

COVID-19 detection and disease progression visualization: Deep learning on chest X-rays for classification and coarse localization

Tahmina Zebin (✉ t.zebin@uea.ac.uk)

University of East Anglia <https://orcid.org/0000-0003-0437-0570>

Shahadate Rezvy

Middlesex University London <https://orcid.org/0000-0002-2684-7117>

Wei Pang

Heriot-Watt University <https://orcid.org/0000-0002-1761-6659>

Research Article

Keywords: Deep learning, Covid19, chest X-ray

Posted Date: June 15th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-34534/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Applied Intelligence on September 12th, 2020. See the published version at <https://doi.org/10.1007/s10489-020-01867-1>.

COVID-19 detection and disease progression visualization: Deep learning on chest X-rays for classification and coarse localization

Tahmina Zebin · Shahadate Rezvy ·
Wei Pang

Received: 17 May 2020 / Accepted: DD Month YEAR

Abstract Chest X-rays are playing an important role in the testing and diagnosis of COVID-19 disease in the recent pandemic. However, due to the limited amount of labelled medical images, automated classification of these images for positive and negative cases remains the biggest challenge in their reliable use in diagnosis and disease progression. We applied and implemented a transfer learning pipeline for classifying COVID-19 chest X-ray images from two publicly available chest X-ray datasets^{1,2}. The classifier effectively distinguishes inflammation in lungs due to COVID-19 and pneumonia (viral and bacterial) from the ones with no infection (normal). We have used multiple pre-trained convolutional backbones as the feature extractor and achieved an overall detection accuracy of 91.2% , 95.3%, 96.7% for the VGG16, ResNet50 and EfficientNetB0 backbones respectively. Additionally, we trained a generative adversarial framework (a cycleGAN) to generate and augment the minority COVID-19 class in our approach. For visual explanations and interpretation purposes, we visualized the regions of input that are important for predictions and a gradient class activation mapping (Grad-CAM) technique is used in the pipeline to produce a coarse localization map of the highlighted regions in the image. This activation map can be used to monitor affected lung regions during disease progression and severity stages.

T. Zebin
School of Computing Sciences, University of East Anglia, UK
E-mail: t.zebin@uea.ac.uk

S. Rezvy
School of Science and Technology, Middlesex University London, UK
E-mail: s.rezvy@mdx.ac.uk

W. Pang
School of Mathematical & Computer Sciences, Heriot-Watt University, UK
E-mail: W.Pang@hw.ac.uk

¹ <https://github.com/ieee8023/covid-chestxray-dataset>

² <https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>

Keywords Activation Maps · COVID-19 · Deep Neural networks · Transfer Learning

1 Introduction

The 2019 novel coronavirus (COVID-19) has become a serious public health problem across the world and is approaching approximately 4.597 million cases worldwide according to the statistics of European Centre for Disease Prevention and Control on May 17th, 2020. The COVID-19 infection may manifest itself as a flu-like illness potentially progressing to an acute respiratory distress syndrome. Despite the worldwide research efforts over the past few months, early detection of COVID-19 remains a challenging problem due to limited resources and the amount of data available for research. The gold standard screening method in COVID-19 is the reverse-transcription polymerase chain reaction (RT-PCR). Chest radiography imaging is being used as an alternative screening method and done in parallel to PCR viral testing [20]. Additionally, false negatives have been reported in PCR results due to insufficient cellular content in the sample or time-consuming nature and inadequate detection when there were positive radiological findings [2]. The accuracy of Chest X-ray (CXR) diagnosis of COVID-19 infection strongly relies on radiological expertise due to the complex morphological patterns of lung involvement which can change in extent and appearance over time. If these patterns are detected with high accuracy, it can enable rapid triaging for screening, diagnosis, and management of patients with suspected or known COVID-19 infection [14].

However, the limited number of trained thoracic radiologists limits the reliable interpretation of complex chest examinations, especially in developing countries. Deep learning techniques, in particular convolutional neural networks (CNNs), have been beating humans in various tasks of computer vision and other video processing tasks in recent years. Deep learning algorithms have already been applied for the detection and classification of pneumonia [15, 22] and other diseases on radiography. Hence, it has become the natural candidate for the analysis of CXR images to address the automated COVID-19 Screening. Some recent transfer learning approaches presented in [3, 5, 7, 14, 20] applied to CXR's of patients has been showing promising results in the identification of COVID-19.

In this paper, as an effort to improve the current COVID-19 detection using a limited number of publicly available CXR dataset, we devise and implement a CXR based COVID-19 disease detection and classification pipeline using a modified VGG-16, ResNet50 [9] and a recent EfficientNetB0 [19] architecture. Following the trend from the literature, for our research, we have assembled a three-class labelled dataset with x-ray images from 'normal', 'COVID-19', 'pneumonia' classes. "COVID-19 Image Data Collection" [6] is currently serving as the main source of COVID-19 CXR's at this stage. To enhance the under-represented COVID-19 class, we train a generative adversarial framework to generate synthetic COVID-19 images during our experiments. Our

choice for the convolutional backbone for this research is mostly driven by their lightweight nature and their performance measures in terms of accuracy, precision and recall performances to accurately detect COVID-19.

The remaining sections of this paper are organized as follows. In Section 2, we review current literature on COVID-19 CXR image analysis using deep learning methods. Design insights are derived from the review of the related work and we provide a description of the dataset for the implemented network in this section. Section 3 gives details on the proposed transfer learning architecture and discusses the necessary settings, pre-trained backbones and procedural stages. The influence of model backbones on the training time, loss and model accuracy are also discussed in this section. The model performance is evaluated in Section 4 where classification results in terms of recall, precision and overall accuracy are compared and contrasted with concurrent methods reported in the literature. We also present a gradient class activation mapping (Grad-CAM) technique to monitor affected lung regions during disease progression for visual explanations and interpretation purposes. Finally, conclusions are drawn in Section 5.

2 Related work

2.1 Deep learning for Chest X-ray and COVID-19 Diagnosis

All the related work found so far addressing the COVID-19 chest X-rays are arxiv pre-prints and are still not peer-reviewed. We provide a summary of the recent literature on coronavirus X-ray classification based on deep learning techniques in this subsection.

In [5], a database of 190 COVID-19, 1345 viral pneumonia, and 1341 normal chest x-ray images was introduced. Training and validation on four different pre-trained networks, namely, Resnet18, DenseNet201, AlexNet and SqueezeNet for the classification of two different schemes (normal and COVID-19 pneumonia; normal, viral and COVID-19 pneumonia). The classification accuracy for both the schemes were 98.3%, 96.7% respectively. The sensitivity, specificity and precision value were also reported. In [10], a comparison among seven different well-known deep learning neural networks architectures was presented. In the experiments, they use a small data set with only 50 images in which 25 samples are from healthy patients and 25 from COVID-19 positive patients. In their experiments, the VGG19 and the DenseNet201 were the best performing architectures. In [20], an architecture called COVID-net is created to classify X-ray images into normal, pneumonia, and COVID-19. Differently from the previous work, they use a much larger dataset consisting of 16,756 chest radiography images across 13,645 patient cases. The authors report an accuracy of 92.4% overall and sensitivity of 80% for COVID-19.

In [7], a pre-trained ResNet50 model is fine-tuned for the problem of classifying X-ray images into normal, COVID-19, bacterial-pneumonia and viral pneumonia. The authors report better results when compared with the

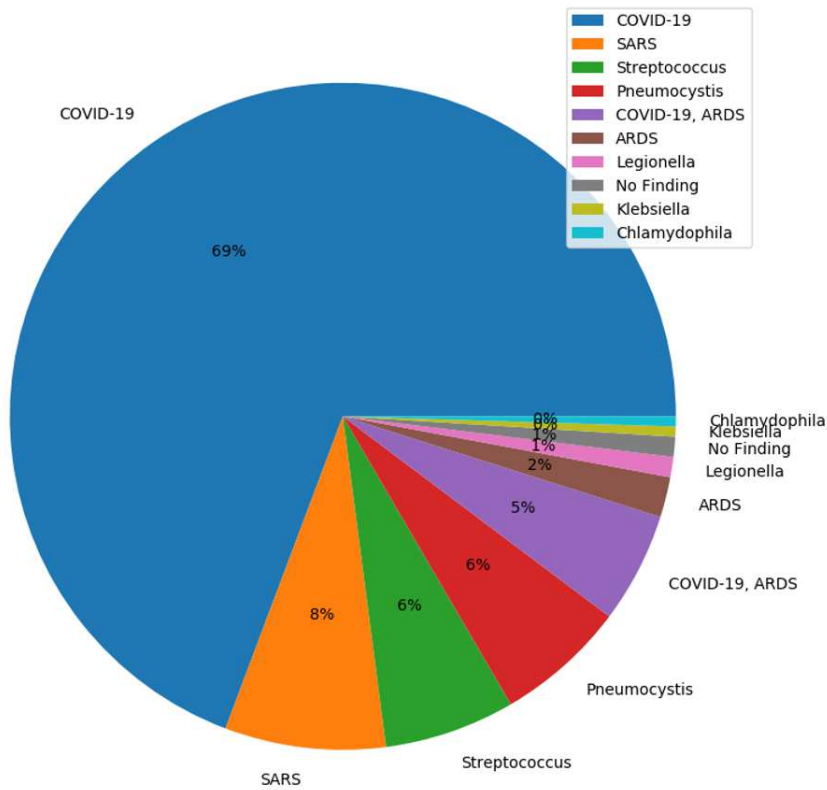


Fig. 1: COVID-19 Image Data Collection: Image distribution as per diagnosis (69% COVID)

COVID-net, 96.23% accuracy overall, and 100% sensitivity for COVID-19. Nevertheless, it is important to highlight that the difference in [7] that it has an extra class than [20] and the dataset consists of 68 COVID-19 radiographs from 45 COVID-19 patients, 1,203 healthy patients, 931 patients with bacterial pneumonia and 660 patients with nonCOVID-19 viral pneumonia. Additionally, the test set has only 8 COVID-19 instances for the claim of 100% sensitivity to be generalized for a larger cohort.

In a very recent paper [14], the authors aimed towards a light-weight implementation of a COVID-19 classifier and with an accuracy of 93.9%, COVID-19 Sensitivity of 96.8% and positive predictive value of 100% using a flat version of EfficientNet backbone. A hierarchical version of EfficientNet was also reported with 93.5% accuracy and COVID-19 sensitivity of 80.6%.

Table 1: Dataset and other settings

Settings	Description
Original Chest X-ray (CXR)	COVID-19: 142; Normal: 300; Pneumonia: 300
Pre-processing	Intensity normalization, class-label encoding
Training set division (80%)	COVID-19: 113; Normal: 240; Pneumonia: 240
Test set division(20%)	COVID-19: 29; Normal: 60; Pneumonia: 60
Augmentation	version1 (v1): Random rotation, width shift, height shift, horizontal flip version2 (v2): 100 CycleGAN synthesized image for COVID-19, followed by augmentation steps in v1
Validation set	5-fold cross-validation on the augmented training set
Pre-trained base models	
VGG16	Fixed-size kernel; parameter: 138M, Input shape: 224, 224,3
Resnet50 [9]	Residual connections; 26M, Input shape: 224, 224,3
EfficientNetB0 [19]	Mobile inverted bottleneck Convolution with depth, width, and resolution; parameter: 5.3M, Input shape: 224, 224,3

2.2 Dataset Description

Following the trend of possible classes found in the literature, we have assembled a three-class dataset with labels, normal - for healthy patients; COVID-19 - for patients with COVID-19; and pneumonia- for patients with viral and bacterial pneumonia. Learning with an unbalanced dataset could produce a biased prediction model towards the classes with more samples. Hence, we have separated all the 142 Antero-posterior (AP) view of COVID-19 positive x-ray images from the ‘‘COVID-19 Image Data Collection’’ [6]. This dataset is a collection of anonymized COVID-19 images, acquired from websites of medical and scientific associations and COVID related research papers. This is a constantly growing dataset and at the time of writing this paper, the dataset had in total 354 X-ray and CT images of patients who were affected by COVID-19 and other diseases, such as MERS, SARS, and ARDS [2]. Fig. 1 shows the percentage of image distribution as per the diagnosis, where 69% of the images had some form of COVID-19 findings. The age distribution of the COVID-19 positive cases. So far, the age group that concentrates most cases is from 50 to 80 years old and for 30 individuals age metadata is not recorded in the dataset.

Chest radiography is widely used for the detection and classification of pneumonia and other pulmonary diseases, so there are several large publicly available datasets of CXR images for normal, viral and bacterial pneumonia datasets such as the NIH Chest X-ray Dataset [11], RSNA Pneumonia Detection Challenge dataset available at <https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/> and a more recent COVIDx dataset from [20]. For assembling the normal and pneumonia classes for our research, we randomly chose 300 images for each of the normal and pneumonia classes from

the Chest-xray-pneumonia dataset (available at the link of footnote 2). We opted for a small number of images for these classes to avoid drastic class imbalance for the model training scenario. This made our original dataset to be consisting of 742 CXR images. 80% of the data is then separated as the training set, the remaining 20% of the dataset contributing as the test data. A detailed division of the dataset can be found in Table 1.

2.3 Image Augmentation

To achieve robust and generalized deep learning models large amounts of data are needed, however, medical imaging data is scarce and labelling the dataset is expensive. We applied two different versions of the augmentation technique on the dataset. In the first version, we applied image augmentation techniques [17] such as random rotation, width shift, height shift, horizontal and vertical flip operations using the ImageDataGenerator functionality from the TensorFlow Keras framework [4, 8].

Generative adversarial networks (GAN) offer a novel method for data augmentation used nowadays. Hence, we have used a CycleGAN [23] architecture for increasing the under-represented COVID-19 class images (described as version 2 for augmentation in Table 1). Utilizing the normal class from our dataset, we trained the CycleGAN to transform Normal images into COVID-19 images. As a proof-of-concept at this stage, we have generated 100 synthetic COVID-19 images to add to our original training dataset. Fig. 2 shows a few examples of the original and generated images side-by-side. After 5000 iterations of the generator and discriminator training, we have achieved near realistic generated CXR images, though there are shape deformations seen in some cases. To be noted, the dataset after augmentation is still quite small, so we employed five-fold cross-validation during training to avoid the over-fitting of the model and the validation set served as a checkpoint for us to the trained model's performance to unseen data.

3 Model Architecture

We implemented the COVID-19 disease detection pipeline using an adapted Convolutional Neural Network architecture and trained it in the feature-representation transfer learning mode. We effectively used a pre-trained VGG-16, ResNet50, and EfficientNet-B0 as our feature extractor. As all these backbones were pre-trained on huge ImageNet dataset, it has learned a good representation of low-level features like spatial, edges, rotation, lighting, shapes and these features can be shared across to enable the knowledge transfer and act as a feature extractor for new images in different computer vision problems. As in our case, the new images have different categories from the source dataset, the pre-trained model is used to extract relevant features from these images based on the principles of transfer learning. We used TensorFlow, Keras, PyTorch,

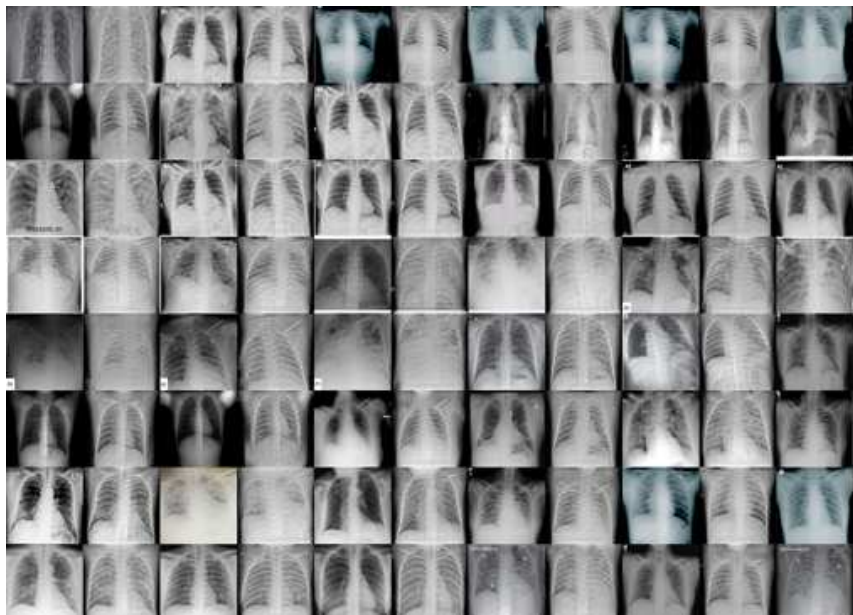


Fig. 2: Generated images from cycleGAN for the underrepresented COVID-19 class

scikit-learn and OpenCV libraries in python for generating various functionalities of the pipeline. Fig. 3 shows an illustration of our proposed pipeline for COVID-19 chest X-ray classification.

3.1 Pre-trained model backbone and network head removal

We removed the network head or the final layers of the pre-trained model (e.g. VGG-16, ResNet50, and EfficientNetB0 backbone in our case) that was initially trained on the ImageNet dataset. This stage is crucial as the pre-trained model was trained for a different classification task. The removal of network head removed weights and bias associated with the class score at the predictor layers. It is then replaced with new untrained layers with the desired number of classes for the new data. We adjusted a three-class network head for the COVID-19 dataset for three possible labels, namely, normal - for healthy patients, COVID-19 - for patients with COVID-19 and pneumonia - for patients with non-COVID-19 pneumonia.

3.2 Transfer learning stages

At the initial stage, we froze the weights of the earlier layers of the pre-trained backbone to help us extract the generic low-level descriptors or patterns from

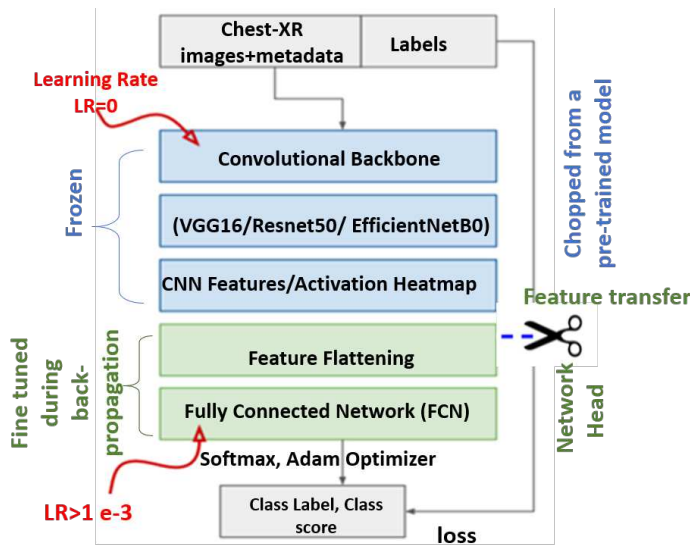


Fig. 3: Transfer learning architecture with pre-trained convolutional backbone for COVID-19 chest X-ray classification

the chest X-ray image data. Later layers of the CNN become progressively more specific to the details of the output classes of the new data-set. Then a newly added network head is trained for adapting the weights according to the patterns and distribution of the new dataset. The network head is updated and fine-tuned during model training. The training of the model has been done offline on an Ubuntu machine with Intel(R) Core i9-9900X CPU @ 3.50GHz, 62GB memory and a GeForce RTX 2060 GPU. The final model was fine-tuned with an Adam optimizer with a learning rate of 0.0001 and a categorical cross-entropy. To be noted, five-fold cross-validation is used during training to avoid the over-fitting of the model.

3.3 Training Loss

Fig. 4 shows the change in loss function for the three convolutional models we experimented during this research. We trained each model for 50 epochs. When the model was trained with the originally assembled three-class dataset, after traditional augmentation the model with VGG16(v1) took the longest during training to reach the stopping loss criteria with a categorical cross-entropy to reach a training dataset accuracy of 0.93. The VGG16(v2) is the same model trained with an enhanced version of the original dataset, where the under-represented COVID-19 class is enhanced by 100 more synthetic images generated with a CycleGAN. The training loss seemingly reached the threshold loss value within 10 epochs in this case. The realistic augmentation

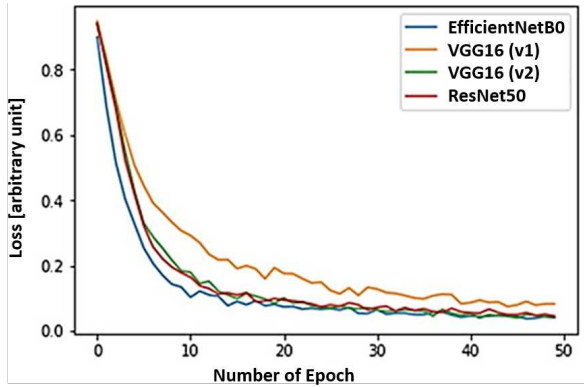


Fig. 4: Comparative loss function on the training dataset

Table 2: Class-wise precision performance comparison with other deep learning techniques in literature with our findings for COVID-19 detection

Backbone	Accuracy	COVID-19	Normal	Pneumonia
Our results:				
VGG16 (v1 Augmentation)	0.88	0.82	0.84	0.98
VGG16 (v2 GAN Augmentation)	0.912	0.93	0.87	0.96
Resnet50	0.953	0.96	0.96	0.93
EfficientNetB0	0.967	0.950	1	0.96
Concurrent proposed approach:				
COVIDNet-CXR Large [20]	0.943	0.909	0.917	0.989
COVIDNet-CXR Small [20]	–	0.964	0.898	0.947
VGG16 [14]	0.77	0.636	–	–
Flat - EfficientNetB0 [14]	0.90	1	–	–
Flat - EfficientNetB3 [14]	0.939	1	–	–

in the COVID-19 class definitely has increased the model’s accuracy by almost 3%. A further improvement is achieved when the backbone was replaced with ResNet50 and EfficientNetB0, with the EfficientNetB0 being the fastest. To be noted, each epoch for the given training dataset and computational setup took about 18 seconds with 232 ms/step for a batch size of 8 and learning rate of 0.0001. The EfficientNetB0 also achieved the best accuracy with the squeeze-and-excitation(SE) optimization stage included in its architecture.

4 Results and Evaluation

4.1 Model evaluation matrices

If True Positive (T_P) is the number of COVID-19 classified rightly as COVID; True Negative (T_N) is the number of normal CXR’s rightly classified normal; False Positive (F_P) is the number of normal events misclassified as COVID-19

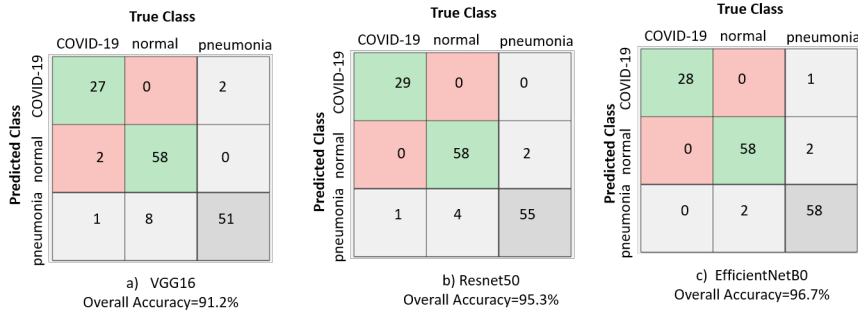


Fig. 5: Confusion matrix and overall accuracy of three backbone models used in this research

and noCovid pneumonia and False Negative (F_N) is the number of COVID-19's misclassified as normal or pneumonia, we can define accuracy, recall and precision of a model can be defined using the following equations [18].

- Accuracy: It is an indicator of the total number of correct predictions provided by the model and defined as follows:

$$\text{Accuracy} = \frac{T_P + T_N}{T_P + T_N + F_P + F_N}. \quad (1)$$

- Recall and precision: Two of the most commonly used performance measures, recall and precision measures are defined as follows:

$$\text{Recall or Sensitivity} = \frac{T_P}{T_P + F_N}. \quad (2)$$

$$\text{Precision or positive predictive value} = \frac{T_P}{T_P + F_P}. \quad (3)$$

Our results show reasonably accurate performance with an overall detection accuracy of 91.2%, 95.3%, 96.7% for our exemplar VGG16, ResNet50 and EfficientB0 backbones respectively on the fixed test set of 29 COVID-19, 60 Normal and 60 CXR images for the Pneumonia class. We presented the confusion matrix plot for the three backbone models under consideration in Fig. 5. The rows correspond to the predicted class (Output Class) and the columns correspond to the true class (Target Class). The diagonal cells in the confusion matrix correspond to observations that are correctly classified (T_P and T_N 's). The off-diagonal cells correspond to incorrectly classified observations (F_P and F_N 's). The number of observations is shown in each cell.

4.2 Comparison with other approaches

We have summarized a class-wise precision and recall performances from various experiments in Table 2. Additionally, we compared our experimental results with some concurrent proposed approach. As can be seen from the results presented in Table 2, for the base VGG16 model, when the under-represented COVID-19 class is enhanced by 100 more synthetic images generated with a CycleGAN, referred on the Table 1 as version 2 augmentation, there was a 3.2% improvement in overall accuracy from the model, and the precision performance for the COVID-19 class has been improved from 0.88 to 0.93 through the addition of these realistically augmented COVID-19 data. When comparing this to the VGG16 model performance with a COVID-19 class precision value of 0.636 presented in Luz et al. [14], this showed a clear improvement, though the dataset used for training are not directly comparable. The VGG16 model, when saved for the inference stage, has a memory footprint of 57 megabytes with 14.7 million parameters. For the ResNet50 base model, the overall accuracy has improved to 95.3% due to a larger number of features extracted by the model, leading to a better distinction between class. This model, when serialized and saved, has a memory footprint of 97 megabytes with 23.7 million parameters. In the approach presented in [7] with ResNet50, the accuracy achieved is 96.23%, which is slightly higher than the value we achieved. However, in their test dataset, there were only 8 instances for the COVID-19 class in a four-class classification scenario, the value may not be robust and generalized for different class distribution.

Our experimentation with the EfficientNetB0 base model has achieved a 96.7% overall accuracy, with a COVID-19 class precision and recall value of 0.95 and 0.965 respectively. When compared to the COVIDNet-CXR model proposed by Wang et al [20], the values were 0.909 and 0.968 respectively. Our version of EfficientNetB0 has higher precision, which is critical as the goal is to be able to detect as many positive COVID-19 cases to reduce the community spread. Using the same backbone, the EfficientNetB3 proposed by Luz et al [14] has shown a precision of 100% for the COVID-19 class, while the overall accuracy is lower than the version we have. To be noted, the EfficientNetB3 model has 12.3 million parameters whereas EfficientNetB0 has 5.3 million parameters, contributing a lighter memory footprint (21 megabytes) than its scaled version. Additionally, the depth, width and resolution scaling in the EfficientNet architecture with squeeze-and-excitation(SE) optimization seemingly outperformed both VGG and ResNet architecture in our experiments.

4.3 Coarse region localization map with gradient class activation

For visual explanations and interpretation purposes, we visualized the regions of input that are important for predictions and a gradient class activation mapping (Grad-CAM) technique [16] is used in the pipeline to produce a

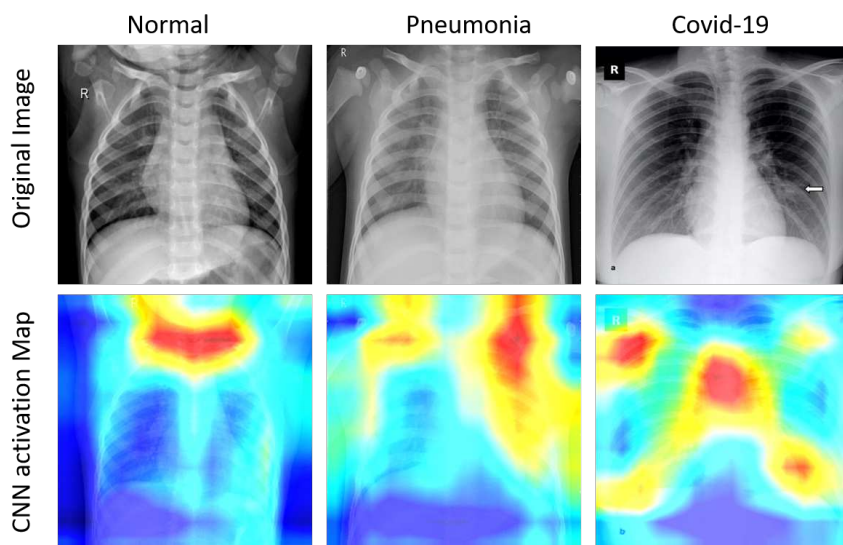


Fig. 6: Activation map visualization for the three classes under consideration. The First column presents a healthy chest x-ray sample, the second, from a patient with pneumonia, and the third one, from a patient with COVID-19, visualizing affected regions in lungs.

coarse localization map of the highlighted regions in the image. In Fig. 6, activation map visualization for the three classes under consideration. The first row represents the original images, and the second, the activation maps. The first column presents a healthy chest x-ray sample, the second, from a patient with pneumonia, and the third one, from a patient with COVID-19. The rightmost CXR taken on the patient shows bilateral patchy ground-glass opacities. These activation maps can be used to monitor affected lung regions during disease progression and severity stages. In Fig. 7, for a patient's x-ray in ICU-care at day 3, 7 and 9, the coarse localization map showed increased inflammation indicating disease severity. There are multi-focal patch/nodular consolidations and ground-glass opacities around the right mid to lower lung zone observed on day 9.

Though clinical symptoms such as consolidations and ground-glass opacities [13] are more accurately recognizable in Computed Tomography (CT) scans, CXR's could still provide a coarse and cheap bed-side indication of such symptoms if these visualizations are enhanced by labels and clinical notes from radiologists and domain experts.

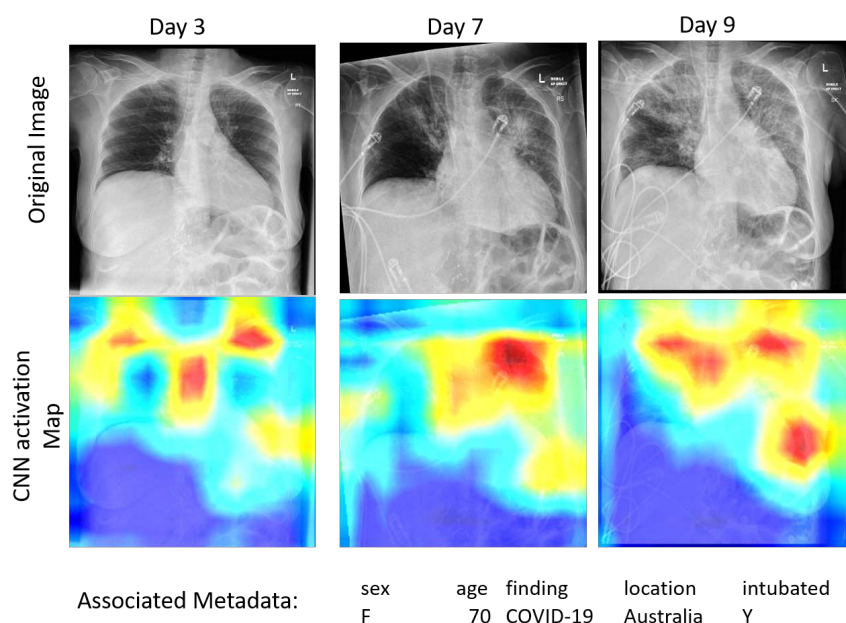


Fig. 7: Activation map visualization for a patient's x-ray in ICU-care at day 3, 7 and 9, visualising increased inflammation indicating disease severity.

5 conclusion

Deep learning applied to chest X-rays of patients has been showing promising results in the identification of COVID-19. In this research, we experimented on lightweight convolutional network architecture with three backbones (VGG-16, ResNet50, and EfficientNetB0 pre-trained on ImageNet dataset) for detecting COVID-19 and associated infection from chest X-ray images. Experiments were conducted to evaluate the convolutional neural networks performance on the generally augmented dataset and on an extended version of the dataset that utilized application of generative adversarial network-based augmentation using CycleGAN. Even with a limited number of images in the COVID-19 class, promising results achieved by the network on the test dataset with a recall value over 90% and a precision value over 93% for all the three models. We would like to emphasize on the fact that, with more images and new data collected for the COVID-19 class, it will be possible to improve the training and to improve sensitivity and detection rate. Our results also indicated the application of generative adversarial network-based augmentation techniques can contribute to accuracy improvement and can produce a more generalized and robust model.

In future, provided the clinical notes and metadata related to survival, need for intubation, need for supplemental oxygen, it is possible to train mixed

image and metadata models aiming to provide prognostic and severity predictions [6, 20]. These models could be highly useful for risk stratification, patient management, and personalized care planning in this critical resource-constrained pandemic scenario.

All models developed in this work have a memory footprint below 100 megabytes. Hence, another future direction from this research will be to extend the trained models implementation on conventional smartphone processor to do fast and cheap on-device inference to provide a proof of concept of transferring the capability of deep learning models on mobile devices [12]. We would like to build on our previous experience in transferring such models using the TensorFlow lite (TFLite) library [1, 21].

References

1. Tensorflow lite for mobile and embedded learning. <https://www.tensorflow.org/lite/microcontrollers/overview> (2019)
2. Araujo-Filho, J.d.A.B., Sawamura, M.V.Y., Costa, A.N., Cerri, G.G., Nomura, C.H.: Covid-19 pneumonia: what is the role of imaging in diagnosis? *Jornal Brasileiro de Pneumologia* **46**(2) (2020)
3. Butt, C., Gill, J., Chun, D., Babu, B.A.: Deep learning system to screen coronavirus disease 2019 pneumonia. *Applied Intelligence* p. 1 (2020)
4. Chollet, F.: Introduction to keras. March 9th (2018)
5. Chowdhury, M.E., et al.: Can AI help in screening viral and covid-19 pneumonia? arXiv preprint arXiv:2003.13145 (2020)
6. Cohen, J.P., Morrison, P., Dao, L.: Covid-19 image data collection. arXiv preprint arXiv:2003.11597 (2020)
7. Farooq, M., Hafeez, A.: Covid-resnet: A deep learning framework for screening of covid19 from radiographs. arXiv preprint arXiv:2003.14395 (2020)
8. Gulli, A., Pal, S.: Deep learning with Keras. Packt Publishing Ltd (2017)
9. 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, pp. 770–778 (2016)
10. 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 preprint arXiv:2003.11055 (2020)
11. Jaeger, S., Candemir, S., Antani, S., Wang, Y.X.J., Lu, P.X., Thoma, G.: Two public chest x-ray datasets for computer-aided screening of pulmonary diseases. *Quantitative imaging in medicine and surgery* **4**(6), 475 (2014)
12. Li, X., Li, C., Zhu, D.: COVID-MobileXpert: on-device covid-19 screening using snapshots of chest x-ray (2020)
13. Li, X., et al.: CT imaging changes of corona virus disease 2019 (COVID-19: a multi-center study in southwest china. *Journal of Translational Medicine* **18**, 1–8 (2020)
14. Luz, E., Silva, P.L., Silva, R., Moreira, G.: Towards an efficient deep learning model for covid-19 patterns detection in x-ray images (2020)
15. Rajpurkar, P., et al.: Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv preprint arXiv:1711.05225 (2017)
16. 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, pp. 618–626 (2017)
17. Shorten, C., Khoshgoftaar, T.M.: A survey on image data augmentation for deep learning. *Journal of Big Data* **6**(1), 60 (2019)
18. Sokolova, M., Lapalme, G.: A systematic analysis of performance measures for classification tasks. *Information Processing & Management* **45**(4), 427 – 437 (2009)

19. Tan, M., Le, Q.V.: Efficientnet: Rethinking model scaling for convolutional neural networks (2019)
20. Wang, L., Wong, A.: Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest radiography images. arXiv preprint arXiv:2003.09871 (2020)
21. Zebin, T., Scully, P.J., Peek, N., Casson, A.J., Ozanyan, K.B.: Design and implementation of a convolutional neural network on an edge computing smartphone for human activity recognition. *IEEE Access* **7**, 133509–133520 (2019)
22. Zech, J.R., Badgeley, M.A., Liu, M., Costa, A.B., Titano, J.J., Oermann, E.K.: Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. *PLoS medicine* **15**(11) (2018)
23. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision, pp. 2223–2232 (2017)

Figures

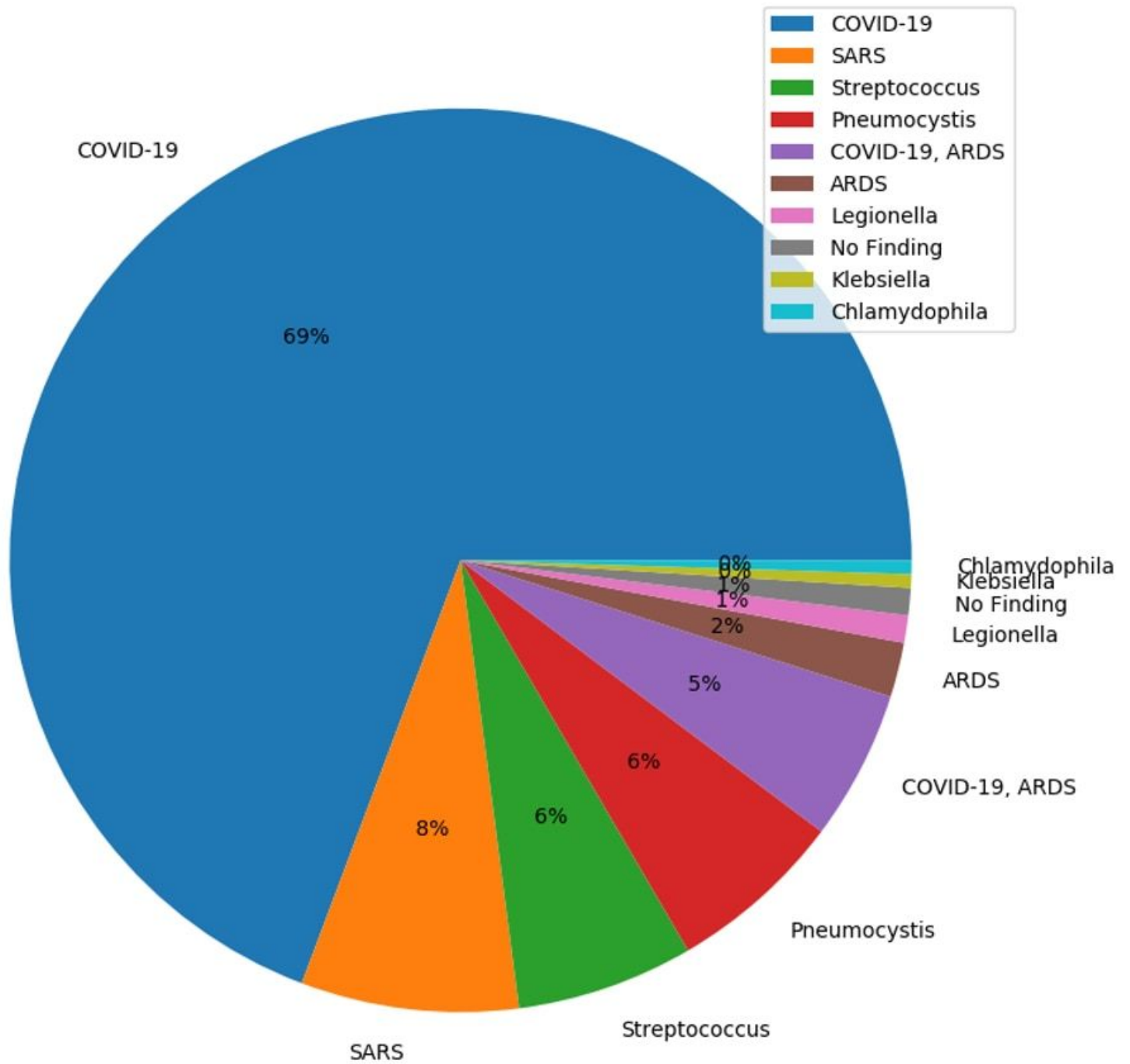


Figure 1

COVID-19 Image Data Collection: Image distribution as per diagnosis (69% COVID)

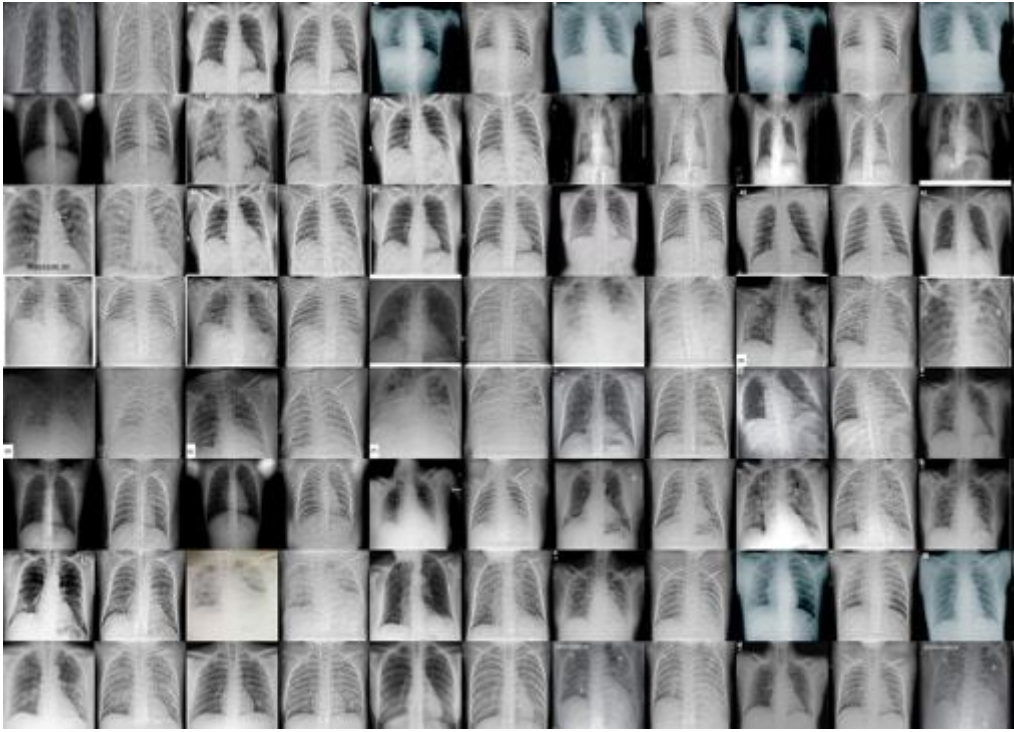


Figure 2

Generated images from cycleGAN for the underrepresented COVID-19 class

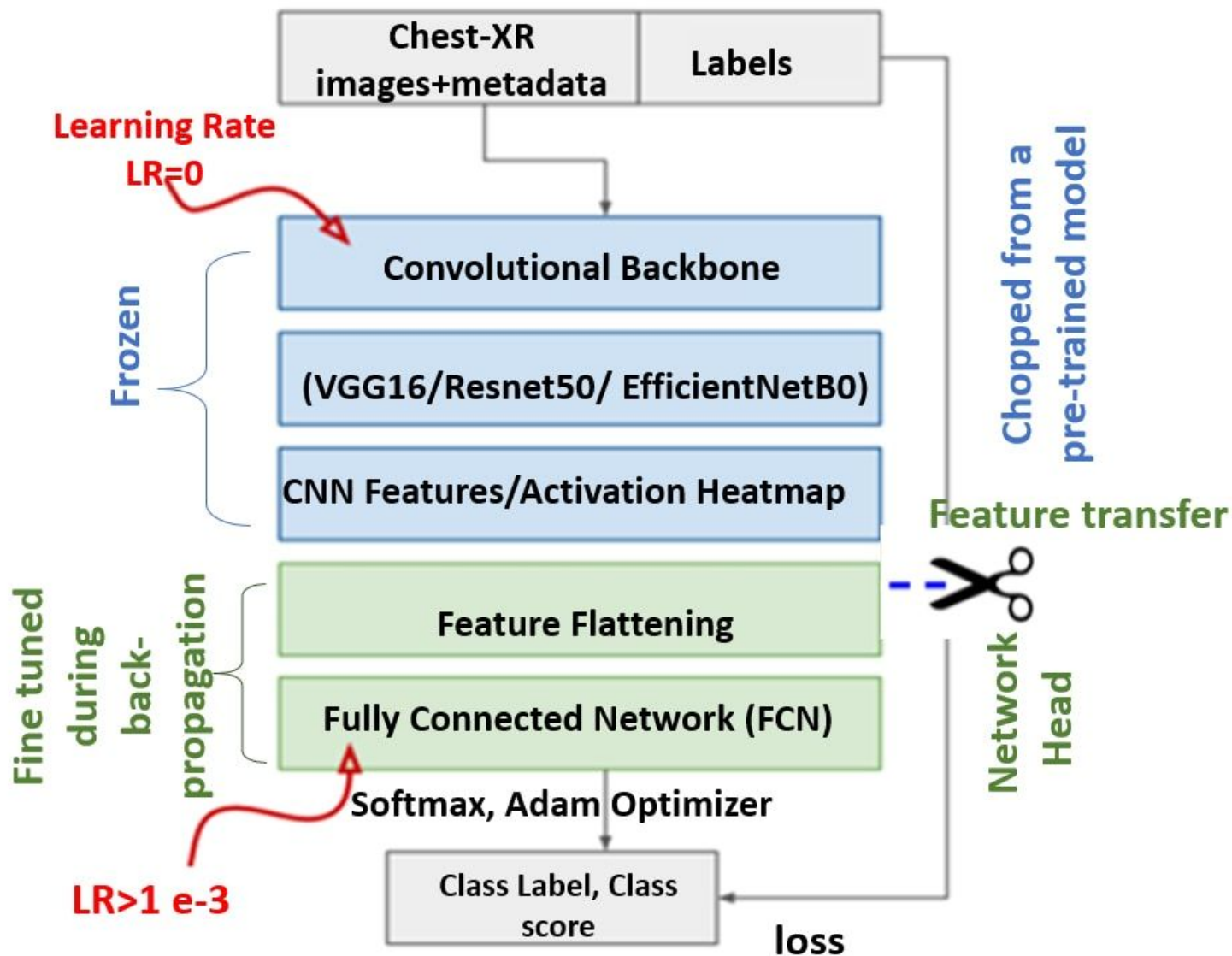


Figure 3

Transfer learning architecture with pre-trained convolutional backbone for COVID-19 chest X-ray classification

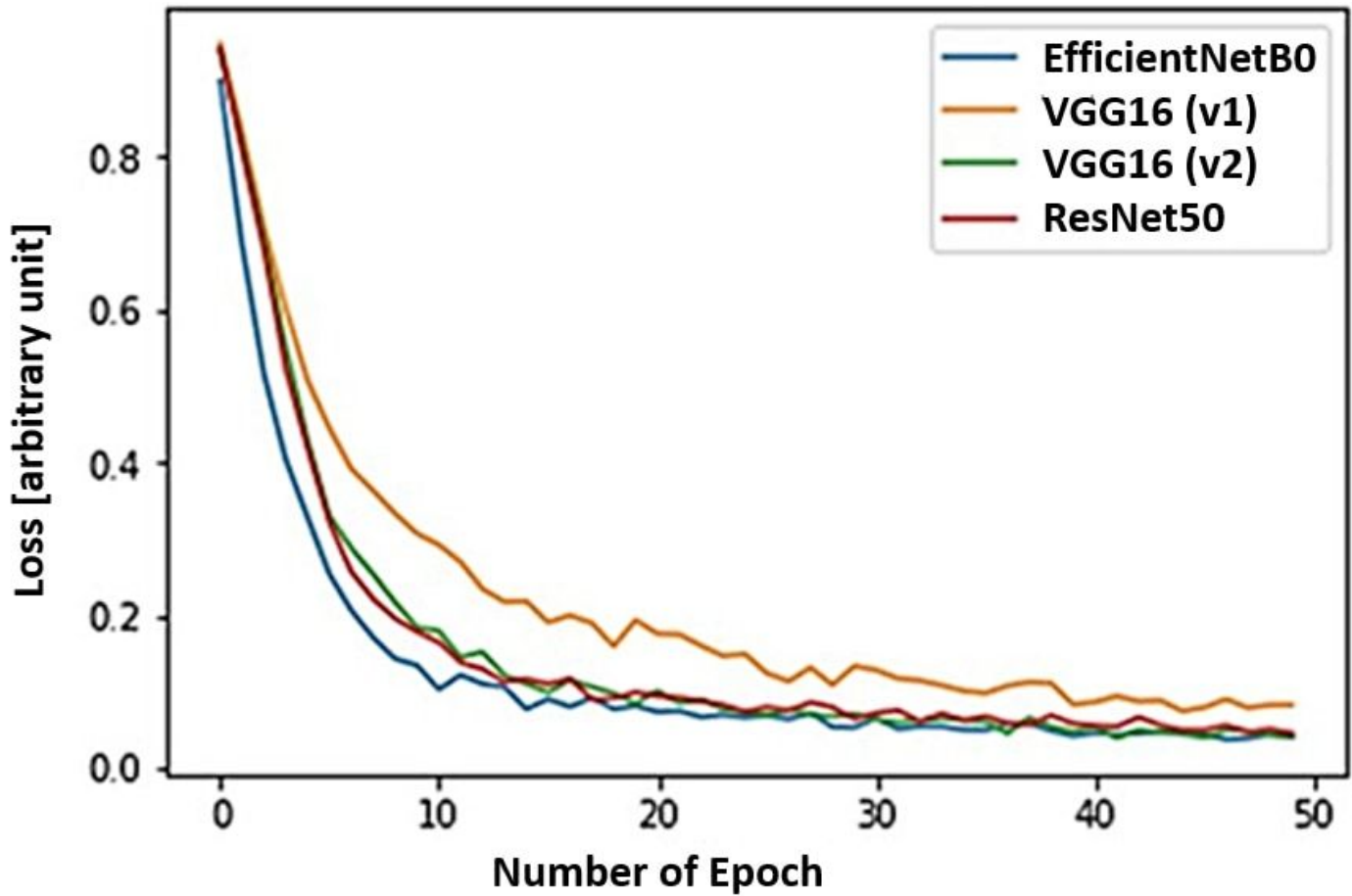


Figure 4

Comparative loss function on the training dataset

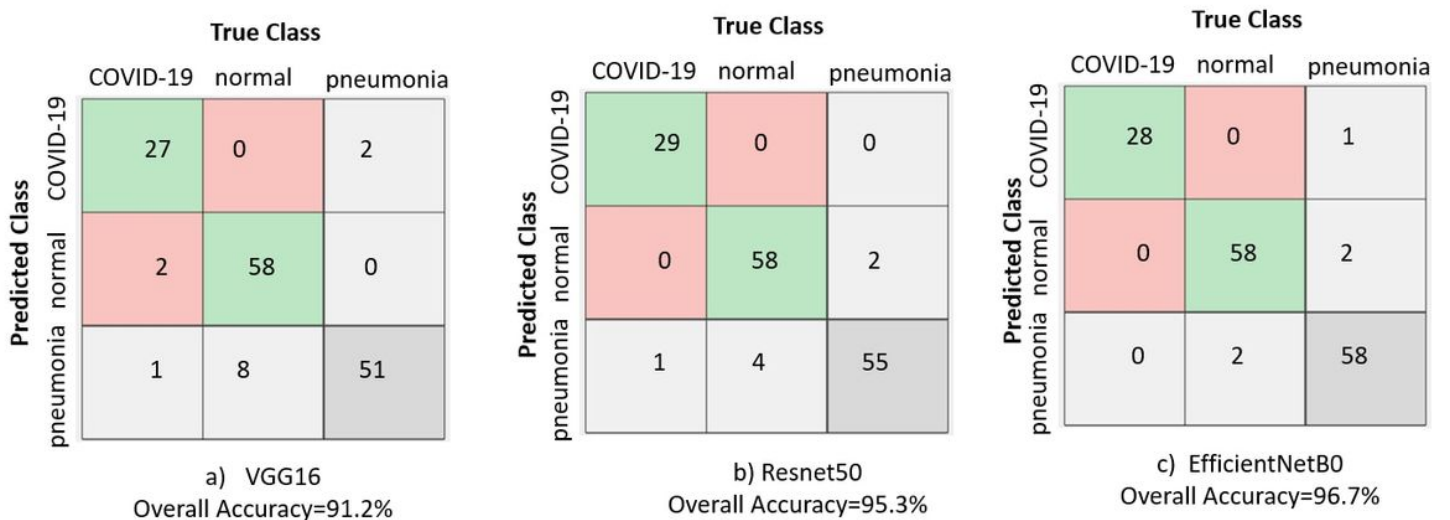


Figure 5

Confusion matrix and overall accuracy of three backbone models used in this research

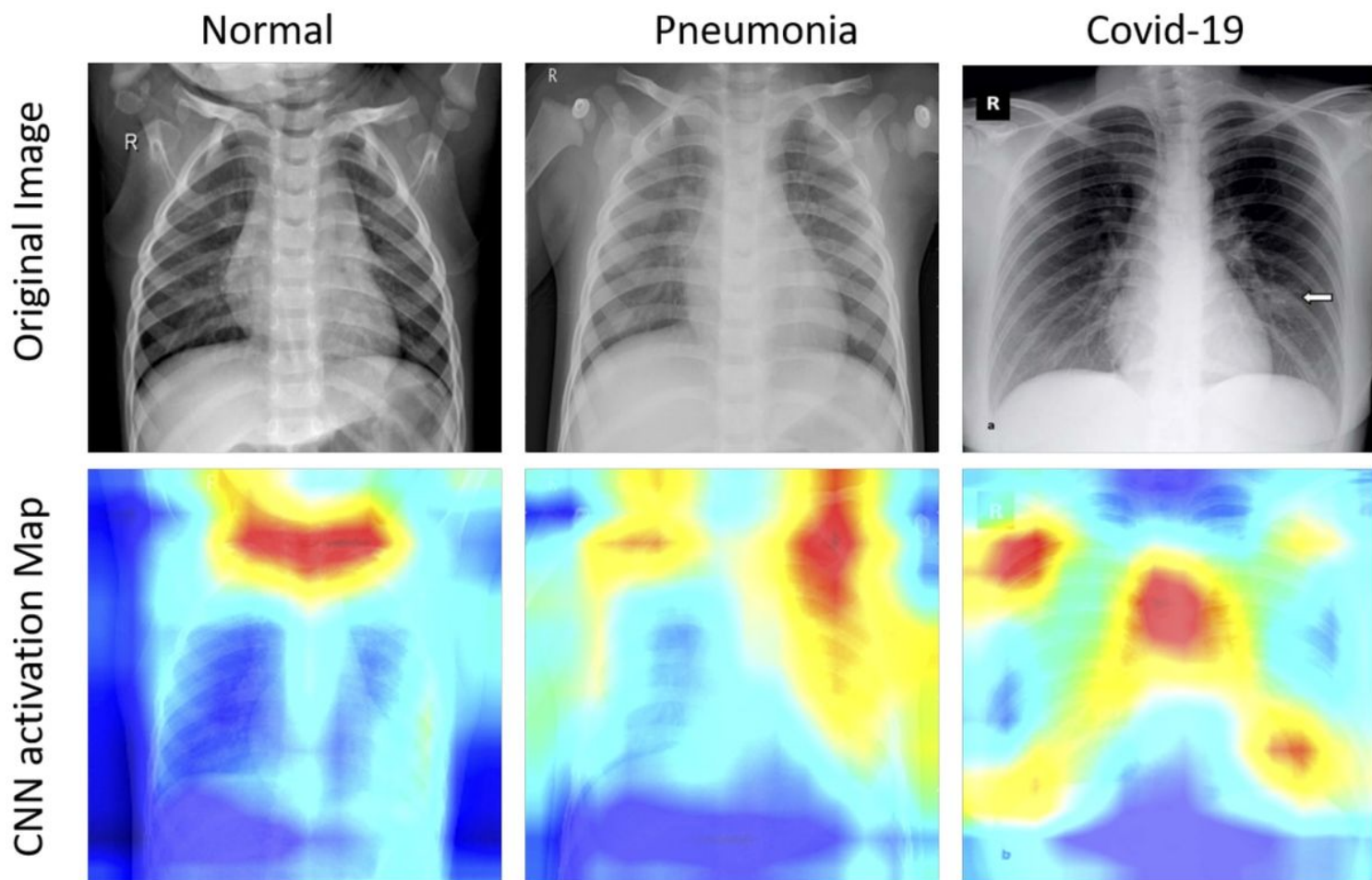


Figure 6

Activation map visualization for the three classes under consideration. The First column presents a healthy chest x-ray sample, the second, from a patient with pneumonia, and the third one, from a patient with COVID-19, visualizing affected regions in lungs.

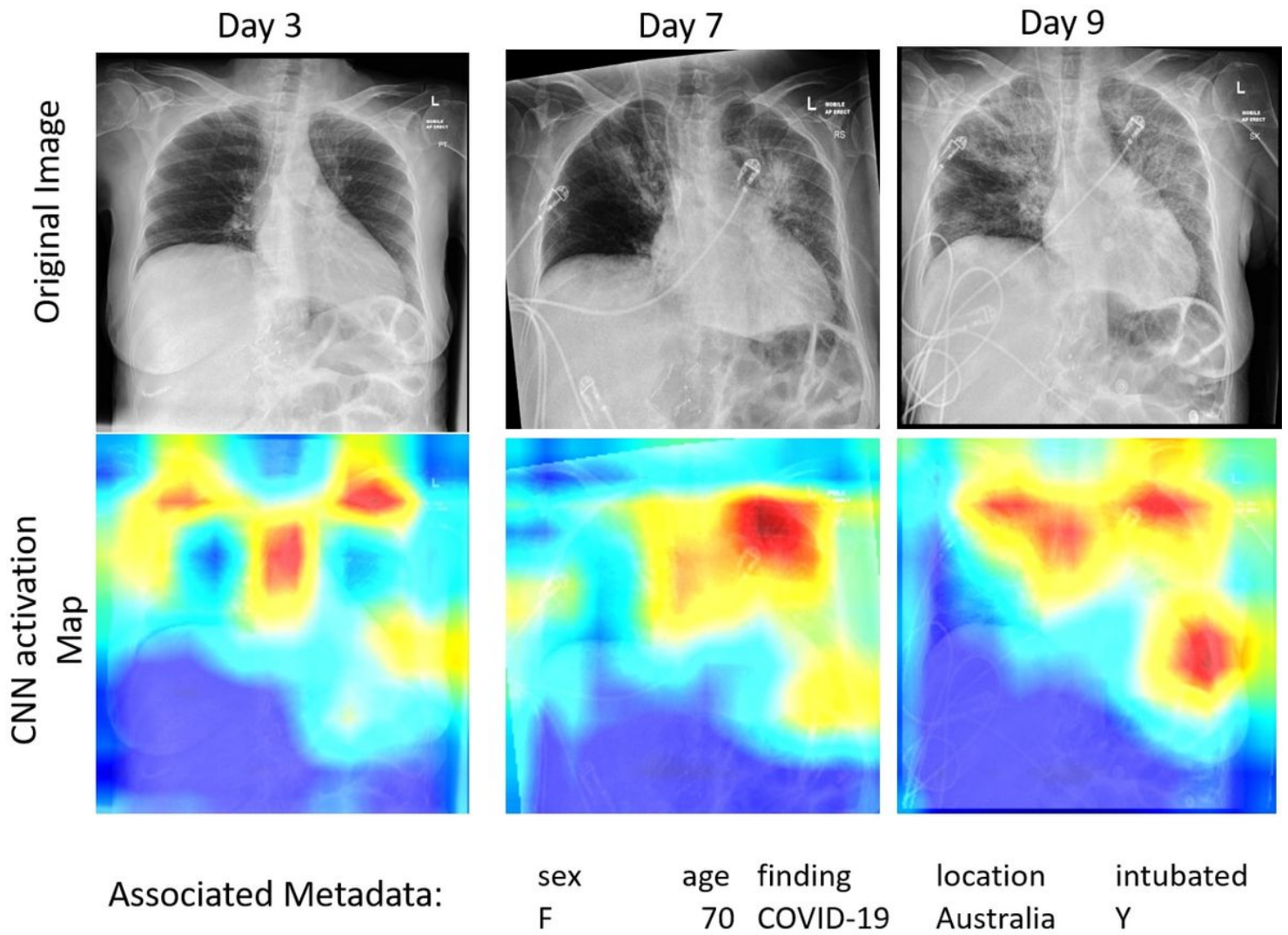


Figure 7

Activation map visualization for a patient's x-ray in ICU-care at day 3, 7 and 9, visualising increased inflammation indicating disease severity.