# COVID-Nets: deep CNN architectures for detecting COVID-19 using chest CT scans

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

[1] Faculty of Computers and Information, South Valley University, Qena, Egypt

[2] Institut für Neuro- und Bioinformatik, University of Lübeck, Lübeck, Germany

[3] Mathematics Department, Faculty of Science, King Khalid University, Abha, Saudi Arabia

[4] Mathematics Department, Faculty of Science, South Valley University, Qena, Egypt

Corresponding Author: Hammam Alshazly
Email address: alshazly@inb.uni-luebeck.de

In this paper we propose two novel deep convolutional network architectures, CovidResNet and CovidDenseNet, to diagnose COVID-19 based on CT images. The models enable transfer learning between different architectures, which might significantly boost the diagnostic performance. Whereas novel architectures usually suffer from the lack of pretrained weights our proposed models can be partly initialized with larger baseline models like ResNet50 and DenseNet121, which is attractive because of the abundance of public repositories. 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 into three different classes. The models differentiate between COVID-19, non-COVID-19 viral pneumonia, and healthy samples. We also investigate their performance under three binary classification scenarios where we distinguish COVID-19 from healthy, COVID-19 from non-COVID-19 viral pneumonia, and non-COVID-19 from healthy, respectively. Our proposed models achieve up to 93.87% accuracy, 99.13% precision, 92.49% sensitivity, 97.73% specificity, 95.70% F1-score, and 96.80% AUC score for binary classification, and up to 83.89% accuracy, 80.36% precision, 82.04% sensitivity, 92.07% specificity, 81.05% F1-score, and 94.20% AUC score for the three-class classification tasks. We also validated our models on the COVID19-CT dataset to differentiate COVID-19 and other non-COVID-19 viral infections, and our CovidDenseNet model achieved the best performance with 81.77% accuracy, 79.05% precision, 84.69% sensitivity, 79.05% specificity, and 81.77% F1-score, and 87.50% AUC score. The experimental results reveal the effectiveness of the proposed networks in automated COVID-19 detection where they outperform standard models on the considered datasets while being more efficient.

# 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

In this paper we propose two novel deep convolutional network architectures, CovidResNet and CovidDenseNet, to diagnose COVID-19 based on CT images. The models enable transfer learning between different architectures, which might significantly boost the diagnostic performance. Whereas novel architectures usually suffer from the lack of pretrained weights our proposed models can be partly initialized with larger baseline models like ResNet50 and DenseNet121, which is attractive because of the abundance of public repositories. 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 into three different classes. The models differentiate between COVID-19, non-COVID-19 viral pneumonia, and healthy samples. We also investigate their performance under three binary classification scenarios where we distinguish COVID-19 from healthy, COVID-19 from non-COVID-19 viral pneumonia, and non-COVID-19 from healthy, respectively. Our proposed models achieve up to 93.87% accuracy, 99.13% precision, 92.49% sensitivity, 97.73% specificity, 95.70% F1-score, and 96.80% AUC score for binary classification, and up to 83.89% accuracy, 80.36% precision, 82.04% sensitivity, 92.07% specificity, 81.05% F1-score, and 94.20% AUC score for the three-class classification tasks. We also validated our models on the COVID19-CT dataset to differentiate COVID-19 and other non-COVID-19 viral infections, and our CovidDenseNet model achieved the best performance with 81.77% accuracy, 79.05% precision, 84.69% sensitivity, 79.05% specificity, and 81.77% F1-score, and 87.50% AUC score. The experimental results reveal the effectiveness of the proposed networks in automated COVID-19 detection where they outperform standard models on the considered datasets while being more efficient.

## 1 INTRODUCTION

**97.28 87.45%** 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[16]. By June 13, 2021, there have been 175,306,598 confirmed COVID-19 cases including 3,792,777 deaths, reported to the WHO, and a total of 2,156,550,767 vaccine doses have been administered[69].

The virus is still spreading widely at the time and the number of cases is increasing per day in several countries even with the extensive and accelerated pace of vaccinations. The appearance of worrisome new variants that are much more contagious causes several countries to struggle in reducing the number of new cases and necessitates urgent vaccinations. Epidemiologists define the current stage of the pandemic as a race between vaccinations and newly reported cases, particularly the infections caused by more infectious variants. The main challenges in controlling the pandemic severity are the

⁴⁷ rapid and wide person-to-person transmission of the virus and the spreading of new viral variants across
⁴⁸ all countries[47]. The standard approach to detect SARS-CoV-2 is performed through a virus-specific
⁴⁹ real-time reverse transcription polymerase chain reaction (RT-PCR) testing. However, RT-PCR testing has
⁵⁰ several shortcomings, including a low sensitivity rate in the range of $60\% - 70\%$, long turnaround times,
⁵¹ variabilities in testing techniques, high expenses, and a limited testing capacity in many countries[20,40].
⁵² Therefore, the development of other effective and scalable diagnostic tools with higher sensitivity to
⁵³ COVID-19 are of crucial importance and urgently required.

⁵⁴ Recent studies have reported that medical imaging of the lungs can be exploited as a suitable alternative
⁵⁵ testing method for COVID-19. The most widely used imaging modalities for the lungs are the chest radiog-
⁵⁶ raphy (X-ray) and computed tomography (CT). Beside their wide availability in hospitals worldwide, their
⁵⁷ usage has improved the diagnostic performance and sensitivity for COVID-19 detection[37]. Nevertheless,
⁵⁸ comparing the diagnostic accuracy of X-ray and CT in detecting COVID-19, it has been reported that the
⁵⁹ sensitivity of X-ray is poor, whereas CT scanning has demonstrated higher sensitivity[12]. Moreover, CT
⁶⁰ screening has shown to be more sensitive even than RT-PCR testing while being significantly faster and
⁶¹ cheaper[2,20]. According to a study conducted on 1014 COVID-19 patients[2], RT-PCR could only detect
⁶² 601/1014 (59%) patients as positives, while the CT-Scan detected 888/1014 (88%) patients as positives.
⁶³ The initial testing for some patients had negative RT-PCR results, whereas the confirmation was inferred
⁶⁴ based on their CT findings. Furthermore, chest CT screening has been strongly recommended for patients
⁶⁵ with specific symptoms compatible with viral infections, and their PCR test results are negative[35].

⁶⁶ While it might be easy to differentiate patients with COVID-19 from healthy individuals based on
⁶⁷ CT, it is very challenging to differentiate COVID-19 from non-COVID-19 viral lung infections such as
⁶⁸ the community acquired pneumonia (CAP) due to two main reasons. First, COVID-19 and other viral
⁶⁹ infections share similar common patterns and features[71]. Patients with COVID-19 usually manifest
⁷⁰ several CT radiological features at different locations and distribution patterns such as ground glass
⁷¹ opacities (GGO), consolidation, bilateral infiltration and crazy paving[23,74]. Second, the CT images
⁷² may present appearance differences for patients with COVID-19 across different severity[75]. For these
⁷³ reasons, COVID-19 diagnosis from CTs requires interpretation of the CT images by expert physicians
⁷⁴ and is a labor-intensive, time-consuming, and often subjective. The CTs are first annotated by a practicing
⁷⁵ radiologist to report the radiographic findings. Then, the findings are analyzed against specific clinical
⁷⁶ factors to obtain the final diagnosis. During the current pandemic, checking every CT image is not a
⁷⁷ feasible option as the frontline physicians are faced with a lack of time and massive workload, which
⁷⁸ increases the physical burden on the staff and might affect the diagnostic quality and efficiency.

⁷⁹ Artificial intelligence (AI) techniques and deep convolutional neural networks (CNNs) have the
⁸⁰ potential to automate COVID-19 detection in patients and assist in the rapid evaluation of CT scans[14].
⁸¹ The powerful representational capability of the deep CNNs can be exploited to differentiate patients
⁸² with COVID-19 from healthy subjects or others with non-COVID-19 viral infections. Therefore, our
⁸³ study introduces two deep CNN architectures that operate end-to-end to enable automated detection and
⁸⁴ effective diagnosis of COVID-19 patients based on CT images. The proposed networks have been tailored
⁸⁵ and validated to differentiate patients with COVID-19, patients with other viral infections, and healthy
⁸⁶ individuals from the SARS-CoV-2 CT-scan dataset[57]. We also investigated the networks effectiveness
⁸⁷ in binary classification with all possible class combinations from the considered dataset. Moreover,
⁸⁸ we validated our models on the COVID19-CT dataset [78], and our models provided promising results
⁸⁹ outperforming state-of-the-art models.

⁹⁰ One common issue when using new, non-canonical network architectures is the lack of models that
⁹¹ have been pretrained on large-scale datasets like ImageNet[18]. Using pretrained models in transfer
⁹² learning approaches is attractive when having only small amounts of data to train the model and training
⁹³ a model from scratch would suffer from poor generalization. This is the case especially for image
⁹⁴ classification tasks for emerging or rare diseases and CNNs with millions of parameters. Starting to train
⁹⁵ with pretrained models offers initial filters, which are already adapted for visual recognition and only
⁹⁶ some modifications have to be made to solve the new task. This promotes the eagerness to benefit from
⁹⁷ transfer learning for COVID-19 detection. Novel architectures, which might be better suited for a specific
⁹⁸ task, have to compete with the pretrained canonical architectures (i.e. ResNet or DenseNet) and might
⁹⁹ show inferior performance because of the lack of pretrained weights. Nevertheless, in order to benefit
¹⁰⁰ from transfer learning, we designed CovidResNet and CovidDenseNet with parameter compatibility as
¹⁰¹ 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 standard's architecture and adding appropriate adapter layers at certain key positions. As a result, weights from the well known ResNet50[26] or DenseNet121[28] architectures can be used to partly initialize CovidResNet and CovidDenseNet, which leads to a boost in performance.

Since the CT images in the considered datasets 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. This procedure has proved to be very effective and results in an improved overall performance[5]. Extensive experiments and analysis on the diagnostic accuracy using standard evaluation metrics were conducted against five standard CNN models. The experimental results show the superior performance for the proposed models over the standard models and are 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 based on chest CT images. The models enable transfer learning between different architectures and they can be partly initialized with larger pretrained models like ResNet50 and DenseNet121, to boost the diagnostic performance. The models have been tailored 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 two benchmark datasets. First, the SARS-CoV-2 CT-scan dataset, 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. Second, the COVID19-CT dataset, which shares similar visual characteristics and is available in a subject-wise data split.
- In contrast to most of the developed systems that were trained and tested on CT images where the same individuals appear in the training and test splits, which is definitely not appropriate. We followed a subject-wise splitting approach, where we choose 60% of the subjects for training and 40% for testing.
- We conduct extensive experiments and a comprehensive analysis to evaluate the performance of the proposed models for the multi-class and binary classification tasks using various evaluation metrics including accuracy, precision, sensitivity, specificity, F1-score, confusion matrix, ROC curve, and area under the ROC curve (AUC).
- Our experimental results reveal the validity of the proposed networks to achieve very promising results with average accuracies of 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 boosts the performance of the single models and achieves the best results.

The remainder of the paper is structured as follows. Section 2 highlights the related work. Our proposed CovidResNet and CovidDenseNet architectures are described in Section 3. The datasets, data splitting and preprocessing, performance evaluation metrics, and the training methodology are detailed in Section 4. Section 5 provides the experimental results. Finally, the paper is concluded in Section 6.

## 2 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[1,8,13,31,49]. For instance, COVID-Net[65] 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[17] 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[36] is a stacked ensemble model for detecting COVID-19 patients from X-ray images. The model combined two pretrained deep CNNs (VGGNet[54] and DenseNet[28]) 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)[24] 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[21] 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[62], 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[25,39,63,70,72]. 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[51] or V-Net[41]. While in the second stage deep CNN models such as ResNet[26] and Attention ResNet[64] 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[34]. 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[67] 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[22], 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[59], 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[57]. 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[73]. 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[57] and achieved sensitivity and specificity rates of 85% and 96.2% respectively. The authors in[55] 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[77]. The network was trained and test on the COVID-CT dataset[78], and achieved accuracy, precision, and sensitivity of 91.33%, 91.30%, and 90%, respectively. In[43], 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.[33] 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[57]. The model achieved accuracy score of 96.25% and a sensitivity rate of 96.21%. Alshazly et al.[6] 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[48], 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[6].

Ensemble learning and deep ensembles were also explored in COVID-19 detection to improve the performance of single models. The authors in[56], proposed an ensemble based on three deep networks including: VGGNet[54], ResNet[26], and DenseNet[28], 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.[79] proposed an ensemble of three pretrained deep CNN models, namely AlexNet[38], GoogleNet[60], and ResNet[26] 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[9] 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 healthy individuals. The 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, which represents a very challenging task. The reasons are the potential overlap and high visual similarities between the radiographic findings of COVID-19 and non-COVID-19 viral infections. In our study, we develop and test two deep network architectures, which can be partly initialized from standard pretrained models to boost the diagnostic performance. The models are evaluated to differentiate patients with COVID-19 from other non-COVID-19 viral infections as well as healthy individuals. The networks were developed and tested for the multi-class and binary classification tasks. The obtained results are promising and validate the effectiveness of our models. Extensive experiments were conducted on two challenging CT image dataset, which contain images of varying sizes and visual characteristics. The obtained results indicate the success of our models to achieve the best perfromance on the considerd datasets.

Table 1 summarizes some of the recently published studies on detecting COVID-19 from chest radiographical images using various techniques.

**Table 1.** A summary of recently published studies on COVID-19 detection. Sens.: sensitivity; Spec.: specificity; Prec.: precision; Acc.: accuracy; and AUC: area under the curve.

| Literature | Modality | Dataset | Model | Results |
|---|---|---|---|---|
| Wang et al.[68] | 3583 X-ray images | Open image data + Xiangya Hospital | Discrimination-DL | Acc. (93.65%) Sens. (90.92%) Spec. (92.62%) AUC (95.5%) |
| Bahgat et al.[10] | 12,933 X-ray images | Several combined datasets | Optimized DenseNet121 | Acc. (98.47%) Prec. (98.50%) Sens. (98.47%) Spec. (99.50%) F1-score (98.49%) AUC (99.83%) |
| Ozturk et al.[46] | 495 X-ray images | COVID-19 image data collection[15] | Shrunken features + PCA + SVM | Acc. (94.2%) Prec. (96.7%) Sens. (93.3%) Spec. (98.5%) F1-score (93.9%) |
| Alshazly et al.[6] | 2482 CT images | SARS-CoV-2 CT-scan dataset[57] | ResNet101 | Acc. (99.4%) Prec. (998.6%) Sens. (99.1%) Spec. (99.6%) F1-score (99.4%) |
| Alshazly et al.[6] | 746 CT images | COVID19-CT dataset[78] | DenseNet201 | Acc. (92.9%) Prec. (91.3%) Sens. (93.7%) Spec. (92.2%) F1-score (92.5%) |
| Barstugan et al.[11] | 150 CT images | COVID-19 database[19] | DWT + SVM | Acc. (97.8%) Prec. (98.4%) Sens. (96.8%) Spec. (98.6%) F1-score (97.6%) |
| Ozturk et al.[44] | 150 CT images | COVID-19 database[19] | Fusion of CNN features | Acc. (95.6%) Prec. (97.7%) Sens. (93.3%) Spec. (97.8%) F1-score (95.5%) |
| Ozturk et al.[45] | 2482 CT images | SARS-CoV-2 CT-scan dataset[57] | Conv. Support Vector Machine (CSVM) | Acc. (94.0%) Prec. (92.1%) Sens. (96.0%) Spec. (92.0%) F1-score (94.1%) |
| Ragab et al.[50] | 2482 CT images | SARS-CoV-2 CT-scan dataset[57] | Fusion of handcrafted and CNN features | Acc. (99%) Prec. (99%) Sens. (99%) F1-score (99%) AUC (100%) |
| Wang et al.[66] | 1065 CT images | Private CT images coll. from 3 hospitals | Modified Inception model | Acc. (89.5%) Sens. (87%) Spec. (88%) F1-score (77%) AUC (93%) |
| Song et al.[58] | 1990 CT images | Private CT images coll. from hospitals | DRE-Net | Acc. (86%) Prec. (79%) Sens. (96%) Spec. (77%) F1-score (87%) AUC (95%) |
| Ardakani et al.[7] | 1020 CT images | Private data | Xception | Acc. (99.02%) Sens. (98.04%) Spec. (100%) F1-score (77%) AUC (93%) |
| Shi et al.[53] | 1658 patients with COVID-19 1027 patients with CAP | Private CT images collected from hospitals | Handcrafted features | Acc. (89.4%) Sens. (90.7%) Spec. (87.2%) AUC (95.5%) |

# 3 COVID-NETS ARCHITECTURES

Here we describe our proposed CovidResNet and CovidDenseNet models for the automated COVID-19 detection. Inspired by the outstanding performance of the well-designed ResNet[26] and DenseNet[28] architectures, we build our networks by following similar construction patterns to get the benefits of both architectures.

The novelty of CovidResNet and CovidDenseNet lies in the feature-interusability with their standard counterparts, i.e. ResNet50 and DenseNet121. The idea is to initialize the models with pretrained weights from the standard models and benefit from the advantages of transfer learning without the costly step to pre-train models on large-scale datasets like the ImageNet dataset. Nevertheless, one benefits from the

272  tiny model size, parameter efficiency and speed of our architectures, which are designed to have much
273  less weights in total.
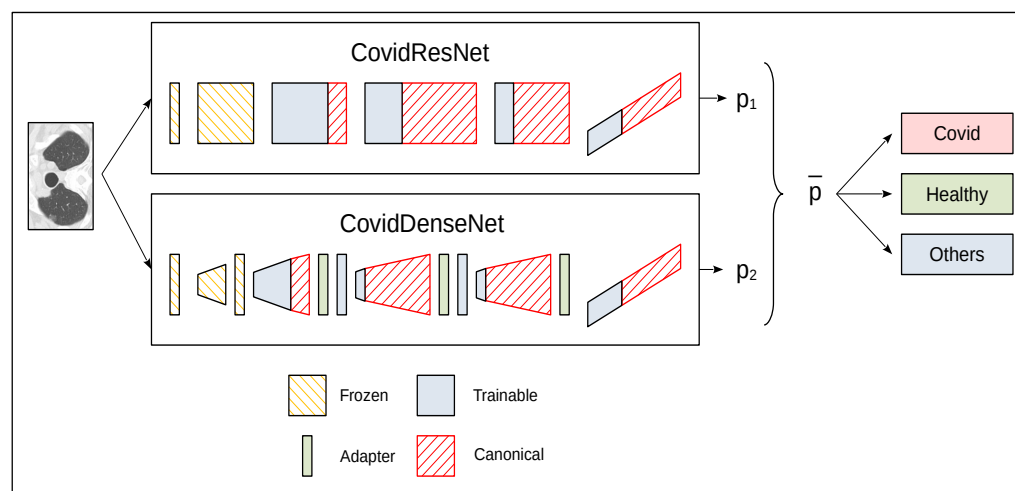
**Figure 1.** A schematic diagram for the ensemble prediction process for the three-class problem. Both networks accept the same input CT image and each network outputs an independent class probability vector. The probability vectors are then averaged to obtain the final predicted class with highest probability.

274  Figure 1 illustrates the approach to build and train CovidResNet and CovidDenseNet and how the
275  feature-interusability is enabled. The diagram shows the architectures as a sequence of stacks. CovidRes-
276  Net contains a convolutional layer, four residual blocks and a fully connected layer. CovidDenseNet
277  also starts with a convolutional layer, followed by four Denseblocks, which have transition layers, and it
278  ends with a fully connected layer. The red faces in Figure 1 indicate the missing weights in comparison
279  to the standard architectures. As illustrated in the diagram, CovidResNet and CovidDenseNet share
280  subnetworks of their counterparts. More precisely, the first convolutional layer and the first stack are
281  identical to ResNet50 and DenseNet121 and the weights are frozen during training. The subsequent stacks
282  contain much less blocks than the standard models to decrease the total number of features and model
283  complexity. The pruning of the layers leads to a problem for the CovidDenseNet model. A shortening of
284  the dense block is bound to a reduction in the number of channels of the output. This leads to a misfit at
285  the subsequent transition layer because more channels are expected. To solve this problem, we insert an
286  adapter layer after a dense block. The adapter layers consist of a $1 \times 1$ convolution to adjust the number
287  of channels accordingly. More details about our architectures are described in the following subsections.

### 3.1 CovidResNet

289  Our CovidResNet architecture is based on the deep residual networks (ResNets)[26]. ResNet is considered
290  a very deep CNN architecture and the winner of the 2015 ImageNet challenges[52]. The main problems
291  that have been addressed by the ResNet models are the vanishing gradients and performance degradation,
292  which occur during training deep networks. A residual learning framework was proposed, which promotes
293  the layers to learn residual functions with respect to the layer input. While conventional network layers
294  are assumed to learn a desired underlying function $y = f(x)$ by some stacked layers, the residual layers
295  attempt to approximate $y$ via $f(x) + x$. The residual layers start with the input $x$ and evolve to a more
296  complex function as the network learns. This type of residual learning allows training very deep networks
297  and attains an improved performance from the increased depth.
298  The basic building block for CovidResNet is the bottleneck residual module depicted in Figure 2. The
299  input signal to the module passes through two branches. The left branch is a stack of three convolutional
300  layers. The first $1 \times 1$ convolution is used for reducing the depth of the feature maps before the costly $3 \times 3$
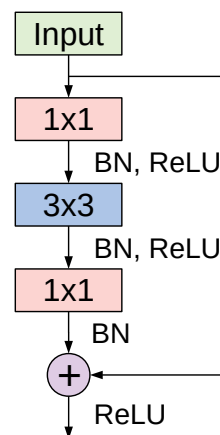
**Figure 2.** The bottleneck residual module used in CovidResNet. The module was first introduced in[26].

301 convolutions, whereas the second $1 \times 1$ is used for increasing the depth to match the input dimensions. The
302 convolutions are followed by batch normalization (BN)[32] and rectified linear unit (ReLU)[42] activation.
303 The right branch is a shortcut connection that connects the module's input with the output of the stacked
304 layers, which are summed up before applying a final ReLU activation.
305      CovidResNet is considered a deep model that consists of 29 layers. The first layer is made of $7 \times 7$
306 convolutional filters with a stride of 2. Following is a max pooling layer to downsample the spatial
307 dimensions. The architecture continues with a stack of four ResNet blocks, where each block has a
308 number between one and three bottleneck residual modules. When moving from a ResNet block to the
309 next one, the spatial dimension is reduced by max pooling and the number of the learned filters is doubled.
310 In the first block, we stack three modules, each having three convolutional layers with 64, 64 and 256
311 filters, respectively. After another max pooling layer, we stack three more bottleneck modules with a
312 configuration of 128, 128 and 512 filters, which forms the second block. The same procedure is repeated
313 for the third and fourth blocks, where the former has two stacked modules and the later has only one. The
314 network ends with an adaptive average pooling step and a fully connected layer. Table 2 summarizes the
315 CovidResNet architecture and a visualization is given in Figure 1. As can be seen in the diagram, the
316 first convolutional layer and the entire first block are frozen during transfer learning. Only the weights of
317 deeper layers are adjusted. The diagram also indicates the complimentary layers that exist in the canonical
318 ResNet50 model but not in CovidResNet.

### 3.2 CovidDenseNet

320 Our CovidDenseNet model is based on the densely connected network (DenseNet) architectures introduced
321 in[28]. DenseNet addressed the notorious problem of vanishing gradients with a different approach
322 compared to ResNet. Instead of using skip connections to combine the feature maps through summation
323 before passing them to the next layer, the feature maps from all preceding layers are considered as the
324 input to the next layer, and its feature maps are passed to all subsequent layers. The advantages of the
325 dense connectivity are the improved flow of information throughout the network, where each layer has a
326 direct access to the gradients from the input and the loss function. DenseNets have shown an improved
327 performance for image recognition tasks and are computationally efficient.
328      The basic building block for the CovidDenseNet model is the DenseNet block. A simplistic form
329 of the dense connectivity of a dense block is shown in Figure 3. The block has three layers and each
330 layer performs a series of batch normalization, ReLU activation, and $3 \times 3$ convolution operations. The
331 concatenated feature maps from all preceding layers are the input to the subsequent layer. Each layer
332 generates $k$ feature maps, where $k$ is the growth rate. So, if $k_0$ is the input to layer $x_0$, then there are
333 $3k + k_0$ feature maps at the end of the 3-layer dense block. However, two main issues arise as the network
334 depth increases. First, as each layer generates $k$ feature maps, the inputs to layer $l$ will be $(l-1)k + k_0$,
335 and with deep networks this number can grow rapidly and slow down computation. Second, when the
336 network gets deeper, we need to reduce the feature maps size to increase the kernel's receptive field. So,

**Table 2.** Description of our CovidResNet architecture for COVID-19 detection. The network accepts an RGB-input of size $253 \times 349$ pixels. The residual modules are placed in brackets multiplied by the number of modules stacked per block.

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

when concatenating feature maps of different sizes we need to match the dimensions. The first issue is addressed by introducing a bottleneck layer of $1 \times 1$ 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 $1 \times 1$ convolution followed by an average pooling operation.
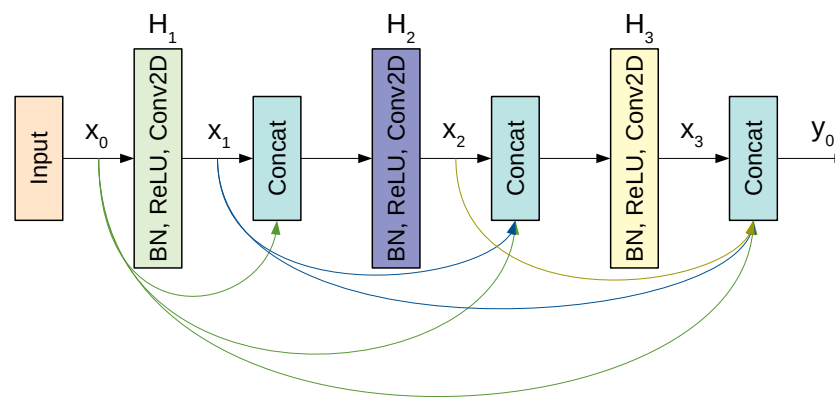
**Figure 3.** A schematic diagram of a 3-layer dense block used in the DenseNet architecture.

To ensure the inter-usability of the weights, CovidDenseNet contains a set of adapter layers consisting of a $1 \times 1$ convolution to increase the number of channels to the size required by the subsequent layer. The number of channels can be seen in Table 3. The last adapter layer is optional. 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 $7 \times 7$ filters and uses a stride of 2, followed by a max pooling operation. Then we stack 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. All details about the CovidDenseNet architecture including the number of blocks as well as the input and output volumes are summarized in Table 3. 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.

**Table 3.** Our CovidDenseNet architecture for COVID-19 detection. The network accepts an RGB-input of size $253 \times 349$ pixels.

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

## 4 METHODOLOGY

Figure 4 illustrates the entire process on how we conduct our experiments. We show that on the SARS-CoV-2 CT-scan dataset, which has three main classes. The dataset is split into training and test sets comprising 60% and 40% of the images from each class, respectively. Suitable preprocessing steps for normalization and data augmentation are carried out before training the models. The evaluation and testing are performed using various standard performance evaluation metrics. In order to reduce the variance of deep neural network models and improve the performance of the single models we consider the ensemble learning approach. An ensemble is built using several independently trained deep models and then combining the predictions from these models. At test time, each CT image is passed to the considered models where each model returns a vector of class scores known as the posterior probabilities. The resulting vectors of posterior probabilities from all models are then averaged and the actual prediction is given as the class with the highest probability. This is a commonly used ensembling approach for neural networks and is referred to as a committee of networks. More details about each step are described in the following subsections.

### 4.1 Datasets

In order to evaluate our proposed models we used two benchmark CT image datasets, which are described below.

#### 4.1.1 SARS-CoV-2 CT-scan dataset

The SARS-CoV-2 CT-scan dataset[57] 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 4. As the dataset contains patients with other pulmonary diseases and the CT images have variable sizes, the dataset is challenging. Figure 5 shows 12 CT images from the
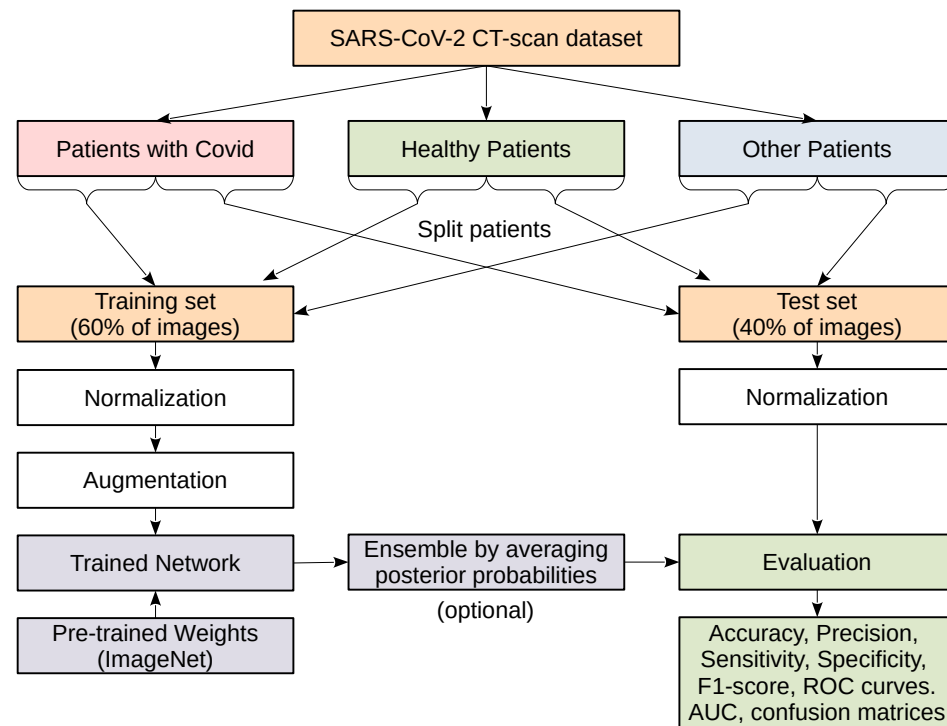
**Figure 4.** A flowchart representing the various steps in the process of models evaluation.

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.

### 4.1.2 COVID19-CT dataset

The COVID19-CT dataset[78] contains a total of 746 CT images. 349 CT images for patients with confirmed COVID-19 and 397 CTs for patients with other non-COVID-19 viral pneumonia. The CT images for COVID-19 were gathered from preprints on medRxiv and bioRxiv and they show various manifestations of COVID-19 pneumonia. The CTs for non-COVID19 cases were collected from different open sources. Due to the heterogeneity of sources, the images have different visual characteristics and varying sizes between $124 \times 153$ and $1485 \times 1853$, which make the dataset very challenging. Figure 6 shows sample CT images from the COVID19-CT dataset.

**Table 4.** Number of subjects and CT scans for each category in the SARS-CoV-2 CT-scan dataset.

| Dataset | No. | COVID-19 | Healthy | Others | Total |
|---|---|---|---|---|---|
| SARS-CoV-2 CT-scan | subjects | 80 | 50 | 80 | 210 |
|  | images | 2168 | 758 | 1247 | 4173 |
| COVID19-CT | subjects | 216 | - | 171 | 337 |
|  | images | 349 | - | 397 | 746 |

### 4.2 Data Preprocessing and Splitting

Wide variations in the CT image sizes in the SARS-CoV-2 CT-scan and COVID19-CT datasets 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 image into a fixed-sized canvas. We apply padding with the average color of the ImageNet
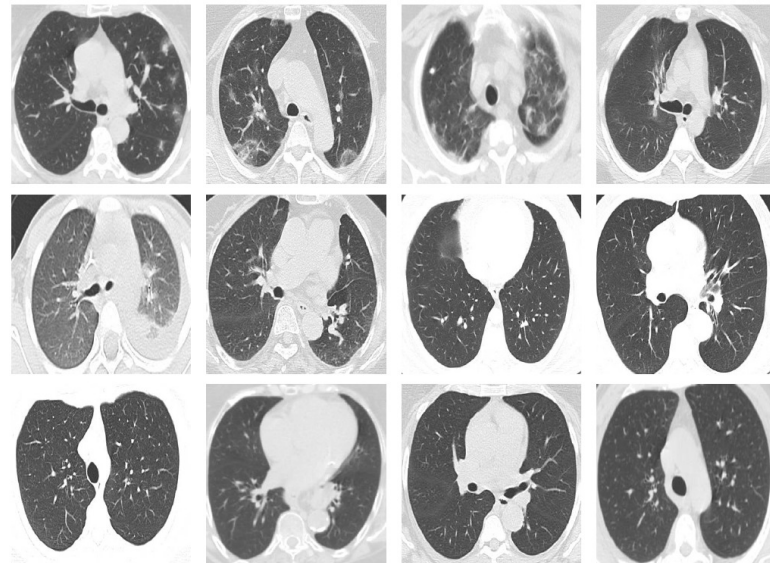
**Figure 5.** Sample CT images from the SARS-CoV-2 CT scan dataset. The CTs represent four images of COVID-19 (first row), four images of Others class (second row), and four images from the Healthy class (third row).
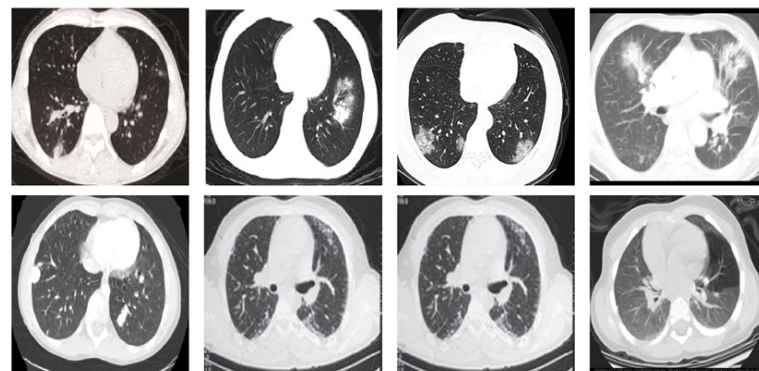
**Figure 6.** Sample CT images from the COVID19-CT dataset. The first row shows four CT images of COVID-19 and the second row shows four CTs for non-COVID19 viral pneumonia.

394  dataset[18] when necessary to match the target shape. We empirically tried different input sizes and
395  found that a canvas with a spatial dimension of $253 \times 349$ works best for CT images from the considered
396  datasets and our architectures. Due to the limited amount of training data and the fact that deep neural
397  networks require large amounts of data to optimize millions of parameters, we recompense the lack of
398  data by implementing different augmentation steps to improve the network's ability to generalize. The
399  augmentation steps include random rescaling, random cropping, Gaussian noise, brightness and contrast
400  changes and random horizontal flipping. Finally, the images are normalized according to the mean and
401  standard deviation of the ImageNet dataset.
402      To conduct our experiments we split the SARS-CoV-2 CT-scan dataset into training and test sets. We
403  follow the subject-wise structure of the dataset, such that the two sets of persons in the training and test set
404  are disjunct. Hence, it is assured, that we evaluate our models on unseen persons. However, the number
405  of CT images per person vary. We divide the subjects into training and test sets such that the amounts
406  of training and test images are 60% and 40%, respectively. The same ratio of persons is used for both
407  scenarios of multi-class and binary classification tasks. Within one scenario we choose the same split for
408  each architecture for the sake of consistency and comparability. However, for the COVID19-CT dataset
409  we use the original split provided by the authors in[78], where they split the dataset into 60% for training,

410  15% for validation, and 25% for testing and reporting results.

411  The class representations are imbalanced for the SARS-CoV-2-CT-scan and COVID19-CT datasets.
412  We apply undersampling to make sure that our models are not biased towards class frequencies. In this
413  procedure a new balanced subset is drawn from the training set every epoch. The images from this subset
414  are augmented and used to update the model weights. Thus it is assured that the model was updated with
415  the same amount of images of every class.

416  ## 4.3 Performance Evaluation Metrics

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

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

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

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

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

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

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

423  Models are difficult to compare when the performance assessment is based on a single evaluation
424  metric only. Hence, we provide multiple evaluation metrics to enable a profound analysis. We plot the
425  ROC curves to visualize the diagnostic ability of the models to differentiate between the different classes.
426  We also compute the area under the ROC curve (AUC) for each model. The ROC curves show the trade-off
427  between the true positive rate (sensitivity) and the false negative rate (1-specificity) at various threshold
428  values. The AUC summarizes the ROC curve and measures the ability of a model to distinguish between
429  the different classes. A high AUC value indicates better performance of the model at distinguishing
430  between the classes.

431  We also provide the confusion matrix for detailed class-wise results. The confusion matrix clearly
432  tells about the exact numbers of correctly detected positive and negative cases, as well as the type of error
433  a model makes (see Table 5).

**Table 5.** Confusion matrix for binary classification

|  | Predicted Values | |
|---|---|---|
|  | **True Positive** | **False Negative** |
| **Actual Values** | **False Positive** | **True Negative** |

434  ## 4.4 Transfer Learning

435  Transfer learning is a method in deep learning, which has become quite popular in the computer vision
436  community because it might significantly boost recognition performance[4]. The idea bases on the
437  transferability of network weights between related image recognition tasks and relies on the universal
438  validity of the visual filters learned during training. Usually, transfer learning occurs in a two-step

439 procedure. First, a model's weights are trained for one task on a dataset, which is typically large.
440 Subsequently, a model is initialized with the weights to solve the actual task and often it is also fine-tuned.
441 As the size of the considered CT image datasets is still limited, we opt for transfer learning to
442 benefit from the pretrained image filters. We initialize SqueezeNet[30], VGG-16[54], Inception-V3[61],
443 ResNet50[26] and DenseNet121[28] with weights, which have been optimized for the ImageNet dataset[52].
444 Parts of our proposed architectures exhibit compatible weight configurations such that we can initialize
445 many weights with ResNet50 and DenseNet121 models that have been pretrained on ImageNet. In
446 CovidResNet all weights are pretrained, but the last layer. In CovidDenseNet the adapter layers and the
447 last layer are randomly initialized and all other weights are copied from the DenseNet121 model that was
448 pretrained on ImageNet.
449 We empirically found that it is not necessary to adjust all weights to the COVID-19 detection problem.
450 We assume that the filters from the first layers in a computer vision network provide somewhat generic
451 filters that can be used for the SARS-CoV-2 CT-scan and COVID19-CT datasets. The idea is to reduce
452 the risk of overfitting by lowering the amount of trained weights. Thus, we freeze the first convolutional
453 layer and the first convolutional block of CovidResNet and only adapt the remaining weights. The first
454 convolutional layer, the first dense block and the first transition layer of CovidDenseNet are also frozen.
455 All weights in the models ResNet50 and DenseNet101 are fine-tuned to enable the comparison between
456 standard models and our novel architectures together with our specifically designed fine tuning strategy.
457 An overview of the CovidResNet and CovidDenseNet architectures can be seen in Figure 1. The layers
458 with frozen weights are highlighted in orange. The trainable layers are colored in blue. See Table 6 for
459 important characteristics of the proposed CovidResNet and CovidDenseNet models compared with the
460 standard architectures.

**Table 6.** Characteristics of CovidResNet and CovidDenseNet compared with the standard architectures. The training time is given per epoch for the SARS-CoV-2 CT-scan dataset.

| Model | Network characteristics | | | | | |
|---|---|---|---|---|---|---|
| | Default input size | Custom input size | Layers | Total Parameters (M) | Trainable Parameters (M) | Training time (s) |
| SqueezeNet | $227 \times 227$ | $253 \times 349$ | 18 | 0.72 | 0.72 | 11.4 |
| VGG-16 | $227 \times 227$ | $253 \times 349$ | 16 | 134.28 | 134.28 | 27.4 |
| Inception-V3 | $299 \times 299$ | $253 \times 349$ | 48 | 21.79 | 21.79 | 16.9 |
| ResNet50 | $224 \times 224$ | $253 \times 349$ | 50 | 23.51 | 23.51 | 19 |
| DenseNet121 | $224 \times 224$ | $253 \times 349$ | 121 | 6.96 | 6.96 | 22 |
| CovidResNet | $253 \times 349$ | $253 \times 349$ | 29 | 9.84 | 9.61 | 11 |
| CovidDenseNet | $253 \times 349$ | $253 \times 349$ | 43 | 3.13 | 2.75 | 12 |

461 ## 4.5 Model Training
462 All standard models are initialized using pretrained weights that have been optimized for the ImageNet
463 dataset. The models are trained using the LAMB optimizer[76] with an initial learning rate of 0.0003 and
464 cross-entropy loss. The standard models are trained for 100 epochs until convergence. Our proposed
465 architectures and the SqueezeNet models need more epochs to converge and we stop training after 150
466 epochs. The reason for the higher number of epochs is the finding that it takes more iterations to train
467 efficient networks that have fewer weights. The learning rate is step-wise reduced until a value of $10^{-6}$ is
468 reached at the end of training. The batch size is 32. We also apply weight decay to regularize the training
469 process. Optimization is performed within the PyTorch framework using an Nvidia GTX 1080 GPU.

470 # 5 RESULTS AND DISCUSSION

471 This section presents and discusses the experimental results obtained by our proposed COVID-Nets
472 architectures for COVID-19 detection. We begin our discussion by the obtained results for the SARS-
473 CoV-2 CT-scan dataset, and then we discuss the obtained results for the COVID19-CT dataset.

474 ## 5.1 Results for SARS-CoV-2 CT-scan dataset
475 Here we report the obtained results on the SARS-CoV-2 CT-scan dataset for the three-class and the binary
476 classification tasks. We also compare the performance of our models with state-of-the-art deep networks.
477 First, we evaluate the ability of our models to differentiate patients with COVID-19, other non-COVID-19
478 viral lung infections, and non-infected healthy individuals. Second, we discuss the results obtained for all
479 three possible binary classification scenarios from the same dataset.

### 5.1.1 Three-class Classification Results

Table 7 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 all standard fine-tuned models. Among the single network architectures, our CovidDenseNet model achieves the best overall performance with an average 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 indicates 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 all other models.

**Table 7.** Comparison of our models against state-of-the-art models for the multi-class problem. The results are given in percentages and the best values are written in bold.

| Model | Class | Evaluation Metrics | | | | |
|---|---|---|---|---|---|---|
| | | Accuracy | Precision | Sensitivity | Specificity | F1-score |
| SqueezeNet | COVID-19 | | 86.99 | 87.30 | 85.95 | 87.15 |
| | Healthy | | 67.05 | 75.08 | 91.62 | 70.84 |
| | Others | | 68.35 | 62.83 | 87.74 | 65.47 |
| | Macro average | 77.78 | 74.13 | 75.06 | 88.44 | 74.49 |
| VGG-16 | COVID-19 | | 90.70 | 87.87 | 90.30 | 89.27 |
| | Healthy | | 64.66 | 79.94 | 90.08 | 71.49 |
| | Others | | 76.61 | 69.49 | 91.06 | 72.88 |
| | Macro average | 80.96 | 77.33 | 79.10 | 90.48 | 77.88 |
| Inception-V3 | COVID-19 | | 88.55 | 90.18 | 87.44 | 89.36 |
| | Healthy | | 66.95 | 76.05 | 91.48 | 71.21 |
| | Others | | 77.35 | 68.28 | 91.57 | 72.53 |
| | Macro average | 81.08 | 77.62 | 78.17 | 90.16 | 77.70 |
| 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 |
| **Our 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 |
| **Our 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** |
| **Proposed 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 |
| **Proposed 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** |

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 7. 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. Building an ensemble through a combination of our independently trained CovidResNet and CovidDenseNet models and their baselines increases the classification accuracy for all classes. When combining the prediction from CovidResNet and CovidDenseNet models (Ensemble 1) we notice a slight improvements over any of the single models in almost all evaluation metrics. Whereas combining CovidDenseNet and DenseNet121 models (Ensemble 2) improved the detection rate of CovidDenseNet for the COVID-19 class with 3%.

Figure 7 shows the confusion matrix for each of our proposed model and standard fine-tuned models, as well as the proposed ensembles that achieve the best 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.
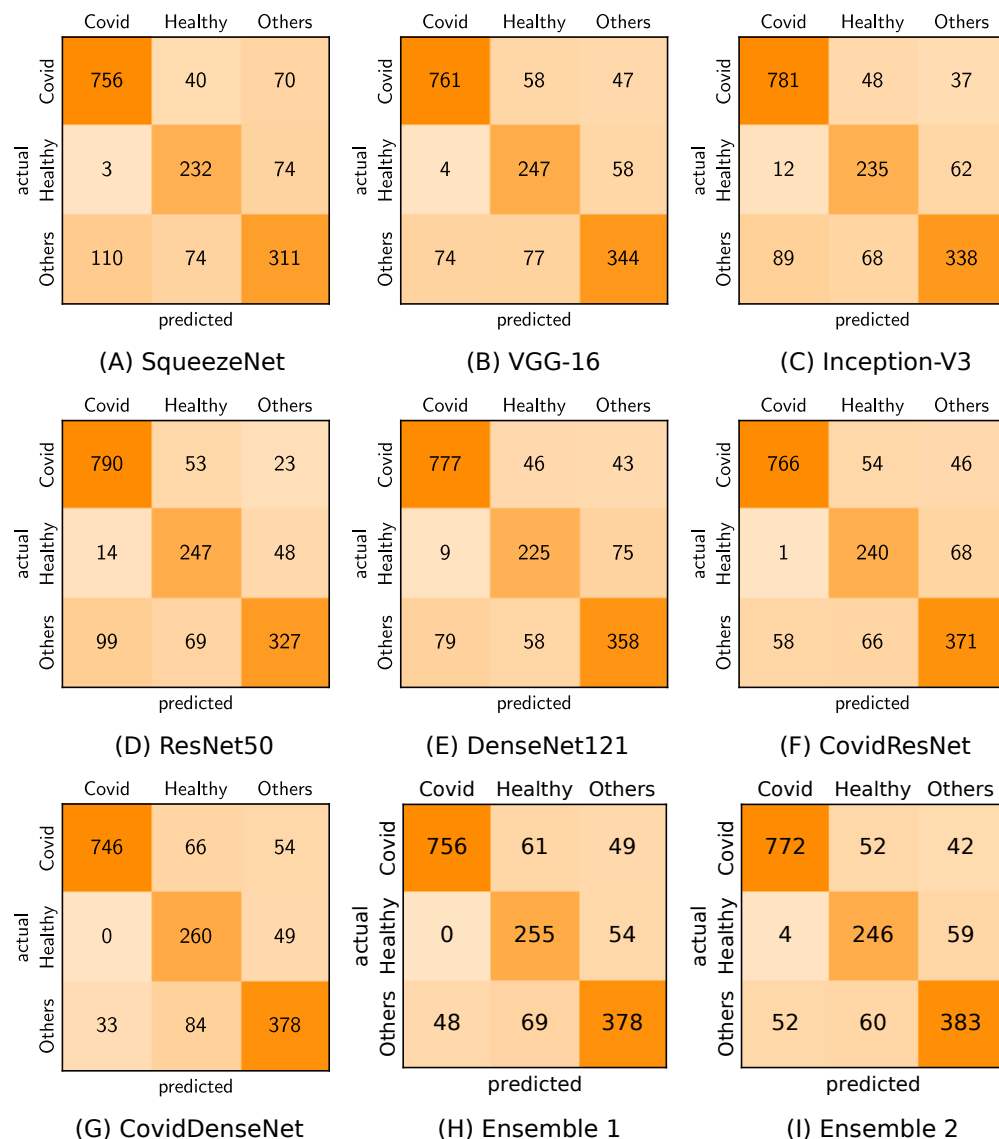
**Figure 7.** Confusion matrices generated by the different models for the three-class classification task.
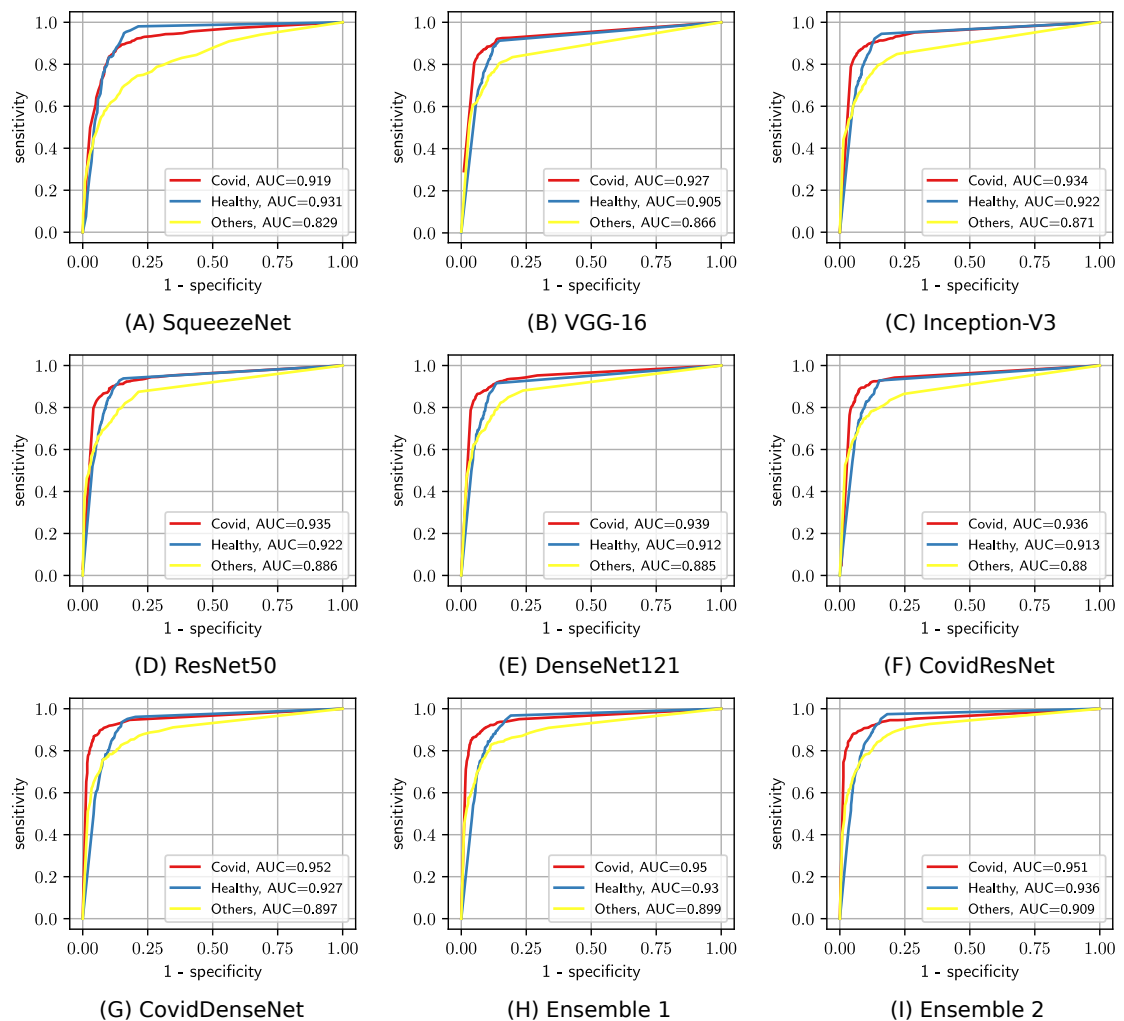
**Figure 8.** The ROC curves and their AUC scores for the different models showing their ability to differentiate between the three classes.

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 8. Our CovidResNet and CovidDenseNet models 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 models 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. 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, Ensemble 1, we notice an improvement in the AUC score for the all three classes within 2%. Similar results are achieved when we combine the CovidDenseNet and DenseNet121, Ensemble 2, even though the models were trained on the same training split of the used dataset.

### 5.1.2 Two-class Classification Results

We tested our proposed architectures on binary classification tasks to validate their ability to distinguish between CT images of all possible classes, as well as to investigate the difficulty of these subtasks from the considered dataset. We investigate three experimental scenarios. First, we tested our models to differentiate patients with COVID-19 from healthy individuals (COVID-19 vs. Healthy). Then, we tested the models to distinguish COVID-19 cases from non-COVID-19 patients infected by other lung diseases (COVID-19 vs. Others). Finally, we tested our models to differentiate non-COVID-19 patients infected by other pulmonary diseases from healthy subjects (Others vs. Healthy). Table 8 presents the results

530 obtained by each model under each of these scenarios.

**Table 8.** The obtained results under three binary classification scenarios.

| Task | Model | Evaluation Metrics | | | | |
|------|-------|----------|-----------|-------------|-------------|----------|
| | | **Accuracy** | **Precision** | **Sensitivity** | **Specificity** | **F1-score** |
| COVID-19 vs. Healthy | SqueezeNet | 92.85 | 95.78 | **94.46** | 88.35 | 95.12 |
| | VGG-16 | 93.96 | 97.95 | 93.76 | 94.50 | 95.81 |
| | Inception-V3 | 93.19 | 96.90 | 93.76 | 91.59 | 95.30 |
| | 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 | SqueezeNet | 81.25 | 86.88 | 83.372 | 77.43 | 85.09 |
| | VGG-16 | 85.24 | 88.38 | 88.68 | 79.09 | 88.53 |
| | InceptionV3 | 81.76 | 84.68 | 87.41 | 71.63 | 86.02 |
| | 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 | SqueezeNet | 86.15 | **91.04** | 86.44 | **85.67** | 88.68 |
| | VGG-16 | 85.13 | 87.78 | 88.66 | 79.18 | 88.22 |
| | Inception-V3 | 85.13 | 86.32 | **90.69** | 75.77 | 88.45 |
| | 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 |
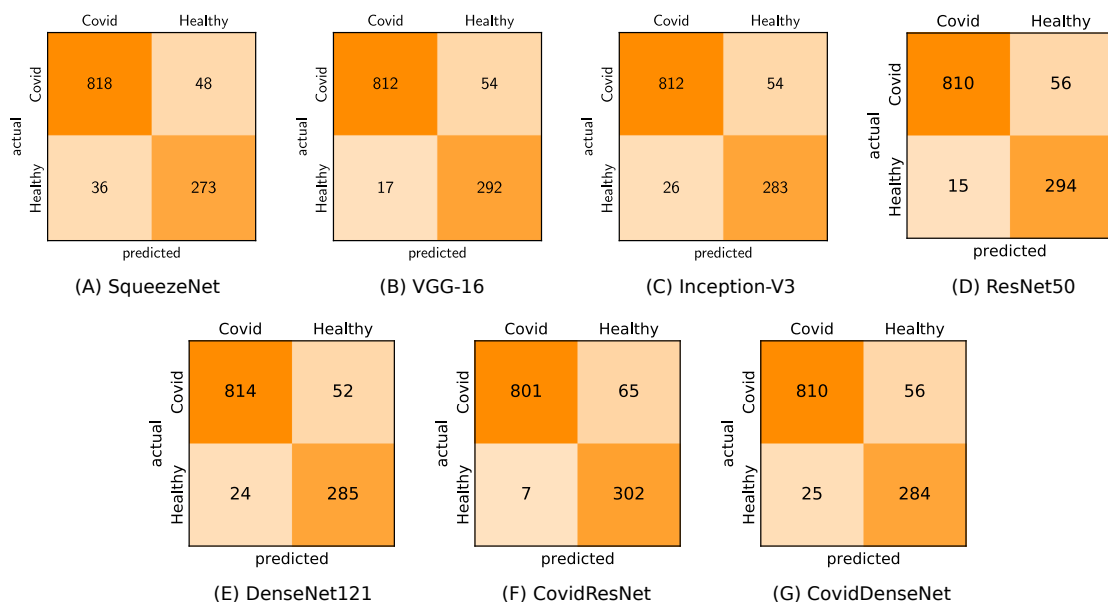
**Figure 9.** Confusion matrices generated by all models for COVID-19 vs. Healthy classification.

531     In the first scenario (COVID-19 vs. Healthy) we used 866 CT images of COVID-19 and 309 CT
532 images from the healthy class for testing. As we can see from Table 8 and under this scenario, the seven
533 models achieve very competitive performance with accuracy above 93% and F1-score above 95%. The
534 models also achieve high precision values above 96%, where our proposed CovidResNet model achieves
535 the highest precision score of 99.13%, indicating that almost all the predicted subjects as COVID-19
536 are correct and only 7 out of 309 healthy CT images were incorrectly classified as COVID-19 positive.
537 CovidResNet also attains the highest specificity score of 97.73%, which indicates its ability to correctly
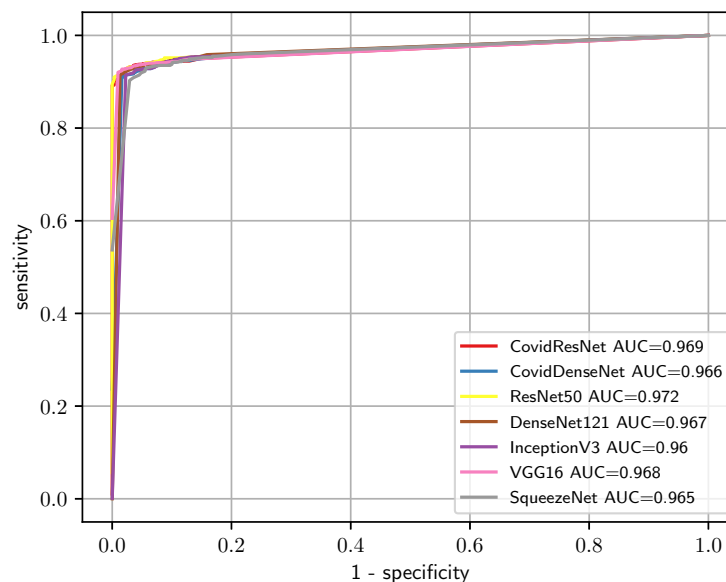
**Figure 10.** Comparison of the predictive performance for CovidResNet and CovidDenseNet and the standard models for COVID-19 vs. Healthy classification. The ROC curves and AUC scores show the competitive performance for all models.

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. For a detailed class-wise results, the confusion matrix for each specific model under the considered scenario is presented in Figure 9.

Figure 10 shows the ROC curves for all evaluated models. Looking at the ROC curves and the AUC scores we can see that all tested models perform on a similar level. The ROC curves look identical and the AUC scores vary within a range less than 1%, with ResNet50 achieving a slightly higher AUC sore of 97.2%.

In the second scenario (COVID-19 vs. Others) we investigated 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 non-COVID-19 viral infections. The obtained results in Table 8 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 standard models, 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 standard models with 12%. The lower specificity of the standard models 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 11
shows the confusion matrix for each model and what type of error each specific model makes.
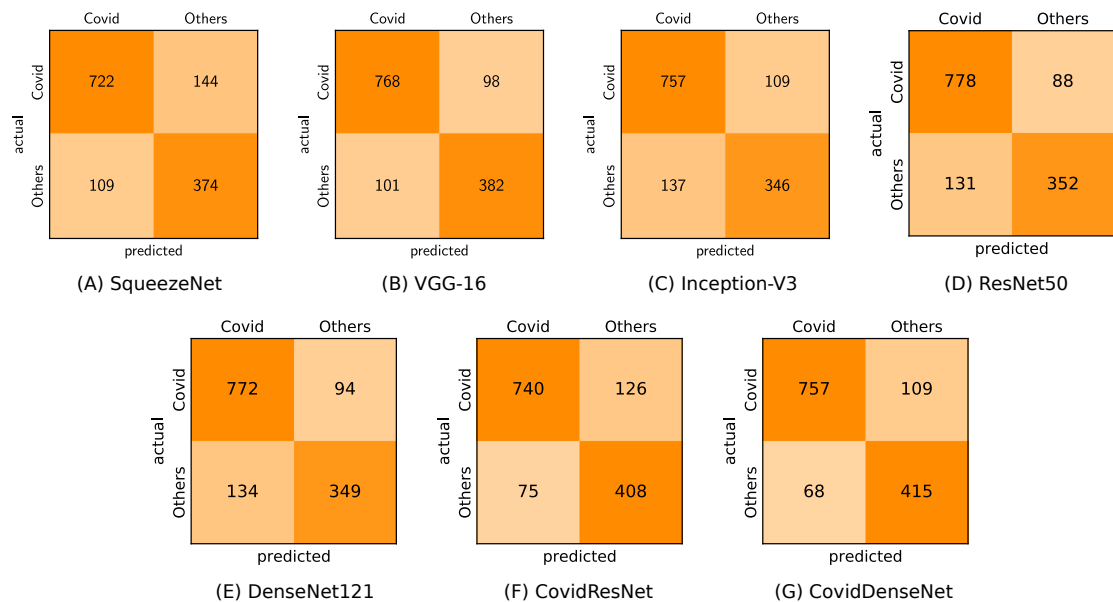
**Figure 11.** Confusion matrices for COVID-19 vs. Others classification for all tested models.
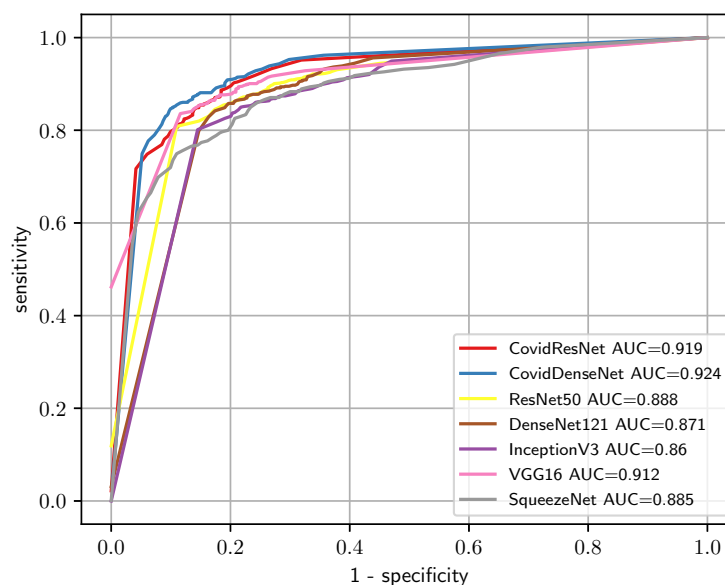
**Figure 12.** Predictive performance of our proposed CovidResNet and CovidDenseNet models against
the standard models for COVID-19 vs. Others classification. The ROC curves show powerful a predictive
power for the CovidDenseNet model.

We also compare the performance of the different models under this scenario by plotting the ROC
curve and computing the AUC score for each model. Figure 12 shows the ROC curves, where we clearly
see that our CovidResNet and CovidDenseNet model are superior to the other models as their ROC curves
are closer to the top-left corner and they achieve higher AUC values. The highest AUC score of 92.4% is
achieved by our CovidDenseNet model exceeding its deeper counterpart DenseNet121 model with more

than 5%.

In our third scenario (Others vs. Healthy) we tested 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 F1-score of 89.29%. With a sensitivity score of 90.28% CovidResNet indicates its ability to detect above 90% of the infected cases. Figure 13 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.
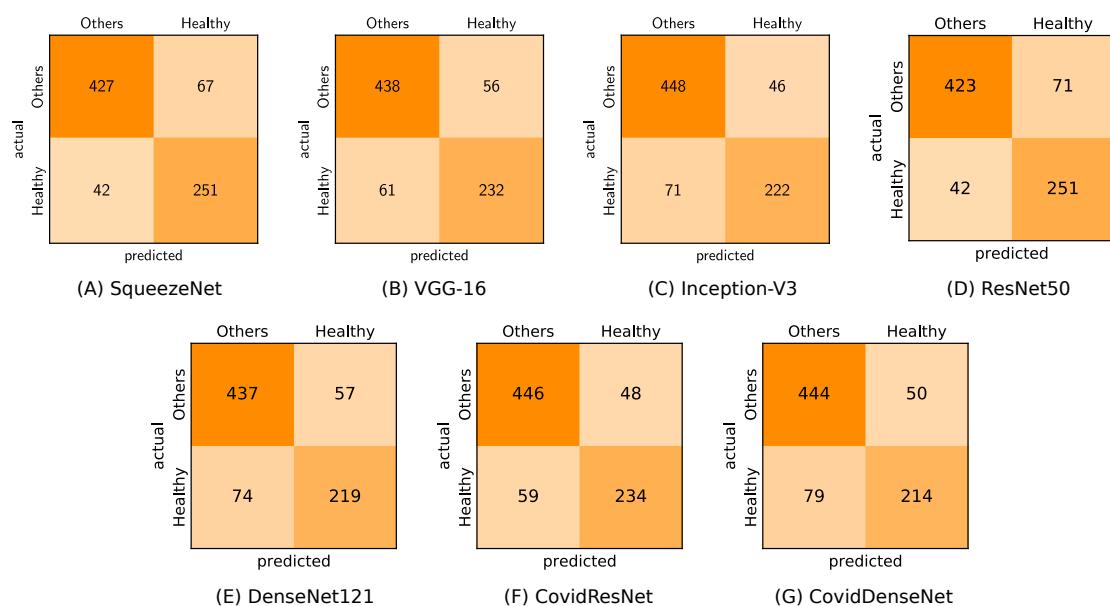
**Figure 13.** Confusion matrices generated by all tested models for Others vs. Healthy classification.

Figure 14 presents the ROC curves and their corresponding AUC scores for all tested models. Again, our proposed models show superior performance compared with other deeper models. Our CovidResNet model achieves a high AUC score of 92.8% comparable to the best model (SqueezeNet) and its ROC curve appears closer to the top-left corner. Our CovidDenseNet model also has a comparable performance to the standard models with AUC score of 89.8%, which indicates approximately 2% improvement compared to its deeper DenseNet121 model.

## 5.2 Results for the COVID19-CT dataset

To validate the effectiveness of our proposed architectures we use the COVID19-CT dataset, which shares similar visual characteristics with the SARS-CoV-2 CT-scan dataset and is also available in a subject-wise structure. We consider the binary classification scenario of differentiating COVID-19 cases from other non-COVID viral pneumonia. We report our results based on the test split provided by the authors in [78] where 98 COVID-19 and 105 non-COVID CT images are used. Table 9 shows the performance of our CovidResNet and CovidDenseNet models compared with the standard models.

From the table, we observe the competitive performance of our proposed models in terms of accuracy, sensitivity, and F1-score. Moreover, our models achieved much better results compared with the reported results from the literature. Looking closely to the results in Table 9 we can see a similarity with the second scenario in Table 8 when differentiating COVID-19 and other non-COVID CT images. Under both
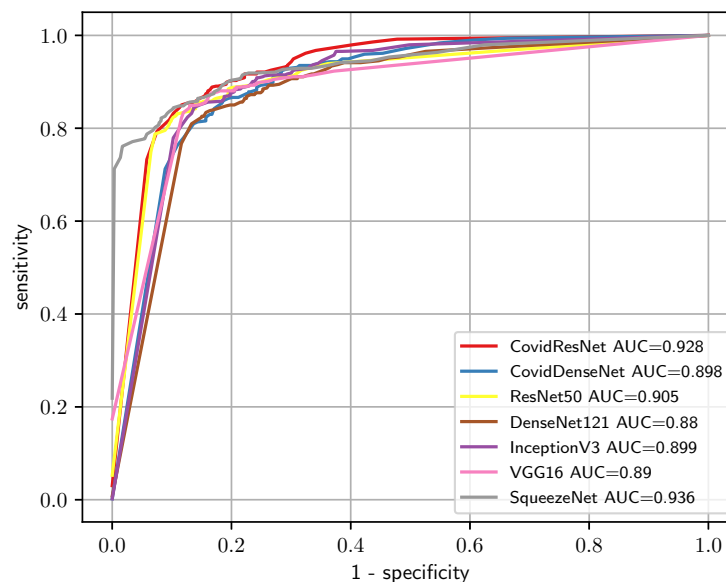
**Figure 14.** Comparison of CovidResNet and CovidDenseNet against the standard models using the ROC curves and AUC scores for the Others vs. Healthy classification task.

**Table 9.** Performance comparison of our proposed networks against all standard models and previous work on the COVID19-CT dataset.

| Task | Model | Evaluation Metrics | | | | |
|------|-------|----------|-----------|-------------|-------------|----------|
| | | **Accuracy** | **Precision** | **Sensitivity** | **Specificity** | **F1-score** |
| COVID vs. Non-COVID | SqueezeNet | 73.89 | 79.22 | 62.24 | 84.76 | 69.70 |
| | VGG-16 | 80.79 | 81.05 | 78.57 | 82.85 | 79.79 |
| | Inception-V3 | 79.80 | 78.21 | 80.61 | 79.05 | 79.40 |
| | ResNet50 | **81.77** | 83.52 | 77.55 | 85.71 | 80.42 |
| | DenseNet121 | 81.28 | **84.09** | 75.51 | **86.67** | 79.57 |
| | CovidResNet | 81.28 | 79.41 | 82.65 | 80.00 | 81.00 |
| | CovidDenseNet | **81.77** | 79.05 | **84.69** | 79.05 | **81.77** |
| Previous Work | ResNet50[27] | 69.0 | - | - | - | 72.0 |
| | DenseNet121[27] | 76.0 | - | - | - | 77.0 |
| | CRNet[27] | 72.0 | - | - | - | 76.0 |
| | Redesigned COVID-Net[67] | 79.0 | - | - | - | 79.0 |
| | M-Inception[66] | 81.0 | - | - | - | 82.0 |
| | Evidential Covid-Net[29] | 81.0 | - | - | - | 81.0 |

608 scenarios our CovidDenseNet model achieved the best performance, which indicates the effectiveness of
609 our models to work on different datasets and to differentiate COVID-19 from other non-COVID cases.
610      In order to give a holistic view on how well our proposed models perform on the COVID19-CT dataset
611 and to show what types of errors they are making, we present the confusion matrices for each model in
612 Figure 15. We also validate the ability of our models to distinguish between classes by plotting the ROC
613 curves and computing the corresponding AUC for each model. Figure 16 depicts the ROC curves for the
614 different models. We can see that our proposed models have the superior performance over the standard
615 models. Our CovidDenseNet achieved the highest AUC score of 87.5% and its ROC curve is higher than
616 the ROC curves for other models. Therefore, we can say that our CovidDenseNet model performed better
617 than all models in detecting the positive class (i.e. COVID-19) in the COVID19-CT dataset.

## 6 CONCLUSION

619 We proposed two deep CNN architectures (CovidResNet and CovidDenseNet) for the automated detection
620 of COVID-19 using chest CT scans. The models were developed and validated on two benchmark CT
621 image datasets. We also presented the first experimental study on the large-scale multi-class SARS-CoV-2
622 CT-scan dataset, which has more than 4000 CT scans. Extensive experiments have been conducted to
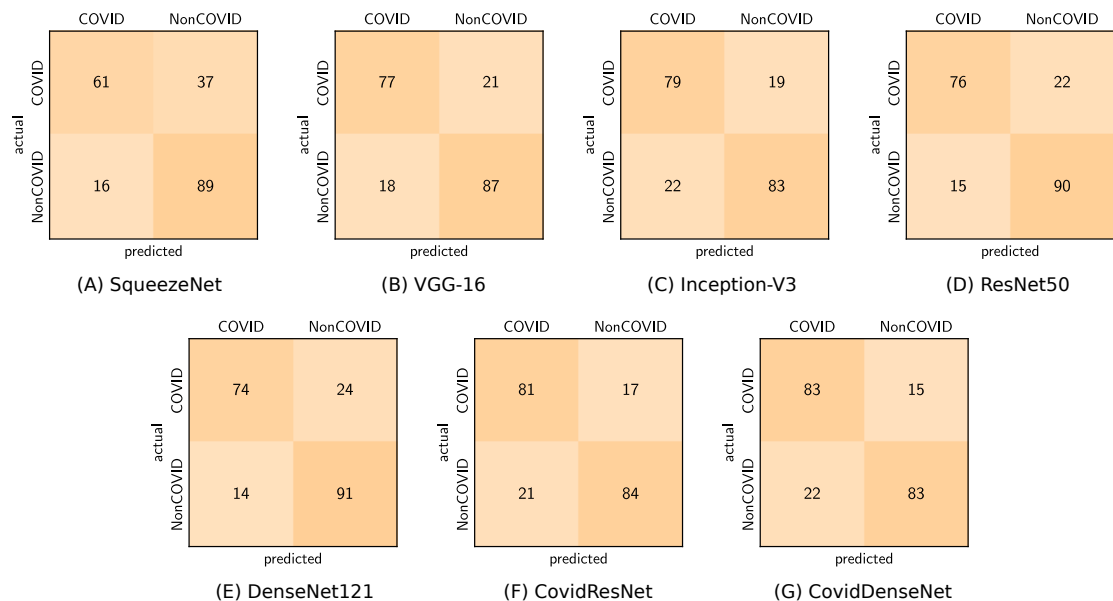
**Figure 15.** Confusion matrices obtained by the different models for COVID vs. non-COVID classification from the COVID19-CT dataset.
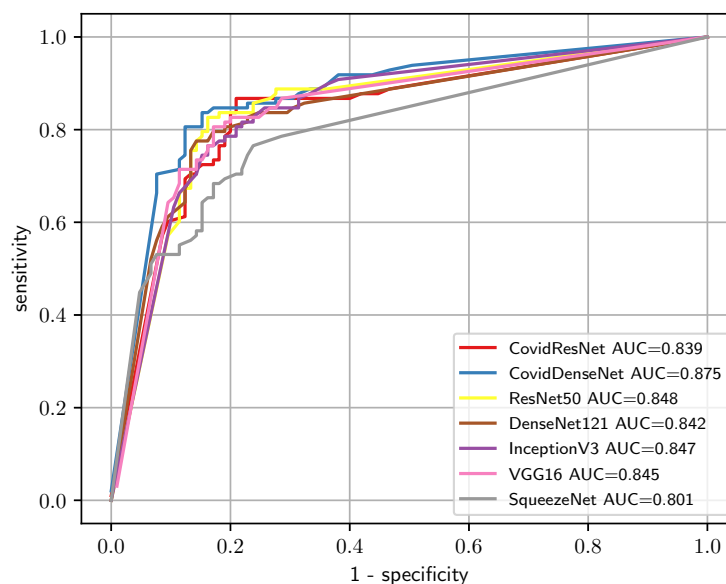
**Figure 16.** Comparison of the diagnostic ability of our CovidResNet and CovidDenseNet with the standard models to distinguish COVID-19 and non-COVID class in the COVID19-CT dataset.

evaluate our models in the multi-class and binary classification tasks from the SARS-CoV-2 CT-scan dataset. 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 results compared with the standard architectures, while being more computationally efficient. Second, we conducted three binary classification scenarios 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 revealed the superior performance of our proposed models over the baseline models. Finally, we tested our models on the COVID19-CT dataset to differentiate COVID-19 patients from others with non-COVID-19 viral infection. Our CovidDenseNet

632 model achieved the best overall detection performance exceeding the state-of-the-art models.

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

# 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. (2019a). 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. (2019b). Handcrafted versus CNN Features for Ear Recognition. *Symmetry*, 11(12):1493.

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

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

[7] Ardakani, A. A., Kanafi, A. R., Acharya, U. R., Khadem, N., and Mohammadi, A. (2020). Application of deep learning technique to manage covid-19 in routine clinical practice using ct images: Results of 10 convolutional neural networks. *Computers in Biology and Medicine*, 121:103795.

[8] 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.

[9] 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.

[10] Bahgat, W. M., Balaha, H. M., AbdulAzeem, Y., and Badawy, M. M. (2021). An optimized transfer learning-based approach for automatic diagnosis of COVID-19 from chest X-ray images. *PeerJ Computer Science*, 7:e555.

[11] Barstugan, M., Ozkaya, U., and Ozturk, S. (2020). Coronavirus (COVID-19) classification using CT images by machine learning methods. *arXiv preprint arXiv:2003.09424*.

[12] 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.

[13] 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.

[14] 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., Reaz, M. B. I., and Islam, M. T. (2020). Can AI help in screening viral and COVID-19 pneumonia? *IEEE Access*, 8:132665–132676.

[15] Cohen, J. P., Morrison, P., Dao, L., Roth, K., Duong, T. Q., and Ghassemi, M. (2020). Covid-19 image data collection: Prospective predictions are the future. *arXiv preprint arXiv:2006.11988*.

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

[17] 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.

[18] 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.

[19] di Radiologia Medica e Interventistica, S. I. (2021). Covid-19 database.

[20] 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.

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

[22] 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.

[23] 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.

[24] Hasan, K., Alam, A., Elahi, T. E., Roy, S., and Wahid, S. R. (2020). CVR-Net: A deep convolutional neural network for coronavirus recognition from chest radiography images. *arXiv preprint arXiv:2007.11993*.

[25] 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*.

[26] 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.

[27] He, X., Yang, X., Zhang, S., Zhao, J., Zhang, Y., Xing, E., and Xie, P. (2020). Sample-Efficient Deep Learning for COVID-19 Diagnosis Based on CT Scans. *medrxiv*.

[28] 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.

[29] Huang, L., Ruan, S., and Denoeux, T. (2021). Covid-19 classification with deep neural network and belief functions. *arXiv preprint arXiv:2101.06958*.

[30] Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., and Keutzer, K. (2017). Squeezenet: Alexnet-level accuracy with 50x fewer parameters and¡ 0.5 mb model size. In *Proceedings of the International Conference on Learning Representations (ICLR)*.

[31] 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*.

[32] 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.

[33] 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*.

[34] 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.

[35] 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.

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

[37] 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.

[38] 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.

[39] 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.

[40] 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*,

741  126:108961.

[41] 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.

[42] 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.

[43] 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.

[44] Özkaya, U., Öztürk, Ş., and Barstugan, M. (2020a). Coronavirus (COVID-19) classification using deep features fusion and ranking technique. In *Big Data Analytics and Artificial Intelligence Against COVID-19: Innovation Vision and Approach*, pages 281–295. Springer.

[45] Özkaya, U., Öztürk, Ş., Budak, S., Melgani, F., and Polat, K. (2020b). Classification of COVID-19 in Chest CT Images using Convolutional Support Vector Machines. *arXiv preprint arXiv:2011.05746*.

[46] Öztürk, Ş., Özkaya, U., and Barstuğan, M. (2021). Classification of Coronavirus (COVID-19) from X-ray and CT images using shrunken features. *International Journal of Imaging Systems and Technology*, 31(1):5–15.

[47] Pachetti, M., Marini, B., Benedetti, F., Giudici, F., Mauro, E., Storici, P., Masciovecchio, C., Angeletti, S., Ciccozzi, M., Gallo, R. C., and Zella, D. (2020). Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *Journal of Translational Medicine*, 18:1–9.

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

[49] 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).

[50] Ragab, D. A. and Attallah, O. (2020). FUSI-CAD: Coronavirus (COVID-19) diagnosis based on the fusion of CNNs and handcrafted features. *PeerJ Computer Science*, 6:e306.

[51] 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.

[52] 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.

[53] Shi, F., Xia, L., Shan, F., Song, B., Wu, D., Wei, Y., Yuan, H., Jiang, H., He, Y., Gao, Y., Sui, H., and Shen, D. (2021). Large-scale screening to distinguish between COVID-19 and community-acquired pneumonia using infection size-aware classification. *Physics in Medicine & Biology*, 66(6):065031.

[54] 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.

[55] 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.

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

[57] 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*.

[58] Song, Y., Zheng, S., Li, L., Zhang, X., Zhang, X., Huang, Z., Chen, J., Zhao, H., Jie, Y., Wang, R., Chong, Y., Shen, J., Zha, Y., and Yang, Y. (2021). Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*.

[59] 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.

[60] 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*

796     *Computer Vision and Pattern Recognition*, pages 1–9.

[61]   Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2818–2826.

[62]   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.

[63]   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. (2021a). AI-assisted CT imaging analysis for COVID-19 screening: Building and deploying a medical AI system. *Applied Soft Computing*, 98:106897.

[64]   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.

[65]   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.

[66]   Wang, S., Kang, B., Ma, J., Zeng, X., Xiao, M., Guo, J., Cai, M., Yang, J., Li, Y., Meng, X., and Xu, B. (2021b). A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19). *European Radiology*, pages 1–9.

[67]   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.

[68]   Wang, Z., Xiao, Y., Li, Y., Zhang, J., Lu, F., Hou, M., and Liu, X. (2021c). Automatically discriminating and localizing COVID-19 from community-acquired pneumonia on chest X-rays. *Pattern Recognition*, 110:107613.

[69]   WHO (2021). Who coronavirus (covid-19) dashboard.

[70]   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.

[71]   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.

[72]   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*.

[73]   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*.

[74]   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.

[75]   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.

[76]   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)*.

[77]   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*.

[78]   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*.

[79]   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.