Deep Learning for Detection and Localization of Thoracic Diseases Using Chest X-Ray Imagery

Somnath Rakshit[1,2], Indrajit Saha[1(✉)], Michal Wlasnowolski[2,3], Ujjwal Maulik[4], and Dariusz Plewczynski[2,3]

[1] Department of Computer Science and Engineering, National Institute of Technical Teachers' Training and Research, Kolkata, India
indrajit@nitttrkol.ac.in
[2] Centre of New Technologies, University of Warsaw, Warsaw, Poland
[3] Faculty of Mathematics and Information Science, Warsaw University of Technology, Warsaw, Poland
[4] Department of Computer Science and Engineering, Jadavpur University, Kolkata, India

Abstract. Classification of diseases from biomedical images is a fast growing emerging field of research. In this regard, chest X-Rays (CXR) are one of the most widely used medical images to diagnose common heart and lung diseases where previous works have explored the usage of various pre-trained deep learning models to perform the classification. However, these models are very deep, thus use large number of parameters. Moreover, it is still not possible to find readily available access to a practicing radiologist for proper diagnosis from an X-Ray image of chest. Hence, this fact motivated us to conduct this research with the aim to classify CXR images in an automated manner with smaller number of parameters during training for 14 different categories of thoracic diseases and produce heatmap for the corresponding image in order to show the location of abnormality. For the purpose of classification, transfer learning is used with the pre-trained network of Resnet18, while the heatmaps are generated using pooling along the channel dimension and then computing the average of class-wise features. The proposed model contains less parameters to train and provides better performance than the other models present in the literature. The trained model is then validated both quantitatively and visually by producing localized images in the form of heatmaps of the CXR images. Moreover, the dataset and code of this work are provided online (http://www.nitttrkol.ac.in/indrajit/projects/deeplearning-chestxray/).

Keywords: Deep learning · Chest X-Ray · Transfer learning · Class activation map

L. Rutkowski et al. (Eds.): ICAISC 2019, LNAI 11509, pp. 271–282, 2019.
https://doi.org/10.1007/978-3-030-20915-5_25

1 Introduction

Chest X-Ray (CXR) is one of the most widely accessible radiology examinations. They are generally the first choice of radiologists because of their non-invasiveness, ability to reveal information which may often go unnoticed, such as pathological alterations, low cost and low dose of radiation. They are widely used as a preliminary diagnosis tool. With the increase in the number of patients, the radiologists have been facing increasing workload. This makes it essential to develop an automated and computerized way of understanding the contents of the CXR images.

Recently, deep learning has revolutionized the field of computer vision. Deep learning has been widely used to solve image classification tasks. This has naturally led to the application of deep learning to biomedical images. However, deep learning needs large datasets in order to be successful. The absence of this type of large datasets previously made it tough to use deep learning to solve various problems. The recent advancements in biomedical research leads to many large datasets being released by various organizations. Thus, the availability of high volume datasets make it possible to apply deep learning in order to solve various computational intelligence tasks using biomedical images. Recently, large datasets like CXR dataset, which are generated by using ionized radiation in the form of X-ray images, from Open-i[1], ChestX-ray8 by Wang et al. [24] and ChestX-ray14 [24] by National Institute of Health (NIH) make it possible to apply deep learning effectively for computer aided diagnosis of heart and lung diseases. However, recent research shows that the methods that have been used to perform CXR image classification are successful, while these models are very deep and use large number of parameters to train the models. Apart from this, it is still not possible to get readily available access to a practising radiologist to diagnose diseases from CXR images.

To address the above facts, in this paper, transfer learning is used to classify CXR images effectively by means of easily available models that have been pre-trained on the ImageNet dataset. In this regard, a pre-trained ResNet18 [8] model is used to perform such classification using transfer learning on the Chest X Ray14 database. This database contains 112,117 frontal-view CXR images of 32,717 unique patients. They are classified into 14 types of diseases viz. Atelectasis, Cardiomegaly, Effusion, Infiltration, Mass, Nodule, Pneumonia, Pneumothorax, Consolidation, Edema, Emphysema, Fibrosis, Pleural Thickening and Hernia. The CXR images are preprocessed, augmented and then passed on to the ResNet18 model for encoding of features using deep learning and finally, classification using a fully connected network. Here ten-crops technique is then used during testing in order to increase the accuracy. The results of ResNet18 model is compared with the existing works such as Yao et al. [27], Li et al. [11], Wang et al. [24], DNetLoc [7] and Rajpurkar et al. [17] where DenseNet121, DenseNet169 [9] and deeper architectures are used respectively. This is to be noted that the advancement of deep neural networks motivates the researcher

[1] https://openi.nlm.nih.gov/.

to apply DenseNet in most of the cases while the shallow model like ResNet18 has not yet been evaluated for similar task to the best of our knowledge. The performance of the trained model is demonstrated quantitatively by means of Area Under Curve (AUC) of Receiver Operating Characteristic (ROC) curve and visually by generating heatmaps of CXR images pointing out the locations containing abnormalities. This model can be of great help to a radiologist in order to prioritize the severely abnormal cases first and to increase the confidence of the radiologist.

2 Background

This section contains a brief background of various diseases viz. Atelectasis, Cardiomegaly, Effusion, Infiltration, Mass, Nodule, Pneumonia, Pneumothorax, Consolidation, Edema, Emphysema, Fibrosis, Pleural Thickening and Hernia which are classified from the CXR images in this paper along with the various deep learning models used for classification and localization.

2.1 Brief Description of the Thoracic Diseases

Atelectasis is a disease in which closure of a lung occurs. It is a condition where the alveoli get deflated to reduced or even zero volume. On the other hand, Cardiomegaly [23] is caused by the enlargement of the heart. It is generally congenital in nature and may also be caused by heart valve diseases, thyroid disorders, pregnancy, kidney related diseases, diabetes and HIV infection among others. While effusion [12], or in this case, pleural effusion, is the condition in which excess fluid gets accumulated in the pleural cavity. This can hamper breathing since the expansion of the lung is limited by the presence of this unwanted liquid.

On the other hand, infiltration (more specifically, pulmonary infiltration) [18] is the presence of foreign objects, which accumulate gradually and can even fill the entire lung, causing life-threatening situation. While pulmonary mass is any opaque area of more than 30 mm size in the CXR image. If the opaque area measures less than 30 mm, it is known as lung nodule. Pulmonary nodules [14] itself do not show any symptoms. However, if the lung nodule was formed due to cancer, then certain symptoms like shortness of breath, pain in chest area, coughing up blood and weight loss may take place. While pneumonia [6] is a disease that is caused by the inflammation in lungs. The primarily affected parts are the alveoli. Apart from the earlier seven thoracic diseases, Pneumothorax is the presence of abnormal air space in the pleural space between the lung and the chest wall. While consolidation, or in this case, pulmonary consolidation [22], is a region of lung tissue that contains liquid instead of air. This disease is associated with the swelling and hardening of normally soft tissue of a normal lung. On the other hand, pulmonary edema [19] is a disease in which excess fluids accumulate in the lungs making breathing difficult. This may occur due to a congestive heart failure or acute lung injury. Emphysema [5] is a disease which results in

shortness of breath. Rupturing of alveoli is one of the causes of this disease. In this condition, a large air space gets created inside the lungs instead of many small sized air spaces. While pulmonary fibrosis [3] is a lung disease in which the tissues present around and between the alveoli get hardened, thus making it difficult for the lung to function properly. Pleural thickening [26] is a condition in which the surrounding tissues of the lungs get scarred and start getting hardened. It is similar to pulmonary fibrosis in the sense that its symptoms cannot be felt early and is felt only when more and more scar tissues grow around the lungs. Lung Hernia [25] occurs when a portion of the lung extends outside the thoracic wall. In most cases, its symptoms are not felt. However, in advanced cases, the patient feels pain while inhaling and swelling in a particular area over the chest and fever is observed.

2.2 Brief Description of Deep Learning Models

ResNet is an image classification model that uses deep learning techniques. It was developed by He et al. [8] and was ranked first in the ImageNet Large Scale Visual Recognition Competition 2015 [4]. The ResNet model solves one of the most notorious problems in the field of machine learning, i.e. the vanishing gradient problem. In this problem, as the model gets deeper, its accuracy becomes saturated and rapidly starts decreasing. To overcome this problem, ResNet introduced the concept of identity shortcut. By reusing the activations of one of the previous layers, this model ensures that the layer next to the current one learns its weights properly. Here, an additive merge operation is performed, thus, forcing the network to learn residual errors. Densenet model was later developed by Huang et al. [9] in which instead of addition, concatenation of the output of previous layers to the output of the current layer was proposed. While the traditional convolutional neural networks have L connections (between each layer and its next one), the Densenet model has $L * (L + 1)/2$ connections. Here, for each layer, the feature maps of all previous layers are used as its inputs while the feature map of this layer is also used in the input of all subsequent layers. Densenet model overcomes the vanishing gradient problem in this fashion and is able to achieve feature reuse. Thus, counter to intuition, it is able to reduce the number of training parameters in spite of having a large network. Here, since each layer has direct access to the original input, better supervision is achieved while training.

Apart from these models, many other models such as Inception [21] and VGG [20] are also present to tackle the problem of efficient image classification using convolutional neural networks.

3 Method

This section describes the deep learning model/architecture used for classification of CXR images and generation of class activation maps (CAM) or heatmaps so as to obtain the localized CXR image. Apart from this, it is also described how

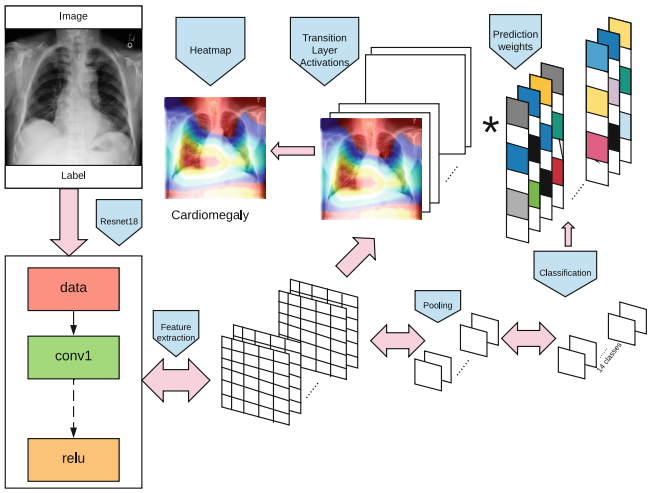

Fig. 1. Pipeline of the proposed model

transfer learning, preprocessing, data augmentation and localization of abnormalities are performed.

3.1 Architecture

The method used in this work can be divided into two stages. First, a pre-trained Resnet18 model is used for fine-tuning and training. It is a convolutional neural network that extracts the features present in the dataset of CXR images. Each of the feature vectors is a list of numbers that can be used by the dense layer in the next step for classification. It is a common practice to use a pre-trained network and fine-tune it for classification of images whose nature is different from the images that were used for training it during the ImageNet challenge. The dense layer thus takes the feature vectors from the Resnet18 model as input and generates a vector whose size is equal to the number of classes (14) as output. Further, a softmax layer is present that adds a non linearity and makes each of the 14 values to fall between 0 and 1. Now, each element of the vector contains a probability between 0 and 1. These are the probabilities of an image to fall in each of the 14 classes. This model is seen to contain 11,183,694 trainable parameters. Here, the objective is to use images and their corresponding diseases as input and generate classification for thoracic diseases for new CXR images. Later on, the CAM of the new images are also generated. Second, using the weights of the trained model, CAM is plotted for new CXR images. While training the model, the class averaged binary cross entropy loss is used. The loss function is described through Eqs. 1 to 3.

$$L = \{l_1, \ldots, l_N\}^\top \tag{1}$$

$$l_n = -w_n \left[y_n \cdot \log x_n + (1 - y_n) \cdot \log(1 - x_n) \right] \tag{2}$$

$$\ell(x, y) = mean(L) \tag{3}$$

In order to train the model, an optimizer is needed. Here, the Adam optimizer [10] is used for this purpose. Adam has low memory requirements and is invariant to the diagonal rescaling of the gradients. The Adam optimizer also has an efficient regret bound on the convergence rate. It is also well suited for problems involving large amount of data. The pipeline of the proposed model is shown in Fig. 1.

3.2 Transfer Learning

Training of the deep learning model can be done in two ways. It can either be done from scratch or it can be done by using the pre-trained weights in a different classification problem. The second method is known as transfer learning. In this method, knowledge obtained from a previous task is used to perform a different task. Thus, the knowledge from the source task is used to predict the outcome of the target task. Training from scratch requires significantly large datasets and hardware resources. However, a smaller dataset can be used in case of transfer learning.

In the present work, transfer learning is adopted. A Resnet18 model and its pre-trained weights based on the ImageNet challenge are used here as a feature extractor. Only the last layer of the classifier is modified. Instead of 1000 neurons in the case of ImageNet challenge, here 14 neurons have been used where each neuron correspond to each of the classes.

3.3 Preprocessing and Data Augmentation

In this work, the images present in the dataset are preprocessed and subjected to data augmentation. In the preprocessing step, the images are normalized after subtracting the mean of all pixels in the dataset from each pixel and then dividing the result by the standard deviation of the dataset. Data augmentation is the process of synthetically increasing the size of the training dataset. It is needed to decrease the chances of overfitting. In data augmentation, new images are synthetically generated from the images that are already present. Here, 2 ways are used to generate images synthetically. Firstly, images are randomly cropped and then resized using bilinear interpolation to form $224 * 224$ images. Secondly, new images are generated by randomly flipping them horizontally. Horizontal flipping is performed to augment the training set so as to simulate images where the abnormality may be present in any of the lung, since lung diseases may take place in either lung. This step is important as it makes the model robust to abnormalities in CXR images.

3.4 Localization of Abnormalities

After obtaining a trained model, heatmaps are generated on test images that point out the exact location of the abnormality. This helps to visualize the

most indicative areas used by the model during classification. The heatmaps are generated by using pooling along the channel dimension and then computing the average of class-wise features. The heatmaps show that the network that is trained on this architecture generalizes well and demonstrates good interpretation ability in localizing the areas of interest.

4 Experimental Results

This section describes the dataset, testbed and results.

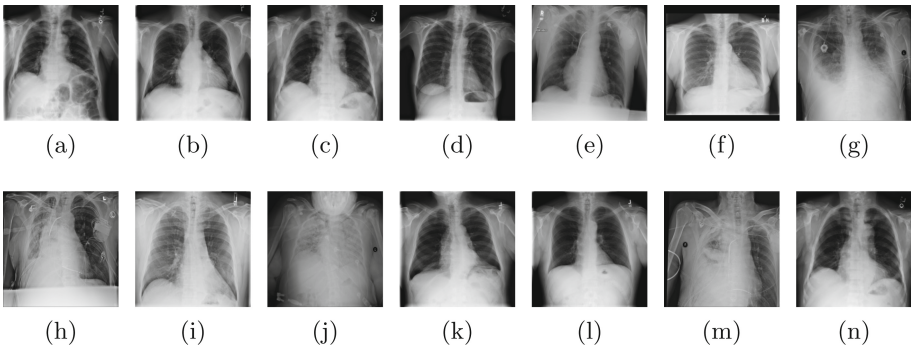

Fig. 2. Examples of the images present in the dataset. All 14 classes are represented here. (a) Atelectasis (b) Cardiomegaly (c) Effusion (d) Infiltration (e) Mass (f) Nodule (g) Pneumonia (h) Pneumothorax (i) Consolidation (j) Edema (k) Emphysema (l) Fibrosis (m) Pleural Thickening (n) Hernia

4.1 Dataset

To perform this experiment, the ChestX-ray14 dataset from National Institute of Health (NIH) has been used. This dataset contains 112,117 frontal images from 32,717 unique patients. The detailed breakdown of the dataset is provided in Table 1. All of the images have a resolution of 1024 * 1024 with 8 bits of gray scale. The dataset has been divided in 3 parts namely train, test and validation consisting of 70%, 20% and 10% of the total dataset respectively. Each image may have one or more labels viz. Atelectasis, Cardiomegaly, Effusion, Infiltration, Mass, Nodule, Pneumonia, Pneumothorax, Consolidation, Edema, Emphysema, Fibrosis, Pleural Thickening and Hernia. These labels were not created manually. Rather, they have been generated by analyzing medical records of patients using optical character recognition. The authors of the dataset have ensured that there is no patient overlap in the test and train data. CXR images from each class of data has been shown in Fig. 2.

Table 1. Table showing the class-wise number of images

Disease	Train	Test	Validation
Atelectasis	7996	2420	1119
Cardiomegaly	1950	582	240
Effusion	9261	2754	1292
Infiltration	13914	3938	2018
Mass	3988	1133	625
Nodule	4375	1335	613
Pneumonia	978	242	133
Pneumothorax	3705	1089	504
Consolidation	3263	957	447
Edema	1690	413	200
Emphysema	1799	509	208
Fibrosis	1158	362	166
Pleural Thickening	2279	734	372
Hernia	144	42	41
No disease	42404	11927	6078
No of images	78467	22432	11218

4.2 Testbed

The entire work has been implemented using PyTorch [15] and executed on a machine having 256 GB RAM and Intel Xeon CPU E5-2680 v4 processor with 28 cores. It has a NVIDIA GeForce GTX 1070 GPU. The Resnet18 model was chosen experimentally after trying out different other models like VGG19, VGG11 and DenseNet121. The training for the Resnet18 model took around 5 days in this configuration. The images have been generated using OpenCV [2] in Python. Also, scikit-learn [16] and numpy [13] libraries have been used for evaluation of the model. The model is trained on a batch size of 16 for 1000 epochs. In the Adam optimizer, the values of the learning rate is set as 0.0001 whereas the range of beta values are set in between 0.9 to 0.999. A weight decay of 10^{-5} and epsilon value of 10^{-8} have been used also.

4.3 Results

In this work, the performance of the model has been tested by means of Area Under Curve (AUC) metric. To test the model, the TenCrop technique [1] has been utilised. This helped in increasing the accuracy of the model slightly. The value of AUC of the model is compared with other models as shown in Table 2 and Fig. 3. It is observed that the Resnet18 model performs better than the other models that have been presented previously in literature. In 8 out of the 14 classes viz. Atelectasis, Effusion, Nodule, Pneumonia, Consolidation, Edema, Fibrosis

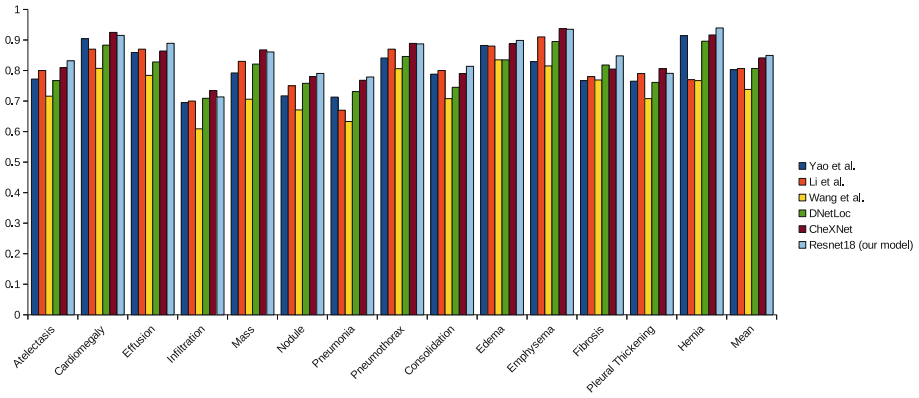

Fig. 3. Bar plot comparing the AUC scores of each of the models for each of the 14 classes along with the mean score

Table 2. Comparison of the proposed model with other models present in literature. The best score for each disease is formatted in bold.

Pathology	Yao et al. [27]	Li et al. [11]	Wang et al. [24]	DNetLoc [7]	CheXNet [17]	Resnet18 (our model)
Atelectasis	0.7720	0.8000	0.7160	0.7670	0.8094	**0.8318**
Cardiomegaly	0.9040	0.8700	0.8070	0.8830	**0.9248**	0.9151
Effusion	0.8590	0.8700	0.7840	0.8280	0.8638	**0.8890**
Infiltration	0.6950	0.7000	0.6090	0.7090	**0.7345**	0.7136
Mass	0.7920	0.8300	0.7060	0.8210	**0.8676**	0.8608
Nodule	0.7170	0.7500	0.6710	0.7580	0.7802	**0.7904**
Pneumonia	0.7130	0.6700	0.6330	0.7310	0.7680	**0.7788**
Pneumothorax	0.8410	0.8700	0.8060	0.8460	**0.8887**	0.8872
Consolidation	0.7880	0.8000	0.7080	0.7450	0.7901	**0.8138**
Edema	0.8820	0.8800	0.8350	0.8350	0.8878	**0.8985**
Emphysema	0.8290	0.9100	0.8150	0.8950	**0.9371**	0.9351
Fibrosis	0.7670	0.7800	0.7690	0.8180	0.8047	**0.8481**
Pleural Thickening	0.7650	0.7900	0.7080	0.7610	**0.8062**	0.7907
Hernia	0.9140	0.7700	0.7670	0.8960	0.9164	**0.9394**
Mean	0.8027	0.8064	0.7381	0.8066	0.8414	**0.8494**

and Hernia, the Resnet18 model shows superiority over other models. In terms of the mean AUC, it is seen that the Resnet18 model obtains the highest score of 0.8494. Also, the images obtained after localization are presented in Fig. 4. It is confirmed visually that the Resnet18 model is able to identify the areas of interest well and thus, can be of assistance to practicing radiologists. Hence, it is possible to detect thoracic diseases from CXR images with less expensive hardware.

In clinical procedures, automated detection and correct spatial localization of thoracic diseases will be of big help to radiologists and will assist them in decision-making as well as prioritization of cases. This work presents a model

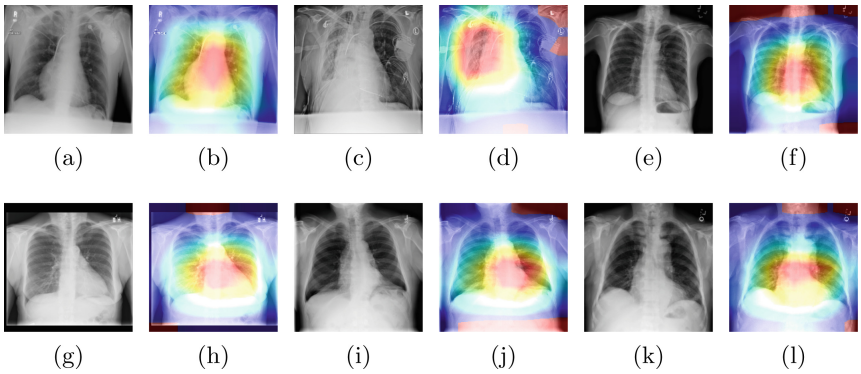

Fig. 4. Heatmaps generated using the trained model. The heatmaps of the following classes are shown: (a) and (b) Mass, (c) and (d) Pneumothorax (e) and (f) Infiltration (g) and (h) Nodule (i) and (j) Emphysema (k) and (l) Effusion

that is both efficient with respect to training time as well as produces better performance than the present state-of-the-art models. It only trains on labeled images and is able to classify as well as localize the areas of interest.

5 Conclusion

The emergence of large biomedical datasets have made it possible to apply deep learning models to perform various computational intelligence tasks. Detection of thoracic diseases from Chest X-Ray images is a challenging task since the machine does not have access to the patient's past medical history. Also, since the number of categories of disease is relatively high, it becomes difficult and tedious for a radiologist to correctly diagnose all diseases from Chest X-Ray images. Some prior works have attempted to solve this problem by using various deep learning models. However, comparatively shallow architectures like Resnet18 model has not been investigated previously. In this regard, it is shown that through transfer learning, Resnet18 model performs better on the ChestX-ray 14 dataset than all other deeper architectures used previously. The performance of this work has been evaluated quantitatively by means of AUC of ROC curve and visually by generating heatmaps. These heatmaps can help in detecting the areas of abnormality so that radiologist can prioritize the severe cases. Finally, we hope that this work will promote future work in the domain of detection of diseases using biomedical images.

Acknowledgment. This work has been co-supported by the Polish National Science Centre (2014/15/ B/ST6/05082), Foundation for Polish Science (TEAM to DP) and by the grant from the Department of Science and Technology, India under Indo-Polish/Polish-Indo project No.: DST/INT/POL/P-36/2016. The work was co-supported by grant 1U54DK107967-01 Nucleome Positioning System for Spatiotemporal Genome Organization and Regulation within 4DNucleome NIH program.

References

1. Krizhevsky, A., Sutskever, I., Hinton, G.: Imagenet classification with deep convolutional neural networks. In: Proceedings of the Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
2. Bradski, G.: The OpenCV Library. Dr. Dobb's J. Softw. Tools **25**, 120–125 (2000)
3. Collard, H., et al.: Acute exacerbations of idiopathic pulmonary fibrosis. Am. J. Respir. Crit. Care Med. **176**(7), 636–643 (2007)
4. Deng, J., Dong, W., Socher, R., Li, L., Li, K., Li, F.: Imagenet: a large-scale hierarchical image database, pp. 248–255 (2009)
5. Eriksson, S.: Pulmonary emphysema and alpha1-antitrypsin deficiency. Acta Med. Scand. **175**(2), 197–205 (1964)
6. Fine, M., et al.: A prediction rule to identify low-risk patients with community-acquired pneumonia. N. Engl. J. Med. **336**(4), 243–250 (1997)
7. Guendel, S., et al.: Learning to recognize abnormalities in chest x-rays with location-aware dense networks. arXiv preprint arXiv:1803.04565 (2018)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition, pp. 770–778 (2016)
9. Huang, G., Liu, Z., Maaten, L.V.D., Weinberger, K.: Densely connected convolutional networks, pp. 2261–2269 (2017)
10. Kinga, D., Adam, J.: A method for stochastic optimization (2015)
11. Li, Z., et al.: Thoracic disease identification and localization with limited supervision. arXiv preprint arXiv:1711.06373 (2017)
12. Light, R., Macgregor, M.I., Luchsinger, P., Ball, W.: Pleural effusions: the diagnostic separation of transudates and exudates. Ann. Internal Med. **77**(4), 507–513 (1972)
13. Oliphant, T.: Python for scientific computing. Comput. Sci. Eng. **9**(3), 10–20 (2007)
14. Ost, D., Fein, A.M., Feinsilver, S.: The solitary pulmonary nodule. N. Engl. J. Med. **348**(25), 2535–2542 (2003)
15. Paszke, A., et al.: Automatic differentiation in pytorch. In: NIPS-W (2017)
16. Pedregosa, F., et al.: Scikit-learn: machine learning in python. J. Mach. Learn. Res. **12**(Oct), 2825–2830 (2011)
17. Rajpurkar, P., et al.: Chexnet: radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv preprint arXiv:1711.05225 (2017)
18. Reeder, W., Goodrich, B.: Pulmonary infiltration with eosinophilia (PIE syndrome). Ann. Internal Med. **36**(5), 1217–1240 (1952)
19. Robin, E., Cross, C., Zelis, R.: Pulmonary edema. N. Engl. J. Med. **288**(6), 292–304 (1973)
20. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
21. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision, pp. 2818–2826 (2016)
22. Targhetta, R., Chavagneux, R., Bourgeois, J., Dauzat, M., Balmes, P., Pourcelot, L.: Sonographic approach to diagnosing pulmonary consolidation. J. Ultrasound Med. **11**(12), 667–672 (1992)
23. Tavora, F., et al.: Cardiomegaly is a common arrhythmogenic substrate in adult sudden cardiac deaths, and is associated with obesity. Pathology **44**(3), 187–191 (2012)

24. Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summer, R.: Chestx-ray8: hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases, pp. 3462–3471 (2017)
25. Weissberg, D., Refaely, Y.: Hernia of the lung. Ann. Thoracic Surg. **74**(6), 1963–1966 (2002)
26. Wu, R., Yang, P., Kuo, S., Luh, K.: "Fluid color" sign: a useful indicator for discrimination between pleural thickening and pleural effusion. J. Ultrasound Med. **14**(10), 767–769 (1995)
27. Yao, L., Poblenz, 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)