



Article

Explainable COVID-19 Detection on Chest X-rays Using an End-to-End Deep Convolutional Neural Network Architecture

Mohamed Chetoui ¹, **Moulay A. Akhloufi** ^{1,*}, **Bardia Yousefi** ^{2,3} and **El Mostafa Bouattane** ⁴¹ Perception, Robotics, and Intelligent Machines Research Group (PRIME), Department of Computer Science, Université de Moncton, Moncton, NB E1A 3E9, Canada; mohamed.chetoui@primeai.ca² Fischell Department of Bioengineering, University of Maryland, College Park, MD 20742, USA; byousefi@umd.edu³ Department of Electrical and Computer Engineering, Laval University, Québec, QC G1V 0A6, Canada⁴ Montfort Academic Hospital & Institut du Savoir Montfort, Ottawa, ON K1G 0A6, Canada; mostafabouattane@montfort.on.ca

* Correspondence: moulay.akhloufi@umanitoba.ca



Citation: Chetoui, M.; Akhloufi, M.A.; Yousefi, B.; Bouattane, E.M. Explainable COVID-19 Detection on Chest X-rays Using an End-to-End Deep Convolutional Neural Network Architecture. *Big Data Cogn. Comput.* **2021**, *5*, 73. <https://doi.org/10.3390/bdcc5040073>

Academic Editor: Min Chen

Received: 16 October 2021

Accepted: 24 November 2021

Published: 7 December 2021

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



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

1. Introduction

Since its appearance at the end of 2019, the coronavirus pandemic (COVID-19), caused by extreme coronavirus (SARS-CoV-2) acute respiratory syndrome, has spread worldwide, causing hundreds of millions of infected people and millions of deaths [1]. As the World Health Organization (WHO) deemed the outbreak to be a public health emergency of international significance, this has raised significant public health concerns in the international community, and on 11 March 2020, it was proclaimed a pandemic [2,3].

The reverse transcription polymerase chain reaction (RT-PCR) test serves as the gold standard to test patients for COVID-19 [4]. However, RT-PCR appears to be inadequate

for large-scale tests in many severely affected areas, particularly during the early outbreak of this disease. As reported in Fang et al. [4], RT-PCR also suffers from its low sensitivity, which can be close to 71%. This is due to many factors, such as sample preparation and quality control [5]. Medical imaging [6] was also proposed as a tool for examining the effects of COVID-19 in the lungs. With the help of CT scans [7] and CXR images, healthy people and COVID-19 infected patients can be detected (See Figure 1 for CXR image examples of COVID-19, other pneumonia and normal). Artificial Intelligence (AI) applied to medical imaging has successfully contributed to the fight against COVID-19 and other diseases [8–10]. AI allows for more stable, accurate and effective imaging solutions compared to the conventional imaging workflow that relies heavily on human labor.

Multiple research studies for COVID-19 detection using medical imaging were published recently. Deep learning methods were successfully used on CXR images leading to interesting results in term of various metrics such as Accuracy, Sensitivity, Specificity and Area Under Curve (AUC).

Apostolopoulos et al. [11] presented five pre-trained CNN models for detecting COVID-19 on CXR images. The deep learning model was trained and validated on 224 positive COVID-19, 700 pneumonia, and 504 normal CXR images. VGG19 [12] and MobileNetv2 [13] achieved the best results with an accuracy score of 0.93 and 0.92, respectively.

A hybrid approach was proposed by Sethy et al. [14], and the authors used CNN models for feature extraction and support vector machines (SVM) for classification. They achieved an accuracy of 0.95 using ResNet50 [15] and SVM on a test set of 50 COVID-19 images.

Wang et al. [16] proposed a CNN model named COVID-Net for detecting COVID-19 using CXR images. The authors used an architecture based on a lightweight residual design pattern named PEPX (projection-expansion-projection-extension). In addition, an explainability algorithm was developed to localize the signs of COVID-19. The results showed a sensitivity of 0.91, an accuracy of 0.93 and a precision of 0.96 for detected COVID-19 cases.

The work presented by Farooq et al. [17], for COVID-19 detection using deep CNN, uses a three-step technique to fine-tune a ResNet-50 [15] architecture and increase the model's performance. At each level, they performed progressive enlargement of input CXR (image = $128 \times 128 \times 3$, $224 \times 224 \times 3$ and $229 \times 229 \times 3$) and fine-tuning of the network. The authors were able to reach an accuracy of 0.96.

Hemdan et al. [18] also proposed a system based on seven CNN models called COVIDX-Net. The best performance was achieved by VGG-19 [12] and DenseNet-201 [19] with an accuracy of 0.90.

Wehbe et al. [20] presented an ensemble of CNN architectures called DeepCOVID-XR using DenseNet-121, ResNet50 [15], Inception-V3 [21], InceptionResNetV2 [22], Xception [23], and EfficientNet-B2 [24] for binary prediction (COVID-19 vs. No-COVID-19). The algorithm was trained and tested on 14,788 images including 4253 of COVID-19 positive cases. On the test set, DeepCOVID-XR obtained an accuracy of 0.83 with an AUC of 0.90 on 300 test images (134 of COVID-19). DeepCOVID-XR accuracy was 0.82 compared to the results obtained by individual radiologists with 0.76–0.81 and the consensus of all five radiologists with 0.81. DeepCOVID-XR achieved a sensitivity of 0.71 better than one radiologist with 0.60 and a specificity of 0.92 better than two radiologists with 0.75 and 0.84.

Chetoui and Akhloufi [25] used a deep CNN based on EfficientNet-B7 for the binary classification of CXR images to detect COVID-19 vs. Healthy. The authors used 2385 CXR images of COVID-19 obtained from multiple datasets. The proposed algorithm achieved an ACC of 0.95 and an AUC of 0.95.

Minaee et al. [26], used four CNN models for binary prediction (COVID-19 vs. No-COVID-19). They used ResNet18 [15], ResNet50, SqueezeNet [27] and DenseNet-121 to identify COVID-19 in the analyzed CXR images. The authors prepared a dataset composed of 5000 CXR images collected from publicly available datasets including 250 COVID-19 images. They selected 100 COVID-19 images to include in the test set and 84 COVID-19

images for the training set. Data augmentation was applied to the training set in order to increase the number of COVID-19 samples to 420. SqueezeNet networks achieved a sensitivity of 0.98 and a specificity of 0.90, and an average AUC of 0.986 was obtained for ResNet18, ResNet50, SqueezeNet and DenseNet-121.

In all these studies, we see very interesting scores. This is due to many reasons. First, the low number of images in the datasets. Second, the optimization of hyper-parameters applied to CNN. Third, the performance of the CNN architecture. However, there is a lack of research dealing with the study of the degree of generalization of the deep learning algorithms on heterogeneous datasets. The degree of generalization is important in the medical field. For example, the CXR images have different image quality, resolution, non-uniform lighting and contrast, orientation, etc. The main contribution of this study is to develop a deep neural network model to detect COVID-19, pneumonia and normal cases using several heterogeneous datasets with different image quality, projection, resolution, lung capture, etc. Since COVID-19 is a pneumonia, another challenge is adapting the network to distinguish between other pneumonia and COVID-19 classes (there are very strong similarities between the two categories in terms of signs). Moreover, we did not apply pre-processing and data augmentation techniques on the CXR images; we kept the original quality.

The model used in this study is a fine-tuned EfficientNet-B5, which was selected based on its performance in ImageNet [28] classification. The EfficientNet family are DCNN and their dimensions are balanced (width, resolution and depth). The model was tested on two scenarios using CXR images: (i) detecting COVID-19 vs. pneumonia vs. normal cases (DeepCCXR-Multi) and (ii) a binary model for COVID-19 vs. normal (DeepCCXR-Bin). During training and testing, several datasets were employed: RSNA [29], CIDE from GitHub [30], BIMCV COVID19+ [31], CHEST X-RAY IMAGES PNEUMONIA (CXRIIP) from Kaggle [32], MONTGOMERY [33], SHENZHEN [33] and NIH [34] which are available publicly, and a proprietary dataset from Montfort hospital (Ottawa, ON, Canada) [35]. Testing with numerous datasets allows the proposed algorithm's generalization performance to be benchmarked and made more robust to real-life conditions [36]. We compare our findings to current techniques and show that the suggested method achieves a higher score of COVID-19 detection across several datasets. Several measures were used to validate the performance of our proposed model in terms of Accuracy, AUC, Sensitivity and Specificity. An explainability model was developed and adapted to visualize the signs of COVID-19 and pneumonia detected by the model.

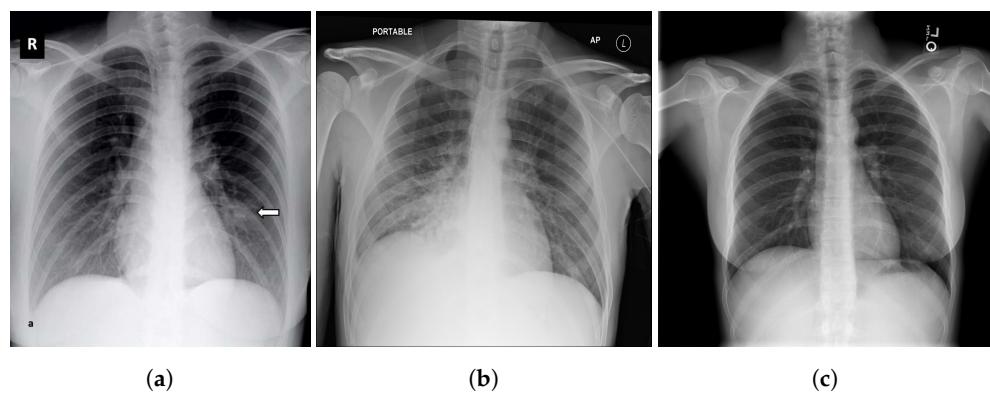


Figure 1. Examples of CXR images: (a) COVID-19, the arrow shows the location of the infection; (b) pneumonia; (c) normal/healthy.

2. Proposed Approach

In this work, we propose a robust system for detecting COVID-19, pneumonia and normal cases, and in the binary problem (COVID-19 and normal) on several datasets using CXR images, the system is based on the fine-tuning of recent CNN called EfficientNet-B5. An explainability algorithm is also developed to visually show the infected regions

identified by the network. The proposed architecture called DeepCCXR (Deep Covid-19 CXR detection) is illustrated in Figure 2.

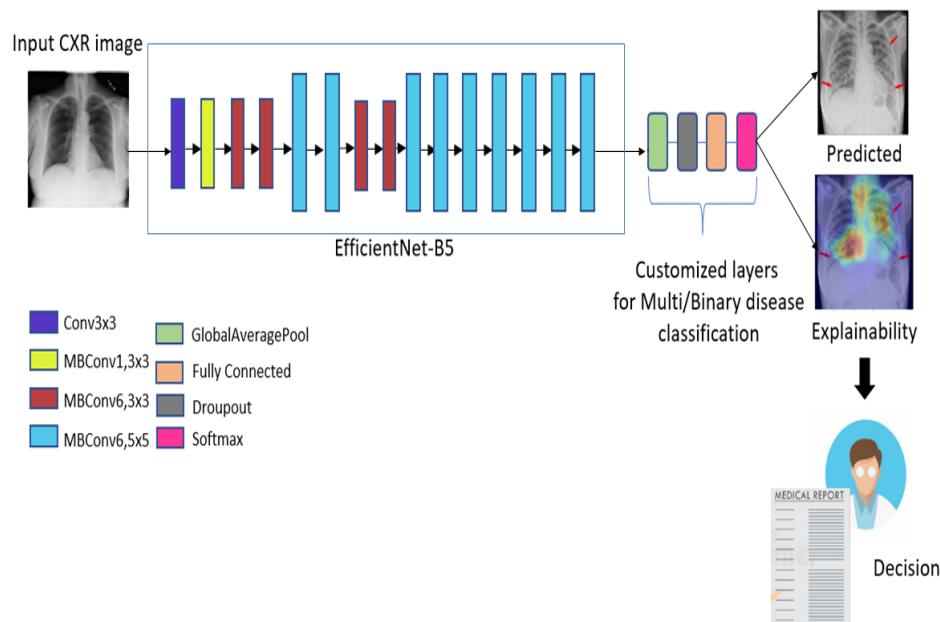


Figure 2. Proposed DeepCCXR architecture for COVID-19 detection.

Deep Learning Model

Tan and Le [24] have investigated the relationship between CNN model width and depth and devised a method for creating CNN models with fewer parameters and higher classification accuracy. The authors called the networks EfficientNet and proposed eight models (EfficientNet-B0 to EfficientNet-B7). Authors tested the performance of the networks on ImageNet dataset [28] and showed that the EfficientNet networks surpass all CNN models in term of accuracy.

The EfficientNet network is based on a novel scaling strategy for CNN models. It employs a straightforward compound coefficient that is quite successful. Unlike existing approaches that scale network parameters such as width, depth and resolution, EfficientNet evenly scales each dimension with a given set of scaling factors. Scaling individual dimensions increases model performance in practice, but balancing all network dimensions in relation to available resources significantly enhances overall performance.

The primary building component of the EfficientNet model family is mobile inverted bottleneck convolution (MBConv). The MobileNet models [37] provided the inspiration for MBConv. The use of depthwise separable convolutions, which combine depthwise and pointwise convolution layers after each other, is one of the main ideas. Then, from MobileNet-V2 (a second upgraded version of MobileNet), two further ideas are borrowed: (1) inverted residual connections and (2) linear bottlenecks.

Figure 3 presents an illustration of inverted residual blocks. The skip connections exist between layers with wide number of channels (64 in Figure 3a).

The number of channels in the residual block is lowered or compressed to 16, reducing the number of parameters required by the next layer's 3×3 convolutions. The sizes of the connected channels are inverted in the inverted residual block illustrated in Figure 3b, so the skip connections now take place between narrower layers with fewer channels. This explains the name of residual blocks. Because we employ depthwise convolutions, even if the number of channels in the layer inside the block increases to 64, the number of parameters is actually lower than in the original ResNet residual block.

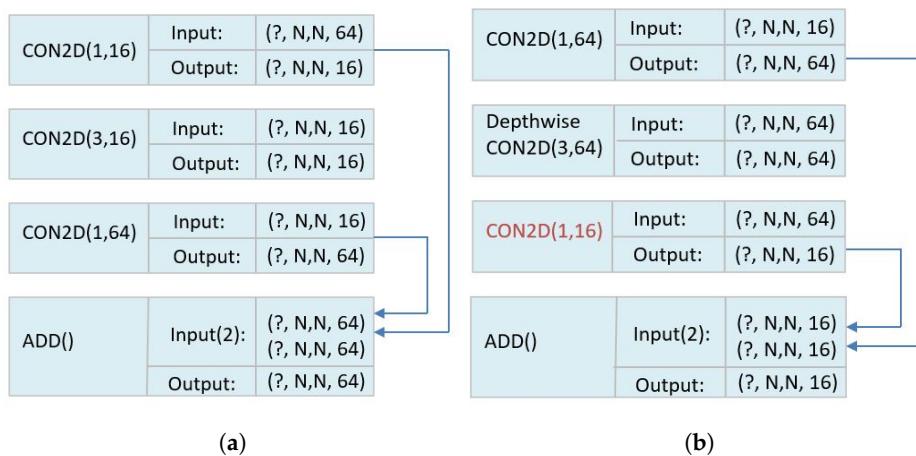


Figure 3. Inverted residual block example; (a) regular residual block where channel size changes from 64→16→64; (b) inverted residual block where channels change from 16→64→16.

The second concept in MobileNetV2 is linear bottlenecks, which implies that for the layer shown in red in Figure 3b, we employ a linear activation function. Because the number of channels is constrained at various network locations, this layer is referred to as a bottleneck layer. According to the authors of MobileNetV2 [37], the ReLU activation function, which is often employed in CNN architectures, does not operate well with inverted residual blocks since it discards values less than zero. The layer with decreased channels (bottleneck channel) performed better when using a linear activation function.

In addition, instead of using the ReLU activation function, our proposed network utilizes a new activation function called Swish [38]. The Swish activation function is comparable in shape to the ReLU and LeakyReLU functions and so shares some of their performance advantages. It has a smoother activation function than the other two. The Swish activation formulation is presented in the following equation:

$$f_{\text{Swish}}(x) = \frac{x}{1 + e^{-\beta x}} \quad (1)$$

where $\beta \geq 0$ is a parameter that can be learned during training of the CNN model. Note, if $\beta = 0$, f_{Swish} becomes the linear activation function and as $\beta \rightarrow \infty$, f_{Swish} looks more and more like the ReLU function except it is smoother [38].

The success of the previously described model scaling idea is highly dependent on the baseline network. To do this, the automatic machine learning (AutoML) MNAS framework is used to generate a new baseline network, which automatically searches for a CNN model that optimizes both precision and efficiency (in FLOPS). EfficientNet-B0 is the name of the baseline network, and its main architecture is shown in Table 1.

The first thing to notice is that this baseline model is made up of MBConv1, MBConv3 and MBConv6 blocks that are repeated multiple times. MBConv blocks come in a variety of shapes and sizes. The second point to note is that the number of channels within each block is increasing or decreasing (through a larger number of filters). The inverted residual connections between the model's narrow layers constitute the model's third observation.

The squeeze-and-excitation (SE) technique was also added in the MBConv blocks by the authors in [24], which helps to increase performance even further.

Remember that the number of filter parameters determines the number of channels that a convolutional layer produces. In most cases, subsequent operations will give these channels equal weight. The SE block is a strategy that weights each channel differently rather than treating them all equally.

The SE block outputs a shape ($1 \times 1 \times \text{channels}$) that specifies the weights for each channel, and the best part is that these weights values, like other parameters, are learned during training. Finally, in Figure 4, we show an example of MBConv blocks that accepts a feature map of size ($128 \times 128 \times 40$) as input and incorporates all of the preceding

principles, including (1) depthwise separable convolutions, (2) inverted residual blocks, (3) linear bottlenecks, (4) Swish activation functions and (5) the squeeze-and-excitation block. MBConv blocks come in a variety of shapes and sizes.

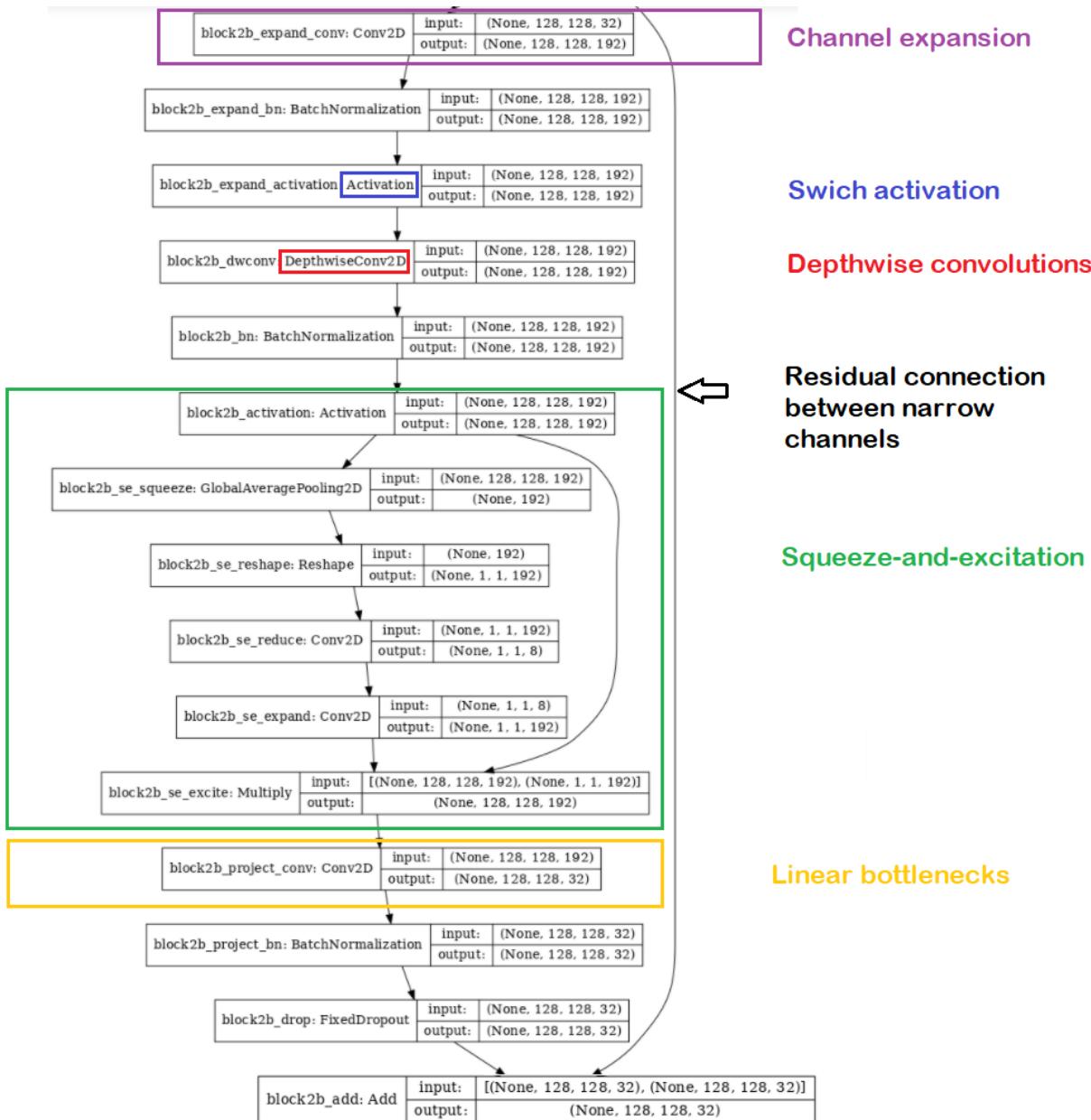


Figure 4. Illustration of MBConv6 with Squeeze-and-Excitation block and inverted residual connection.

In general, EfficientNet models outperform existing CNN such as AlexNet, DenseNet and GoogleNet in terms of accuracy and efficiency and have been widely used in the medical field [39–41].

Finally, to increase accuracy and avoid overfitting, we modified the final convolution layer by adding a Global Average Pooling (GAP) to the network. Following GAP, we added a dense layer with size 1024 and a 50% Dropout. After dense layers, we added a Softmax layer with three neurons to give the probability prediction scores for detecting one of the three classes. We kept the same architecture for the binary classification, and

only the Softmax layer was changed to give the probability prediction scores for detecting COVID-19 vs. normal.

Table 1. The EfficientNet-B0 general architecture

Stage	Operator	Resolution H × W	#Channels	Layers
1	Conv3 × 3	244 × 224	32	1
2	MBConv1, K3 × 3	112 × 112	16	1
3	MBConv6, K3 × 3	112 × 112	24	2
4	MBConv6, K5 × 5	56 × 56	40	2
5	MBConv6, K3 × 3	28 × 28	80	3
6	MBConv6, K5 × 5	14 × 14	112	3
7	MBConv6, K5 × 5	14 × 14	192	4
8	MBConv6, K3 × 3	7 × 7	320	1
9	Conv1 × 1 and pooling and FC	7 × 7	1280	1

3. Datasets

In this work, we used nine datasets for training and testing, which are presented in the following subsections.

3.1. COVID-19 Image Data Collection (C IDC)

Cohen et al. [30] provided an open dataset of CXR and CT scan images of patients with COVID-19 and other viral/bacterial pneumonia who were positive or suspected (MERS, SARS and ARDS). The information was primarily scraped from medical websites that collected publicly available COVID-19 CXR images from hospitals and clinicians. There are 654 COVID-19 CXR images in the dataset, which were gathered from various sources. The purpose of this dataset is to create artificial intelligence-based tools to anticipate and comprehend illness. COVID-19 CXR images from this dataset are shown in Figure 5.

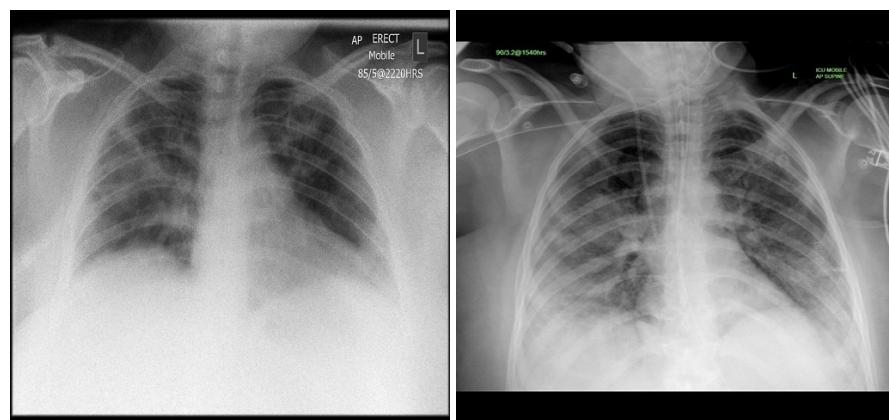


Figure 5. Examples of COVID-19 image data collection (C IDC) images [42].

3.2. COVID-19 Radiography

The COVID-19 RADIOGRAPHY [42] database contains 219 CXR COVID-19 positive images obtained from Kaggle [42] and created by a team of researchers from Qatar University (Doha, Qatar), the University of Dhaka (Bangladesh) and their collaborators from Pakistan and Malaysia with the help of various medical doctors who created a database of CXR images for COVID-19 positive cases. Figure 6 shows examples of COVID-19 RADIOGRAPHY images from this dataset.

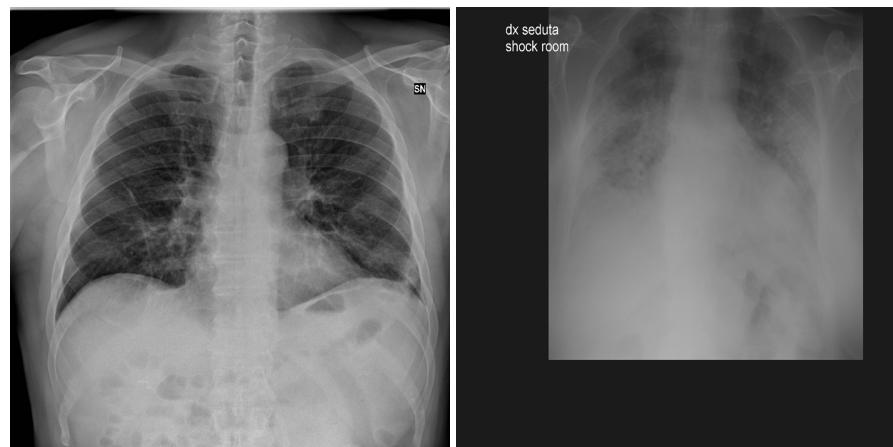


Figure 6. Examples of COVID-19 RADIOGRAPHY images [42].

3.3. BIMCV COVID19+

COVID19+ is a large CXR dataset of COVID-19 positive patients and computed tomography (CT) images, as well as their radiography findings, pathologies, polymerase chain reaction (PCR) test results, immunoglobulin G (IgG) and immunoglobulin M (IgM) diagnostic antibody tests and radiography reports, from the Medical Imaging Databank in Valencia Region Medical Image Bank (BIMCV). A team of expert radiologists annotates the images and stores them in high resolution. Moreover, extensive information is provided, including the patient's demographic information, type of projection (PA-AP) and acquisition parameters for the imaging study, among others. The database includes 1380 CXR, 885 DX (Digital X-ray) and 163 computed tomography images. Figure 7 shows examples of BIMCV COVID-19+ CXR images.

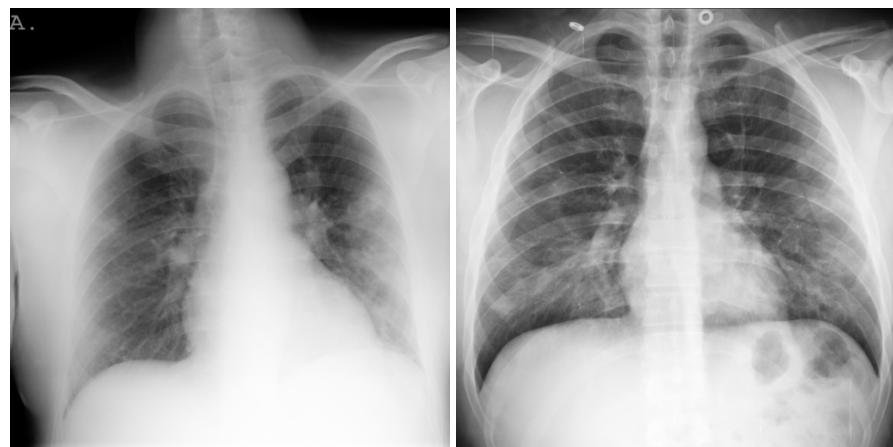


Figure 7. Examples of BIMCV COVID19+ images [31].

3.4. RSNA

The RSNA [29] dataset is a dataset of CXR images with patients metadata. This dataset was provided for a challenge in Kaggle by the US National Institutes of Health Clinical Center and is available on Kaggle competition website [43]. It contains 26,684 CXR images for unique patients, and each image is labeled with one of three different classes from the associated radiology reports: 'Normal', 'No Lung Opacity/Not Normal', 'Lung Opacity'. Figure 8 shows image examples from the RSNA dataset.

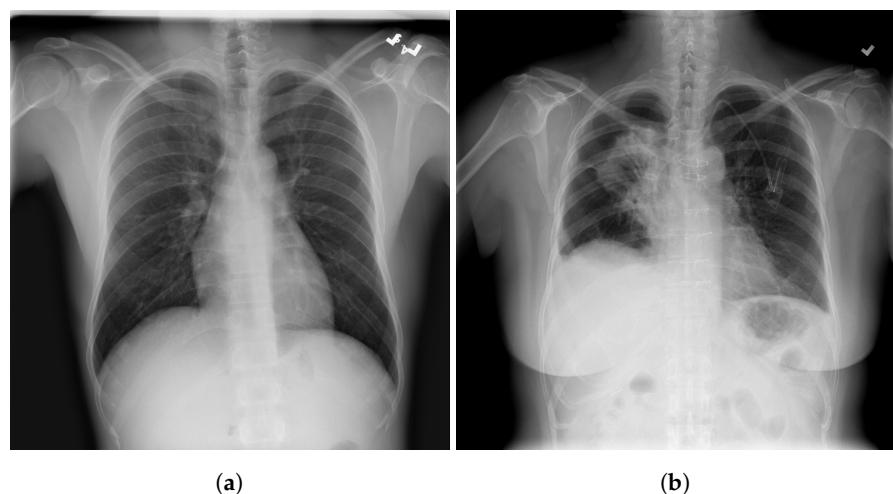


Figure 8. Examples of RSNA images: (a) normal; (b) lung opacity [29].

3.5. Chest X-ray Images Pneumonia (CXRIP)

Chest X-ray images (anterior–posterior) [32] were selected from retrospective cohorts of pediatric patients of one to five years old from Guangzhou Women and Children's Medical Center (Guangzhou, China). All CXR imaging was performed as part of patients' routine clinical care. All CXR were initially screened for quality control by removing all low-quality images. The diagnoses for the images were then graded by two expert physicians before being cleared for training the AI system. The dataset is partitioned into three folders for training, testing and validation and contains sub-folders for each image category (pneumonia or normal). The dataset contains 5863 CXR images and two classes, pneumonia and normal. Figure 9 shows examples of CXR images from patients with pneumonia and normal cases.

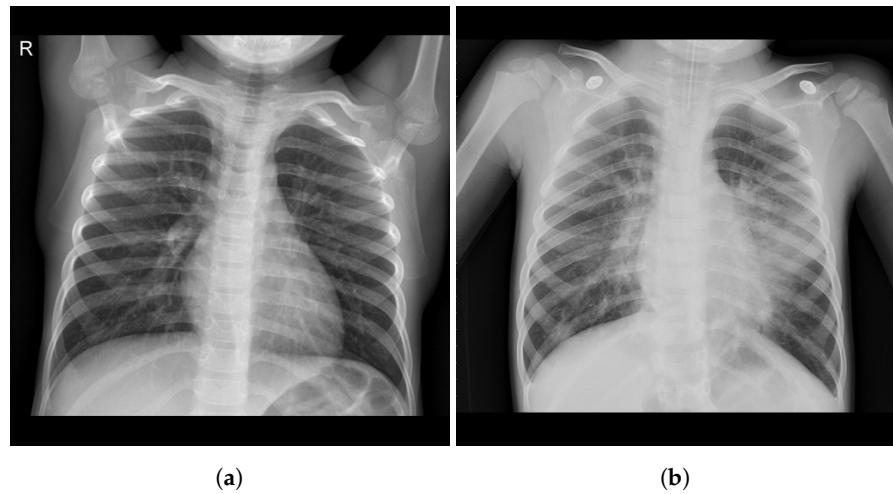


Figure 9. Examples images from CXRIP dataset. (a) Normal and (b) bacterial pneumonia [32].

3.6. Montgomery County X-ray

The MONTGOMERY County CXR dataset [33] has been acquired from the tuberculosis control program of the Department of Health and Human Services of MONTGOMERY County (Rockville, MD, USA). The dataset contains 138 posterior–anterior CXR, of which 80 are normal and 58 are abnormal with manifestations of tuberculosis. All CXR images are identified and available in DICOM format. The dataset includes radiology readings available as a text file. Figure 10 shows example images from the MONTGOMERY dataset.

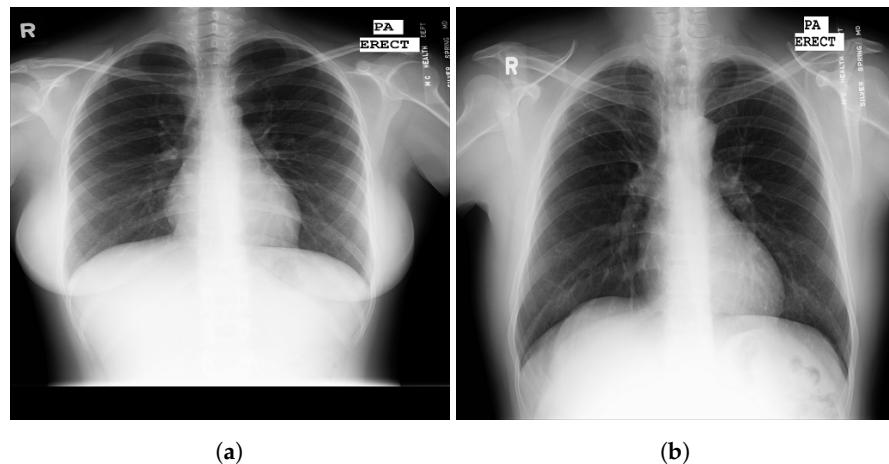


Figure 10. Examples of MONTGOMERY images. (a) Normal patient; (b) patient with tuberculosis [33].

3.7. Shenzhen Hospital X-ray

The SHENZHEN [33] Hospital X-ray dataset has been collected by SHENZHEN Hospital (Shenzhen, Guangdong, China). The CXR were acquired as part of the routine care at SHENZHEN Hospital. The dataset contains 326 normal CXR and 336 abnormal CXR showing various manifestations of tuberculosis. Figure 11 shows some CXR images from SHENZHEN dataset.

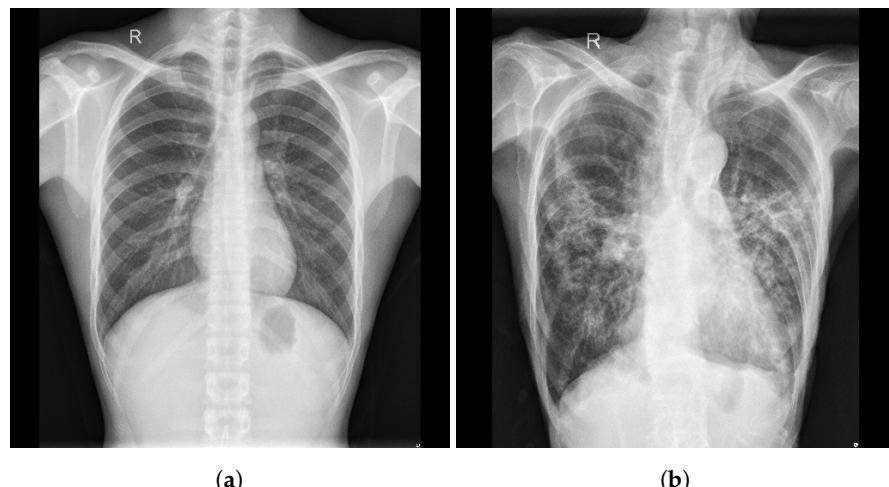


Figure 11. Examples of SHENZHEN images. (a) Normal patient; (b) patient with tuberculosis [33].

3.8. National Institute of Health (NIH)

The NIH [34] Chest X-ray dataset is comprised of 112,120 CXR images with disease labels from 30,805 unique patients. This dataset was obtained from the National Institute of Health (Bethesda, MD, USA). There are 15 classes in the dataset (14 diseases and one for 'No findings'). Infiltration, edema, atelectasis, pneumothorax, consolidation, emphysema, effusion, fibrosis, pneumonia, cardiomegaly, pleural thickening, mass, nodule and hernia are some of the disease groups that images can be classed as. Expert physicians assigned grades to the CXR images. Figure 12 shows example images from the NIH dataset.

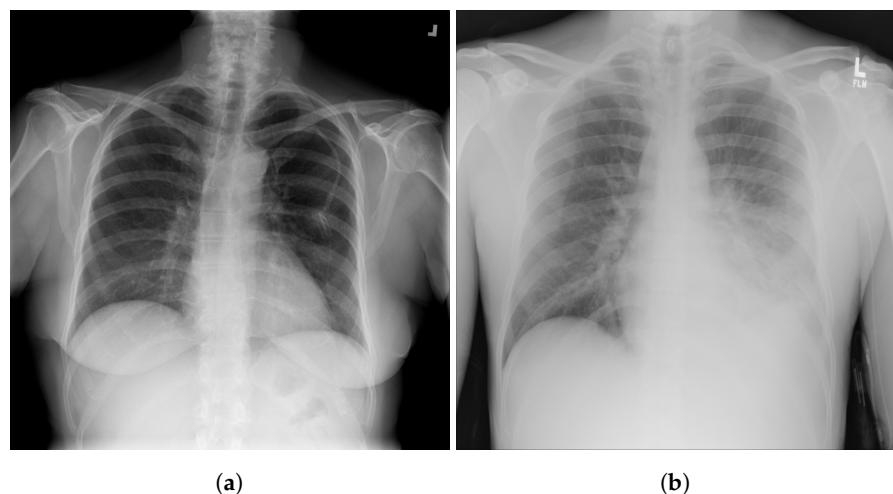


Figure 12. Examples of NIH images. (a) No findings (normal); (b) patient with pneumonia [34].

3.9. Montfort Dataset

In addition to the above datasets, we collected more images in collaboration with health professionals from Montfort hospital (Ottawa, ON, Canada) and built the Montfort dataset [35]. This dataset contains 236 CXR images, with 150 COVID-19, 29 pneumonia (other than COVID-19) and 57 normal patients (no findings). Radiology reports and RT-PCR testing are used to label CXR images.

4. Experimental Results

4.1. Data Distribution for Multi-Class and Binary Models

To train the DeepCCXR-Multi (COVID-19 vs. other pneumonia vs. normal), we used NIH dataset for the pneumonia and normal sets, with 14,226 CXR images (8551 for normal and 5675 for other pneumonia). To build the COVID-19 set, we combined CXR COVID-19 positive images from multiple datasets (Montfort Dataset, BIMCV COVID19+, CIDC and COVID-19 RADIOGRAPHY dataset), leading to a total of 3288 CXR COVID-19 positive images. We used 2060 images for training and 1228 for testing. For the test set of pneumonia and normal cases, we kept samples from each dataset CXRIP, RSNA, NIH, SHENZHEN and MONTGOMERY) in order to validate the generalization performance of our model on all these datasets (since the quality and technology used are different). Our test set is comprised of 1128 normal, 1228 COVID-19 and 1072 pneumonia images.

For DeepCCXR-Bin, we used 10,611 training images. 2060 COVID-19 images (Same as DeepCCXR-Multi) and 8551 normal images from the NIH dataset. The test set contains 1128 normal cases and 1228 COVID-19 cases. We divided the bases manually, because the random division may give bad results according to [20].

Figure 13 gives an overview for the datasets distribution.

For all datasets used in this study, we kept image quality, without applying any pre-processing. As we can see in Figure 14, an example of images in CIDR datasets shows the unbalanced color (blue and gray). Figure 15 shows the capture of lungs includes the abdominal part. Furthermore, the images are different in the orientation as shown in Figure 16. Moreover, we kept the CXR images with poor quality as shown in Figure 17.

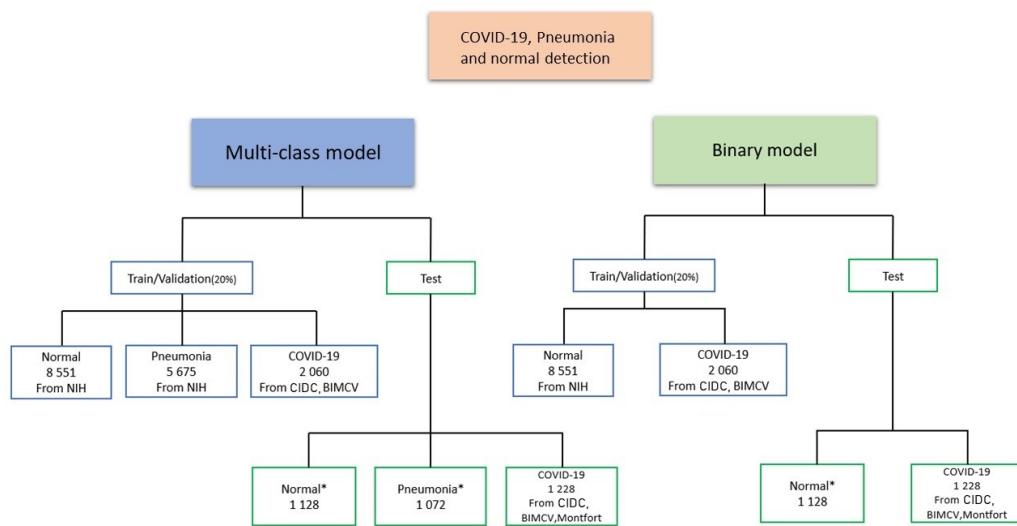


Figure 13. Datasets distribution for DeepCCXR-Multi and DeepCCXR-Bin (*normal and pneumonia CXR images in the test set are combination of samples from NIH, RSNA, CXRIP, MONTGOMERY and SHENZHEN).

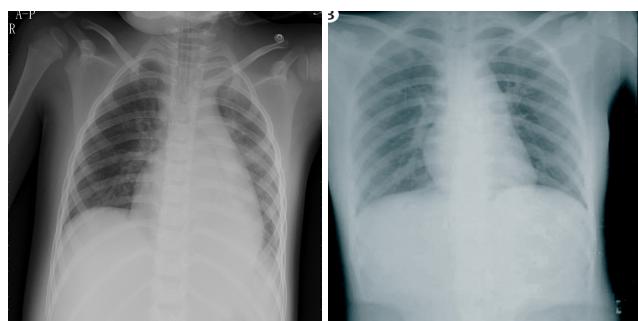


Figure 14. Example of unbalanced colors.

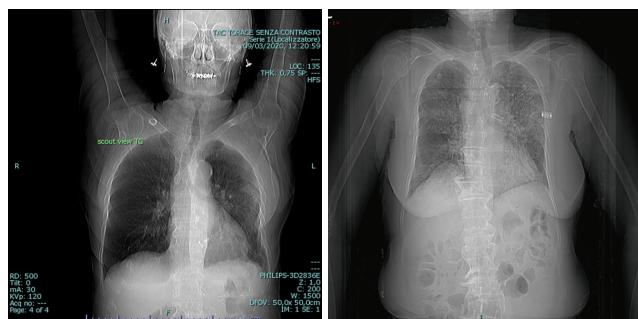


Figure 15. Example of bad projection of lungs.

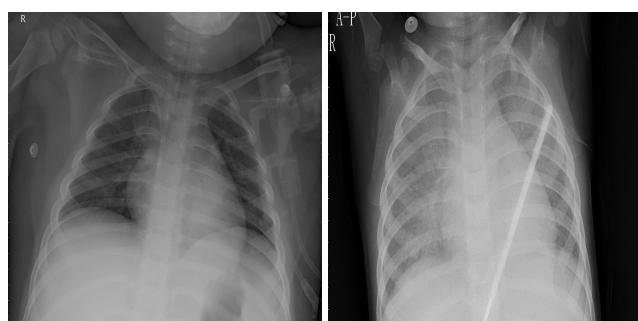


Figure 16. Example of different orientation.

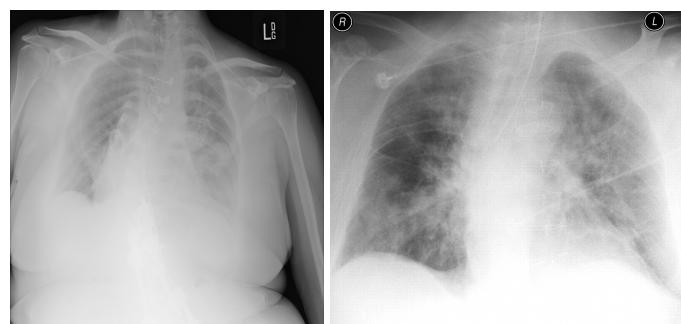


Figure 17. Example of low quality in CIDR and RSNA datasets.

4.2. Training Parameters

The models were developed using Keras (Tensorflow) Library [44], and training was conducted on Microsoft Azure servers [45]. We must use an optimization function to optimize the learning parameters. Because different optimizers have distinct effects on parameter training, we studied the effects of SGD [46] and Adam [47] on model performance. Multiple comparative experiments were conducted under the same settings. SGD is determined to be much superior to Adam in terms of convergence and training time reduction. When we utilized Adam as an optimizer, the gradient of each sample was modified every time, which increased the noise. Each iteration is not moving in the same direction as the overall optimization, and it may converge to a local minimum, lowering accuracy. We used 200 epochs for training and a batch size of 16. All CXR images were resized to 512×512 . We used the pre-trained model with ImageNet weights. To prevent the class imbalance issue, we applied the class-weight approach [48] during model training on the NIH, CIDC and BIMCV datasets. The technique directly accounts for the asymmetry of cost errors. Hyperparameters optimization was conducted on the validation set, and the best results were kept for testing. Table 2 summarizes the hyperparameters used in our deep learning models.

Table 2. Hyperparameters configuration.

Configuration	Value
Optimizer	SGD
Epoch	200 complete training
Batch size	16
Learning rate	0.003
Batch normalization	True
Dropout	50% after dense layers
Model checkpoint	Monitor = 'val_acc', save_best_only = True, mode = 'auto'

4.3. Metrics

For this work, we used the following metrics: accuracy (ACC), sensitivity (SE), specificity (SP) and area under curve (AUC) [9]. The SE and SP show the performance of the proposed approach. AUC computed using ROC curve is a performance measure widely used for medical classification problems to highlight the compromise between good and bad classifications by the model. In medical applications and published papers about COVID-19, the main used metrics are sensitivity and specificity; therefore, we have used these metrics. This way, we can compare our work with past published work. These metrics are defined in the following:

$$TPR/SE = \frac{TP}{TP + FN} \quad (2)$$

$$SP = \frac{TN}{TN + FP} \quad (3)$$

$$ACC = \frac{TP + TN}{TP + FN + TN + FP} \quad (4)$$

$$FPR = 1 - SP \quad (5)$$

where TP means true positives for samples classified as positive and FN means false negatives for the samples incorrectly classified as negative, FP means false positives for the samples incorrectly classified as positive and TN means true negatives for the samples correctly classified as negative.

4.4. Results

The proposed fine-tuned EfficientNet-B5 gives interesting results for the DeepCCXR-Multi (COVID-19 vs. other pneumonia vs. normal). The model achieved an average SP and SN of 0.97 and 0.94, respectively. The obtained AUC was 0.973 on 3,428 CXR images (normal, COVID-19 and pneumonia) used in the test set. The details of the obtained SN and SP for each class are given in Table 3. The DeepCCXR-Bin also achieves high scores with an AUC of 0.985, an average SP of 0.94 and an SN of 0.98. Figure 18a,b shows the ROC curves for both models. We can see the high performance achieved by the proposed models. The confusion matrices for DeepCCXR-Multi and DeepCCXR-Bin are given in Figure 19a,b. As we can see, only 11 from 1228 images were misclassified in the DeepCCXR-Multi. For the DeepCCXR-Bin, only 19 images from 1228 were misclassified. This shows the high performance obtained by our models.

Table 3. Sensitivity and specificity for each infection type for DeepCCXR-Multi (best results highlighted in bold).

Metrics	COVID-19	Normal	Pneumonia
Sensitivity	0.99	0.88	0.92
Specificity	0.99	0.96	0.94

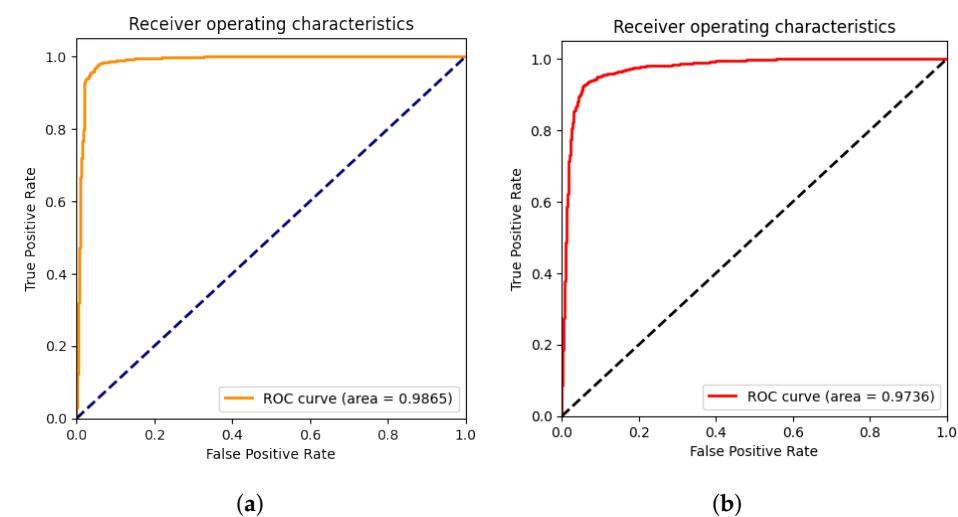


Figure 18. ROC curve of both models: (a) DeepCCXR-Bin (COVID-19 vs. normal); (b) DeepCCXR-Multi (COVID-19 vs. other pneumonia vs. normal).

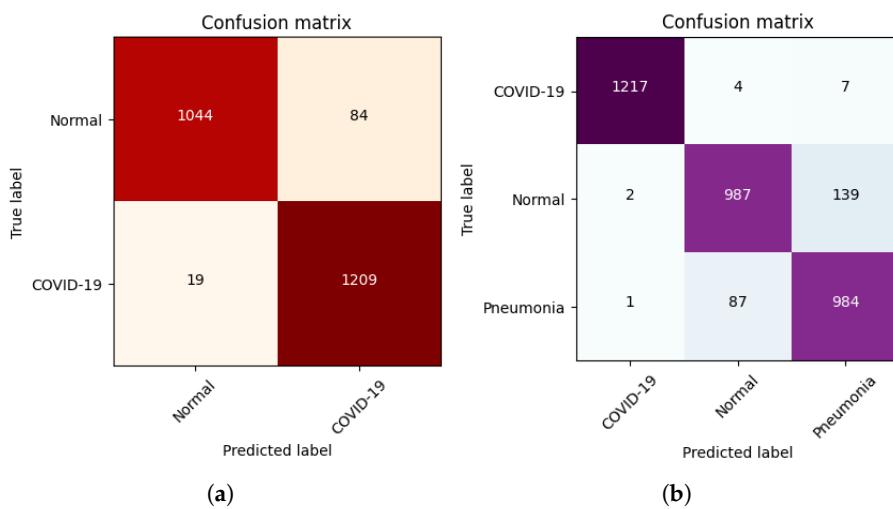


Figure 19. Confusion matrix of both models: (a) DeepCCXR-Bin (COVID-19 vs. normal); (b) DeepCCXR-Bin (COVID-19 vs. other pneumonia vs. normal).

4.5. Explainability

To understand how the model learned to detect the signs of pneumonia pathology including COVID-19 signs, we developed an explainability algorithm based on the use of Gradient-weighted Class Activation Mapping (Grad-CAM) [49]. This algorithm provides a visual output of the most interesting areas found by the proposed CNN models. Grad-CAM uses the gradients of any target, flowing into the final convolutional layer to generate a coarse localization map that highlights important regions in the predictive image. The proposed technique blends Grad-CAM with fine-grained visualizations to construct a high-resolution class-discriminative visualization [50–52].

Figure 20 shows samples of TP and TN using Grad-CAM to localize the signs of COVID-19, pneumonia and normal regions on CXR images. As we can see in Figure 20a,b in the example of TP cases of COVID-19 detected in CIDR dataset, the heatmap localized the signs in the lungs. In Figure 20c,d as well, we can see a sample of true positive cases in COVID-19 RADIOGRAPHY dataset, and the green color in the lungs indicates there is something abnormal detected by the model, which classifies them as COVID-19. Our model identified the dense homogeneous opacity regions as the most significant signs for COVID-19 which correlates well with radiology findings in COVID-19 medical research studies [53].

Other samples of TP cases of pneumonia are presented in Figure 20e,f for RSNA dataset, as we can see in the CXR images that the heatmap localizes the opacity in the lungs, and despite the poor quality and the bad projection on the lungs, the model classified the images correctly. Similarly in Figure 20g,h, we can see the positive cases of pneumonia in NIH dataset, the heatmap focus on the opacity area on the lungs.

The TN cases for the MONTGOMERY datasets are presented in Figure 20k,l. Generally, the heatmap focuses on something outside of the lungs or near the heart to distinguish between the normal and the other cases.

This explains the efficiency of the proposed deep learning approach in detecting COVID-19 and pneumonia signs and its high performance in COVID-19 classification.

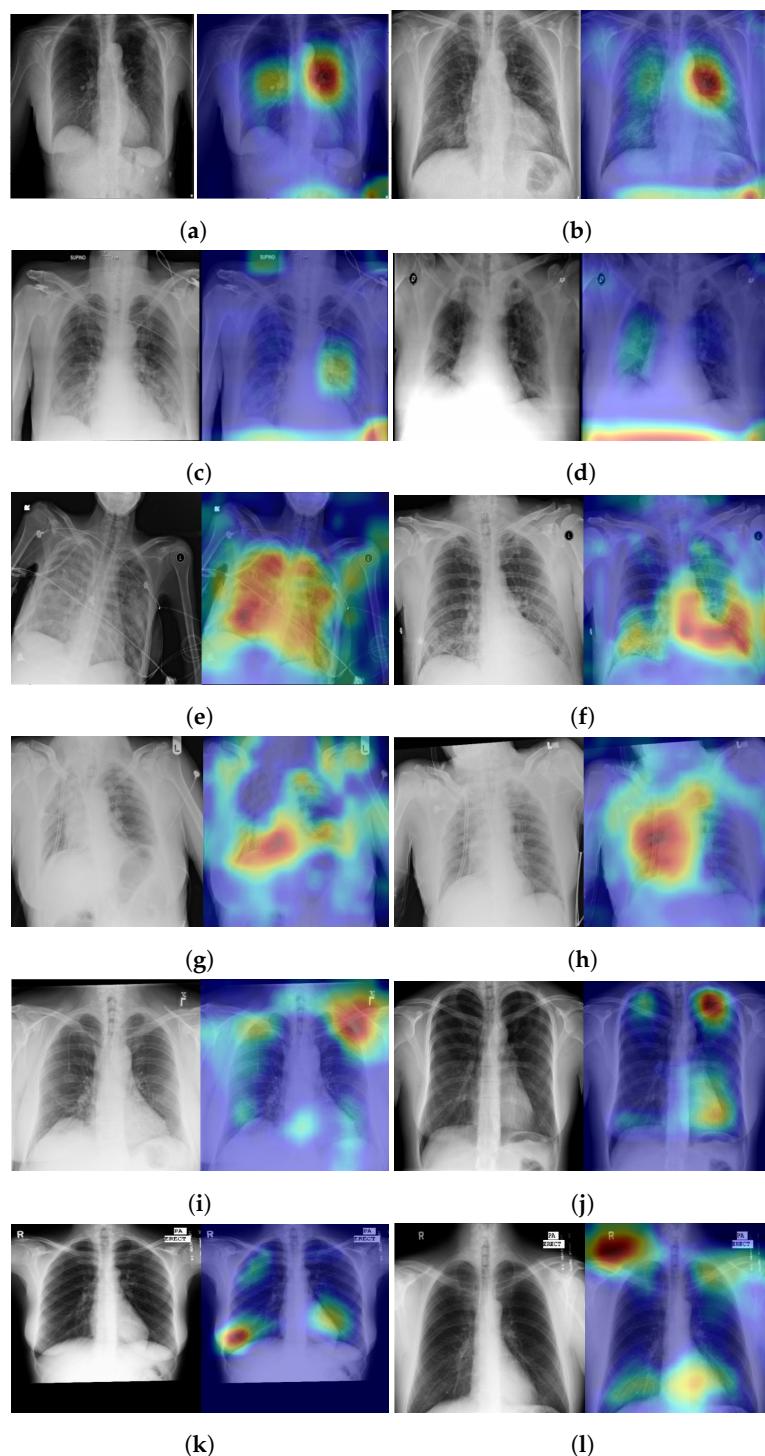


Figure 20. Explainability of TP and TN cases. The heatmap highlights important areas detected by our deep learning model in the CXR images: (a,b) CIDE-C; (c,d) COVID-19 RADIOGRAPHY; (e,f) RSNA; (g,h) NIH; (i,j) CIDE-C; (k,l) MONTGOMERY.

4.6. Performance Comparison

In Table 4, the performance of COVID-19 detection is compared to that of contemporary approaches. We can see that DeepCCXR outperforms most of the recently published work for detecting COVID-19 on CXR images. In addition, the number of COVID-19 images used in this study is higher compared to other studies, which confirms the degree of generalization for detecting COVID-19 with the proposed model.

Our previous model [54], which was generated using a small number of COVID-19 positive CXR images, has been improved by the DeepCCXR model (192 images). Based on ResNet50, our prior model had an AUC of 0.97, an SP of 0.96 and an SN of 0.95.

DeepCCXR obtained the same scores as Minaee et al. [26] in terms of AUC and sensitivity; still, our proposed models perform best for the specificity. However, the work of Minaee et al. [26] used only 203 COVID-19 images, which does not help in evaluating the generalization of the algorithm. Moreover, this number represents only 5% of a small dataset (203 vs. 4797), which can make the model sensitive to overfitting and bias the results.

The only work using a large number of COVID-19 images was published recently by Wehbe et al. [20]. Our results outperform the results in this last work [20]. The authors tested their models on a number of COVID-19 images close to ours (1192 in Wehbe et al. [20] vs. 1228 in this work). According to their paper, their model surpassed radiologists, which leads us to think that our model will have an even better performance if compared to a manual interpretation by radiologists. This shows that DeepCCXR is robust and able to detect COVID-19 with a high sensitivity (0.99). In addition, our developed explainability (EXP) model shows a very precise localization of the COVID-19 signs and can be used as a CAD tool to further help physicians in their diagnosis.

Table 4. Performance comparison with state-of-the-art methods using CXR images for COVID-19 detection.

Ref.	Dataset	#COVID-19 Images	ACC	AUC	SN	SP	EXP
Apostolopoulos et al. [11]	CIDC	224	0.93	-	0.98	0.96	No
Sethy et al. [14]	CIDC	25	0.95	-	0.97	0.93	No
Chetoui et al. [54]	CIDC	192	-	0.97	0.95	0.96	Yes
Chetoui and Akhloufi [25]	Multiple datasets	2385	0.95	0.95	0.97	0.90	Yes
Wang et al. [16]	ActualMed, CIDC	226	0.93	-	0.91	-	Yes
Hemdan et al. [18]	CIDC	25	0.90	-	1.00	0.83	No
Wehbe et al. [20]	Multiple institutions	4253	0.83	0.90	0.71	0.92	Yes
Minaee et al. [26]	CIDC	203	-	0.98	0.98	0.90	Yes
DeepCCXR-Multi	Multiple datasets	3288	0.93	0.97	0.97	0.94	Yes
DeepCCXR-Bin	Multiple datasets	3288	0.96	0.98	0.94	0.98	Yes

5. Individual Tests

To confirm the degree of generalization, we conducted individual tests with each of the nine datasets.

5.1. DeepCCXR-Bin for Individual Datasets

We validated DeepCCXR-Bin model on the nine datasets separately.

- CIDC dataset: The CIDC dataset does not contain the 'normal' class. We added 1128 CXR normal images from RSNA dataset (not in the training set), and we kept all the CXR images of COVID-19 (654 CXR images).
- CXRIP dataset: The original CXRIP dataset contains pneumonia and normal classes. We added 1228 COVID-19 CXR images in place of the pneumonia class, and we kept the 1435 CXR images for normal class.
- MONTGOMERY, SHENZHEN, NIH and RSNA datasets: As with the CXRIP dataset, the original MONTGOMERY, SHENZHEN, NIH and RSNA datasets contain two categories, pneumonia and normal. We kept their normal classes and we replaced the pneumonia class with COVID-19 CXR images.
- COVID-19 RADIOGRAPHY dataset: The COVID-19 RADIOGRAPHY dataset contains only 218 CXR images of COVID-19. We added 1,128 normal CXR images from the RSNA dataset.

- The Montfort dataset contain three categories (COVID-19, normal and pneumonia). We removed the pneumonia class and we kept the normal and COVID-19 classes.

The results of each dataset are presented in Table 5. Very interesting AUC scores of 0.999, 0.997 and 0.999 were obtained in the CXRIP, RSNA and MONTGOMERY datasets, respectively. DeepCCXR-Bin gives a high AUC score of 0.986 in the NIH and 0.978 in SHENZHEN datasets. Scores of 0.965 and 0.961 were obtained in BIMCV COVID19+ and COVID-19 RADIOGRAPHY. SP and SN are equal to 0.99 in RSNA, MONTGOMERY and CXRIP. This shows the degree of generalization of our binary model (DeepCCXR-Bin) on the heterogeneous data. The confusion matrices for the nine datasets are shown in Figure 21. Only 74 images from 654 CXR of COVID-19 were misclassified in the CIDC dataset (see Figure 21b), and only six of 218 of COVID-19 CXR were misclassified in the COVID-19 RADIOGRAPHY dataset (see Figure 21f). Figure 21c represents the confusion matrix for CXRIP, this dataset contains 1435 normal cases, and only 15 were misclassified. An SP of 1 was obtained in the MONTGOMERY dataset because all 33 patients were classified correctly (see Figure 21d). The Montfort dataset obtained a good classification with only five patients from each category misclassified (see Figure 21i). Figure 22 shows the ROC curves for detecting COVID-19 on the nine datasets, and we can see that the curve of CXRIP is the highest because with an AUC of 0.999, followed by the curve of RSNA which obtained 0.999. The lowest is the curve of CIDC also which gives 0.945 for the AUC.

5.2. DeepCCXR-Multi for Individual Datasets

Similar to DeepCCXR-Bin, we tested DeepCCXR-Multi on individual datasets.

- CIDC dataset: This dataset provides the pneumonia class with 369 CXR images and 654 of COVID-19, to build a dataset with three categories. We added 1128 CXR normal images from RSNA dataset.
- The COVID-19 RADIOGRAPHY dataset contains the COVID-19 CXR class. We added normal and pneumonia classes of 1,128 CXR images and 1498 from RSNA dataset, respectively.
- The remaining datasets come with normal and pneumonia classes, and we added COVID-19 CXR images to build the datasets with three categories. The Montfort contains the three classes (normal, pneumonia and COVID-19).

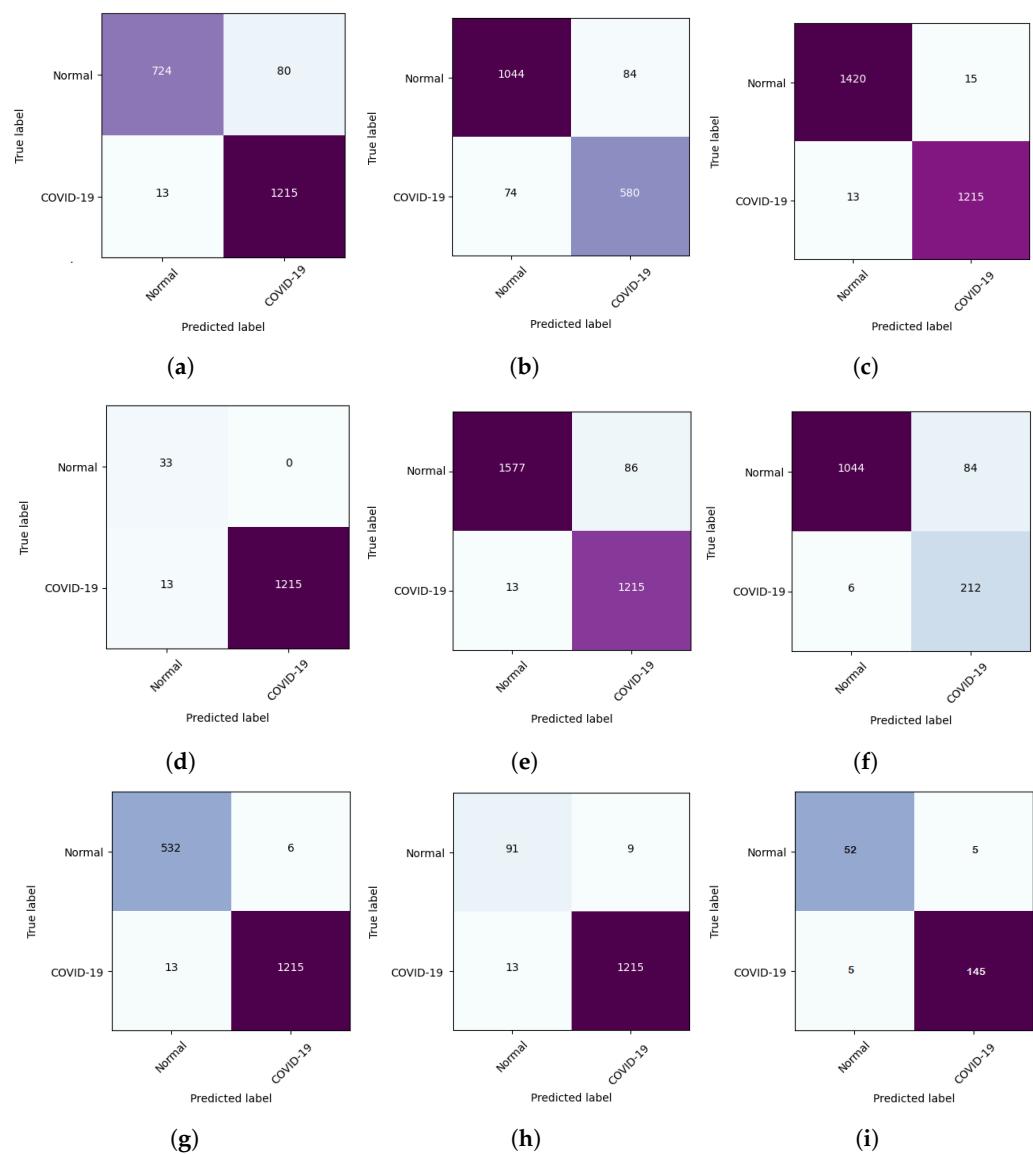
The results of each dataset are presented in Table 6. A high score was obtained for CXRIP with an AUC of 0.994. The CIDC, MONTGOMERY and SHENZHEN obtained an AUC of 0.987, 0.988 and 0.981, respectively. RSNA obtained an AUC score of 0.959 by using 2790 CXR images. An AUC of 0.908 was obtained by the NIH dataset using 4389 CXR images (1228 for COVID-19, 1663 for normal and 1498 for pneumonia). The AUC score for COVID-19 RADIOGRAPHY is 0.849, an SN of 0.92 and an SP of 0.92. The Montfort dataset obtained an AUC score of 0.801 because we have a few CXR images in the pneumonia cases (24 CXR images) and 58 CXR images in the normal cases.

The ROC curves of DeepCCXR-Multi for detecting COVID-19 in nine datasets are presented in Figure 23. As we can see, the curves of CIDC, CXRIP, SHENZHEN and MONTGOMERY are the highest because the AUC score is between 0.980 and 0.999, followed by the curve of RSNA dataset with an AUC of 0.950.

The confusion matrices for the multi-class classification are presented in Figure 24. Figure 24b gives the confusion matrix for CIDC dataset, we can see the model achieve good classification results because 358 CXR images from 369 in pneumonia class are correctly classified, and 944 CXR images in normal class are also correctly classified. As with the CXRIP dataset, we can see the model gives a better classification of the three classes, only four patients from 1435 are misclassified, and only 129 from 2123 in pneumonia class are misclassified (see Figure 24c). The confusion matrix for Montfort shows the COVID-19 patients are correctly classified (see Figure 24g). This confirms the robustness of our model in the classification of three classes.

Table 5. Performance measures of DeepCCXR-Bin for COVID-19 detection in nine datasets.

Dataset	ACC	AUC	SP	SN
CIDC	0.91	0.94	0.90	0.91
CXRIP	0.98	0.99	0.99	0.99
RSNA	0.98	0.99	0.99	0.99
NIH	0.96	0.98	0.96	0.97
MONTGOMERY	0.98	0.99	0.86	0.99
SHENZHEN	0.98	0.97	0.93	0.95
Montfort	0.95	0.98	0.91	0.97
BIMCV COVID19+	0.95	0.96	0.94	0.93
COVID-19 RADIOGRAPHY	0.93	0.96	0.86	0.95

**Figure 21.** Confusion matrices of DeepCCXR-Bin (COVID-19 vs. normal): (a) BIMCV COVID19+; (b) CIDC; (c) CXRIP; (d) MONTGOMERY; (e) NIH; (f) COVID-19 RADIOGRAPHY; (g) RSNA; (h) SHENZHEN; (i) Montfort.

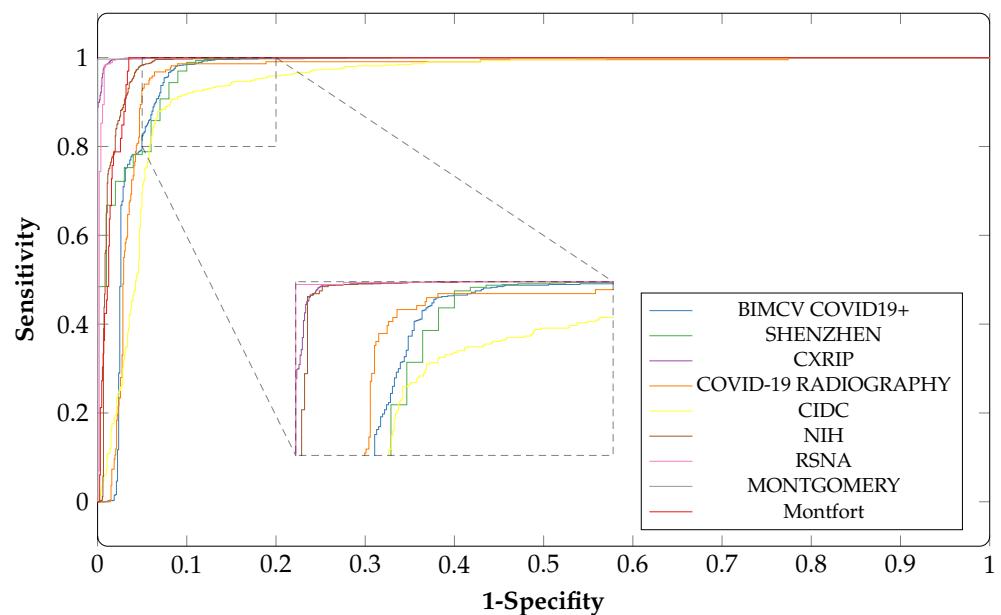


Figure 22. ROC curves of DeepCCXR-Bin for detecting COVID-19 in nine datasets.

Table 6. Performance measures of DeepCCXR-Multi for COVID-19 detection in nine datasets.

Dataset	ACC	AUC	SP	SN
CIDC	0.85	0.98	0.83	0.87
CXRIP	0.93	0.99	0.95	0.94
RSNA	0.92	0.95	0.92	0.93
NIH	0.90	0.90	0.91	0.89
MONTGOMERY	0.88	0.98	0.54	0.75
SHENZHEN	0.90	0.98	0.78	0.92
Montfort	0.70	0.80	0.50	0.51
BIMCV COVID19+	0.87	0.86	0.88	0.86
COVID-19 RADIOGRAPHY	0.90	0.84	0.92	0.92

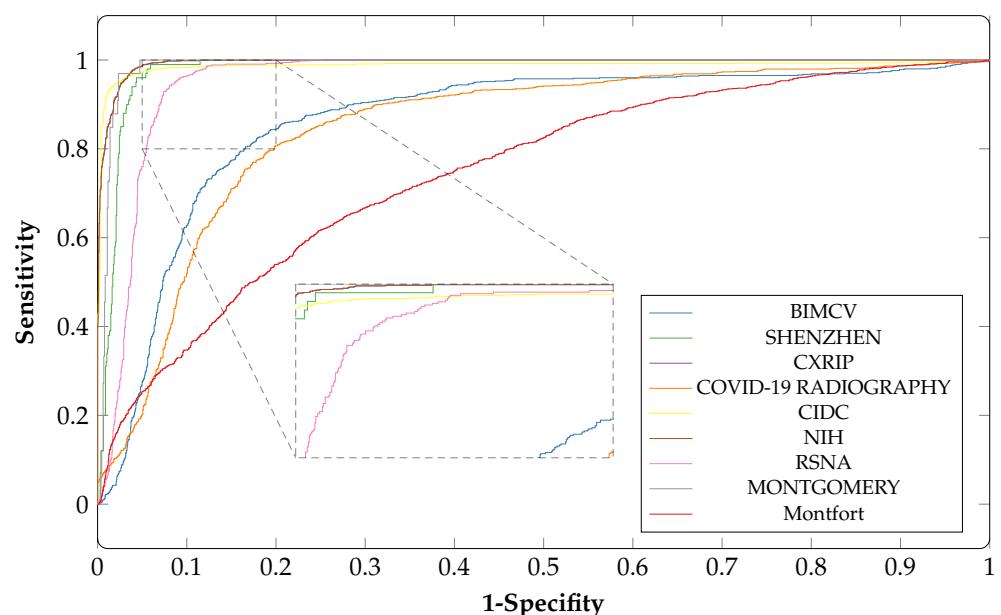


Figure 23. ROC curves of DeepCCXR-Multi for detecting COVID-19 in nine datasets.

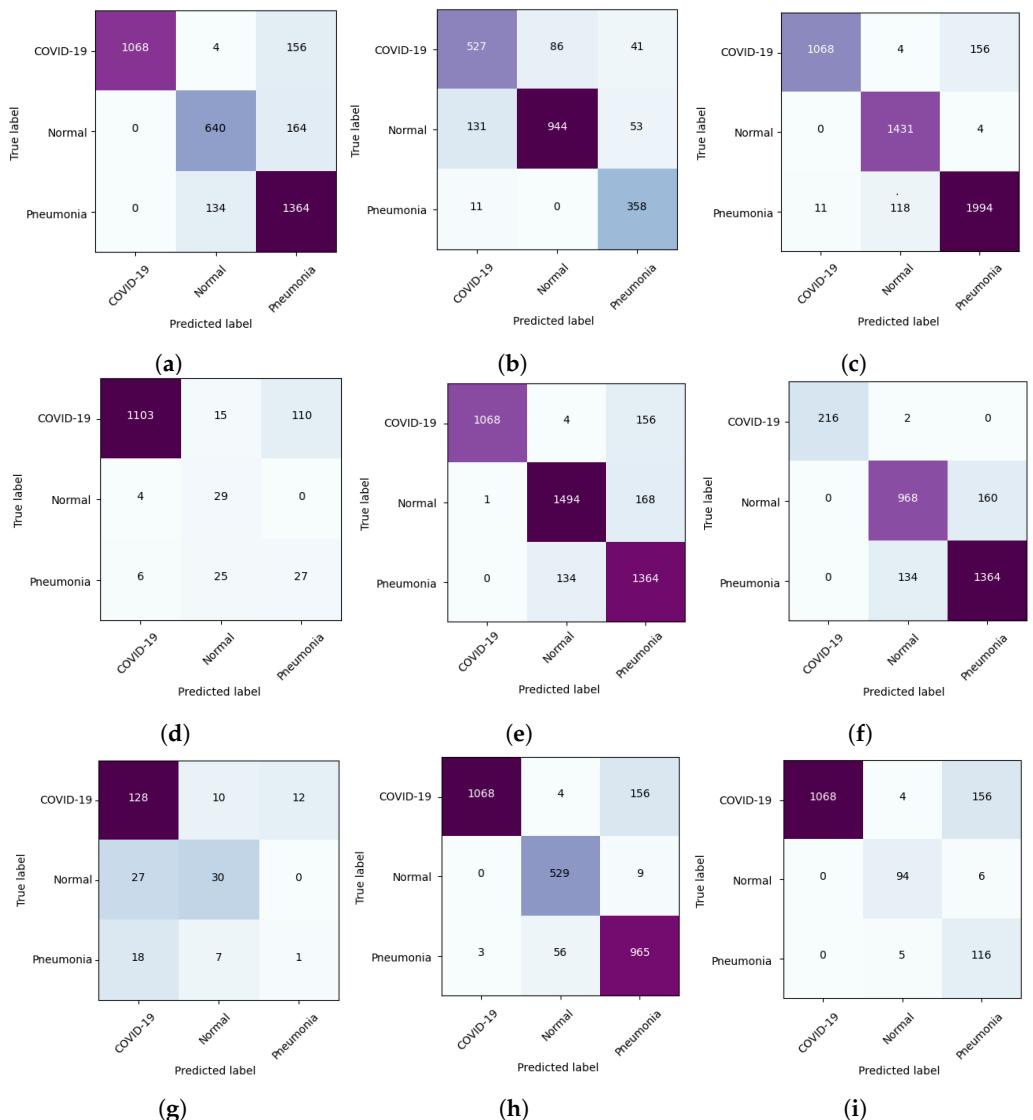


Figure 24. Confusion matrices of DeepCCXR-Multi (COVID-19 vs. pneumonia vs. normal): (a) BIMCV COVID19+; (b) CIDC; (c) CXRIP; (d) MONTGOMERY; (e) NIH; (f) COVID-19 RADIOGRAPHY; (g) Montfort; (h) RSNA; (i) SHENZHEN.

6. Model Limitations

Despite the good results obtained, both models have some limitations in classifying CXR images correctly; this sometimes happens in the presence of poor quality images which contain artifacts and noise resembling the opacity in the lungs. Figure 25a,b shows an example of FP cases in RSNA dataset, the heatmap localizes the opacity in image Figure 25a,b in the right lung, which is probably just a noise in the CXR images and the model to classify these images as COVID-19 instead of normal. The same can be said for Figure 25c in the SHENZHEN dataset, as the models classified the normal case as pneumonia. An example of FN is presented in Figure 25d from the RSNA dataset: the model detects this image as normal because the lungs appear clear on the top which is similar to the normal cases.

Some CXR images also contain cables, text and medical objects which are captured with the thorax. This makes the detection of the region of interest and disease signs more difficult for the model. Figure 25e shows an example of FP in the RSNA dataset: the image is labeled normal by the authors of the dataset and it is classified COVID-19 by the deep model, and the opacity on the top left lung leads to this error (same in Figure 25f).

Figure 25g shows an example of an FP case in the NIH dataset: the cables hide some parts of lung which gives pneumonia as the classification result.

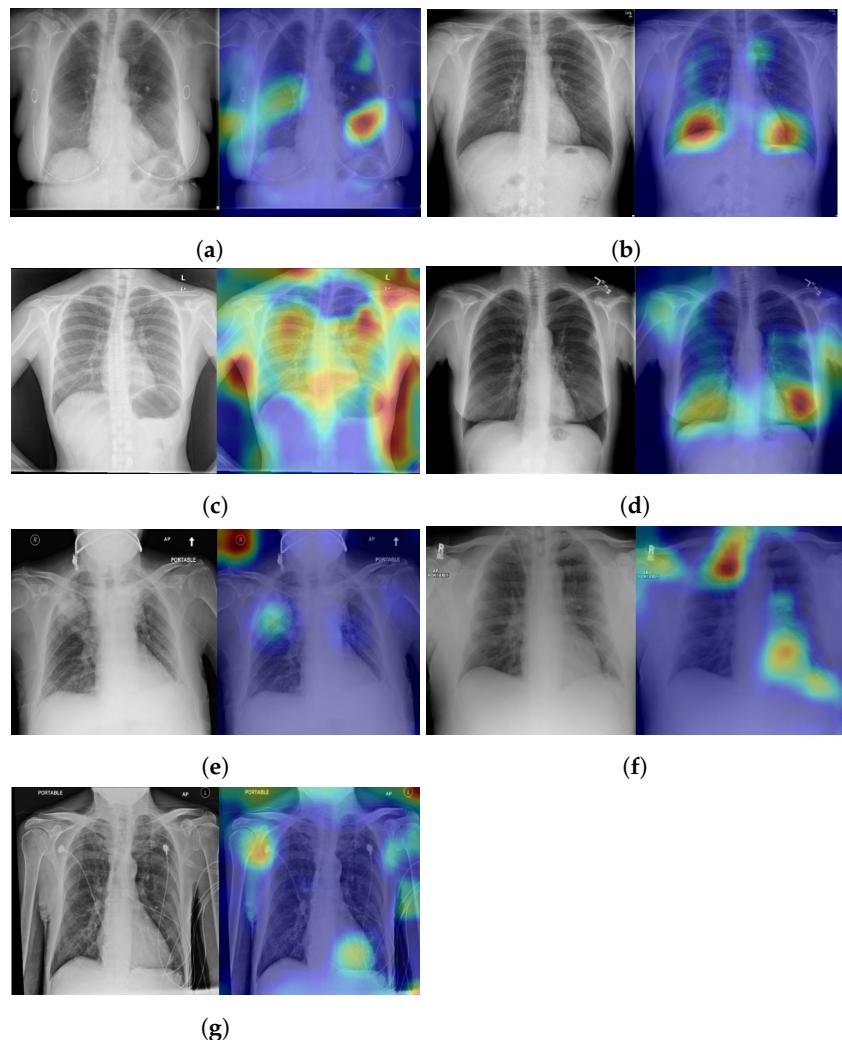


Figure 25. Examples of false positive and false negative cases detected by our deep learning models on CXR images: (a) and (b) CIDC; (c) SHENZHEN; (d), (e) and (f) RSNA; (g) NIH.

7. Conclusions

In this work, we developed DeepCCXR, a deep convolutional neural network (CNN) architecture for COVID-19 detection on Chest X-ray images. The proposed model is based on a recent architecture called EfficientNET-B5. DeepCCXR was fine-tuned and trained to detect COVID-19 using 1228 COVID-19 positive CXR images.

The obtained results show that our model outperforms recent deep learning approaches for COVID-19 detection on CXR. The model achieved high AUCs scores with 0.973 and 0.986 for DeepCCXR-Multi and DeepCCXR-Bin, respectively. The sensitivity reaches 0.97 and the specificity around 0.98 in average. For the COVID-19 class, we obtained a sensitivity of 0.99 (tested on 1228 COVID-19 positive images), meaning that we have a better performance in measuring the proportion of actual positives that are correctly identified as such (e.g., the percentage of people who are correctly identified as having COVID-19). An explainability algorithm was also developed and showed that DeepCCXR is efficient in identifying the most important pathology regions (the signs of COVID-19).

Previously published works used a limited number of COVID-19 images for testing. The large number of test images in our work shows that DeepCCXR is robust in detecting

COVID-19 and other pneumonia cases. Moreover, multiple datasets with varying quality were used in this work, which confirms the good degree of generalization of our model.

The proposed technique is an interesting contribution in the development of a CAD system able to detect COVID-19 and other pneumonia cases in CXR images. The model was deployed online and can be used freely by researchers and health professionals [55].

Future work includes extending our architecture to process CT scan images to detect COVID-19 and adapting the model to identify other types of diseases with radiography images.

Author Contributions: Conceptualization, M.C. and M.A.A.; methodology, M.C. and M.A.A.; software, M.C.; validation, M.C., M.A.A., E.M.B. and B.Y.; formal analysis, M.C., M.A.A. and E.M.B.; investigation, M.A.A. and M.C.; resources, M.C.; data curation, M.C.; writing—original draft preparation, M.C. and M.A.A.; writing—review and editing, M.C., M.A.A., B.Y. and E.M.B.; supervision, M.A.A. and E.M.B.; project administration, M.A.A. and E.M.B.; funding acquisition, M.A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by the Microsoft AI For Health program and by support from Atlantic Canada Opportunities Agency (ACOA), Regional Economic Growth through Innovation - Business Scale-Up and Productivity (project 217148), Natural Sciences and Engineering Research Council of Canada (NSERC), Alliance Grants (ALLRP 552039-20), New Brunswick Innovation Foundation (NBIF) and COVID-19 Research Fund (COV2020-042).

Institutional Review Board Statement: Université de Moncton IRB waived the approval requirements since the data used in this work was anonymized.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this work come mainly for public datasets. Please see the section describing the datasets.

Acknowledgments: The authors would like to acknowledge Joseph Abdulnour for his assistance with transferring and organizing anonymized CXR images.

Conflicts of Interest: The authors declare no conflict of interest.

References

- WHO. Coronavirus Disease 2020. 2020. Available online: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports> (accessed on 16 October 2021).
- WHO. Statement on the Second Meeting of the International Health Regulations (2005) Emergency Committee Regarding the Outbreak of Novel Coronavirus (2019-nCoV). 2020. Available online: <https://www.who.int/> (accessed on 16 October 2021).
- WHO. WHO Director-General's Opening Remarks at the Media Briefing on COVID-19. 2020. Available online: <https://www.who.int> (accessed on 16 October 2021).
- Fang, Y.; Zhang, H.; Xie, J.; Lin, M.; Ying, L.; Pang, P.; Ji, W. Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR. *Radiology* **2020**, *296*, E115–E117. [[CrossRef](#)]
- Fan, K.S.; Ghani, S.A.; Machairas, N.; Lenti, L.; Fan, K.H.; Richardson, D.; Scott, A.; Raptis, D.A. COVID-19 prevention and treatment information on the internet: A systematic analysis and quality assessment. *BMJ Open* **2020**, *10*, e040487. [[CrossRef](#)]
- Chowdhury, M.E.; Rahman, T.; Khandakar, A.; Mazhar, R.; Kadir, M.A.; Mahbub, Z.B.; Islam, K.R.; Khan, M.S.; Iqbal, A.; Al Emadi, N.; et al. Can AI help in screening viral and COVID-19 pneumonia? *IEEE Access* **2020**, *8*, 132665–132676. [[CrossRef](#)]
- Domingues, I.; Pereira, G.; Martins, P.; Duarte, H.; Santos, J.; Abreu, P.H. Using deep learning techniques in medical imaging: A systematic review of applications on CT and PET. *Artif. Intell. Rev.* **2020**, *53*, 4093–4160. [[CrossRef](#)]
- Shi, F.; Wang, J.; Shi, J.; Wu, Z.; Wang, Q.; Tang, Z.; He, K.; Shi, Y.; Shen, D. Review of Artificial Intelligence Techniques in Imaging Data Acquisition, Segmentation, and Diagnosis for COVID-19. *IEEE Rev. Biomed. Eng.* **2021**, *14*, 4–15. [[CrossRef](#)]
- Chetoui, M.; Akhloufi, M.A. Explainable end-to-end deep learning for diabetic retinopathy detection across multiple datasets. *J. Med. Imaging* **2020**, *7*, 044503. [[CrossRef](#)]
- Chetoui, M.; Akhloufi, M.A. Explainable Diabetic Retinopathy using EfficientNET. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 20–24 July 2020; pp. 1966–1969.
- Apostolopoulos, I.D.; Mpesiana, T.A. Covid-19: Automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. *Phys. Eng. Sci. Med.* **2020**, *43*, 635–640. [[CrossRef](#)] [[PubMed](#)]
- Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
- Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4510–4520.

14. Ghaderzadeh, M.; Asadi, F. Deep Learning in the Detection and Diagnosis of COVID-19 Using Radiology Modalities: A Systematic Review. *J. Healthc. Eng.* **2021**, *2021*, 6677314. [CrossRef]
15. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
16. Wang, L.; Lin, Z.Q.; Wong, A. Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest X-ray images. *Sci. Rep.* **2020**, *10*, 19549. [CrossRef] [PubMed]
17. Farooq, M.; Hafeez, A. Covid-resnet: A deep learning framework for screening of covid19 from radiographs. *arXiv* **2020**, arXiv:2003.14395.
18. Hemdan, E.E.D.; Shouman, M.A.; Karar, M.E. Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images. *arXiv* **2020**, arXiv:2003.11055.
19. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
20. Wehbe, R.M.; Sheng, J.; Dutta, S.; Chai, S.; David, A.; Barutcu, S.; Wu, Y.; Cantrell, D.R.; Xiao, N.; Allen, B.D.; et al. DeepCOVID-XR: An Artificial Intelligence Algorithm to Detect COVID-19 on Chest Radiographs Trained and Tested on a Large U.S. Clinical Data Set. *Radiology* **2021**, *299*, E167–E176. [CrossRef]
21. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
22. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A. Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; Volume 31.
23. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
24. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the 36th International Conference on Machine Learning, PMLR, Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
25. Chetoui, M.; Akhloufi, M.A. Deep Efficient Neural Networks for Explainable COVID-19 Detection on CXR Images. In *Advances and Trends in Artificial Intelligence, Artificial Intelligence Practices*; IEA/AIE: Kuala Lumpur, Malaysia, 2021; Chapter 29. [CrossRef]
26. Minaee, S.; Kafieh, R.; Sonka, M.; Yazdani, S.; Soufi, G.J. Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning. *Med. Image Anal.* **2020**, *65*, 101794. [CrossRef] [PubMed]
27. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
28. imageNet. Large Scale Visual Recognition Challenge (ILSVRC). Available online: <https://image-net.org/> (accessed on 16 October 2021).
29. Pan, I.; Cadrian-Chênevert, A.; Cheng, P.M. Tackling the radiological society of north america pneumonia detection challenge. *Am. J. Roentgenol.* **2019**, *213*, 568–574. [CrossRef]
30. Cohen, J.P.; Morrison, P.; Dao, L.; Roth, K.; Duong, T.Q.; Ghassemi, M. COVID-19 Image Data Collection: Prospective Predictions Are the Future. *arXiv* **2020**, arXiv 2006.11988.
31. de la Iglesia Vayá, M.; Saborit, J.M.; Montell, J.A.; Pertusa, A.; Bustos, A.; Cazorla, M.; Galant, J.; Barber, X.; Orozco-Beltrán, D.; García-García, F.; et al. BIMCV COVID-19+: A large annotated dataset of RX and CT images from COVID-19 patients. *arXiv* **2020**, arXiv:2006.01174.
32. Kermany, D.S.; Goldbaum, M.; Cai, W.; Valentim, C.C.; Liang, H.; Baxter, S.L.; McKeown, A.; Yang, G.; Wu, X.; Yan, F.; et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell* **2018**, *172*, 1122–1131.e9. [CrossRef] [PubMed]
33. National Library of Medicine. Tuberculosis Chest X-ray Image Data Sets. Available online: <https://lhncbc.nlm.nih.gov/publication/pub9931> (accessed on 16 October 2021).
34. Malhotra, A.; Mittal, S.; Majumdar, P.; Chhabra, S.; Thakral, K.; Vatsa, M.; Singh, R.; Chaudhury, S.; Pudrod, A.; Agrawal, A. Multi-Task Driven Explainable Diagnosis of COVID-19 using Chest X-ray Images. *arXiv* **2020**, arXiv:2008.03205.
35. Montfort, H. Hopital Montfort. 2020. Available online: <https://hopitalmontfort.com/> (accessed on 16 October 2021).
36. Heaven, W.D. Google’s Medical AI Was Super Accurate in a Lab. Real Life Was a Different Story. 2020. Available online: <https://www.technologyreview.com/2020/04/27/1000658/google-medical-ai-accurate-lab-real-life-clinic-covid-diabetes-retina-disease/> (accessed on 2 July 2020).
37. Sandler, M.; Howard, A.G.; Zhu, M.; Zhmoginov, A.; Chen, L. Inverted Residuals and Linear Bottlenecks: Mobile Networks for Classification, Detection and Segmentation. *arXiv* **2018**, arXiv:1801.04381.
38. Kızırk, A. Comparison of activation functions for deep neural networks, Available online: <https://towardsdatascience.com/comparison-of-activation-functions-for-deep-neural-networks-706ac4284c8a> (accessed on 16 October 2021).
39. Tamaki, Y.; Akiyama, F.; Iwase, T.; Kaneko, T.; Tsuda, H.; Sato, K.; Ueda, S.; Mano, M.; Masuda, N.; Takeda, M.; et al. Molecular detection of lymph node metastases in breast cancer patients: Results of a multicenter trial using the one-step nucleic acid amplification assay. *Clin. Cancer Res.* **2009**, *15*, 2879–2884. [CrossRef] [PubMed]
40. Marques, G.; Agarwal, D.; de la Torre Díez, I. Automated medical diagnosis of COVID-19 through EfficientNet convolutional neural network. *Appl. Soft Comput.* **2020**, *96*, 106691. [CrossRef] [PubMed]

41. Khan, I.U.; Aslam, N.; Anwar, T.; Aljameel, S.S.; Ullah, M.; Khan, R.; Rehman, A.; Akhtar, N. Remote Diagnosis and Triaging Model for Skin Cancer Using EfficientNet and Extreme Gradient Boosting. *Complexity* **2021**, *2021*, 5591614. [[CrossRef](#)]
42. Rahman, T. COVID-19 Radiography Database. 2020. Available online: <https://www.kaggle.com/tawsifurrahman/covid19-radiography-database> (accessed on 16 October 2021).
43. Shih, G.; Wu, C.C.; Halabi, S.S.; Kohli, M.D.; Prevedello, L.M.; Cook, T.S.; Sharma, A.; Amorosa, J.K.; Arteaga, V.; Galperin-Aizenberg, M.; et al. Augmenting the National Institutes of Health chest radiograph dataset with expert annotations of possible pneumonia. *Radiol. Artif. Intell.* **2019**, *1*, e180041. [[CrossRef](#)] [[PubMed](#)]
44. Chollet, F. Keras. 2015. Available online: <https://keras.io> (accessed on 16 October 2021).
45. Microsoft. Microsoft Azure. Available online: <https://azure.microsoft.com/> (accessed on 16 October 2021).
46. Ruder, S. An overview of gradient descent optimization algorithms. *arXiv* **2016**, arXiv:1609.04747.
47. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
48. Huang, W.; Song, G.; Li, M.; Hu, W.; Xie, K. Adaptive weight optimization for classification of imbalanced data. In Proceedings of the International Conference on Intelligent Science and Big Data Engineering, Beijing, China, 31 July–2 August 2013; Springer: Berlin/Heidelberg, Germany, 2013; pp. 546–553.
49. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 618–626.
50. Panwar, H.; Gupta, P.; Siddiqui, M.K.; Morales-Menendez, R.; Bhardwaj, P.; Singh, V. A deep learning and grad-CAM based color visualization approach for fast detection of COVID-19 cases using chest X-ray and CT-Scan images. *Chaos Solitons Fractals* **2020**, *140*, 110190. [[CrossRef](#)] [[PubMed](#)]
51. Kim, M.; Janssens, O.; Park, H.m.; Zuallaert, J.; Van Hoecke, S.; De Neve, W. Web applicable computer-aided diagnosis of glaucoma using deep learning. *arXiv* **2018**, arXiv:1812.02405.
52. Kowsari, K.; Sali, R.; Ehsan, L.; Adorno, W.; Ali, A.; Moore, S.; Amadi, B.; Kelly, P.; Syed, S.; Brown, D. Hmic: Hierarchical medical image classification, a deep learning approach. *Information* **2020**, *11*, 318. [[CrossRef](#)] [[PubMed](#)]
53. Stephanie, S.; Shum, T.; Cleveland, H.; Challa, S.R.; Herring, A.; Jacobson, F.L.; Hatabu, H.; Byrne, S.C.; Shashi, K.; Araki, T.; et al. Determinants of chest x-ray sensitivity for covid-19: A multi-institutional study in the united states. *Radiol. Cardiothorac. Imaging* **2020**, *2*, e200337. [[CrossRef](#)] [[PubMed](#)]
54. Chetoui, M.; Traoré, A.; Akhloufi, M.A. Deep Learning for COVID-19 Detection on Chest X-Ray and CT Scan. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 20–24 July 2020.
55. Couturier, A.; Chetoui, M.; Akhloufi, M.A. COVID-19 Detection Using Deep Learning. 2020. Available online: <https://covid19.primeai.ca/> (accessed on 16 October 2021).