COVID-Nets: Deep CNN Architectures for Detecting COVID-19 Using Chest CT Scans

Hammam Alshazly[1,2], Christoph Linse[1], Mohamed Abdalla[3,4], Erhardt Barth[1], and Thomas Martinetz[1]

[1]**Institute for Neuro- and Bioinformatics, University of Lübeck, 23562 Lübeck, Germany**
[2]**Faculty of Computers and Information, South Valley University, Qena 83523, Egypt**
[3]**Mathematics Department, Faculty of Science, King Khalid University, Abha 62529, Saudi Arabia**
[4]**Mathematics Department, Faculty of Science, South Valley University, Qena 83523, Egypt**

Corresponding author:
Hammam Alshazly

Email address: alshazly@inb.uni-luebeck.de

ABSTRACT

This paper introduces two novel deep convolutional neural network (CNN) architectures for automated detection of COVID-19. The first model, CovidResNet, is inspired by the deep residual network (ResNet) architecture. The second model, CovidDenseNet, exploits the power of densely connected convolutional networks (DenseNet). The proposed networks are designed to provide fast and accurate diagnosis of COVID-19 using computed tomography (CT) images for the multi-class and binary classification tasks. The architectures are utilized in a first experimental study on the SARS-CoV-2 CT-scan dataset, which contains 4173 CT images for 210 subjects structured in a subject-wise manner for three different classes. First, we train and test the networks to differentiate COVID-19, non-COVID-19 viral infections, and healthy. Second, we train and test the networks on binary classification with three different scenarios: COVID-19 vs. healthy, COVID-19 vs. other non-COVID-19 viral pneumonia, and non-COVID-19 viral pneumonia vs. healthy. Our proposed models achieve up to 93.96% accuracy, 99.13% precision, 94% sensitivity, 97.73% specificity, and a 95.80% F1-score for binary classification, and up to 83.89% accuracy, 80.36% precision, 82% sensitivity, 92% specificity, and a 81% F1-score for the three-class classification tasks. The experimental results reveal the validity and effectiveness of the proposed networks in automated COVID-19 detection. The proposed models also outperform the baseline ResNet and DenseNet architectures while being more efficient.

Coronavirus disease 2019 (COVID-19), a highly infectious disease that affects primarily the respiratory system, is caused by the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2). The disease has presented massive public health crises and has been declared by the World Health Organization (WHO) as a global pandemic [11]. A major challenge in controlling the pandemic severity is the rapid spread of the virus and its wide person-to-person transmission [16]. The standard approach to detect SARS-CoV-2 is performed through a virus-specific real-time reverse transcription polymerase chain reaction (RT-PCR) testing. However, RT-PCR testing has several shortcomings, including a low sensitivity rate in the range of $60\% - 70\%$, long turnaround times, variabilities in testing techniques, high expenses, and a limited testing capacity in many countries [14,32]. Therefore, other diagnostic methods with higher sensitivity to COVID-19 are of crucial importance and urgently required.

Recent studies have reported that medical imaging of the lungs can be exploited as a suitable alternative testing method for COVID-19. The most widely used imaging modalities for the lungs are the chest radiography (X-ray) and computed tomography (CT). Beside their wide availability in hospitals worldwide, their usage has improved the diagnostic performance and sensitivity for COVID-19 detection [29]. Nevertheless, comparing the diagnostic accuracy of X-ray and CT in COVID-19, it has been reported that the sensitivity of X-ray is poor, whereas CT scanning has demonstrated higher sensitivity [8]. Moreover, CT screening has shown to be more sensitive even than RT-PCR testing while being significantly faster and

cheaper[2,14]. According to a study conducted on 1014 COVID-19 patients[2], RT-PCR could only detect 601/1014 (59%) patients as positives, while the CT-Scan detected 888/1014 (88%) patients as positives. The initial testing for some patients had negative RT-PCR results, whereas the confirmation was inferred based on their CT findings. Furthermore, chest CT screening has been strongly recommended for patients with specific symptoms compatible with viral infections, and their PCR test results are negative[27].

While it might be easy to differentiate patients with COVID-19 from healthy individuals based on CT, it is very challenging to differentiate COVID-19 from non-COVID-19 viral lung infections such as the community acquired pneumonia (CAP) due to two main reasons. First, COVID-19 and other viral infections share similar common patterns and features[52]. Patients with COVID-19 usually manifest several CT radiological features at different locations and distribution patterns such as ground glass opacities (GGO), consolidation, bilateral infiltration and crazy paving[18,55]. Second, the CT images may present appearance differences for patients with COVID-19 across different severity[56]. For these reasons, COVID-19 diagnosis from CTs requires interpretation of the CT images by expert physicians and is a labor-intensive, time-consuming, and often subjective. The CTs are first annotated by a practicing radiologist to report the radiographic findings. Then, the findings are analyzed against specific clinical factors to obtain the final diagnosis. During the current pandemic, checking every CT image is not a feasible option as the frontline physicians are faced with a lack of time and massive workload, which increases the physical burden on the staff and might affect the diagnostic quality and efficiency.

Artificial intelligence (AI) techniques and deep convolutional neural networks (CNNs) have the potential to automate COVID-19 detection in patients and assist in the rapid evaluation of CT scans[10]. The powerful representational capability of the deep CNNs can be exploited to differentiate patients with COVID-19 from healthy subjects or others with non-COVID-19 viral infections.

The present study introduces two deep CNN architectures that operate end-to-end to enable automated detection and effective diagnosis of COVID-19 patients based on CT images. The proposed networks have been tailored and validated to differentiate patients with COVID-19, patients with other viral infections, and healthy individuals from the SARS-CoV-2 CT-scan dataset[43]. We also investigated the networks effectiveness in binary classification with all possible class combinations from the considered dataset.

One common issue when using new, non-canonical network architectures is the lack of models that have been pretrained on large-scale datasets like ImageNet[13]. Using pretrained models in transfer learning approaches is attractive when having only small amounts of data to train the model and training a model from scratch would suffer from poor generalization. This is the case especially for image classification tasks for emerging or rare diseases and CNNs with millions of parameters. Starting to train with pretrained models offers initial filters, which are already adapted for visual recognition and only some modifications have to be made to solve the new task. This promotes the eagerness to benefit from transfer learning for COVID-19 detection. Novel architectures, which might be better suited for a specific task, have to compete with the pretrained canonical architectures (i.e. ResNet50 or DenseNet121) and might show inferior performance because of the lack of pretrained weights. Nevertheless, in order to benefit from transfer learning, we designed CovidResNet and CovidDenseNet with parameter compatibility as a key design feature. The idea is to make some network weights compatible with those of pretrained models, which can be found in public repositories. The inter-usability of the weights is realized by sharing some parts of the baseline's architecture and adding appropriate adapter layers at certain key positions. As a result, weights from the well known ResNet50 or DenseNet121 architectures can be used to partly initialize CovidResNet and CovidDenseNet, which leads to a boost in performance.

Since the CT images in the dataset have different sizes, scaling them to match a fixed input size will probably distort them. We opted for a different preprocessing procedure and experimentally investigated an approach to preserve the aspect ratios of the CT images in the SARS-CoV-2 CT-scan dataset. This procedure has proved to be very effective and results in an improved overall performance[4]. Extensive experiments and analysis on the diagnostic accuracy using standard evaluation metrics are performed against two baseline CNN models. The experimental results show superior performance for the proposed networks over the baseline models while having less layers and being computationally more efficient.

The main contributions of this work are summarized as follows:

- We propose two novel deep CNN architectures (CovidResNet and CovidDenseNet) for automated COVID-19 detection using chest CT scans. The models have been developed and validated for the multi-class and binary classification tasks to differentiate COVID-19 patients from non-COVID-19 viral infections as well as healthy subjects.

- The networks are trained and tested on CT images from the SARS-CoV-2 CT-scan dataset[43], which contains 4173 CT images for 210 subjects distributed into three different classes. To the best of our knowledge, this is the first experimental study to be conducted on the SARS-COV-2 CT-scan dataset with a subject-wise data split. Therefore, our models and the reported results may serve as a baseline to benchmark and compare any future work on this dataset.
- Most of the developed systems were trained and tested on CT scans from datasets where the same individuals appear in the training and test splits. This is definitely not appropriate, especially when construing diagnostic systems. Therefore, we followed a subject-wise splitting approach where 60% of the subjects are considered for training and 40% for testing.
- Extensive experiments and a comprehensive analysis are conducted to evaluate the performance of the proposed models against baseline models using standard evaluation metrics of accuracy, precision, sensitivity, specificity, F1-score, confusion matrix, ROC curve, and area under the ROC curve (AUC).
- Experimental results reveal the validity of our proposed networks to achieve very promising results with an average accuracy above 93% and 82% for the binary and multi-class classification tasks, respectively. Our CovidResNet and CovidDenseNet models have shown to be effective in differentiating COVID-19 patients from other non-COVID-19 and healthy individuals. Moreover, constructing an ensemble of the proposed networks further boosts the performance of the single models and achieves the best results on the considered dataset.

The remainder of the paper is structured as follows. Section 1 highlights the related work. Our proposed CovidResNet and CovidDenseNet architectures are described in Section 2. The dataset, data splitting and preprocessing, performance evaluation metrics, and the training methodology are detailed in Section 3. Section 4 provides the experimental results. Finally, the paper is concluded in Section 5.

1 RELATED WORK

This section explores the extensive work on constructing computer-aided diagnostic (CAD) systems for COVID-19 detection based on AI techniques and more specifically deep convolutional networks. Many effective approaches have been proposed to diagnose COVID-19 using chest radiography images including X-rays and CT scans. We hereafter discuss the most relevant work and highlight their success and achieved results.

A considerable number of CAD systems utilise X-ray images to diagnose COVID-19[1,6,9,23,37]. For instance, COVID-Net[49] is a deep CNN model designed specifically for detecting COVID-19 cases from chest X-ray images. COVID-Net was trained and tested on the CVOIDx dataset with a total of 13,975 X-ray images gathered from five different sources of chest radiography images. The network achieved 91.0% sensitivity rate for COVID-19 cases. DeepCoroNet[12] is another deep network approach proposed for automated detection of COVID-19 cases from X-ray images. The experimental analysis was performed on a combined dataset of COVID-19, pneumonia, and healthy X-ray images. The model provided a high success rate for the three-class classification problem exceeding other competitive models. CoVNet-19[28] is a stacked ensemble model for detecting COVID-19 patients from X-ray images. The model combined two pretrained deep CNNs (VGGNet[40] and DenseNet[22]) for feature extraction, and support vector machines (SVMs) for the final classification. The model achieved accuracy of 98.28% and a sensitivity rate above 95% for the COVID-19 class outperforming any of the single models. Coronavirus recognition network (CVR-Net)[19] is a multi-scale CNN-based model proposed to recognize COVID-19 from radiography images including both CT and X-ray images. The model was trained and evaluated for the multi-class and two-class classification tasks. The model achieved promising results with average accuracy ranging from of 82% and 99% for the multi-class and binary classification using X-ray images and 78% for CT images. COVID-ResNet[15] is a deep learning approach to differentiate COVID19 cases from other pneumonia cases based on X-ray images. The model was trained and validated on the COVIDx dataset and achieved an accuracy of 96.23%. In[46], an artificial neural network approach based on capsule networks was introduced to detect COVID-19 from X-ray images. The proposed model was investigated to differentiate COVID-19 cases in the two-class (COVID-19 and no-findings) as well as multi-class (COVID-19, Pneumonia, and normal) classification tasks. The model achieved average accuracy of 97.24%, and 84.22% for the two-class, and multi-class tasks, respectively.

Similar AI-based systems have been developed for automatically analyzing CT images for detecting

COVID-19 pneumonia[20,31,47,51,53]. These systems were constructed through a combination of segmentation and classification models. In the first stage, the lung region or the lesion region are first segmented from the CT scans using segmentation models such as U-Net[38] or V-Net[33]. While in the second stage deep CNN models such as ResNet[21] and Attention ResNet[48] were adopted to perform the diagnosis of COVID-19. For instance, an AI-based system for diagnosing COVID-19 based on CT scans was proposed in[26]. The system was trained and tested on CT-scan dataset consisting of CT scans of different classes including COVID-19, influenza, non-viral pneumonia, and non-pneumonia subjects. A comprehensive analysis was performed on a test cohort to evaluate the performance of the system in the multi-class and binary classification tasks. The model achieved an AUC score of 97.81% and a sensitivity of 91.51% for the multi-class classification task.

At the same time, new deep CNN architectures were designed and adopted to diagnose COVID-19. The authors in[50] redesigned the COVID-Net architecture and its learning methodology to be applied to CT images, and to improve the prediction accuracy and computational complexity. Besides, a joint learning approach was proposed to improve the diagnostic performance of COVID-19 cases and to tackle the data heterogeneity in the used CT scan datasets. Experiments on two CT image datasets show the success of the proposed joint learning approach with 90.83% accuracy and 85.89% sensitivity on the largest dataset. COVIDNet-CT[17], is deep CNN tailored specifically for the detection COVID-19 cases from chest CT images. The network was designed with a high architectural diversity and lightweight design patterns to achieve high representational capacity and computational efficiency. Training and testing were conducted on a collected CT image dataset named COVIDx-CT, which had CT images for three different classes, including: COVID-19 pneumonia, non-COVID-19 infections, and normal controls. The network achieved high sensitivity and specificity scores for the COVID-19 class reaching up to 97.3% and 99.9%, respectively.

Covid CT-Net[44], is a simple deep CNN developed for differentiating COVID-19 CTs from non-COVID-19 CT images. The network was trained and validated on the SARS-CoV-2 CT-scan dataset, which consists of 2492 CT scans for two class: COVID-19 and non-COVID-19[43]. The experimental results confessed an improved accuracy, specificity, and sensitivity of 95.78%, 95.56%, and 96%, respectively. An attentional convolutional network (COVID CT-Net) to predict COVID-19 from CT images was proposed in[54]. The network represented a combination of stacked residual modules empowered with attention-aware units to perform a more accurate prediction. The model was trained and validated on the SARS-CoV-2 CT-scan dataset[43] and achieved sensitivity and specificity rates of 85% and 96.2% respectively. The authors in[41] proposed a classification model for COVID-19 patients using chest CT images. The model adopted multi-objective differential evolution-based convolutional neural networks to differentiate positive COVID-19 cases from others. Experimental results showed that the proposed model was able to classify the CT images with an acceptable accuracy rate. Zhang et al. proposed a residual learning diagnosis detection network to differentiate COVID-19 cases from other heterogeneous CT images[58]. The network was trained and test on the COVID-CT dataset[59], and achieved accuracy, precision, and sensitivity of 91.33%, 91.30%, and 90%, respectively. In[35], an attention network was proposed to diagnose COVID-19 from community acquired pneumonia based on CT images. The network was trained and validated on a large-scale CT image dataset collected from eight hospitals. The network testing was performed on an independent CT data, and achieved an accuracy of 87.5%, a sensitivity of 86.9%, and a specificity of 90.1%.

Jaiswal et al.[25] proposed a deep transfer learning approach using a variant of the DenseNet models. The pretrained 201-layer DenseNet model on the IamgeNet dataset was utilized as a base for feature extraction with three added fully connected layers to perform the classification task. The experiments were conducted on the SARS-CoV-2 CT-scan dataset[43]. The model achieved accuracy score of 96.25% and a sensitivity rate of 96.21%. Alshazly et al.[5] conducted experimental study by adopting 12 pretrained deep CNN models, which were fine-tuned using CT images, to differentiate Patients with COVID-19 and non-COVID-19 subjects. Extensive experiments and analysis were performed on two COVID-19 CT scans datasets. The models were trained using custom-sized inputs for each deep model and achieved state-of-the-art results on the considered datasets. Further, visualization techniques were applied to provide visual explanations and show the ability of the fine-tuned models to accurately localize COVID-19 associated regions. In[36], a similar comprehensive study with 16 pretrained networks was carried out to detect COVID-19 based on CT images. The obtained results were comparable with those achieved in previous reposts as in[5].

Ensemble learning and deep ensembles were also explored in COVID-19 detection to improve the performance of single models. The authors in [42], proposed an ensemble based on three deep networks including: VGGNet [40], ResNet [21], and DenseNet [22], which were pretrained on natural images. The networks were considered for extracting features from the CT images, and a set of fully connected layers were added on top to perform the classification task. Experiments were conducted on a dataset with CT scans collected from different sources for patients with COVID-19, other lung diseases, and healthy subjects. The proposed ensemble achieved better performance than using any single model from the ensembled networks. Tao et al. [60] proposed an ensemble of three pretrained deep CNN models, namely AlexNet [30], GoogleNet [45], and ResNet [21] to improve the classification accuracy of COVID-19. Experiments were conducted on a collected CT image dataset organized in three different classes, including: COVID-19, lung tumors, and normal lungs. The obtained results showed an improved classification performance for the ensemble compared to any single individual model. The authors in [7] proposed a CAD system for distinguishing COVID-19 and non-COVID-19 cases. The system was trained and tested using CT images, where the CT image features were extracted with four pretrained deep CNN models, and then were fused for training support vector machine classifiers. The authors experimented with different fusion strategies to investigate the impact of feature fusion on the diagnostic performance. The system achieved accuracy, sensitivity, and specificity scores of 94.7%, 95.6%, and 93.7%, respectively.

The above-mentioned techniques were trained and tested on chest radiography images from of the same subjects, and were proposed for differentiating between COVID-19 and health individuals. Their obtained results need to be validated on datasets that are structured in a subject-wise level and for differentiating COVID-19 from other non-COVID-19 findings. The reasons are the potential overlap and high visual similarities between the radiographic findings of COVID-19 and non-COVID-19 viral infections, which makes the task very challenging. In our study, we develop and test two deep network architectures to differentiate patients with COVID-19 from other non-COVID-19 viral infections as well as healthy individuals. The networks have been developed and tested for the multi-class and binary classification tasks. The obtained results are promising and validate the effectiveness of our models. We experiment on the SARS-CoV-2 CT scan dataset, which is organized in a patient-wise structure. The CT scans of 60% of subjects are used for training, while CT scans of 40% of subjects are used for testing and reporting results.

2 COVID-NETS ARCHITECTURES

In the following subsections we describe our proposed CovidResNet and CovidDenseNet models for the automated COVID-19 detection on the SARS-CoV-2 CT-scan dataset. Inspired by the outstanding performance of the well-designed ResNet [21] and DenseNet [22] architectures, we build our networks by following similar construction patterns to get the benefits from both architectures.

2.1 CovidResNet

Our CovidResNet architecture is based on the deep residual networks (ResNets) [21]. ResNet is considered a very deep CNN architecture and the winner of the 2015 ImageNet challenges [39]. The main problems that have been addressed by the ResNet models are the vanishing gradients and performance degradation, which occur during training deep networks. A residual learning framework was proposed, which promotes the layers to learn residual functions with respect to the layer input. While conventional network layers are assumed to learn a desired underlying function $y = f(x)$ by some stacked layers, the residual layers attempt to approximate y via $f(x) + x$. The residual layers start with the input x and evolve to a more complex function as the network learns. This type of residual learning allows training very deep networks and attains an improved performance from the increased depth.

The basic building block for CovidResNet is the bottleneck residual module depicted in Figure 1. The input signal to the module passes through two branches. The left branch is a stack of three convolutional layers. The first 1×1 convolution is used for reducing the depth of the feature maps before the costly 3×3 convolutions, whereas the second 1×1 is used for increasing the depth to match the input dimensions. The convolutions are followed by batch normalization (BN) [24] and rectified linear unit (ReLU) [34] activation. The right branch is a shortcut connection that connects the module's input with the output of the stacked layers, which are summed up before applying a final ReLU activation.

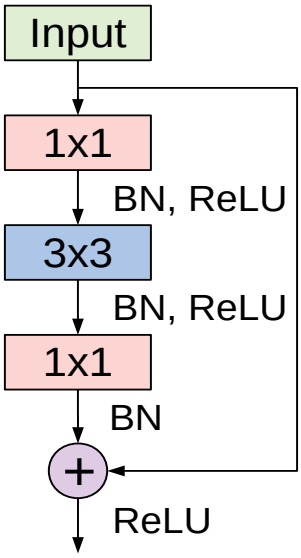


Figure 1. The bottleneck residual module used in CovidResNet. The module was first introduced in [21].

CovidResNet is considered a deep model that consists of 29 layers. The first layer is made of 7×7 convolutional filters with a stride of 2. Following is a max pooling layer to downsample the spatial dimensions. The architecture continues with a stack of four ResNet blocks, where each block has a number between one and three bottleneck residual modules. When moving from a ResNet block to the next one, the spatial dimension is reduced by max pooling and the number of the learned filters is doubled. In the first block, we stack three modules, each having three convolutional layers with 64, 64 and 256 filters, respectively. After another max pooling layer, we stack three more bottleneck modules with a configuration of 128, 128 and 512 filters, which forms the second block. The same procedure is repeated for the third and fourth blocks, where the former has two stacked modules and the later has only one. The network ends with an adaptive average pooling step and a fully connected layer. Table 1 summarizes the CovidResNet architecture and a visualization is given in Figure 2. As can be seen in the diagram, the first convolutional layer and the entire first block are frozen during transfer learning. Only the weights of deeper layers are adjusted using the SARS-COV-2 CT-scan dataset. The diagram also indicates the complimentary layers that exist in the canonical ResNet50 model but not in CovidResNet.

2.2 CovidDenseNet

Our CovidDenseNet model is based on the densely connected network (DenseNet) architectures introduced in [22]. DenseNet addressed the notorious problem of vanishing gradients with a different approach compared to ResNet. Instead of using skip connections to combine the feature maps through summation before passing them to the next layer, the feature maps from all preceding layers are considered as the input to the next layer, and its feature maps are passed to all subsequent layers. The advantages of the dense connectivity are the improved flow of information throughout the network, where each layer has a direct access to the gradients from the input and the loss function. DenseNets have shown an improved performance for image recognition tasks and are computationally efficient.

The basic building block for the CovidDenseNet model is the DenseNet block. A simplistic form of the dense connectivity of a dense block is shown in Figure 3. The block has three layers and each layer performs a series of batch normalization, ReLU activation, and 3×3 convolution operations. The concatenated feature maps from all preceding layers are the input to the subsequent layer. Each layer generates k feature maps, where k is the growth rate. So, if k_0 is the input to layer x_0, then there are $3k + k_0$ feature maps at the end of the 3-layer dense block. However, two main issues arise as the network depth increases. First, as each layer generates k feature maps, the inputs to layer l will be $(l-1)k + k_0$, and with deep networks this number can grow rapidly and slow down computation. Second, when the network gets deeper, we need to reduce the feature maps size to increase the kernel's receptive field. So,

Table 1. Description of our CovidResNet architecture for COVID-19 detection. The network accepts an RGB-input of size 257×353 pixels. The residual modules are placed in brackets multiplied by the number of modules stacked per block.

Layers	Output size	CovidResNet
Convolution	129×177	7×7, 64, stride 2
Pooling	64×88	3×3 max pooling, stride 2
ResNet Block (1)	64×88	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
ResNet Block (2)	32×44	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 3$
ResNet Block (3)	16×22	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 2$
ResNet Block (4)	8×11	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 1$
Classification layer	1×1	8×11 Adaptive average pool
		fully connected, softmax

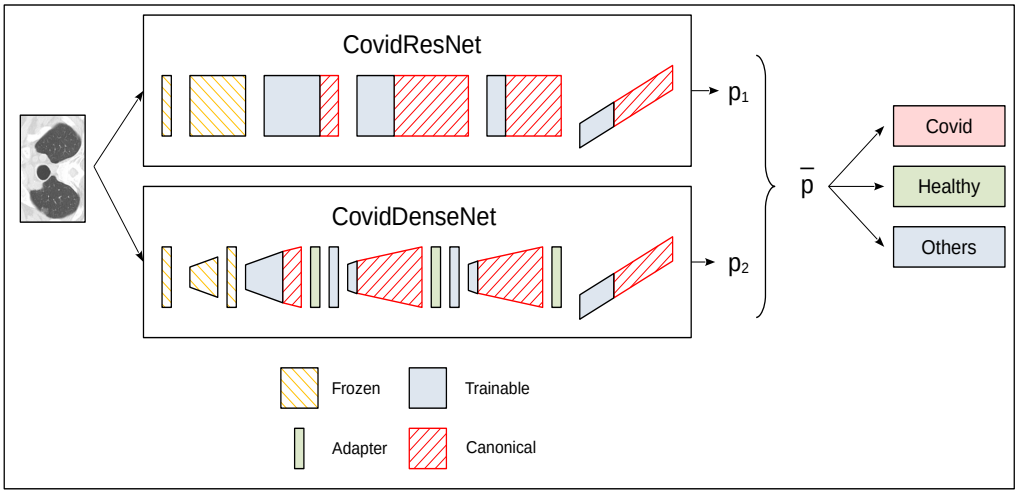


Figure 2. A schematic diagram for the ensemble prediction process for the three-class problem. Both networks accept the same input CT image and each network outputs an independent class probability vector. The probability vectors are then averaged to obtain the final predicted class with highest probability.

when concatenating feature maps of different sizes we need to match the dimensions. The first issue is addressed by introducing a bottleneck layer of 1×1 convolution and $4 \times k$ filters after every concatenation. The second issue is addressed by adding a transition layer between the dense blocks. The layer includes batch normalization and 1×1 convolution followed by an average pooling operation.

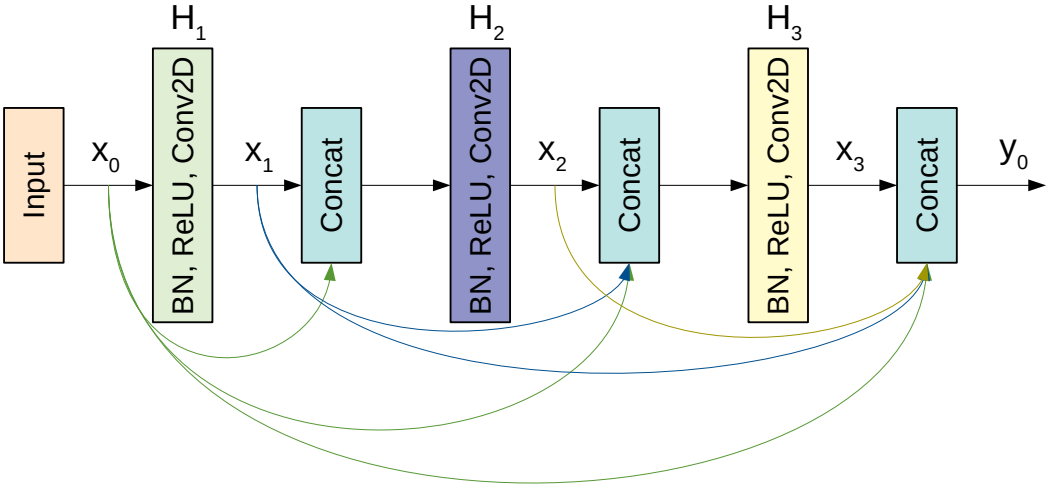


Figure 3. A schematic diagram of a 3-layer dense block used in the DenseNet architecture.

To ensure the inter-usability of the weights, CovidDenseNet contains a set of adapter layers. They consist of a 1×1 convolution to increase the number of channels to the size required by the subsequent layer. The number of channels can be seen in Table 2. The last adapter layer is optional, nevertheless we use it in our experiments to also use the pretrained weights for the last batch-norm. The adapters are inserted between a dense block and the transition layer.

Our CovidDenseNet model consists of 43 weighted layers. The first layer is a convolutional layer with 7×7 filters and uses a stride of 2, followed by a max pooling operation. Then we stacked four dense blocks interspersed by transition layers. After the last dense block we perform an adaptive average pooling and add a fully connected layer with a softmax classifier. The details of the CovidDenseNet architecture on how many dense blocks are stacked for each stage, the input and output volume before and after each specific stage are summarized in Table 2. In order to use CovidDenseNet with transfer learning, we implement the network in a three-step procedure. It involves downloading a pretrained DenseNet121, removing 2, 22 and 15 layers from the second, third and fourth dense block respectively, adding the adapter layers and then freezing the first convolutional layer, as well as the first dense block.

3 METHODOLOGY

3.1 Dataset

The SARS-CoV-2 CT-scan dataset[43] is considered one of the largest CT scan datasets currently available for research that follows a patient-wise structure. The CT scans have been collected in public hospitals in Sao Paulo, Brazil, with a total of 4173 CT scans for 210 different subjects. The CT scans are distributed into three classes, namely COVID-19, Healthy, and Others. The exact number of patients and CT scans for each category is summarized in Table 3. As the dataset contains patients with other pulmonary diseases and the CT images have variable sizes, the dataset is challenging. Figure 4 shows 12 CT images from the SARS-CoV-2 CT-scan dataset, where the first row includes 4 COVID-19 images, the second row shows 4 images from the Healthy class, and the third row illustrates 4 images with other lung diseases from the Others class.

3.2 Data Preprocessing and Splitting

Wide variations in the CT image sizes in the SARS-CoV-2 CT-scan dataset ask for a strategy to resize the images to a consistent input dimension for the network. The most frequently used approach to unify images with different aspect rations involves stretching, which can result in images that look unnatural or distorted. Therefore, we opt for a different procedure to preserve the aspect ratio by embedding the

Table 2. Our CovidDenseNet architecture for COVID-19 detection. The network accepts an RGB-input of size 257×353 pixels.

Layers	Output size	CovidDenseNet
Convolution	129×177	7×7, 64, stride 2
Pooling	64×88	3×3 max pooling, stride 2
Dense Block [1]	64×88	$\begin{bmatrix} 1 \times 1, conv \\ 3 \times 3, conv \end{bmatrix} \times 6$
Transition Layer [1]	64×88	1×1 conv
	32×44	2×2 average pooling, stride 2
Dense Block [2]	32×44	$\begin{bmatrix} 1 \times 1, conv \\ 3 \times 3, conv \end{bmatrix} \times 10$
Adapter Layer [2]	32×44	1×1 conv, 512 channels
Transition Layer [2]	32×44	1×1 conv
	16×22	2×2 average pooling, stride 2
Dense Block [3]	16×22	$\begin{bmatrix} 1 \times 1, conv \\ 3 \times 3, conv \end{bmatrix} \times 2$
Adapter Layer [3]	16×22	1×1 conv, 1024 channels
Transition Layer [3]	16×22	1×1 conv
	8×11	2×2 average pooling, stride 2
Dense Block [4]	8×11	$\begin{bmatrix} 1 \times 1, conv \\ 3 \times 3, conv \end{bmatrix} \times 1$
Adapter Layer [4] (opt.)	8×11	1×1 conv, 1024 channels
Classification layer	1×1	8×11 adaptive average pool
		fully connected, softmax

Table 3. Number of subjects and CT scans for each of the three categories in the SARS-CoV-2 CT-scan dataset.

No.	COVID-19	Healthy	Others	Total
Subjects	80	50	80	210
Images	2168	758	1247	4173

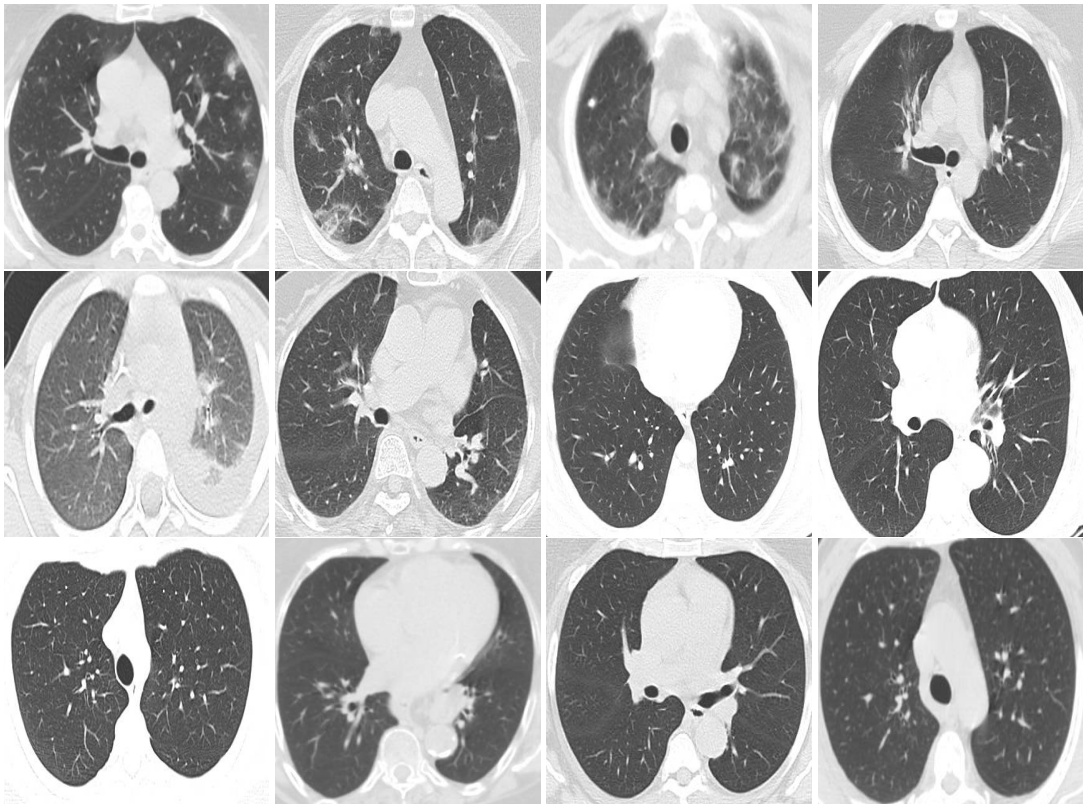


Figure 4. Sample CT images from the SARS-CoV-2 CT scan dataset. The CTs represent four images of COVID-19 (first row), four images of Others class (second row), and four images from the Healthy class (third row).

image into a fixed-sized canvas. We apply padding with the average color of the ImageNet dataset[13] when necessary to match the target shape. We empirically tried different input sizes and found that a canvas with a spatial dimension of 257×353 works best for CT images from the SARS-CoV-2 CT-scan dataset and our architectures. Due to the limited amount of training data and the fact that deep neural networks require large amounts of data to optimize millions of parameters, we recompense the lack of data by implementing different augmentation steps to improve the network's ability to generalize. The augmentation steps include random rescaling, random cropping, Gaussian noise, brightness and contrast changes and random horizontal flipping. Finally, the images are normalized according to the mean and standard deviation of the ImageNet dataset.

To conduct our experiments and analysis we split the dataset into training and test sets. We follow the subject-wise structure of the dataset, such that the two sets of persons in the training and test set are disjunct. Hence, it is assured, that we evaluate our models on unseen persons. However, the number of CT images per person vary. We choose 59.5% of the subjects for training and 40.5% for testing, such that the amount of training images is 60% and 40%, respectively. The same ratio of persons is used for both scenarios of multi-class and binary classification tasks. Within one scenario we choose the same split for each architecture for the sake of consistency and comparability.

3.3 Performance Evaluation Metrics

In order to evaluate the performance of our models we consider a set of standard quantitative evaluation metrics including:

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

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

$$Sensitivity = (\mathrm{TP})/(\mathrm{TP}+\mathrm{FN}) \tag{3}$$

$$Specificity = (\mathrm{TN})/(\mathrm{TN}+\mathrm{FP}) \tag{4}$$

$$F1\text{-}score = (2\times\mathrm{TP})/(2\times\mathrm{TP}+\mathrm{FP}+\mathrm{FN}) \tag{5}$$

where TP and TN refer to the total number of cases that are correctly classified as True Positives (TP) and True Negatives (TN), while FP and FN are the total number of cases that are incorrectly classified as False Positives (FP) and False Negatives (FN), respectively. We also report the macro average scores for the multi-class experiments to show the overall performance across the different classes of the dataset.

Models are difficult to compare when the performance assessment is based on a single evaluation metric only. Hence, we provide multiple evaluation metrics to enable a profound analysis. We plot the ROC curves to visualize the diagnostic ability of the models to differentiate between the different classes. We also compute the area under the ROC curve (AUC) for each model. The ROC curves show the trade-off between the true positive rate (sensitivity) and the false negative rate (1-specificity) at various threshold values. The AUC summarizes the ROC curve and measures the ability of a model to distinguish between the different classes. A high AUC value indicates better performance of the model at distinguishing between the classes. In addition, we provide the confusion matrices for detailed class-wise results. Confusion matrices clearly tell about the exact numbers of correctly detected positive and negative cases, as well as the type of error a model makes.

3.4 Transfer Learning

Transfer learning is a method in deep learning, which has become quite popular in the computer vision community because it might significantly boost recognition performance. The idea bases on the transferability of network weights between related image recognition tasks and relies on the universal validity of the visual filters learned during training. Usually, transfer learning occurs in a two-step procedure. First, a model's weights are trained for one task on a dataset, which is typically large. Subsequently, a model is initialized with the weights to solve the actual task and often it is also fine-tuned. As the size of the SARS-CoV-2 CT-scan dataset is limited, we opt for transfer learning to benefit from the pretrained image filters. We initialize ResNet50 and DenseNet121 with weights, that have been optimized for the ImageNet dataset. Parts of our proposed architectures exhibit compatible weight configurations such that we can initialize many weights with ResNet50 and DenseNet121 models that have been pretrained on ImageNet. In CovidResNet all weights are pretrained, but the last layer. In CovidDenseNet the adapter layers and the last layer are randomly initialized and all other weights are copied from the DenseNet121 model that was pretrained on ImageNet.

We empirically found that it is not necessary to adjust all weights to the COVID-19 detection problem. We assume that the filters from the first layers in a computer vision network provide somewhat generic filters that can be used for the SARS-CoV-2 CT-scan dataset. The idea is to reduce the risk of overfitting by lowering the amount of trained weights. Thus, we freeze the first convolutional layer and the first convolutional block of CovidResNet and only adapt the remaining weights. The first convolutional layer, the first dense block and the first transition layer of CovidDenseNet are also frozen. All weights in the models ResNet50 and DenseNet101 are fine-tuned to enable the comparison between the baseline with our novel architectures together with our specifically designed fine tuning strategy. An overview of the CovidResNet and CovidDenseNet architectures can be seen in Figure 2. The layers with frozen weights are highlighted in orange. The trainable layers are colored in blue. See Table 4 for important characteristics of the proposed CovidResNet and CovidDenseNet models compared with the baselines.

Table 4. Characteristics of the proposed CNN architectures compared to the baseline architectures. Our models have less layers and trainable parameters, which results in smaller model size and shorter execution time.

Model	Network characteristics					
	Default input size	Custom input size	Layers	Total Parameters (M)	Trainable Parameters (M)	Training time per epoch (s)
ResNet50	224×224	257×353	50	23.51	23.51	19
DenseNet121	224×224	257×353	121	6.96	6.96	22
CovidResNet	257×353	257×353	29	9.84	9.61	11
CovidDenseNet	257×353	257×353	43	3.13	2.75	12

3.5 Model Training

The models are initialized using pretrained weights that have been optimized for the ImageNet dataset. Then, we train the models using the LAMB optimizer[57], an initial learning rate of 0.0003 and cross-entropy loss. The baseline models are trained for 100 epochs until convergence. Our proposed architectures need more epochs to converge and we stop training after 150 epochs. The learning rate is step-wise reduced until a value of 1e-6 is reached at the end of training. The batch size is 32. We also apply weight decay to regularize the training process. Optimization is performed within the PyTorch framework using an Nvidia GTX 1080 GPU.

4 RESULTS AND DISCUSSION

This section presents the quantitative results for COVID-19 detection by our proposed COVID-Nets architectures. We compare their performances with two baseline models, ResNet50 and DenseNet121, for the three-class and the binary classification tasks. We begin with discussing the obtained results for differentiating patients with COVID-19, non-COVID-19 other viral lung infections, and non-infected healthy individuals. Then, we discuss the results obtained by each model in all three possible two-class classification scenarios.

4.1 Three-class Classification Results

Table 5 provides the performance metrics, which are computed for each specific class, and the macro-average scores obtained by each model. Our proposed models achieve very promising results and outperform both, the ResNet50 and DenseNet121 models. Among the single network architectures, our CovidDenseNet model achieves the best overall performance with an accuracy of 82.87%. Moreover, the model achieves the highest precision score of 95.76% for the COVID-19 class. Furthermore, the model achieves the best overall specificity score of 95.90% for COVID-19 class, which proves its ability to designate most of the non-COVID-19 subjects as negative. However, the model obtains a sensitivity rate of 86.14% for the COVID-19 cases. The model also has a high sensitivity for the Others class. When considering the macro average scores for all evaluation metrics we observe that CovidDenseNet provides better performance compared to the other models. Similarly, our proposed CovidResNet model achieves better performance with respect to macro average precision, specificity and F1-score compared to the baseline models.

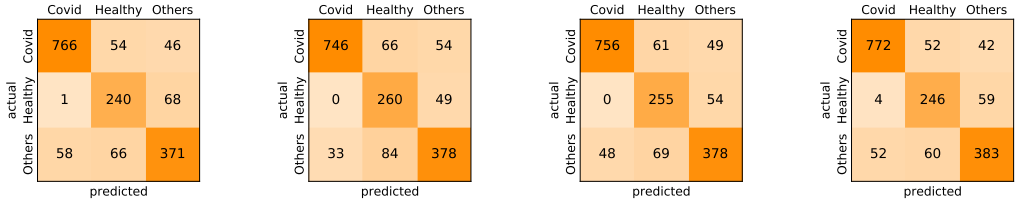
Based on our experimental results, which indicate superior performances for CovidResNet and CovidDenseNet, we considered these models for constructing ensembles for improving the overall diagnostic performance. The idea stems from the stochastic nature of deep networks where each network learns specific features and patterns. Building an ensemble of several independently trained networks and taking the unweighted average of their outputs can generate synergistic effects by exploiting the powerful feature extraction capability of each network[3]. Several ensemble combinations have been tested and we report the results of the best two ensembles in Table 5. We can see that in both cases, the ensemble models achieve better performance with respect to the macro average metrics compared to any individual network. The ensemble of CovidDenseNet and DenseNet121 models has improved the detection rate of CovidDenseNet model for the COVID-19 class with 3%.

Figure 5 shows the confusion matrix for each of our proposed model as well as the ensemble that is achieving the best overall performance. By analyzing the confusion matrix we get insights on the class specific results achieved by each model with respect to the number of correctly classified and misclassified cases.

We also plot the ROC curves and compute the AUC to investigate the diagnostic accuracy of the proposed models for the multi-class problem in Figure 6. Our CovidResNet and CovidDenseNet models

Table 5. Comparison of the different models for the three-class classification task. Results are given in percentages and the best metric values are highlighted in bold.

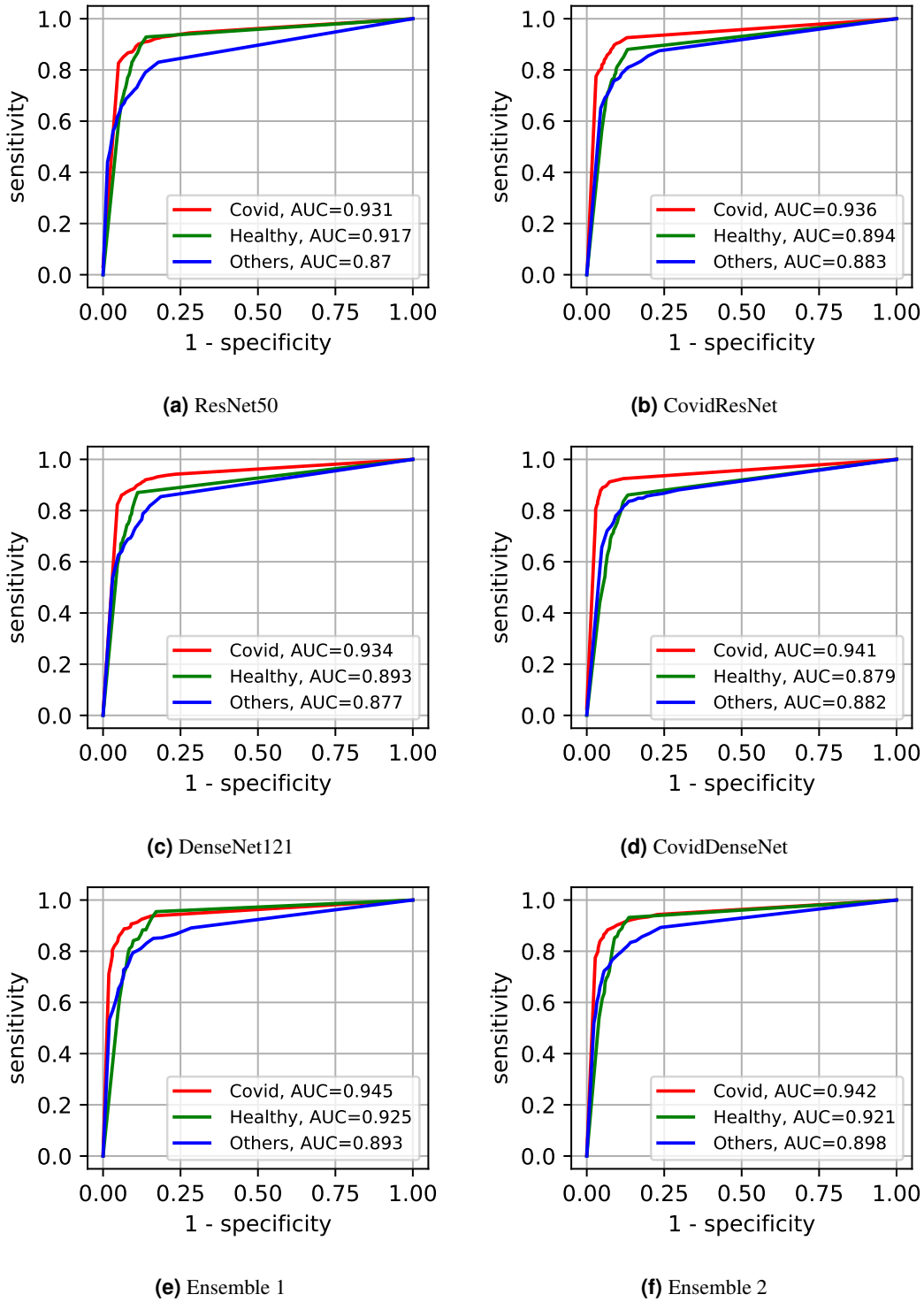
Model	Class	Evaluation Metrics				
		Accuracy	Precision	Sensitivity	Specificity	F1-score
ResNet50	COVID-19		87.49	91.22	85.95	89.32
	Healthy		66.94	79.94	91.04	72.86
	Others		82.16	66.06	93.96	73.24
	Macro average	81.68	78.86	79.07	90.31	78.47
DenseNet121	COVID-19		89.83	89.72	89.05	89.77
	Healthy		68.39	72.82	92.36	70.53
	Others		75.21	72.32	89.96	73.74
	Macro average	81.44	77.81	78.29	90.46	78.02
CovidResNet	COVID-19		92.85	88.45	92.66	90.60
	Healthy		66.67	77.67	91.18	71.75
	Others		76.49	74.95	90.30	75.71
	Macro average	82.46	78.67	80.36	91.38	79.35
CovidDenseNet	COVID-19		95.76	86.14	95.90	90.70
	Healthy		63.41	84.14	88.98	72.32
	Others		78.59	76.36	91.23	77.46
	Macro average	**82.87**	**79.25**	**82.22**	**92.04**	**80.16**
Ensemble 1	COVID-19		94.03	87.30	94.03	90.54
	Healthy		66.23	82.52	90.45	73.49
	Others		76.36	76.36	91.23	77.46
	Macro average	83.17	79.62	82.06	91.90	80.49
Ensemble 2	COVID-19		93.24	89.15	93.03	91.15
	Healthy		68.72	79.61	91.77	73.76
	Others		79.13	77.37	91.40	78.24
	Macro average	**83.89**	**80.36**	**82.04**	**92.07**	**81.05**



(a) CovidResNet **(b)** CovidDenseNet **(c)** Ensemble 1 **(d)** Ensemble 2

Figure 5. Confusion matrices generated by the different models for the three-class classification task.

show superior performance and achieve higher AUC scores for the classes COVID-19 and Others, which indicates that our models detect COVID-19 and the other lung infections better than the deeper baselines of ResNet50 and DenseNet121. The AUC scores for the class Healthy is quite low as it has fewer number of subjects and CT images, which could be insufficient to learn discriminative features for separating this class from the other two classes.



(a) ResNet50 **(b)** CovidResNet

(c) DenseNet121 **(d)** CovidDenseNet

(e) Ensemble 1 **(f)** Ensemble 2

Figure 6. The ROC curves and their AUC scores for the different models showing their ability to differentiate between the three classes.

Building ensembles through a combination of our independently trained CovidResNet and CovidDenseNet models and their baselines increases the classification accuracy for all classes. The superiority of the ensembles over single models is also reflected in the ROC curves and their corresponding AUC scores. When combining CovidDenseNet and CovidResNet, which we refer to it as Ensemble 1, we notice that the AUC score for the Healthy class increased from 87.9% to 92.5% , whereas an increment within 1% in the AUC score is attained for the COVID-19 and Others classes. Similar results are achieved when we combine the CovidDenseNet and its deeper baseline DenseNet121, which we refer to it as Ensemble 2, even though the models were trained on the same training split of the used dataset.

4.2 Two-class Classification Results

We have trained and tested our proposed architectures on binary classification tasks to investigate their ability to distinguish between CT images of all possible classes as well as to investigate the difficulty of these subtasks on the given dataset. We investigate three experimental scenarios. First, we train and test our models to differentiate patients with COVID-19 from healthy individuals (COVID-19 vs. Healthy). Then, we train and test the models to distinguish COVID-19 cases from non-COVID-19 patients infected by other lung diseases (COVID-19 vs. Others). Finally, we train and test our models to differentiate non-COVID-19 patients infected by other pulmonary diseases from healthy subjects (Others vs. Healthy). Table 6 presents the results obtained by each model under each of these scenarios.

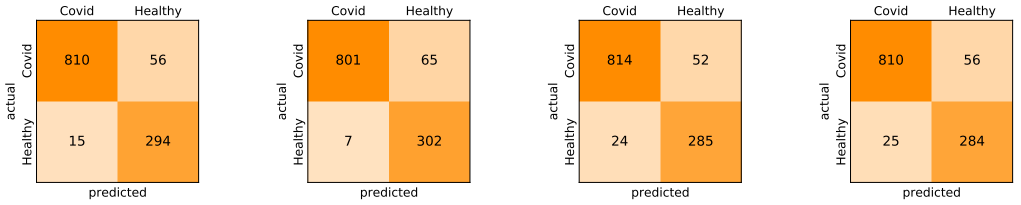
In the first scenario (COVID-19 vs. Healthy) we used 866 CT images of COVID-19 and 309 CT images from the healthy class for testing. As we can see from Table 6 and under this scenario, all four models achieve very competitive performance with accuracy above 93% and F1-score above 95%. The models also achieve high precision values above 97%, where our proposed CovidResNet model achieves the highest precision score of 99.13%, indicating that almost all the predicted subjects as COVID-19 are correct and only 7 out of 309 healthy CT images were incorrectly classified as COVID-19 positive. CovidResNet also attains the highest specificity score of 97.73%, which indicates its ability to correctly identify 302 out of 309 normal CT images as COVID-19 negative. However, CovidResNet has a lower sensitivity rate compared to other models. The model is able to correctly detect 92.49% of COVID-19 cases and 65 COVID-19 CTs were incorrectly detected as non-COVID-19 (false negatives). Nevertheless, this high false negative rate is a common problem among all the tested models and can be attributed to two main reasons. First, in some cases, patients with COVID-19 may show normal chest CT findings at the early days of infection, and therefore it is hard to exclude all COVID-19 cases based only on the chest CT predictive results. Second, the findings on CTs can be very tiny and can barely be detected by the models, as the CT images of COVID-19 patients may manifest different imaging characteristics such as specific patterns progressively with time based on the severity of the infection.

Table 6. The obtained results under three binary classification scenarios.

Task	Model	Evaluation Metrics				
		Accuracy	Precision	Sensitivity	Specificity	F1-score
COVID-19 vs. Healthy	ResNet50	**93.96**	98.18	93.53	95.15	**95.80**
	DenseNet121	93.53	97.14	**94.00**	92.23	95.54
	CovidResNet	93.87	**99.13**	92.49	**97.73**	95.70
	CovidDenseNet	93.11	97.01	93.53	91.91	95.24
COVID-19 vs. Others	ResNet50	83.77	85.59	**89.84**	72.88	87.66
	DenseNet121	83.10	85.21	89.15	72.26	87.13
	CovidResNet	85.10	90.80	85.45	84.47	88.04
	CovidDenseNet	**86.88**	**91.76**	87.41	**85.92**	**89.53**
Others vs. Healthy	ResNet50	85.64	**90.97**	85.63	**85.67**	88.22
	DenseNet121	83.35	85.52	88.46	74.74	86.97
	CovidResNet	**86.40**	88.32	**90.28**	79.86	**89.29**
	CovidDenseNet	83.61	84.89	89.88	73.04	87.32

For a detailed class-wise results, the confusion matrix for each specific model under the considered scenario is presented in Figure 7.

Figure 8 shows the ROC curves for all evaluated models. Looking at the ROC curves and the AUC scores we can see that the four models perform on a similar level. The ROC curves look identical and the AUC scores vary within a range of 1%, with ResNet50 achieving a slightly higher AUC sore of 97%.



(a) ResNet50 **(b)** CovidResNet **(c)** DenseNet121 **(d)** CovidDenseNet

Figure 7. Confusion matrices for COVID-19 vs. Healthy classification.

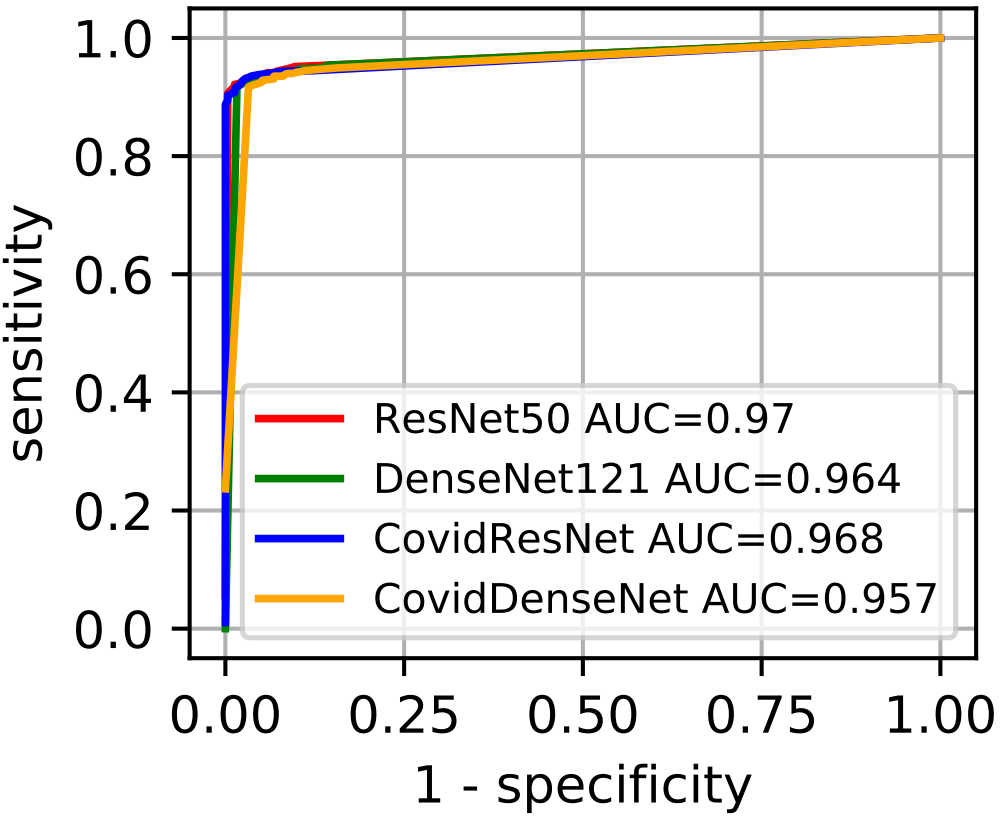
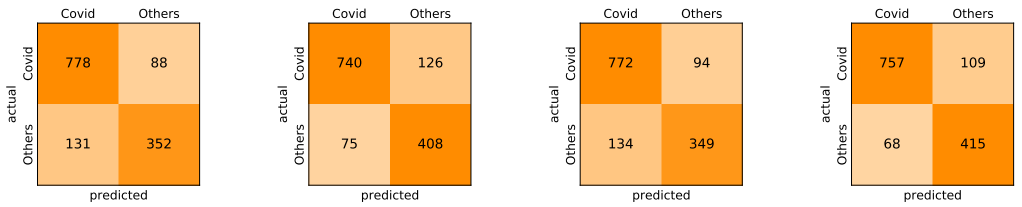


Figure 8. Comparison of the predictive performance for CovidResNet and CovidDenseNet and the baseline models for COVID-19 vs. Healthy classification. The ROC curves and AUC scores show the competitive performance for all models.

In the second scenario (COVID-19 vs. Others) we investigate the effectiveness of our models in differentiating the CTs of COVID-19 from others with viral lung infections. It is worth mentioning that this is a challenging task due to the potential overlap of findings on CT images between COVID-19 and the other lung viral infections. The obtained results in Table 6 clearly show lower performance with respect to all evaluation metrics compared to the obtained results in the first scenario. Nevertheless, our proposed CovidResNet and CovidDenseNet models achieve higher accuracy values compared with the baselines, where our CovidDenseNet model attains an accuracy of 86.88%. Our proposed models also achieve much better results with respect to precision, specificity, and F1-score values. Our CovidDenseNet model achieves the highest precision score of 91.76% indicating its ability to correctly identify CTs with COVID-19. Only 68 out of 483 CT images from the Others class are incorrectly classified as COVID-19 (false positives). It is also worth noting that our CovidResNet and CovidDenseNet models achieve much higher specificity rates above 85% outperforming the baseline models with 12%. The lower specificity of ResNet50 and DenseNet121 may stem from the difficulty to distinguish the CT findings of COVID-19 from findings of other non-COVID-19 viral diseases. On the contrary, our CovidDenseNet model correctly detected 415 out of 483 CT images as other lung diseases. However, our models show slightly lower sensitivity rates compared to the other models due to more false negatives. Nevertheless, a high false negative rate is a common issue for all the tested models due to the potential overlap of the imaging findings.

By investigating the confusion matrix we get a detailed class-wise analysis. Figure 9 shows the confusion matrix for each model and what type of error each specific model makes.



(a) ResNet50 **(b)** CovidResNet **(c)** DenseNet121 **(d)** CovidDenseNet

Figure 9. Confusion matrices for COVID-19 vs. Others classification.

We also compare the performance of the different models under this scenario by plotting the ROC curve and computing the AUC for each model. Figure 10 shows the ROC curves, where we clearly see that our CovidResNet and CovidDenseNet model are superior to their baseline models as their ROC curves are closer to the top-left corner and they achieve higher AUC values. The highest AUC score of 91.7% is achieved by our CovidDenseNet model exceeding its deeper counterpart DenseNet121 model with more than 5%.

In our third scenario (Others vs. Healthy) we test the ability of our architectures to differentiate patients infected with other pulmonary diseases and non-infected healthy individuals. While our main objective in this work is to develop architectures to differentiate patients with COVID-19 from other non-COVID-19 viral infections as well as healthy subjects, we report our results under this scenario for the sake of completeness. In our experiments we treat people infected by other viral infections as the positive class and the healthy individuals as the negative class. Under this scenario, our CovidResNet model achieves the best overall performance with 86.40% accuracy. The model also achieves the highest sensitivity rate of 90.28%, which indicates its ability to detect above 90% of the infected cases.

Figure 11 shows the confusion matrix for each of the tested models. We can observe that all the models have high false positive rates under this scenario compared with the first scenario (COVID-19 vs. Healthy). A possible reason is that we have more CT images in the COVID-19 class to learn fairly discriminative features, whereas the limited amount of CT scans for the Others class makes it difficult to distinguish them from non-infected or normal CT images. Therefore, we need to collect more CT images for both classes to reduce the false positive as well as the false negative rates.

Figure 12 presents the ROC curves and their corresponding AUC scores for all tested models. Again, our proposed models show superior performance compared with their deeper baseline models. Our CovidResNet model achieves the highest AUC score of 92.1% and its ROC curve appears closer to the

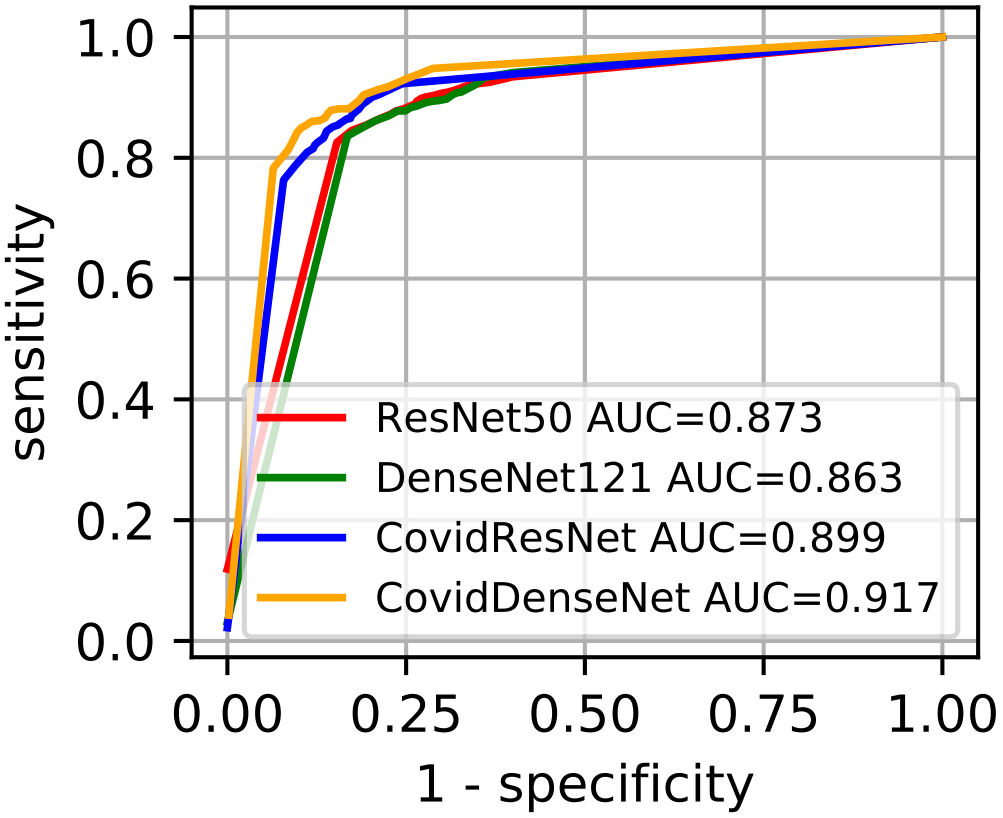
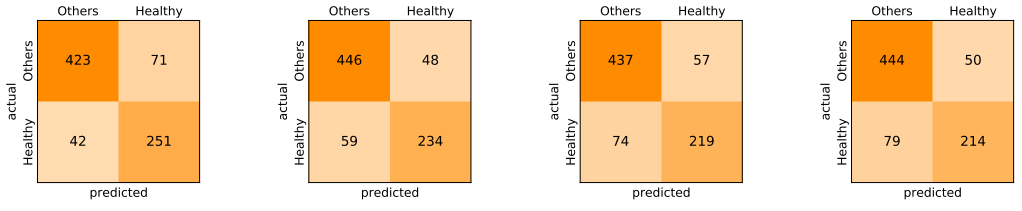


Figure 10. Predictive performance of our proposed CovidResNet and CovidDenseNet models vs. the baseline models for COVID-19 vs. Others classification. The ROC curves show powerful a predictive power for the CovidDenseNet model.



(a) ResNet50 **(b)** CovidResNet **(c)** DenseNet121 **(d)** CovidDenseNet

Figure 11. Confusion matrices for Others vs. Healthy classification.

top-left corner. Our CovidDenseNet model has superior performance within approximately 2% compared to its deeper DenseNet121 model.

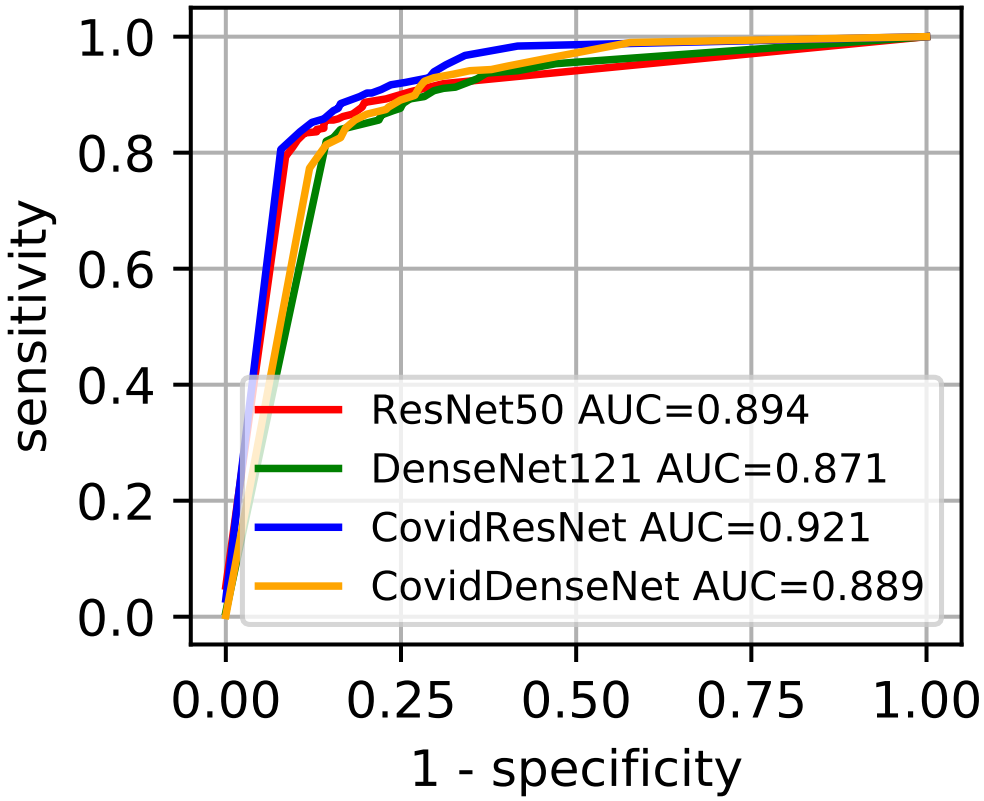


Figure 12. Comparison of the proposed CovidResNet and CovidDenseNet with the baseline models using the ROC curve and AUC for Others vs. Healthy classification.

5 CONCLUSION

We proposed two deep CNN architectures (CovidResNet and CovidDenseNet) for the automated detection of COVID-19 using chest CT scans. The models were developed and validated on the large multi-class SARS-CoV-2 CT-scan dataset, which has more than 4000 CT scans. We conducted extensive experiments to evaluate our models in multi-class and binary classification tasks. First, we trained our models to differentiate COVID-19 cases from other non-COVID-19 infections as well as from healthy subjects. Experimental results show the effectiveness of the proposed architectures to achieve better accuracy compared with the baseline ResNet and DenseNet architectures, while having less layers and being more computationally efficient. Second, we conducted three binary classification to differentiate COVID-19 from healthy individuals, COVID-19 from other non-COVID-19 patients, and non-COVID-19 viral infections from non-infected healthy subjects. The obtained results demonstrate the superior performance of our proposed models over the baseline models.

As to our knowledge, this is the first experimental study on the SARS-CoV-2 CT-scan dataset that considers subject-wise splits for training and testing. Therefore, our models and results can be used as a baseline benchmark for any future experiments conducted on this dataset. Although our experimental results are promising, there is still room for improvement. We assume that experiments conducted on even larger datasets of CT scans will improve the diagnostic accuracy and provide a more reliable estimation of the models' performance. Collecting more CT scans and subjects for all classes and particularly the Healthy and Others categories can further improve the diagnostic performance of the proposed models.

ACKNOWLEDGMENTS

The first and third authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding their work through Research Groups Program under grant number RGP.2/1/42. The work of Christoph Linse was supported by the Bundesministeriums für Wirtschaft und Energie (BMWi) through the Mittelstand 4.0-Kompetenzzentrum Kiel Project.

REFERENCES

[1] Abraham, B. and Nair, M. S. (2020). Computer-aided detection of COVID-19 from X-ray images using multi-CNN and Bayesnet classifier. *Biocybernetics and Biomedical Engineering*, 40(4):1436–1445.

[2] Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., Tao, Q., Sun, Z., and Xia, L. (2020). Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: A report of 1014 cases. *Radiology*, 296(2):E32–E40.

[3] Alshazly, H., Linse, C., Barth, E., and Martinetz, T. (2019). Ensembles of deep learning models and transfer learning for ear recognition. *Sensors*, 19(19):4139.

[4] Alshazly, H., Linse, C., Barth, E., and Martinetz, T. (2020). Deep convolutional neural networks for unconstrained ear recognition. *IEEE Access*, 8:170295–170310.

[5] Alshazly, H., Linse, C., Barth, E., and Martinetz, T. (2021). Explainable COVID-19 Detection Using Chest CT Scans and Deep Learning. *Sensors*, 21:455.

[6] Aslan, M. F., Unlersen, M. F., Sabanci, K., and Durdu, A. (2020). CNN-based transfer learning–BiLSTM network: A novel approach for COVID-19 infection detection. *Applied Soft Computing*, 98:106912.

[7] Attallah, O., Ragab, D. A., and Sharkas, M. (2020). MULTI-DEEP: A novel CAD system for coronavirus (COVID-19) diagnosis from CT images using multiple convolution neural networks. *PeerJ*, 8:e10086.

[8] Borakati, A., Perera, A., Johnson, J., and Sood, T. (2020). Diagnostic accuracy of x-ray versus ct in covid-19: a propensity-matched database study. *BMJ Open*, 10(11):e042946.

[9] Brunese, L., Mercaldo, F., Reginelli, A., and Santone, A. (2020). Explainable deep learning for pulmonary disease and coronavirus COVID-19 detection from X-rays. *Computer Methods and Programs in Biomedicine*, 196:105608.

[10] Chowdhury, M. E., Rahman, T., Khandakar, A., Mazhar, R., Kadir, M. A., Mahbub, Z. B., Islam, K. R., Khan, M. S., Iqbal, A., Al Emadi, N., et al. (2020). Can AI help in screening viral and COVID-19 pneumonia? *IEEE Access*, 8:132665–132676.

[11] Cucinotta, D. and Vanelli, M. (2020). WHO Declares COVID-19 a Pandemic. *Acta Biomed*, 91(1):157—-160.

[12] Demir, F. (2021). DeepCoroNet: A deep LSTM approach for automated detection of COVID-19 cases from chest X-ray images. *Applied Soft Computing*, 103:107160.

[13] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). ImageNet: A Large-scale Hierarchical Image Database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 248–255.

[14] Fang, Y., Zhang, H., Xie, J., Lin, M., Ying, L., Pang, P., and Ji, W. (2020). Sensitivity of chest CT for COVID-19: Comparison to RT-PCR. *Radiology*, 296(2):200432.

[15] Farooq, M. and Hafeez, A. (2020). COVID-ResNet: A deep learning framework for screening of COVID19 from radiographs. *arXiv preprint arXiv:2003.14395*.

[16] Ghinai, I., McPherson, T. D., Hunter, J. C., Kirking, H. L., Christiansen, D., Joshi, K., Rubin, R., Morales-Estrada, S., Black, S. R., Pacilli, M., et al. (2020). First known person-to-person transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in the USA. *The Lancet*, 395(10230):1137–1144.

[17] Gunraj, H., Wang, L., and Wong, A. (2020). COVIDNet-CT: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases From Chest CT Images. *Frontiers in Medicine*, 7.

[18] Hani, C., Trieu, N. H., Saab, I., Dangeard, S., Bennani, S., Chassagnon, G., and Revel, M.-P. (2020). COVID-19 pneumonia: A review of typical CT findings and differential diagnosis. *Diagnostic and Interventional Imaging*, 101(5):263–268.

[19] Hasan, K., Alam, A., Elahi, T. E., Roy, S., and Wahid, S. R. (2020). CVR-Net: A deep convo-

lutional neural network for coronavirus recognition from chest radiography images. *arXiv preprint arXiv:2007.11993*.

[20] Hasan, K., Jawad, T., Hasan, K. N. I., Partha, S. B., and Masba, M. A. (2021). COVID-19 identification from volumetric chest CT scans using a progressively resized 3D-CNN incorporating segmentation, augmentation, and class-rebalancing. *arXiv preprint arXiv:2102.06169*.

[21] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778.

[22] Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4700–4708.

[23] Ibrahim, A. U., Ozsoz, M., Serte, S., Al-Turjman, F., and Yakoi, P. S. (2021). Pneumonia Classification Using Deep Learning from Chest X-ray Images During COVID-19. *Cognitive Computation*.

[24] Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, pages 448–456.

[25] Jaiswal, A., Gianchandani, N., Singh, D., Kumar, V., and Kaur, M. (2020). Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning. *Journal of Biomolecular Structure and Dynamics*.

[26] Jin, C., Chen, W., Cao, Y., Xu, Z., Tan, Z., Zhang, X., Deng, L., Zheng, C., Zhou, J., Shi, H., and Feng, J. (2020). Development and evaluation of an artificial intelligence system for COVID-19 diagnosis. *Nature Communications*, 11(1):5088.

[27] Kanne, J. (2020). Chest ct findings in 2019 novel coronavirus (2019-ncov) infections from wuhan, china: Key points for the radiologist. *Radiology*, 295(1):16–17.

[28] Kedia, P., Katarya, R., et al. (2021). CoVNet-19: A Deep Learning model for the detection and analysis of COVID-19 patients. *Applied Soft Computing*, 104:107184.

[29] Kim, H., Hong, H., and Yoo, S. H. (2020). Diagnostic Performance of CT and Reverse Transcriptase Polymerase Chain Reaction for Coronavirus Disease 2019: A Meta-Analysis. *Radiology*, 296:E145—-E155.

[30] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems*, pages 1097–1105.

[31] Li, Z., Zhong, Z., Li, Y., Zhang, T., Gao, L., Jin, D., Sun, Y., Ye, X., Yu, L., Hu, Z., Xiao, J., Huang, L., and Tang, Y. (2020). From community-acquired pneumonia to COVID-19: A deep learning–based method for quantitative analysis of COVID-19 on thick-section CT scans. *European radiology*, 30(12):6828–6837.

[32] Long, C., Xu, H., Shen, Q., Zhang, X., Fan, B., Wang, C., Zeng, B., Li, Z., Li, X., and Li, H. (2020). Diagnosis of the Coronavirus disease (COVID-19): rRT-PCR or CT? *European journal of radiology*, 126:108961.

[33] Milletari, F., Navab, N., and Ahmadi, S.-A. (2016). V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In *4th International Conference on 3D Vision (3DV)*, pages 565–571.

[34] Nair, V. and Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 807–814.

[35] Ouyang, X., Huo, J., Xia, L., Shan, F., Liu, J., Mo, Z., Yan, F., Ding, Z., Yang, Q., Song, B., Shi, F., Yuan, H., Wei, Y., Cao, X., Gao, Y., Wu, D., Wang, Q., and Shen, D. (2020). Dual-Sampling Attention Network for Diagnosis of COVID-19 From Community Acquired Pneumonia. *IEEE Transactions on Medical Imaging*, 39(8):2595–2605.

[36] Pham, T. D. (2020). A comprehensive study on classification of COVID-19 on computed tomography with pretrained convolutional neural networks. *Scientific Reports*, 10:16942.

[37] Pham, T. D. (2021). Classification of COVID-19 chest X-rays with deep learning: new models or fine tuning? *Health Information Science and Systems*, 9(2).

[38] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241.

[39] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. (2015). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252.

[40] Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In *Proceedings of the International Conference on Learning Representations (ICLR)*, pages 1–14.

[41] Singh, D., Kumar, V., and Kaur, M. (2020). Classification of COVID-19 patients from chest CT images using multi-objective differential evolution-based convolutional neural networks. *European Journal of Clinical Microbiology & Infectious Diseases*, 39:1379–1389.

[42] Singh, D., Kumar, V., and Kaur, M. (2021). Densely connected convolutional networks-based COVID-19 screening model. *Applied Intelligence*.

[43] Soares, E., Angelov, P., Biaso, S., Froes, M. H., and Abe, D. K. (2020). SARS-CoV-2 CT-scan dataset: A large dataset of real patients CT scans for SARS-CoV-2 identification. *medRxiv*.

[44] Swapnarekha, H., Behera, H. S., Nayak, J., and Naik, B. (2021). Covid CT-net: A deep learning framework for COVID-19 prognosis using CT images. *Journal of Interdisciplinary Mathematics*, pages 1–26.

[45] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–9.

[46] Toraman, S., Alakus, T. B., and Turkoglu, I. (2020). Convolutional capsnet: A novel artificial neural network approach to detect covid-19 disease from x-ray images using capsule networks. *Chaos, Solitons & Fractals*, 140:110122.

[47] Wang, B., Jin, S., Yan, Q., Xu, H., Luo, C., Wei, L., Zha, W., Hou, X., Ma, W., Xu, Z., Zheng, Z., Sun, W., Lan, L., Zhang, W., Mu, X., , Wang, Z., Lee, J., Jin, Z., Jin, H., Zhang, L., Zhao, B., Ren, Z., Wang, S., Xu, W., Wang, X., Wang, J., ZhengYou, and Dong, J. (2021). AI-assisted CT imaging analysis for COVID-19 screening: Building and deploying a medical AI system. *Applied Soft Computing*, 98:106897.

[48] Wang, F., Jiang, M., Qian, C., Yang, S., Li, C., Zhang, H., Wang, X., and Tang, X. (2017). Residual attention network for image classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3156–3164.

[49] Wang, L., Lin, Z. Q., and Wong, A. (2020a). COVID-Net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest X-ray images. *Scientific Reports*, 10(1):19549.

[50] Wang, Z., Liu, Q., and Dou, Q. (2020b). Contrastive Cross-site Learning with Redesigned Net for COVID-19 CT Classification. *IEEE Journal of Biomedical and Health Informatics*, 24(10):2806–2813.

[51] Wu, X., Chen, C., Zhong, M., Wang, J., and Shi, J. (2021). COVID-AL: The diagnosis of COVID-19 with deep active learning. *Medical Image Analysis*, 68:101913.

[52] Xu, M., Ouyang, L., Han, L., Sun, K., Yu, T., Li, Q., Tian, H., Safarnejad, L., Zhang, H., Gao, Y., Bao, F. S., Chen, Y., Robinson, P., Ge, Y., Zhu, B., Liu, J., and Chen, S. (2021). Accurately Differentiating Between Patients With COVID-19, Patients With Other Viral Infections, and Healthy Individuals: Multimodal Late Fusion Learning Approach. *Journal of Medical Internet Research*, 23(1):e25535.

[53] Xu, X., Jiang, X., Ma, C., Du, P., Li, X., Lv, S., Yu, L., Ni, Q., Chen, Y., Su, J., Lang, G., Li, Y., Zhao, H., Liu, J., Ruan, L., Qiu, J. S. Y., Wu, W., Liang, T., and Li, L. (2020). A deep learning system to screen novel coronavirus disease 2019 pneumonia. *Engineering*.

[54] Yazdani, S., Minaee, S., Kafieh, R., Saeedizadeh, N., and Sonka, M. (2020). COVID CT-Net: Predicting Covid-19 From Chest CT Images Using Attentional Convolutional Network. *arXiv preprint arXiv:2009.05096*.

[55] Ye, Z., Zhang, Y., Wang, Y., Huang, Z., and Song, B. (2020). Chest CT manifestations of new coronavirus disease 2019 (COVID-19): A pictorial review. *European radiology*, 30:4381–4389.

[56] Yilmaz, P. D., Kadiyoran, C., Bakdik, S., Poyraz, N., and Vatansev, H. (2020). Early Computed Tomography Findings of Novel Coronavirus Disease 2019 (COVID-19) Pneumonia. *Archives of Clinical Infectious Diseases*, 15(5):e106868.

[57] You, Y., Li, J., Reddi, S., Hseu, J., Kumar, S., Bhojanapalli, S., Song, X., Demmel, J., and Hsieh, C.-J. (2020). Large batch optimization for deep learning: Training bert in 76 minutes. In *International Conference on Learning Representations (ICLR)*.

[58] Zhang, M., Chu, R., Dong, C., Wei, J., Lu, W., and Xiong, N. (2021). RLDD: An Advanced Residual

Learning Diagnosis Detection System for COVID-19 in IIoT. *IEEE Transactions on Industrial Informatics*.

[59] Zhao, J., Zhang, Y., He, X., and Xie, P. (2020). COVID-CT-Dataset: A CT scan dataset about COVID-19. *arXiv preprint arXiv:2003.13865*.

[60] Zhou, T., Lu, H., Yang, Z., Qiu, S., Huo, B., and Dong, Y. (2020). The ensemble deep learning model for novel COVID-19 on CT images. *Applied Soft Computing*, 98:106885.