THORACIC DISEASES CLASSIFICATION

Karthikeyan S, Siddharth Iyyapan G, Aravind Kumar N and Dheeraj Chintamreddy

Abstract

In this research paper, we explore the ChestX-ray14 dataset each annotated with fourteen disease labels.Our approach involves enhancing the inherent capabilities of these models by introducing additional dense layers and optimizing their performance through modifications to the choice of optimizers. As we navigate through the complexities of medical image analysis, our primary goal is to improve the accuracy of disease prediction.We leverage two powerful convolutional neural network architectures, InceptionV3 and ResNet-50, for multi-label classification of chest X-ray images. Unlike previous approaches that primarily rely on image-level labels, our models are trained on the rich dataset with detailed annotations for each image, allowing for a more nuanced understanding of thoracic diseases.The models InceptionV3 and ResNet-50 enable us to achieve a comprehensive analysis of chest X-rays, providing valuable insights into the detection of commonly occurring thoracic diseases. While our preliminary results demonstrate promise in terms of quantitative performance, we acknowledge the persistent challenges in fully automating the process of "reading chest X-rays." The intricacies of recognizing and precisely locating common disease patterns necessitate ongoing refinement of deep learning models for enhanced accuracy in computer-aided diagnosis.

Introduction

The intersection of healthcare and technology has paved the way for transformative advancements in medical diagnostics and treatment. Within this dynamic landscape, the field of medical image analysis has emerged as a cornerstone, offering clinicians unprecedented insights into the intricacies of diseases, particularly within the realm of thoracic disorders. The availability of vast datasets, coupled with sophisticated machine learning methodologies, has propelled research endeavors aimed at enhancing diagnostic precision and expediting patient care.

Among the repositories that encapsulate the essence of this convergence is the "ChestX-ray14" dataset to mine disease classifications from associated radiological reports has endowed this dataset with a rich tapestry of information, encompassing diverse disease classes from Atelectasis to Hernia. This corpus, originating from 30,805 unique patients, not only provides a comprehensive snapshot of thoracic pathology but also lays the foundation for leveraging machine learning to augment diagnostic capabilities.

The labels affixed to each image are anticipated to be of high fidelity, surpassing a 90% accuracy threshold, rendering them suitable for robust applications, particularly in the realm of weakly-supervised learning. However, the dataset is not without its intricacies; the absence of publicly available radiology reports adds an additional layer of complexity, emphasizing the importance of methodological transparency and collaborative efforts within the research community.

This research endeavor navigates the complexities of the "ChestX-ray14" dataset with the aim of advancing the state-of-the-art in chest X-ray image classification. Our focal point lies in the application of deep learning models, specifically ResNet50 and InceptionV3, to perform multi-label classification across the 14 disease classes. To harness the full potential of these models, we introduce extra dense layers and strategically modify optimizers to enhance their accuracy in discerning complex patterns within the chest X-ray images.

As we embark on this scientific journey, we acknowledge the dataset's limitations, including the potential presence of erroneous labels and the scarcity of disease region bounding boxes. The inherent challenges presented by the dataset fuel our commitment to innovation, pushing the boundaries of what is achievable in medical image analysis.

This paper unfolds by offering a detailed exploration of the dataset, delineating its unique characteristics, limitations, and the intricate methodology behind label creation. The architecture of our deep learning models, underscored by the incorporation of extra dense layers and optimizer modifications, is elaborated upon. Results from extensive experiments, including quantitative assessments and performance metrics, are presented to validate the efficacy of our proposed enhancements.

In sharing our insights, we aspire to contribute not only to the refinement of diagnostic tools but also to the broader discourse on the fusion of medical sciences and artificial intelligence. By navigating the complex interplay between data, technology, and healthcare, we envision a future where chest X-ray analysis becomes not just a diagnostic tool but a catalyst for proactive and personalized patient care.

Materials and Methods

Dataset Description:

The NIH ChestX-ray14 dataset stands as a pivotal resource in the realm of medical imaging, comprising a vast

collection of 112,120 high-resolution chest X-ray images. These images, sourced from 30,805 unique patients, are meticulously annotated with disease labels, creating a diverse and comprehensive dataset for multi-label classification. The dataset's richness is a result of employing Natural Language Processing (NLP) to extract disease classifications from associated radiological reports, ensuring a nuanced representation of thoracic pathology.The 15 classes within the dataset encapsulate a spectrum of thoracic conditions, from common ailments such as Atelectasis, Pneumonia, and Pleural Thickening to more intricate diagnoses like Nodule Mass and Cardiomegaly. One notable feature is the inclusion of a "No Finding" class, providing a reference for images devoid of discernible abnormalities.

Diseases and their causes:

Atelectasis involves the partial or complete collapse of lung tissue, often caused by airway blockages or compression. Consolidation signifies the solidification of lung tissue due to fluid or inflammatory exudate, commonly associated with conditions like pneumonia.

Infiltration refers to the abnormal presence of substances within lung tissue, such as cells or blood, indicating various pathological processes. Pneumothorax results from air entering the space between the lung and chest wall, leading to lung collapse. It can occur spontaneously or due to trauma. Edema is the abnormal accumulation of fluid in tissues. Pulmonary edema involves fluid buildup in the lungs, often associated with heart failure. Emphysema is a chronic lung condition characterized by air sac destruction, making it challenging to exhale air, commonly linked to long-term smoking.

Fibrosis involves the scarring of lung tissue, limiting its functionality. Causes include environmental exposures and autoimmune disorders. Effusion, pleural effusion is the accumulation of fluid in the pleural cavity, usually a consequence of infections, heart failure, or other medical conditions.

Pneumonia is an inflammatory lung condition, often caused by infections, leading to symptoms such as cough, fever, and difficulty breathing. Pleural Thickening is the abnormal thickening of the pleura, the membrane lining the lungs and chest cavity, often associated with inflammation or scarring. Cardiomegaly signifies an enlarged heart, commonly associated with cardiovascular conditions like hypertension or cardiomyopathy. Nodule, a pulmonary nodule is a small, round lesion in the lung. While often benign, it may also indicate early stages of lung cancer. Mass refers to an abnormal lump or growth in the lung tissue.

Evaluation is crucial to determine whether the mass is benign or malignant. Hernia in chest X-rays typically refers to the protrusion of abdominal organs into the chest cavity through an opening in the diaphragm. No Findings, the "No Findings" category denotes the absence of pathological abnormalities in the chest X-ray images, serving as a reference for normal cases.
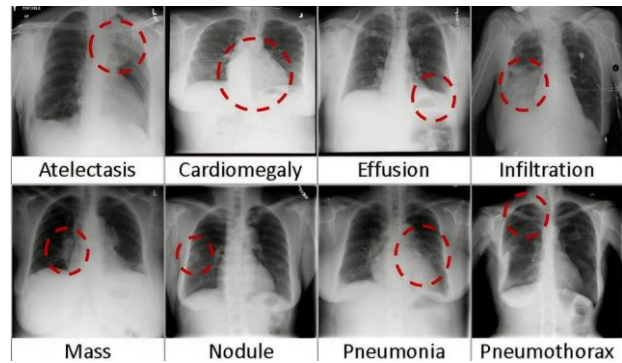

Figure 1. 8 Common thoracic diseases

Data Pre-processing

Data generators, created using the `flow_ from dataframe` method from Keras' `ImageDataGenerator` class, were instrumental in facilitating efficient and dynamic data loading during model training. Specifically, three generators were established for the training, validation, and testing datasets. To enhance dataset diversity and contribute to the robustness of trained models, augmentation techniques such as rotation, zooming, and horizontal flipping were applied to the training dataset using Keras' `ImageDataGenerator`.

The organization of images within the `'./NIH_chest'` directory and the use of DataFrame structures (`X_train` and `X_valid`) played a crucial role in mapping image file paths and associated disease labels for both training and validation datasets. For optimized memory usage during model training, image resizing to a standardized target size of (`img_width, img_height`) and batch size definition were implemented. Given the multi-label nature of the classification task, the `class_mode` parameter in the data generators was set to `'raw'`, allowing the generators to yield arrays of shape (`batch_size, num_classes`) containing raw target values for each sample.

The dataset was further divided into training and validation sets, with respective data generators (`train_ generator` and `validation_generator`) prepared for model training and validation. Additionally, provisions for a test data generator (`test_generator`) were included, although currently commented out, to be employed for evaluating models on an independent test dataset.

Classification:

Resnet 50 Model:

ResNet-50, or Residual network, a deep CNN with 50 layers, is introduced to resolve the vanishing gradient problem. This architecture has many Residual Blocks, and the network uses a skip connection technique and skips some layers during training. The advantage is that if any layer impacts the model performance negatively, it will be skipped in regularization. These residual blocks are stacked together to form ResNets.

When f(Y l) and h(Y l) are identical mappings, it has been shown that the signal can go directly from one unit to another in either direction. Furthermore, both identity mappings achieve a low error rate when used as identity mappings.

Equation for Residual Network is shown in Eq. (1) and (2)

$$Y_l = h(X_l) + \mathcal{F}(X_l, W_l) \qquad (1)$$

$$X_{l+1} = f(Y_l) \qquad (2)$$

Where,

$\mathcal{F}$ = Stacked non-linear layer

f = ReLU activation function

$f(Y_l)$ = Identity mappings

$h(X_l)$ = Identity mappings

Instead of layers learning the underlying mapping, this network allows the network to fit the residual mapping. Here, the ReLU Activation Function is used. The ReLU is a non-linear or piecewise linear function that, if positive, outputs the input directly; else, it returns zero. It is the most often used activation function in neural networks, especially CNN and Multilayer Perceptron's.
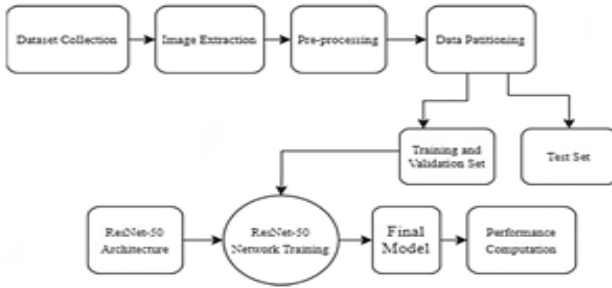

Figure 2. Proposed method block diagram

InceptionV3 Model:

The Inception V3 is a deep learning model based on Convolutional Neural Networks, which is used for image classification. The inception V3 is a superior version of the basic model Inception V1 which was introduced as GoogLeNet in 2014. As the name suggests it was developed by a team at Google. When multiple deep layers of convolutions were used in a model it resulted in the overfitting of the data. To avoid this from happening the inception V1 model uses the idea of using multiple filters of different sizes on the same level. Thus in the inception models instead of having deep layers, we have parallel layers thus making our model wider rather than making it deeper.

Deep learning is the foundation of CNN model for Inception-V3 thoracic chest x-ray images and is essential for improving prediction performance. This system performed certain preparation, including integrating images of a specific size, and employing data augmentation techniques, before fitting the model. After adapting and fine tuning the algorithm, understanding and identifying. A model of loss and accuracy were plotted to show how loss and accuracy change over time. Convolutional neural networks are frequently used for photos categorisation in artificial intelligence. Data convolution, max pooling, and flattened

are maintained of before the inputs is passed through a neural network. It works because weights are configured using different components. Weights are calculated and evaluated after the data have moved through to the hidden levels. The network immediately enters a phase known as back propagation after taking information from the cost function. The input layer weight is then rearranged once more, and the procedure proceeds until it locates the suitable places for weight adjustment here. The block architecture in Figure 1 takes as input a dataset with subsections such as data augmentation, batch normalization, inception module, sigmoid classification. A particularly effective way to do this is through data augmentation. The gap between the sets used for training and validation, as well as any possible future testing sets, will be reduced by the augmented data's representation of a more extensive collection of possible data points for lowering overfitting that has been established. Batch normalisation is a technique for accelerating and stabilising artificial neural networks. Pretrained model construction by rescaling and recentring the inputs of the layers.
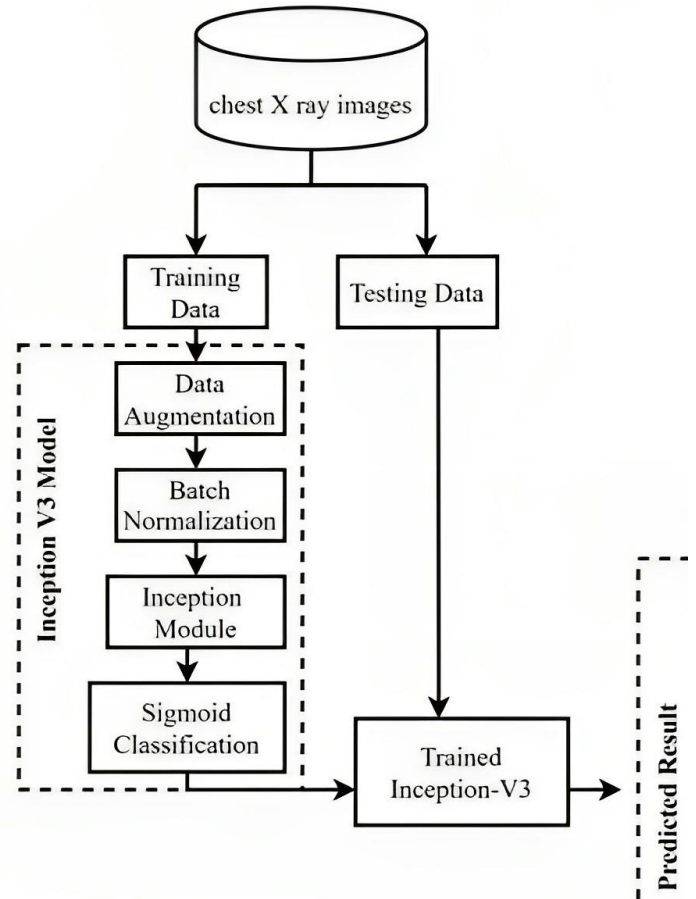

Figure 3. WorkFlow Block Diagram of InceptionV3

Model Architecture:

In the first part of the study, the InceptionV3 architecture pre-trained on the ImageNet dataset served as the foundation for the model. Custom classification layers, including Global Average Pooling, two Dense layers with ReLU activation functions (1024 and 512 units, respectively), and a final Dense layer with sigmoid activation for multi-label

classification, were added. The model was compiled using the Adam optimizer with a learning rate of 0.0001, and binary crossentropy loss was employed for the multi-label classification task. The training process spanned 10 epochs, utilizing both training and validation generators.

In the second part of the study, the ResNet50 architecture, also pre-trained on ImageNet, was adopted as the base model. Following a similar approach to InceptionV3, custom classification layers were appended to the ResNet50 base. The ResNet50 model underwent compilation with the Adam optimizer (learning rate = 0.0001), utilizing binary crossentropy loss. Additional metrics, including accuracy, precision, and recall, were incorporated for evaluation purposes. Similar to the InceptionV3 model, the ResNet50 model underwent training for 10 epochs, making use of both the training and validation generators.

Model Training:

Both InceptionV3 and ResNet50 models were trained using the respective training generators with steps per epoch calculated based on the batch size. The models were validated using the validation generators, and the training process was conducted over 10 epochs.

Model Evaluation:

Training progress and performance metrics were visually assessed through the plotting of loss curves for both training and validation datasets. After training completion, the models were evaluated on a separate test dataset to assess their generalization capabilities. Evaluation metrics included accuracy, precision, and recall.

Save Models:

The trained InceptionV3 and ResNet50 models were saved for future use and deployment.

Visualizing Results:

Visual representation of the training and validation loss curves was generated to facilitate insights into the training dynamics and model convergence.

Sigmoid Function: The sigmoid function is an operational technique that aids a neuron's decision-making capacity. These processes result in a Zero or one. To create a binary output, it perfectly. The result is decided by selecting the individual with the best accuracy value. When it comes to categorisation, this function performs better than the threshold function. For the final dense block, this activation function is frequently used. The sigmoid function's equation is

$$A = \frac{1}{1+e^{-B}}$$

ReLU is activating that employs the assumption of rectification to work. From the beginning through a predetermined point, this function's output remains at 0. The output modifies and rises as the input modifies whenever it crosses or reaches a particular value

Results:

In the results section of our study, we present a detailed analysis of the performance metrics obtained, specifically focusing on per-label Area Under the Curve (AUC) and per-label accuracy. These metrics serve as crucial indicators of the effectiveness of our approach in predicting individual disease labels within the ChestX-ray14 dataset. The per-label AUC values offer insights into the discriminative power of our models for each specific thoracic disease, providing a nuanced understanding of their diagnostic capabilities. Concurrently, the per-label accuracy metrics shed light on the overall precision of our models in correctly classifying individual diseases, considering both true positive and true negative predictions. We aim to highlight the strengths and potential limitations of our deep learning models in the context of multi-label classification for chest X-ray images.

Diseases	AUC(Area under the curve)
Pneumonia	0.5307
Hernia	0.4760
No finding	0.5544
Pneumothorax	0.4624
Nodule	0.4970
Mass	0.4563
Consolidation	0.4703
Edema	0.6053
Effusion	0.4581
Infiltration	0.5481
Emphysema	0.5451
Cardiomegaly	0.5282
Pleural Thickening	0.4991
Fibrosis	0.4570
Atelectasis	0.5049

Table 1. AUC of the InceptionV3 Model

Diseases	Per Label Accuracy
Pneumonia	0.9873
Hernia	0.8639
No finding	0.9953
Pneumothorax	0.9873
Nodule	0.9179
Mass	0.9787
Consolidation	0.9764
Edema	0.9720
Effusion	0.9660
Infiltration	0.9600
Emphysema	0.9560
Cardiomegaly	0.9126
Pleural Thickening	0.6037
Fibrosis	0.9713
Atelectasis	0.9620

Table 2. Per Label Accuracy of the InceptionV3 Model

Diseases	AUC(Area under the curve)
Pneumonia	0.4171
Hernia	0.4494
No finding	0.6692
Pneumothorax	0.5944
Nodule	0.5002
Mass	0.5205
Consolidation	0.4941
Edema	0.5608
Effusion	0.4560
Infiltration	0.4792
Emphysema	0.5185
Cardiomegaly	0.5054
Pleural Thickening	0.4912
Fibrosis	0.5346
Atelectasis	0.4903

Table 3. AUC of the Resnet-50 Model

Diseases	Per Label Accuracy
Pneumonia	0.9873
Hernia	0.8639
No finding	0.9953
Pneumothorax	0.9873
Nodule	0.9179
Mass	0.9787
Consolidation	0.9764
Edema	0.9720
Effusion	0.9660
Infiltration	0.9600
Emphysema	0.9560
Cardiomegaly	0.9126
Pleural Thickening	0.6037
Fibrosis	0.9713
Atelectasis	0.9620

Table 4. Per Label Accuracy of the Resnet-50 Model

InceptionV3 overall AUC : 0.5061
InceptionV3 accuracy : 0.6037
Resnet-50 overall AUC : 0.5121
Resnet-50 accuracy : 0.6037

Conclusion:

In conclusion, our research delves into the exploration of the ChestX-ray14 dataset, leveraging the capabilities of deep learning models, specifically InceptionV3 and ResNet-50, for multi-label classification of chest X-ray images. By enhancing these models with additional dense layers and optimizing their performance through modified optimizers, we aimed to improve the accuracy of disease prediction. Unlike previous approaches, our models were trained on a rich dataset with detailed annotations for each image, providing a more nuanced understanding of thoracic diseases.

The use of InceptionV3 and ResNet-50 allowed for a comprehensive analysis of chest X-rays, offering valuable insights into the detection of commonly occurring thoracic diseases. While our preliminary results show promise in terms of quantitative performance, we acknowledge the persistent challenges in fully automating the process of "reading chest X-rays." The intricate nature of recognizing and precisely locating common disease patterns necessitates ongoing refinement of deep learning models to enhance accuracy in computer-aided diagnosis.

References:

[1] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 2097–2106.

[2] A. I. Aviles-Rivero et al., "GraphXNET chest X-ray classification under extreme minimal supervision," 2019, arXiv:1907.10085. [Online]. Available: https://arxiv.org/abs/1907.10085

[3] J. Cai, L. Lu, A. P. Harrison, X. Shi, P. Chen, and L. Yang, "Iterative attention mining for weakly supervised thoracic disease pattern localization in chest X-rays," in Proc. Int. Conf. Med. Imag. Comput. Comput. Assist. Intervent. (MICCAI). Springer, 2018, pp. 589–598.

[4] Z. Li et al., "Thoracic disease identification and localization with limited supervision," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2018, pp. 8290–8299.

[5] J. Liu, G. Zhao, Y. Fei, M. Zhang, Y. Wang, and Y. Yu, "Align, attend and locate: Chest X-ray diagnosis via contrast induced attention network with limited supervision," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2019, pp. 10632–10641.

[6] P. Rajpurkar et al., "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," 2017, arXiv:1711.05225. [Online]. Available: http://arxiv.org/abs/1711.05225

[7] C. Qin, D. Yao, Y. Shi, and Z. Song, "Computer-aided detection in chest radiography based on artificial intelligence: A survey," BioMed. Eng. OnLine, vol. 17, p. 113, Aug. 2018.

[8] C. Brestel, R. Shadmi, I. Tamir, M. Cohen-Sfaty, and E. Elnekave, "RadBot-CXR: Classification of four clinical finding categories in chest X-ray using deep learning," in Proc. Med. Imag. Deep Learn., 2018, pp. 1–9. [Online]. Available: https://openreview.net/ forum?id=B1tMhcIDM

[9] J. Irvin et al., "CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison," in Proc. AAAI Conf. Artif. Intell., vol. 33, Jul. 2019, pp. 590–597.

[10] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 2921–2929.

[11] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards realtime object detection with region proposal networks," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), 2015, pp. 91–99.

[12] K. Yan, X. Wang, L. Lu, and R. M. Summers, "Deeplesion: Automated mining of large-scale lesion

annotations and universal lesion detection with deep learning," J. Med. Imag., vol. 5, no. 3, 2018, Art. no. 036501.

[13] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 2980–2988.

[14] K. Duan, S. Bai, L. Xie, H. Qi, Q. Huang, and Q. Tian, "CenterNet: Keypoint triplets for object detection," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2019, pp. 6569–6578.

[15] M. Jaderberg et al., "Spatial transformer networks," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 2015, pp. 2017–2025.

[16] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 2117–2125.

[17] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2014, pp. 580–587.

[18] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," Int. J. Comput. Vis., vol. 104, no. 2, pp. 154–171, Sep. 2013.

[19] R. Girshick, "Fast R-CNN," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Dec. 2015, pp. 1440–1448.

[20] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 779–788.

[21] W. Liu et al., "SSD: Single shot multibox detector," in Proc. Eur. Conf. Comput. Vis. (ECCV). Cham, Switzerland: Springer, 2016, pp. 21–37.

[22] H. Law and J. Deng, "CornerNet: Detecting objects as paired keypoints," in Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 734–750.

[23] J. Hung and A. Carpenter, "Applying faster R-CNN for object detection on malaria images," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) Workshops, Jul. 2017, pp. 56–61.

[24] N. M. Noor et al., "Texture-based statistical detection and discrimination of some respiratory diseases using chest radiograph," in Advances in Medical Diagnostic Technology Singapore: Springer, 2014, pp. 75–97.

[25] M. Cicero et al., "Training and validating a deep convolutional neural network for computer-aided detection and classification of abnormalities on frontal chest radiographs," Investigative Radiol., vol. 52, no. 5, pp. 281–287, 2017.