28] proposed FCDenseNet103 as a backbone segmentation network architecture. They claimed that in comparison with Unet, FCDenseNet103 has higher segmentation performance. So using FCDenseNet103 on our dataset could improve the performance of our final proposed model. It would also be very interesting to assess the extent of usefulness of image segmentation in light of using pretrained models, like in our solution.

ACKNOWLEDGMENT

The authors gratefully acknowledge financial support from The Deanship of Scientific Research, King Faisal University (KFU) in Saudi Arabia. The present work was done under research project Number (KFU242418).

REFERENCES

[1] Mete Ahishali, Aysen Degerli, Mehmet Yamac, Serkan Kiranyaz, Muhammad EH Chowdhury, Khalid Hameed, Tahir Hamid, Rashid Mazhar, and Moncef Gabbouj. A comparative study on early detection of covid-19 from chest x-ray images. *arXiv preprint arXiv:2006.05332*, 2020.

[2] Ioannis D Apostolopoulos and Tzani A Mpesiana. Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. *Physical and Engineering Sciences in Medicine*, page 1, 2020.

[3] Sohaib Asif, Yi Wenhui, Hou Jin, Yi Tao, and Si Jinhai. Classification of covid-19 from chest x-ray images using deep convolutional neural networks. *medRxiv*, 2020.

[4] Muhammad EH Chowdhury, Tawsifur Rahman, Amith Khandakar, Rashid Mazhar, Muhammad Abdul Kadir, Zaid Bin Mahbub, Khandaker Reajul Islam, Muhammad Salman Khan, Atif Iqbal, Nasser Al-Emadi, et al. Can ai help in screening viral and covid-19 pneumonia? *arXiv preprint arXiv:2003.13145*, 2020.

[5] Michael Chung, Adam Bernheim, Xueyan Mei, Ning Zhang, Mingqian Huang, Xianjun Zeng, Jiufa Cui, Wenjian Xu, Yang Yang, Zahi A Fayad, et al. Ct imaging features of 2019 novel coronavirus (2019-ncov). *Radiology*, 295(1):202–207, 2020.

[6] Joseph Paul Cohen, Paul Morrison, and Lan Dao. Covid-19 image data collection. *arXiv 2003.11597*, 2020.

[7] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[8] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. pages 2672–2680, 2014.

[9] google colab pro. https://colab.research.google.com/signup. Accessed: 2021-9-20.

[10] Cyril Goutte and Eric Gaussier. A probabilistic interpretation of precision, recall and f-score, with implication for evaluation. In *European conference on information retrieval*, pages 345–359. Springer, 2005.

[11] Gradio. https://gradio.app/. Accessed: 2021-12-3.

[12] Varun Gulshan, Lily Peng, Marc Coram, Martin C Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, Kasumi Widner, Tom Madams, Jorge Cuadros, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*, 316(22):2402–2410, 2016.

[13] Ali Hassani, Steven Walton, Nikhil Shah, Abulikemu Abuduweili, Jiachen Li, and Humphrey Shi. Escaping the big data paradigm with compact transformers. *arXiv preprint arXiv:2104.05704*, 2021.

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[15] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. pages 4700–4708, 2017.

[16] Forrest N Iandola, Song Han, Matthew W Moskewicz, Khalid Ashraf, William J Dally, and Kurt Keutzer. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and¡ 0.5 mb model size. *arXiv preprint arXiv:1602.07360*, 2016.

[17] Jyoti Islam and Yanqing Zhang. Towards robust lung segmentation in chest radiographs with deep learning. *arXiv preprint arXiv:1811.12638*, 2018.

[18] Italian Society of Medical and Interventional Radiology. https://www.sirm.org/en/italian-society-of-medical-and-interventional-radiology/. Accessed: 2020-09-09.

[19] Irvin Jeremy, Rajpurkar Pranav, Ko Michael, Yu Yifan, Ciurea-Ilcus Silviana, Chute Chris, Marklund Henrik, Haghgoo Behzad, Ball Robyn, Shpanskaya Katie, et al. Mong david a. In *Halabi Safwan S., Sandberg Jesse K., Jones Ricky, Larson David B., Langlotz Curtis P., Patel Bhavik N., Lungren Matthew P., Ng Andrew Y. CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison. Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 590–597, 2019.

[20] Kaggle. https://www.kaggle.com/nikhilpandey360/lung-segmentation-from-chest-x-ray-dataset/output?select=cxr_reg_weights.best.hdf5. Accessed: 2021-9-1.

[21] Asif Iqbal Khan, Junaid Latief Shah, and Mohammad Mudasir Bhat. Coronet: A deep neural network for detection and diagnosis of covid-19 from chest x-ray images. *Computer Methods and Programs in Biomedicine*, 196:105581, 2020.

[22] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[23] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. *arXiv preprint arXiv:2103.14030*, 2021.

[24] Mohamed Loey, Gunasekaran Manogaran, and Nour Eldeen M Khalifa. A deep transfer learning model with classical data augmentation and cgan to detect covid-19 from chest ct radiography digital images. *Neural Computing and Applications*, pages 1–13, 2020.

[25] Tanvir Mahmud, Md Awsafur Rahman, and Shaikh Anowarul Fattah. 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, 2020.

[26] Agnieszka Mikołajczyk and Michał Grochowski. Data augmentation for improving deep learning in image classification problem. In *2018 international interdisciplinary PhD workshop (IIPhDW)*, pages 117–122. IEEE, 2018.

[27] Ali Narin, Ceren Kaya, and Ziynet Pamuk. Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. *arXiv preprint arXiv:2003.10849*, 2020.

[28] Yujin Oh, Sangjoon Park, and Jong Chul Ye. Deep learning covid-19 features on cxr using limited training data sets. *IEEE Transactions on Medical Imaging*, 2020.

[29] World Health Organization. Diagnostic testing for sars-cov-2: interim guidance, 11 september 2020. Technical report, World Health Organization, 2020.

[30] Xi Ouyang, Jiayu Huo, Liming Xia, Fei Shan, Jun Liu, Zhanhao Mo, Fuhua Yan, Zhongxiang Ding, Qi Yang, Bin Song, et al. Dual-sampling attention network for diagnosis of covid-19 from community acquired pneumonia. *IEEE Transactions on Medical Imaging*, 2020.

[31] Tulin Ozturk, Muhammed Talo, Eylul Azra Yildirim, Ulas Baran Baloglu, Ozal Yildirim, and U Rajendra Acharya. Automated detection of covid-19 cases using deep neural networks with x-ray images. *Computers in Biology and Medicine*, page 103792, 2020.

[32] Sangjoon Park, Gwanghyun Kim, Yujin Oh, J. Seo, Sang Min Lee, Jin Hwan Kim, Sungjun Moon, Jae-Kwang Lim, and Jong-Chul Ye. Vision transformer using low-level chest x-ray feature corpus for covid-19 diagnosis and severity quantification. *ArXiv*, abs/2104.07235, 2021.

[33] Paul Mooney. Chest x-ray images (pneumonia). https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia. Accessed: 2020-09-22.

[34] Radiopaedia. Cases. https://radiopaedia.org/encyclopaedia/cases/chest?lang=us&modality=X-ray&page=1#collapse-by-diagnostic-certainties. Accessed: 2020-09-10.

[35] Md Mamunur Rahaman, Chen Li, Yudong Yao, Frank Kulwa, Mohammad Asadur Rahman, Qian Wang, Shouliang Qi, Fanjie Kong, Xuemin Zhu, and Xin Zhao. Identification of covid-19 samples from chest x-ray images using deep learning: A comparison of transfer learning approaches. *Journal of X-ray Science and Technology*, (Preprint):1–19, 2020.

[36] Tawsifur Rahman, Muhammad EH Chowdhury, Amith Khandakar, Khandaker R Islam, Khandaker F Islam, Zaid B Mahbub, Muhammad A Kadir, and Saad Kashem. Transfer learning with deep convolutional neural network (cnn) for pneumonia detection using chest x-ray. *Applied Sciences*, 10(9):3233, 2020.

[37] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225*, 2017.

[38] Sadman Sakib, Tahrat Tazrin, Mostafa M Fouda, Zubair Md Fadlullah, and Mohsen Guizani. Dl-crc: Deep learning-based chest radiograph classification for covid-19 detection: A novel approach. *IEEE Access*, 8:171575–171589, 2020.

[39] Suman Kumar Reddy. https://inblog. Accessed:2021-12-5.

[40] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning*, pages 6105–6114. PMLR, 2019.

[41] Enzo Tartaglione, Carlo Alberto Barbano, Claudio Berzovini, Marco Calandri, and Marco Grangetto. Unveiling covid-19 from chest x-ray with deep learning: a hurdles race with small data. *International Journal of Environmental Research and Public Health*, 17(18):6933, 2020.

[42] Thread reader. Chest imaging. https://threadreaderapp.com/thread/1243928581983670272.html. Accessed: 2020-09-22.

[43] Hugo Touvron, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, and Hervé Jégou. Training data-efficient image transformers & distillation through attention. In *International Conference on Machine Learning*, pages 10347–10357. PMLR, 2021.

[44] Towards data science. https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5. Accessed: 2020-12-3.

[45] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.

[46] Abdul Waheed, Muskan Goyal, Deepak Gupta, Ashish Khanna, Fadi Al-Turjman, and Plácido Rogerio Pinheiro. Covidgan: Data augmentation using auxiliary classifier gan for improved covid-19 detection. *IEEE Access*, 8:91916–91923, 2020.

[47] Kun Wang, Xiaohong Zhang, Sheng Huang, Feiyu Chen, Xiangbo Zhang, and Luwen Huangfu. Learning to recognize thoracic disease in chest x-rays with knowledge-guided deep zoom neural networks. *IEEE Access*, 8:159790–159805, 2020.

[48] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M Summers. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2097–2106, 2017.

[49] wikipedia. https://en.wikipedia.org/wiki/Confusion_matrix. Accessed: 2020-12-3.

[50] Wikipedia. https://en.wikipedia.org/wiki/Receiver_operating_characteristic. Accessed: 2020-12-3.

[51] Wikipedia. https://en.wikipedia.org/wiki/Sensitivity_and_specificity. Accessed: 2021-12-5.