

CovidAID: COVID-19 Detection Using Chest X-Ray

Arpan Mangal¹, Surya Kalia¹, Harish Rajagopal², Krithika Rangarajan^{3,1},
Vinay Namboodiri^{2,4}, Subhashis Banerjee¹, and Chetan Arora¹

¹ Indian Institute of Technology Delhi

² Indian Institute of Technology Kanpur

³ All India Institute of Medical Sciences Delhi

⁴ University of Bath, UK

<https://github.com/arpanmangal/CovidAID>

Abstract. The exponential increase in COVID-19 patients is overwhelming healthcare systems across the world. With limited testing kits, it is impossible for every patient with respiratory illness to be tested using conventional techniques (RT-PCR). The tests also have long turn-around time, and limited sensitivity. Detecting possible COVID-19 infections on Chest X-Ray may help quarantine high risk patients while test results are awaited. X-Ray machines are already available in most healthcare systems, and with most modern X-Ray systems already digitized, there is no transportation time involved for the samples either. In this work we propose the use of chest X-Ray to prioritize the selection of patients for further RT-PCR testing. This may be useful in an inpatient setting where the present systems are struggling to decide whether to keep the patient in the ward along with other patients or isolate them in COVID-19 areas. It would also help in identifying patients with high likelihood of COVID with a false negative RT-PCR who would need repeat testing. Further, we propose the use of modern AI techniques to detect the COVID-19 patients using X-Ray images in an automated manner, particularly in settings where radiologists are not available, and help make the proposed testing technology scalable. We present CovidAID: COVID-19 AI Detector, a novel deep neural network based model to triage patients for appropriate testing. On the publicly available covid-chestxray-dataset [2] dataset, our model gives 90.5% accuracy with 100% sensitivity (recall) for the COVID-19 infection. We significantly improve upon the results of Covid-Net [10] on the same dataset.

Keywords: Deep learning, COVID-19 Detection, Chest X-Ray

1 Introduction

The sudden spike in the number of patients with COVID-19, a new respiratory virus, has put unprecedented load over healthcare systems across the world. In many countries, the healthcare systems have already been overwhelmed. There

are limited kits for diagnosis, limited hospital beds for admission of such patients, limited personal protective equipment (PPE) for healthcare personnel and limited ventilators. It is thus important to differentiate which patients with severe acute respiratory illness (SARI) could have COVID-19 infection in order to efficiently utilize the limited resources. In this work we propose the use of chest X-Ray to detect COVID-19 infection in the patients exhibiting symptoms of SARI. Using our tool one can classify a given X-Ray in one of the four classes: normal, bacterial pneumonia, viral pneumonia, and covid pneumonia. The use of X-Ray has several advantages over conventional diagnostic tests:

1. X-Ray imaging is much more widespread and cost effective than the conventional diagnostic tests.
2. Transfer of digital X-Ray images does not require any transportation from point of acquisition to the point of analysis, thus making the diagnostic process extremely quick.
3. Unlike CT Scans, portable X-Ray machines also enable testing within an isolation ward itself, hence reducing the requirement of additional Personal Protective Equipment (PPE), an extremely scarce and valuable resource in this scenario. It also reduces the risk of hospital acquired infection for the patients.

The main contribution of this work is in proposing a novel deep neural network based model for highly accurate detection of COVID-19 infection from the chest X-Ray images of the patients. Radiographs in the current setting are in most cases interpreted by non-radiologists. Further, given the novelty of the virus, many of the radiologists themselves may not be familiar with all the nuances of the infection, and may be lacking in the adequate expertise to make highly accurate diagnosis. Therefore this automated tool can serve as a guide for those in the forefront of this analysis.

We would like to re-emphasize that we are not proposing the use of the proposed model as alternative to the conventional diagnostic tests for COVID-19 infection, but as a triage tool to determine the suitability of a patient with SARI to undergo the test for COVID-19 infection.

To help accelerate the research in this area, we are releasing our training code and trained models publicly for open access at <https://github.com/arpanmangal/CovidAID>. However, we note that both the model and this report merely captures our current understanding of this rapidly evolving problem, that too on very limited data currently available. We will keep updating the model and this report as we get newer understanding and better results.

2 Related Work

2.1 Pneumonia detection in Chest X-Rays

Various deep learning based approaches have been developed to identify different thoracic diseases, including pneumonia [7, 8, 11, 13]. We choose CheXNet [7] to

build upon, which could detect pneumonia from chest X-Rays at a level exceeding practicing radiologists. CheXNet is trained on ChestX-ray14 [11] (the largest publicly available chest X-ray dataset), gives better performance than previous approaches [11, 13], and has a simpler architecture than later approaches [8].

CheXNet [7] is a 121-layer DenseNet [3] based model trained on the ChestX-ray14 [11] dataset comprising of 112,120 frontal-view chest X-Ray images. The model is trained to classify X-Ray images into 14 different thoracic disease classes, including pneumonia. Given the visual similarity of the input samples, we found this to be the closest pre-trained backbone to develop a model for identifying COVID-19 pneumonia.

2.2 COVID-19 detection in Chest X-Rays

Since the recent sudden surge of COVID-19 infections across the world, many alternative screening approaches have been developed to identify suspected cases of COVID-19. However there are only limited such open-source applications available for use [1, 5, 10] which use chest X-Ray images. Publicly available data on chest X-Rays for COVID-19 are also limited.

COVID-Net [10] is the only approach having an open source and actively-maintained tool which has ability to identify COVID-19 as well as other pneumonia while showing considerable sensitivity for COVID-19 detection. COVID-Net uses a machine-driven design exploration to learn the architecture design starting from initial design prototype and requirements. It takes as input a chest X-Ray image and outputs a prediction among three classes: *Normal*, *Pneumonia* and *COVID-19*. We treat this model as our baseline, comparing our results with it.

3 Method

Given the limited amount of available X-Ray samples (including bacterial and viral pneumonia), it is hard to train a deep neural network from scratch. Hence we use a pre-trained backbone trained on a large dataset. Our approaches use the pre-trained model of CheXNet [7], released by Weng *et al.* [12].

3.1 Problem Formulation and Loss Function

We aim to classify a given frontal-view chest X-Ray image into the following classes: *Normal*, *Bacterial Pneumonia*, *Viral Pneumonia* and *COVID-19*. We have trained our model in two configurations, the one which classifies into the above four classes, and the other configuration with three classes (clubbing viral and bacterial pneumonia into one). The motivation behind the four class configuration is to better understand if any confusion between regular pneumonia and COVID-19 is due to the similarity of pathology between COVID-19 and viral pneumonia.

Similar to CheXNet [7], we treat each class as a binary classification problem, with input as a frontal-view chest X-Ray image X and output being a binary

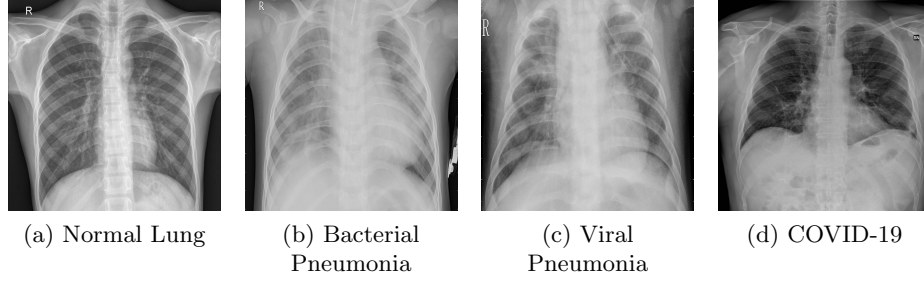


Fig. 1: Class-wise examples of frontal-view chest X-Ray images. The patient with bacterial pneumonia shows consolidation of the right lower lobe. Radiograph of a patient with viral pneumonia shows patchy consolidation in the right middle and lower zone. The last image from a patient with COVID pneumonia shows patchy ground glass opacity in the Left Lower Zone

labels $y_c \in \{0, 1\}$, indicating absence or presence of class c symptoms in the image respectively. We use the weighted binary cross-entropy loss as suggested by CheXNet [7]:

$$\mathcal{L}(X, y; \theta) = \sum_{c=1}^C \left(-w_c^+ \mathbb{I}\{y = c\} \log p_c(\hat{y} = 1 \mid X; \theta) - w_c^- \mathbb{I}\{y \neq c\} \log p_c(\hat{y} = 0 \mid X; \theta) \right). \quad (1)$$

Here, C is the number of classes and y is the ground truth for X . $p_c(\hat{y} = 1 \mid X; \theta)$ and $p_c(\hat{y} = 0 \mid X; \theta)$ are the probability scores for X being and not being in class c , respectively, as assigned by the network based on network weights θ . The two terms are weighted by $w_c^+ = \frac{N_c}{N_c + P_c}$ and $w_c^- = \frac{P_c}{N_c + P_c}$, where P_c and N_c are the number of positive and negative samples of class c , respectively in the training set.

3.2 Model Architecture

Our model contains pre-trained CheXNet [7], with a 121-layer Dense Convolutional Network (DenseNet) [3] backbone, followed by a fully connected layer. We replace CheXNet's [7] final classifier of 14 classes with our classification layer of 4 classes (3 classes for the clubbed pneumonia configuration), each with a sigmoid activation to produce the final output.

3.3 Training

For training we initialize our model with pre-trained weights from CheXNet implementation by Weng *et al.* [12], and then following the two stage training process described below:

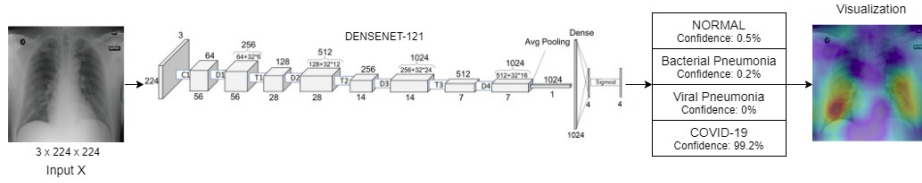


Fig. 2: CovidAID model (DenseNet image from [9])

1. In the first step, DenseNet’s backbone weights are frozen and only the final fully connected layer is trained. Training is performed using Adam optimizer with following parameters: $\beta_1 = 0.9$, $\beta_2 = 0.999$, and learning rate 10^{-4} . We use mini-batches of size 16, and train for about 30 epochs. The model with the lowest validation loss is selected for next stage.
2. In the second stage, the network weights are initialized from above, but the whole network is trained end-to-end (all layers), using the same hyper-parameters. We use mini-batch size of 8 in this stage due to memory constraints, and train for 10 epochs. Again, the model with lowest validation loss is selected for testing.

4 Dataset and Evaluation

We use the covid-chestxray-dataset [2] for COVID-19 frontal-view chest X-Ray images and chest-xray-pneumonia dataset [4] for frontal-view chest X-Ray images with bacterial/viral pneumonia as well as of normal lungs. We use the pre-trained CheXNet model, thus implicitly using robust features obtained after training on ChestX-ray14 [11] dataset.

The covid-chestxray-dataset [2] does not contain proper data split for training purposes, so we perform our own split, as shown in Tables 1 and 2. Since multiple images for the same patient could be found in the dataset, we split the data by patient-IDs to prevent any information leakage. We choose 20% of the images as test set, and 10 images are kept as validation set.

| | Normal | Pneumonia | | COVID-19 | Total |
|-------|--------|-----------|-------|----------|-------|
| | | Bacterial | Viral | | |
| Train | 1341 | 2530 | 1337 | 115 | 5323 |
| Val | 8 | 8 | 8 | 10 | 37 |
| Test | 234 | 242 | 148 | 30 | 654 |

Table 1: Sample-wise Data Split

| | Normal | Pneumonia | | COVID-19 | Total |
|-------|--------|-----------|-------|----------|-------|
| | | Bacterial | Viral | | |
| Train | 1000 | 1353 | 1083 | 80 | 3516 |
| Val | 8 | 7 | 7 | 7 | 29 |
| Test | 202 | 77 | 126 | 19 | 424 |

Table 2: Patient-wise Data Split

4.1 Sampling

The combined dataset (covid-chestxray-dataset and chest-xray-pneumonia) has a high data imbalance due to scarce COVID-19 data. Note that this imbalance is different from the positive-negative class imbalance (for which w^+ and w^- were introduced in the loss function). To ensure that training loss due to COVID-19 does not get masked by training loss due to other classes, we consider only a random subset of pneumonia data in each batch. The size of this subset should neither be too small, which will lead to overfitting on the COVID-19 data, nor too large to mask the COVID-19 loss, and is fixed empirically. In each batch we take data from classes *Normal*, *Bacterial Pneumonia*, *Viral Pneumonia* and *COVID-19* in the ratio 5 : 5 : 5 : 1. In case of the three class classification network, this ratio is 7 : 7 : 1.

5 Results

Our results indicate that this approach can lead to COVID-19 detection from X-Ray images with an AUROC (Area under ROC curve) of 0.9994 for the COVID-19 positive class, with a mean AUROC of 0.9738 (for 4-class classification configuration). Since we have modeled the problem as a binary classification problem for each class, given an input image X , we treat the class with maximum confidence score as the final prediction for calculating Accuracy, Sensitivity (Recall), PPV and confusion matrix.

| Pathology | AUROC | Sensitivity | PPV |
|-------------|--------|-------------|-------|
| Normal Lung | 0.9795 | 0.744 | 0.989 |
| Pneumonia | 0.9814 | 0.995 | 0.868 |
| COVID-19 | 0.9997 | 1.000 | 0.968 |

Table 3: Class-wise results for 3-class classification

| Pathology | AUROC | Sensitivity | PPV |
|-----------------|--------|-------------|-------|
| Normal Lung | 0.9788 | 0.761 | 0.989 |
| Bact. Pneumonia | 0.9798 | 0.961 | 0.881 |
| Viral Pneumonia | 0.9370 | 0.872 | 0.721 |
| COVID-19 | 0.9994 | 1.000 | 0.938 |

Table 4: Class-wise results for 4-class classification

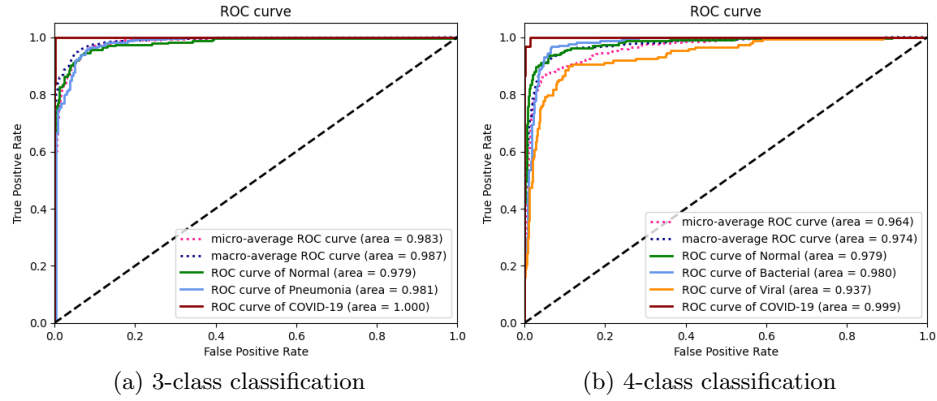


Fig. 3: ROC curves for our two configurations

We obtained an accuracy of 87.2% for 4-class classification configuration and 90.5% for the 3-class classification. The class-wise results for AUROC, Sensitivity (Recall) and PPV (Positive Predictive Value or Precision) are given in Tables 3 and 4. The corresponding ROC curves are shown in Fig. 3.

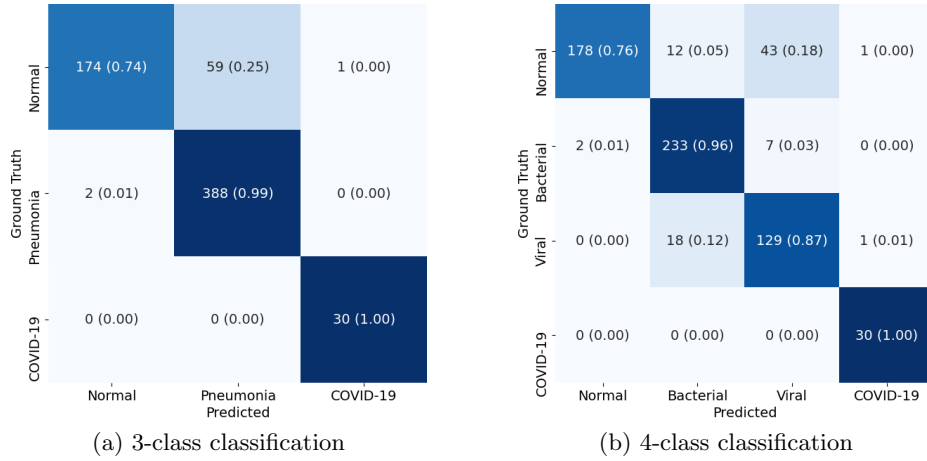


Fig. 4: Confusion matrices for our two configurations

The confusion matrices on the test data for the two configurations of our model are shown in Fig. 4. It can be seen that viral pneumonia, having a sensitivity (recall) value of 0.87 is often confused with bacterial pneumonia. This is likely due to the overlapping imaging characteristics. Our sensitivity (recall) for

COVID-19 positive class is 1.0, which is at par with the sensitivity (recall) for bacterial pneumonia.

5.1 Comparison with COVID-Net

To compare our approach with COVID-Net [10] we evaluate their published pre-trained models on the same test data split as ours. Fig. 5 compares their ROC curves with our model along with the confusion matrix. The COVID-Net inference shown here is made using their 'Small' variant which seemed to perform better among their two variants. It can be seen that our model outperforms COVID-Net by >0.14 AUROC in detecting regular Pneumonia as well as COVID-19. It should be noted that we used a different Pneumonia dataset from that used by COVID-Net, however, the COVID-19 data used is the same.

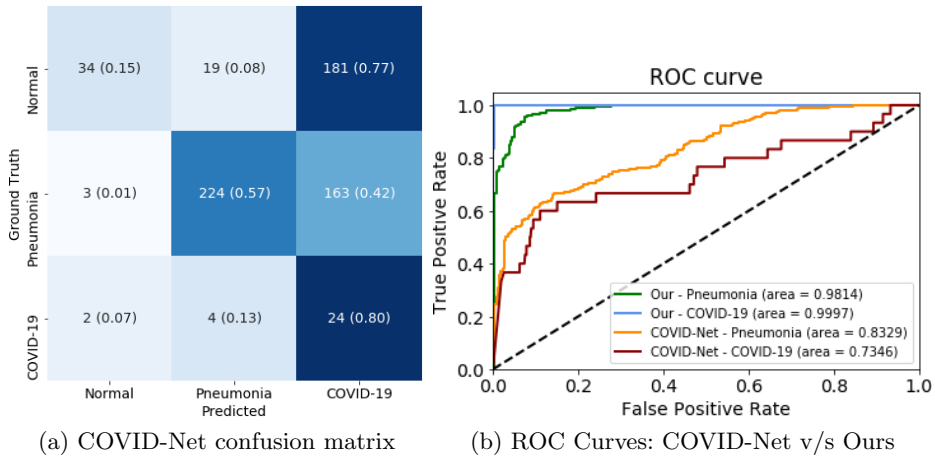


Fig. 5: Comparison with COVID-Net

| Model | F1 Score | 95% Confidence Interval |
|-----------|----------|-------------------------|
| COVID-Net | 0.3591 | 0.3508 - 0.3674 |
| CovidAID | 0.9230 | 0.9178 - 0.9282 |

Table 5: F1 Score (95% CI) comparison

We compute F1 scores for CovidAID and Covid-Net [10] over 10,000 bootstrapped samples from our test set of 654 images. 100 instances are taken (with replacement) of size 100 each for the bootstrap. The F1 scores along with their 95% confidence intervals are shown in Table 5.

The high margin of difference of the F1 scores of the two models clearly establish the superior performance of our model over COVID-Net.

5.2 Visualizations

To demonstrate the results qualitatively, we generate saliency maps for our model’s predictions using RISE [6]. In this approach 1000 randomly masked versions of a given X-ray image are queried and their classification scores are used to create a weighted mask corresponding to each output class. The core idea behind the RISE [6] approach is that masks which preserve semantically important parts of the image will lead to a higher classification score and hence a higher weight in the final mask for the respective class.

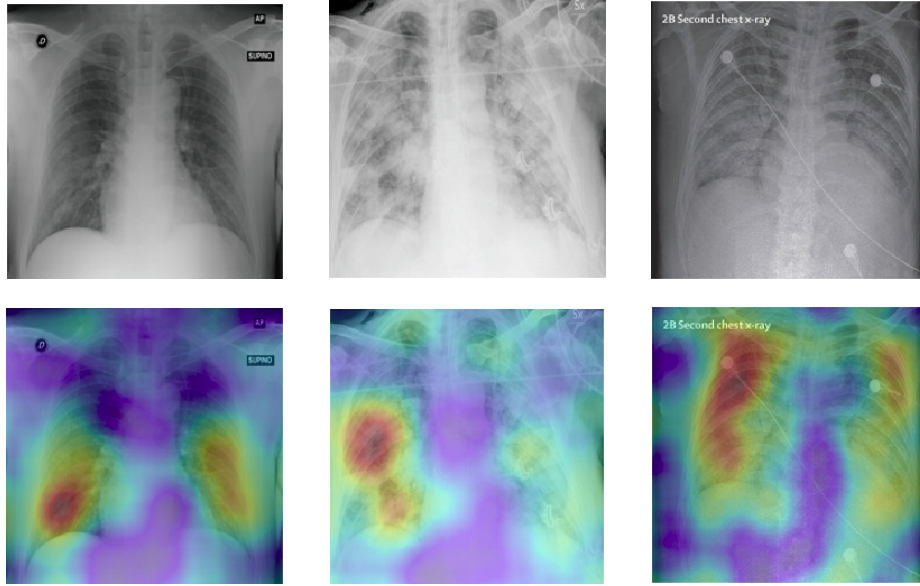


Fig. 6: Saliency map visaulization over COVID-19 positive X-Rays.
Red denotes region of greater importance.

The purpose of these visualizations was to have an additional check to rule out model over-fitting as well as to validate whether the regions of attention correspond to the right features from a radiologist’s perspective. Fig. 6 shows some of the saliency maps on COVID-19 positive X-rays.

6 Conclusion

We have presented some initial results on detecting COVID-19 positive cases from chest X-Rays using a deep-learning model. We have demonstrated sig-

nificant improvement in performance over COVID-Net [10], the only publicly maintained tool for classification of COVID-19 positive X-rays, on the same chest-xray-pneumonia dataset [4]. The results look promising, though the size of the publicly available dataset is small. We plan to further validate our approach using larger COVID-19 X-ray image datasets and clinical trials.

References

1. Bennett, J.M.: Smart ct scan based covid19 virus detector. https://github.com/JordanMicahBennett/SMART-CT-SCAN-BASED-COVID19_VIRUS_DETECTOR (2020)
2. Cohen, J.P., Morrison, P., Dao, L.: Covid-19 image data collection. arXiv 2003.11597 (2020), <https://github.com/ieee8023/covid-chestxray-dataset>
3. 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 (2017)
4. Mooney, P.: Kaggle chest x-ray images (pneumonia) dataset. <https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia> (2018)
5. Nisar, Z.: <https://github.com/zeeshannisar/covid-19>. <https://github.com/zeeshannisar/COVID-19> (2020)
6. Petsiuk, V., Das, A., Saenko, K.: Rise: Randomized input sampling for explanation of black-box models. In: Proceedings of the British Machine Vision Conference (BMVC) (2018)
7. Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K., et al.: Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv preprint arXiv:1711.05225 (2017)
8. Ranjan, E., Paul, S., Kapoor, S., Kar, A., Sethuraman, R., Sheet, D.: Jointly learning convolutional representations to compress radiological images and classify thoracic diseases in the compressed domain (12 2018). <https://doi.org/10.1145/3293353.3293408>
9. Ruiz, P.: Understanding and visualizing densenets. <https://towardsdatascience.com/understanding-and-visualizing-densenets-7f688092391a> (2018)
10. Wang, L., Wong, A.: Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest radiography images (2020)
11. Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R.M.: Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2097–2106 (2017)
12. Weng, X., Zhuang, N., Tian, J., Liu, Y.: Chexnet for classification and localization of thoracic diseases. <https://github.com/arnoweng/CheXNet/> (2017)
13. Yao, L., Poblens, E., Dagunts, D., Covington, B., Bernard, D., Lyman, K.: Learning to diagnose from scratch by exploiting dependencies among labels. arXiv preprint arXiv:1710.10501 (2017)