CTU Journal of Innovation and Sustainable Development
ISSN 2588-1418 | e-ISSN 2815-6412

DOI:10.22144/ctujoisd.2025.053

A hybrid deep learning approach for detecting lung abnormalities from chest X-ray images

Viet Dung Nguyen[1], Van Huan Vu[2], Duc Lam Le[3], Huan Vu[4], and Ngoc Dung Bui[5*]

[1]*School of Electrical and Electronic Engineering, Hanoi University of Science and Technology, Viet Nam*

[2]*Faculty of Information Technology, Hanoi University of Natural Resources and Environment, Viet Nam*

[3]*Vinschool Times City, Viet Nam*

[4]*College of Technology, National Economics University, Viet Nam*

[5]*Faculty of Information Technology, University of Transport and Communications, Viet Nam*

**Corresponding author (dnbui@utc.edu.vn)*

Article info.

Received 11 Jul 2025
Revised 8 Aug 2025
Accepted 1 Oct 2025

Keywords

Chest X-ray, deep learning, lung abnormality

ABSTRACT

This paper proposes a hybrid deep learning model for lung abnormalities detection using X-ray images. To improve the performance and accuracy of the model, we use the transfer learning technique with two pre-trained models VGG16 and DenseNet121. Moreover, to extract deeply the feature of lung abnormal, frontal and lateral views of X-ray images have been trained using ensemble technique. The features extracted by these two models will be combined and passed to the classification layer. The experimental results on three datasets demonstrate the effectiveness of the proposed model, which outperforms the individual performance of the two base models, achieving a higher accuracy rate of 89%. Furthermore, in comparative assessments against several alternative models and datasets from previous research, our method demonstrates its efficiency, boasting an impressive AUC value of 0.95. These results underscore the promise of our approach in advancing the accuracy and effectiveness of lung abnormality detection in chest X-ray images.

1. INTRODUCTION

Early and accurate detection of lung abnormalities in chest X-rays is paramount for timely diagnosis and effective patient management. Timely intervention can significantly improve treatment outcomes and patient prognoses. However, traditional methods often rely on visual assessment by radiologists, making them susceptible to inter-observer variability and human error. Additionally, manual methods can be time-consuming and subjective, impacting diagnostic efficiency and potentially delaying essential treatment. Although some lung abnormalities can be seen visually on X-ray images, such as lung opacities, lung consolidation, atelectasis, bronchial wall thickening, etc., diagnostic methods are based on deep learning models have been proposed to help detect lung diseases early with high accuracy to reduce the overload of doctors. In recent years, deep learning has emerged as a powerful tool for medical image analysis, offering promising solutions for automated abnormality detection. These models can automatically learn complex patterns from large datasets, potentially overcoming the limitations of traditional methods. By leveraging the ability of deep learning to extract subtle features and identify nuanced patterns, we can potentially achieve more accurate and consistent detection of lung abnormalities in X-ray images.

A novel semi-supervised learning algorithm for classifying lung abnormalities in X-ray images based on a holistic philosophy (Ioannis et al., 2019). A multi-label DCNN classification model, using fine-tuning pre-trained networks (e.g., AlexNet, GoogLeNet, VGG16, and ResNet50) by removing fully connected layers and the final classification layer while adding a transition layer, pooling layer, prediction layer, and activation function to detect and segment the lung disease's location (Xiaosong et al., 2017). Several deep learning models, ResNet and ReCoNet, a lightweight model with fewer parameters than modern models but achieved a remarkable 98% accuracy for detecting COVID-19 patient abnormalities in chest X-ray images (Sabbir et al., 2020; Farooq & Hafeez, 2020). A generative CheXNet is a convolutional neural network with 121 layers trained on the ChestX-ray14 dataset containing over 100,000 frontal chest X-ray images with 14 disease labels, which has proposed for detecting pneumonia from chest X-ray images achieved higher accuracy than specialized radiologists (Pranav et al., 2017). Various mod-els DenseNet121, InceptionResNetV2, ResNet152V2 and MobiNet have been employed to classify various lung disease signs in X-ray images (Rahman et al., 2020; Mundher et al., 2021; Vingroup, 2020).

Aiming to overcome the limitations of single-model approaches, this paper proposes a hybrid deep learning model that cooperates capabilities of different models to enhance accuracy in identifying lung abnormalities on chest X-ray images. Our proposed hybrid model is introduced in Section 2. Experimental results are depicted in Section 3, while Section 4 is the conclusion.

2. MATERIALS AND METHOD

As stated in Section 1, we aim to integrate capabilities of different models, especially VGG16 and DenseNet121 to improve lung disease classification accuracy. While VGG possesses feature extraction ability, DenseNet has dense connectivity. Combining them potentially improves accuracy in identifying lung abnormalities on chest X-ray images.

2.1. VGG16

VGG16, invented in 2014, contributed significantly to improving the accuracy of previous models like AlexNet, LeNet, and others (Karen & Andrew, 2014). In terms of architecture, VGG16 introduced several enhancements, including 13 two-dimensional convolutional layers, convolutional blocks, stacking multiple CNN layers and max-pooling layers instead of interleaving just one CNN layer and max-pooling layer, and using ReLU after each CONV layer. VGG16 exclusively employed small 3x3 filters to reduce the number of model parameters, leading to better computational efficiency.

In this paper, we keep the convolutional blocks unchanged and modify the final classification layer, using the convolutional blocks as a feature extraction tool for the model. The features are then classified with the desired output results. The features extracted from the images are passed through a new fully connected layer that we built, consisting of a Flatten layer, a Dense layer with 128 outputs, and finally, a sigmoid classifier. Figure 1 provides a visualization of the VGG16 model used in the paper.

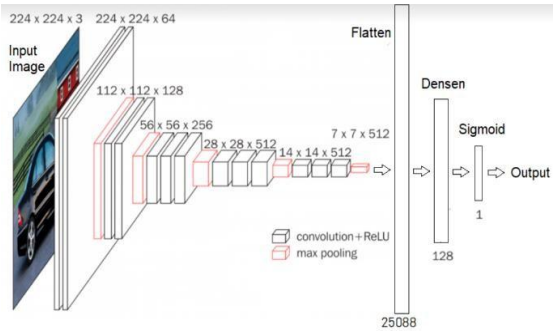

Figure 1. VGG16 model structure

(*Source: Karen & Andrew, 2014*)

2.2. DenseNet121

DenseNet is a convolutional neural network (CNN) architecture for image recognition tasks, first introduced in 2017 (Gao et al., 2017). This architecture stands out in optimizing the flow of information from previous layers to subsequent layers. Instead of connecting only neighboring layers, DenseNet connects each layer to all the subsequent layers. This enhances information flow between layers and encourages feature reuse. In this paper, we improve and utilize the DenseNet121 model for training and detecting lung abnormalities. The convolutional layers of DenseNet121 are used to extract feature maps with a size of 7x7x1024. Average pooling is applied to the feature map to create a 1024-dimensional vector, and then the fully connected layer is modified to have 512 hidden nodes with a ReLU activation function to reduce the number of parameters and computational complexity in the model. The final layer uses a

sigmoid function to compute the output probability score in a single node. The optimization function employs "Adam", and the loss function is "binary cross-entropy" with a learning rate of 1e-5. The number of hidden nodes in the fully connected layer is selected after experimentation with different values. Figure 2 illustrates the structure of DenseNet121 in the paper.

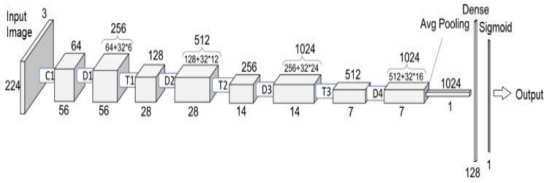

Figure 2. DenseNet model structure

(Gao et al., 2017)

2.3. Proposed method

By employing the strengths of the VGG and DenseNet architectures, we propose a novel hybrid deep learning approach for detecting lung abnormalities in chest X-ray images by combining the VGG16 and DenseNet architectures, each generating distinct features prior to the fully connected layers (Figure 3). VGG16, with its deep convolutional layers, captures essential low- and mid-level features from the images, while DenseNet enhances feature propagation and re-use through its densely connected network structure, enabling the model to learn more intricate and detailed representations. By concatenating the feature maps from both networks before passing them through the fully connected layers, we leverage the complementary strengths of both models, allowing for a more robust and comprehensive feature representation.

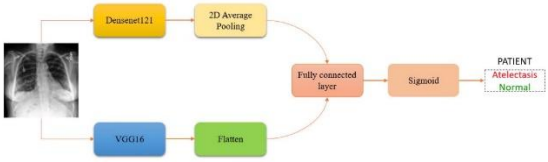

Figure 3. Proposed model structure

The feature map after passing through the VGG16 model will be flattened to become a vector with a size of 25088, while the DenseNet feature vector, after passing through the 2D average pooling layer for feature extraction, will be of size 1024. These vectors will be concatenated after passing through the merging layer of the two models, resulting in a new vector of size 26112. The fully connected layer has 128 hidden nodes with a ReLU activation function. Finally, a sigmoid layer is used. The optimization function "Adam" and the loss function "binary cross-entropy" are applied with a learning rate of 1e-5. This hybrid architecture aims to improve the accuracy and sensitivity of abnormality detection, capturing both broad patterns and subtle abnormalities in the lung region that may be missed by a single model.

3. RESULTS AND DISCUSSION

3.1. Dataset

In this paper, we used the NIH, CheXpert and VinBigData datasets. The NIH dataset includes 112,120 X-ray images with disease labels from 30,805 different patients. To generate these labels, the authors used natural language processing to extract classification information from relevant medical reports. These labels are expected to achieve an accuracy of over 90% and are suitable for weakly supervised learning. CheXpert included X-ray images from 65240 patients. For CheXpert dataset (Figure 4), we employed 24000 images for training and 4000 images for testing. Furthermore, our focus was primarily on the Effusion disease category, the largest number of images within the dataset. Figure 4 illustrates the samples from this dataset. The VinBigData dataset consists of 15,000 chest X-ray images collected and labeled in Vietnam, then evaluated by comparing with doctors' readings on a test dataset of 3,000 additional images.

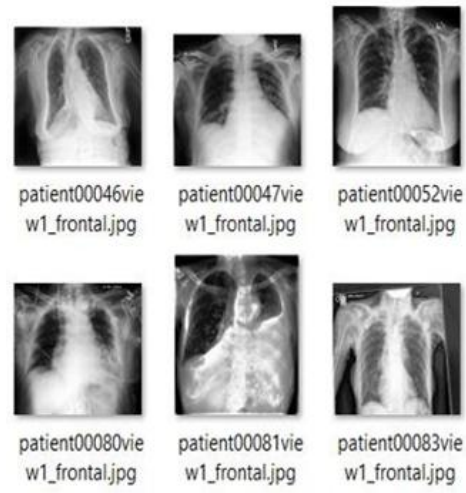

Figure 4. Example images from CheXpert dataset

3.2. Results

NIH dataset

Our proposed model has been evaluated against 2 individual models VGG16, DenseNet121 on different lung diseases such as atelectasis and infiltrating. The comparison results on Atelectasis classification are shown in Table 1.

Table 1. Models comparison on NIH dataset

Model	Accuracy	Loss	AUC
VGG16	74.47%	0.5648	0.8277
DenseNet121	75.20%	0.5191	0.8375
Our model	**76.33%**	**0.5047**	**0.8522**

Table 2. AUC comparision

Pathology	Xiaosong et al., 2017	Li et al., 2017	Pranav et al., 2017	Proposed
Atelectasis	0.716	0.722	0.8094	**0.8522**
Infiltration	0.609	0.695	0.7345	**0.8185**

The highest accuracy was achieved by combining VGG16 and DenseNet121, with this result obtained after just 2 epochs, whereas VGG16 required 3 epochs and DenseNet121 required 4 epochs. This can be attributed to the complementary strengths of the two models: VGG16 excels in image classification tasks, while DenseNet121 effectively handles large datasets and reduces overfitting. Additionally, the ensemble model outperformed both VGG16 and DenseNet121 in terms of accuracy and AUC score for detecting infiltration. Our proposed model is also compared to 3 well-known models by authors (Xiaosong et al., 2017; Li et al., 2017; Pranav et al., 2017). Results in Table 3 indicate that our model is much better than other works in Atelectasis as well as Infiltration classification.

CheXpert database

As the CheXpert composes of frontal and lateral X-ray images of patients therefore we experiment the ability of VGG16 and DenseNet121 in classifying Pleural effusion using just frontal or lateral X-ray images. The Pleural effusion classification accuracies are given in Table 3.

Table 3 indicates that DenseNet121 manifests efficiency a little bit higher than that of VGG14 in both views. Therefore, we have decided to use DenseNet121 for our ensemble model as shown in Figure 5.

Table 3. Pleural effusion classification accuracies

Pathology	Frontal view	Lateral view
VGG16	0.8738	0.8896
DenseNet121	0.8755	0.8950

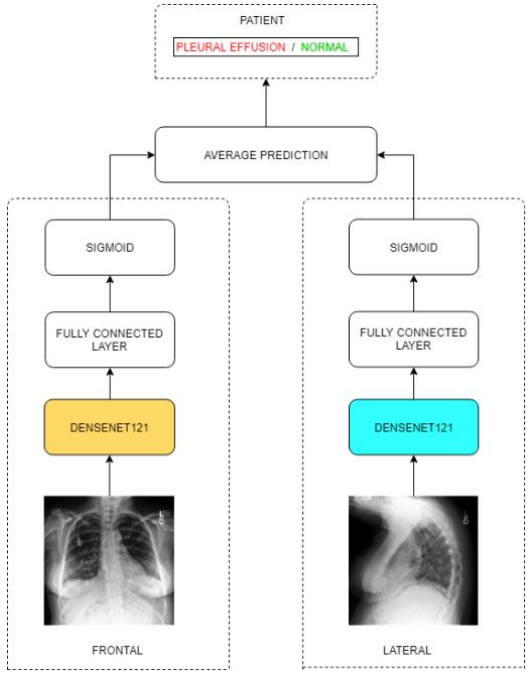

Figure 5. Ensemble model for CheXpert X-ray images

Table 4. Ensemble model vs Single view model

No. of patients	Model	Accuracy
200	Ensemble	0.89
	Frontal	0.8
550	Ensemble	0.882
	Frontal	0.809
2000	Ensemble	0.855
	Frontal	0.79

The number of nodes in each fully connected layer is set as 512. Results shown in Table 4 prove that the ensemble model archives higher performance that single view model no matter how patients may involve in the experiment.

We compared our model to Jeremy Irvin's model for detecting pleural effusion. Jeremy Irvin's model is a modified version of DenseNet121 architecture. Table 5 presents a comparison of the AUC scores between our model and Jeremy Irvin's model.

Table 5. Proposed model and Jeremy Irvin's model comparison

Pathology	Jeremy et al. (2019) (U-Ignore)	Jeremy et al. (2019) (U-Zeros)	Proposed
Pleural effusion	0.928	0.931	**0.9531**

According to Table 5, the ROC score of the proposed model is higher than Jeremy Irvin's model. This result could be attributed to differences in the construction methods of the two models. There are several key differences between the two methods: Jeremy Irvin's method used two types of images (frontal and lateral views) for training and testing, whereas we trained on frontal view images only. In (Jeremy et al., 2019), data labels were divided into "disease" and "the rest," whereas we divided them into two distinct labels: "Normal" and a "Disease" label. These two labels can have significant differences in pathological features, and there would be no overlap between images with these labels. The ratio of the number of training and testing images was different. We used a 6:1 ratio for training and testing, while (Jeremy et al., 2019) used a 7:1 ratio.

For the VinBigData dataset, we integrate it with the CheXpert dataset for training and evaluation. The data from this dataset indicates that a single image may have multiple disease labels, potentially causing overlaps during training. However, this also increases the number of training images, which can help improve the model's accuracy. We propose three scenarios for evaluating our model. In all three scenarios, the disease labels will be No Disease and Effusion, ensuring alignment with the objectives of the CheXpert dataset.

Scenario 1: train on the CheXpert dataset and test on the VinBigData dataset. In this scenario, the CheXpert dataset is used for training at a 6:1 ratio, comprising 30,000 training images and 5,000 testing images, and ensuring an equal number of images for both labels, No Disease and Effusion, in each set. For model training, the proposed ensemble model has been applied to the NIH dataset (NIH dataset, 2020)), where the integrated layers of VGG16 and DenseNet121 extract features independently. When combined in the classification layer, this approach achieves higher accuracy with fewer epochs. The feature map from VGG16 is flattened into a 25,088-dimensional vector, while DenseNet121, after passing through a 2D average pooling layer to extract key features, produces a 1,024-dimensional vector. These vectors are merged after passing through the ensemble layer of both models, resulting in a 26,112-dimensional vector. The fully connected layer consists of 128 hidden units with a ReLU activation function, followed by a sigmoid layer to compute probability scores. The model is optimized using the "Adam" optimizer and the "binary cross-entropy" loss function, with a learning rate of 1e-5. The performance of the model is shown in Figure 5.

Figure 6 shows that the model achieves the highest accuracy of 0.8738 after 2 epochs. After training the model, we tested it on 100, 200, and 500 images from the VinBigData dataset, achieving the results shown in Table 6.

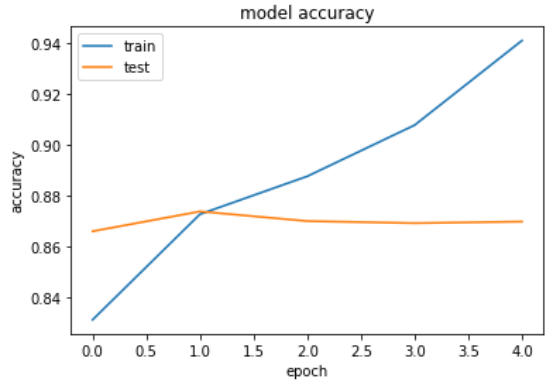

Figure 6. Accuracy of scenario 1

Table 6. Testing accuracy on the VinBigData dataset

Image number	Normal	Effusion
100	0.95	0.56
200	0.95	0.61
500	0.938	0.572

When comparing the accuracy of our model with the frontal and lateral view models trained on the CheXpert dataset, we found that it did not demonstrate a significant improvement over the other two models, both of which were trained using DenseNet121 architecture. In fact, its accuracy was lower than that of the lateral view model. This lower accuracy may be due to our approach of training the model using images from both views combined, rather than separately, as was done for the lateral view model. For this reason, in the previous section, we did not use our proposed model for training on the CheXpert dataset. When evaluating the model on the VinBigData dataset, we observed that predictions for the No Disease label were significantly more accurate than those for the Effusion label. This can be attributed to the fact that

each image in the DICOM dataset may have multiple disease labels. As these pathological conditions often exhibit similar features on X-ray images, distinguishing between them is challenging. Furthermore, employing a single-output classification approach may have further contributed to the reduced accuracy of the test results.

Scenario 2: Train using the VinBigData dataset and test on the CheXpert dataset. In contrast with the first scenario, this scenario uses the VinBigData dataset for training while keeping the model parameters, architecture, and labels identical to those in the first scenario. The ratio of training to validation images remains 6:1, with 1,680 training images and 280 validation images. Once the model weights are obtained, the evaluation will be performed using images from the CheXpert dataset. The performance of the model is shown in Figure 7. The maximum accuracy is 0.9393 after the 16 epochs.

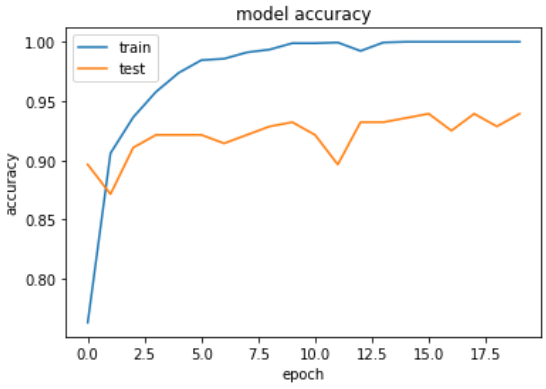

Figure 7. Accuracy of scenario 2

Table 7. Testing accuracy on the CheXpert dataset

Image number	Normal	Effusion
100	47%	90%
200	45.5%	86%
500	46.8%	88%

Similar to scenario one, we tested it on 100, 200, and 500 images from the CheXpert dataset, achieving the results shown in Table 7. As shown in Table 7, after training the model and evaluating it on the CheXpert dataset, contrary to the first scenario, the Effusion label produced significantly better results than the normal label.

Scenario 3: Combine both datasets and use the merged data for both training and testing. In this scenario, we will combine images from both datasets to create a new, balanced dataset, ensuring an equal number of images for each label from both datasets. The training set will consist of 3,360 images, with 1,680 images from CheXpert and 1,680 from VinBigData, while the testing set will follow the same approach with 560 images. Each label is evenly distributed across both the training and testing sets. After splitting the data, we will retain 10% of the remaining Effusion-labeled images for post-training evaluation, maintaining a 1:1 ratio with No Disease images from the CheXpert dataset. The model parameters and architecture will remain unchanged from the previous two scenarios. The performance of the model is shown in Figure 8.

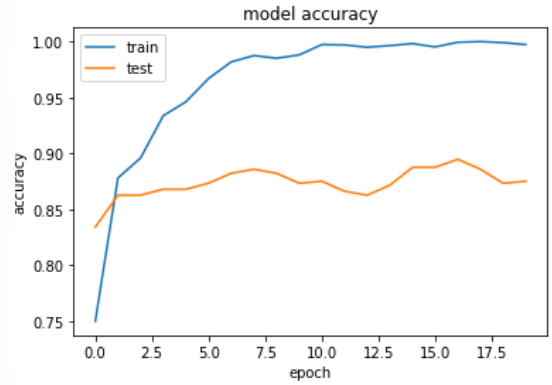

Figure 8. Accuracy of scenario 3

Table 8. Testing accuracy on the two datasets

Dataset	Normal	Effusion
CheXpert	84.37%	73.43%
Dicom	98.43%	84.61%

Figure 8 shows that the model achieved the highest accuracy of 0.8875 after 20 epochs. For testing, we will use 10% of the images after splitting them into training and testing sets. The comparison of the two datasets is shown in Table 5. In this scenario, unlike the previous two, there is no significant discrepancy in the test results between the two output labels. Additionally, the model performed better on the VinBigData dataset compared to the CheXpert dataset.

4. CONCLUSION

This paper proposes a method that combines two popular deep learning models, VGG16 and DenseNet121, for detecting lung abnormalities through chest X-ray images. The integration of VGG16 and DenseNet not only enhances the network's ability to extract diverse features but also

addresses the limitations of individual models. VGG16 excels in capturing basic image patterns, while DenseNet's dense connectivity ensures better feature reuse and gradient flow, crucial for detecting fine-grained details in medical images. The experimental results show that the accuracy of the proposed model with the transfer learning technique is higher than that of the individual model. Additionally, the ensemble technique increases accuracy by combining features from frontal and lateral views of X-ray images. In the future, we aim to further develop and improve the deep learning model as well as use more datasets to achieve higher accuracy and effectively apply it for early detection and diagnosis of lung pathologies.

CONFLICT OF INTEREST

We have no conflicts of interest to disclose.

REFERENCES

Farooq, M., & Hafeez A. (2020). COVID-resnet: A deep learning framework for screening of COVID19 from radiographs. *ArXiv*, abs/2003.14395.

Gao, H., Zhuang, L., Laurens van der, M., Kilian, Q. W. (2017). Densely connected convolutional networks. *CVPR*, arXiv:1608.06993.

Ioannis, L., Andreas, K., Panagiotis, P. (2019). Detecting lung abnormalities from X-rays using an improved SSL algorithm. *Electronic Notes in Theoretical Computer Science*, *343*, 19-33.

Jeremy, I., Pranav, R., Michael, K., Yifan, Y., Silviana, C., Chri,s C., Henrik, M., Behzad, H., Robyn, B., Katie, S., Jayne, S., David, A. Mong, Safwan, S. H., Jesse, K. S., Ricky, J., David, B. L., Curtis, P. L., Bhavik, N. P., Matthew, P. L., Andrew, Y. N. (2019). CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison. *Proceedings of the 33rd AAAI Conference on Artificial Intelligence/31st IAAI Conference/9th AAAI Symposium on Educational Advances in Artificial Intelligence*, *33*, 590–597.

Karen, S., Andrew, Z. (2014). Very deep convolutional networks for large-scale image recognition. *CVPR*, arXiv:1409.1556.

Li, Y., Eric, P., Dmitry, D., Ben, C., Devon, B., Kevin, L. (2017) Learning to diagnose from scratch by exploiting dependencies among labels. *CVPR*, arXiv:1701.10501

Mundher, M. T., Ningbo, Z., Talal, A. A., Mohammed, A., Asaad, S. H., Modhi, L. M. (2021). KL- MOB: Automated COVID-19 recognition using a novel approach based on image enhancement and a modified MobileNet CNN. *PeerJ Computer Science*, *7*, e694. https://peerj.com/articles/cs-694

NIH dataset. (2020). *CXR8.* https://nihcc.app.box.com/v/ChestXray-NIHCC

Pranav, R., Jeremy, I., Kaylie, Z., Matthew, P. L., Andrew, Y. Ng. (2017), CheXNet: Radiologist- level pneumonia detection on chest X-rays with deep learning. *ArXiv*, 1-7. https://arxiv.org/abs/1711.05225

Rahman, T., Chowdhury, M., Khandakar, A., Islam, K., Islam, K., Mahbub, Z. (2020). Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray. *Applied Sciences*, *10*, 3233.

Sabbir, A., Moi, H. Y., Maxine, T., Md.Kamrul, H. (2020). ReCoNet: multi-level preprocessing of Chest X-rays for COVID-19 detection using convolutional neural networks. *Medrxiv*, 1-9. https://doi.org/10.1101/2020.07.11.20149112

VinBigData. (2020). *Chest X-ray abnormalities detection - automatically localize and classify thoracic abnormalities from chest radiographs*. https://www.kaggle.com/competitions/vinbigdata-chest-xray-abnormalities-detection

Xiaosong, W., Yifan, P., Le L., Zhiyong, L., Mohammadhadi, B., & Ronald, M. S. (2017). Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. *IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2097-2106).