

A Review of AI-Based Diagnosis of Multiple Thoracic Diseases in Chest Radiography



Roshan Shetty, Prasad Narasimha Sarappadi, K. M. Sudarshan,
and Raveendra Gudodagi

1 Introduction

Diseases seen in the thoracic cavity of the human body are called thoracic diseases. Millions of people suffer from these thoracic diseases worldwide. The chest X-ray is a recommended initial procedure for thoracic diseases in hospitals. Figure 1 shows chest X-ray images of some common thoracic diseases. Interpretation of X-rays and radiologists' decision plays a vital role during the treatment of such diseases. Medical images such as chest X-rays provide valuable information, and AI-based diagnosis of medical images is very much essential due to the shortage of radiologists. So artificial intelligence is alternative for solving such diagnosis problems and improved health care for the public. The AI-based computer-aided diagnosis is dominated by deep learning (DL) algorithms, implemented with the help of the convolution neural network (CNN).

Deep learning falls under neural networks which are not new to the scientific world. Deep learning introduced the deep aspect of neural networks and is termed deep neural network (DNN). The advantage of DNN over neural networks is its ability to discover hidden features from input data. In deep learning, CNNs are specialized in the detection and classification of any images. Today, deep learning with CNN applications is increasing rapidly in computer-aided diagnosis of medical images. It will help the medical community in terms of decreased workload, quick diagnosis, and overall improvement in the health-care system [1].

R. Shetty (✉)

Alva's IET, Moodubidire, India

P. N. Sarappadi · K. M. Sudarshan · R. Gudodagi

REVA University, Bengaluru, India

e-mail: prasadsn@reva.edu.in; sudarshankm@reva.edu.in; raveendra.g@reva.edu.in

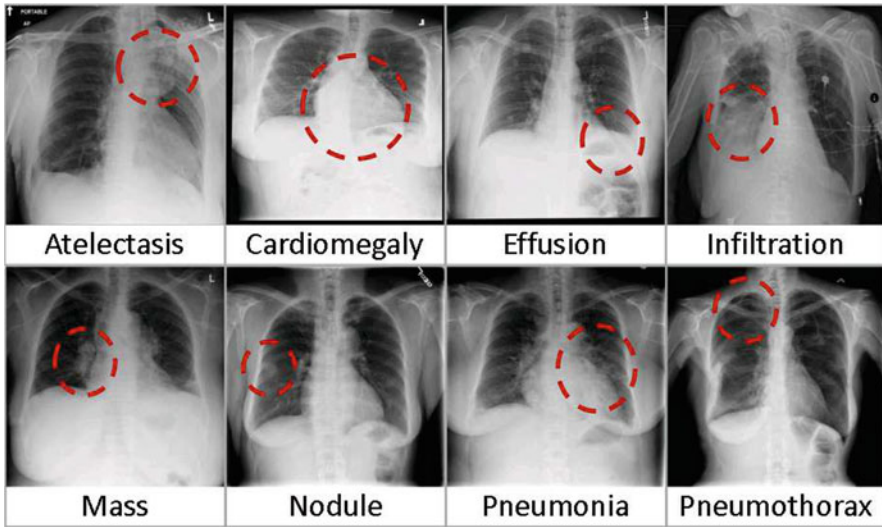


Fig. 1 Eight common thoracic diseases

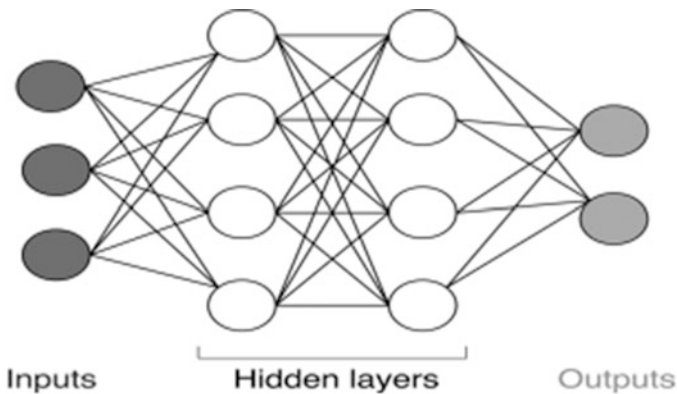


Fig. 2 Multilayer architecture of deep neural network [2]

Neural networks are inspired by the activity of the human brain, particularly the activities of neurons, which led to the development of a new area called artificial neural networks (ANNs). These neural networks and backpropagation training algorithms have a long history in the scientific world [2]. Although neural networks have been used for many years, in recent years, three factors triggered the learning of large neural networks. They are called deep neural networks (DNNs). Those factors are (1) powerful, massive parallel computing hardware called ultrafast graphics processing units (GPUs), (2) availability of large labeled datasets, and (3) advancements in training technologies and architecture. Figure 2 shows the “deep”

aspect of deep learning which refers to the multilayer network architecture. Here, information flows from input to output nodes through many hidden layers of nodes.

A type of DNN called as CNN makes use of layers of kernels or filters to extract features from the entire image without anyone's supervision. In 2012, for the first time, human competitive results were reported with CNN such as AlexNet on a widely used computer vision benchmark [3], which had only 8 layers and 62 million parameters. This encouraged many researchers to use CNN on medical images. Important design factors to be considered for building a CNN-based chest X-ray thoracic disease interpretation model are (1) learning forms, (2) network architecture, and (3) training type.

In the next section, we will discuss briefly on these important design factors one by one. Section 3 explains different deep learning models built to tackle multiple thoracic disease detection and classification problem. Section 4 contains a discussion and future trends. Section 5 has concluding remarks.

2 Important Design Factors

2.1 Learning Forms

Supervised learning This is one of the leading learning forms where machines are fed with a lot of labeled input data and output is already known for it. It is predictive in nature with high accuracy. The model considers image labels as ground truths. For instance, in the learning system for classifying thoracic diseases, a particular X-ray image may be labeled as a "hernia." A deep model during supervised training continuously adjusts network parameters or weights in all layers. These adjustments are done by comparing deep model prediction with the ground truth (hernia).

Unsupervised learning The machine is fed with unlabeled input data and output is not mapped with input. It is an independent learning procedure, mainly used to detect hidden patterns without any ground truths.

Semi-supervised A hybrid approach makes use of large unlabeled data and less labeled data. Scarcity of labeled data made semi-supervised learning one of the problems of significant importance in modern-day data analysis. A semi-supervised model is always expected to do more than that in the label, for example, preparing a model to execute many tasks such as detection, classification, and also localization of pathology in the case of medical images.

Weakly supervised learning This is a different form of training method, in which learning happens with inaccurate labels. Supervised and weakly supervised learning are popular, among deep learning-based models on diagnosis of thoracic diseases.

2.2 Network Architecture

CNN: This is the most popular choice with impressive results compared to others in image classification tasks [4]. The CNN model consists of two major components:

1. Convolution filters and pooling layers: Convolution operation is performed by sliding the filter over the entire input image to produce feature maps. The ReLU activation function is used to add some nonlinearity to convolution results. So values in the final feature maps are not actually the sum. The pooling process reduces dimensions of image and number of parameters. This indirectly reduces training time of a model and overfitting.
2. Fully connected layer as classification part: It assigns a probability for the object on the image as what the algorithm predicts. Most of the time, softmax function is the common choice in fully connected layers for classification tasks, but the sigmoid function is popularly used in chest X-ray-based CAD systems. The fully connected layer works with one-dimensional vector numbers. So the three-dimensional output of predecessors needs to be converted to a one-dimensional vector. Figure 3 shows deep CNN architecture for thoracic disease classification.

The performance of CNN is dependent on the following:

1. Network architecture adopted: Different CNNs for image feature learning are AlexNet, VGG, GoogLeNet, ResNet, and DenseNet.
2. Amount of training data: Less training data results in overfitting. So we need to produce more input data from the current dataset. With the help of some common transformations, more input data are created; such a process is called data augmentation. In this way, we artificially boost the size of the training set to reduce overfitting. Dropout is also used to prevent overfitting during training

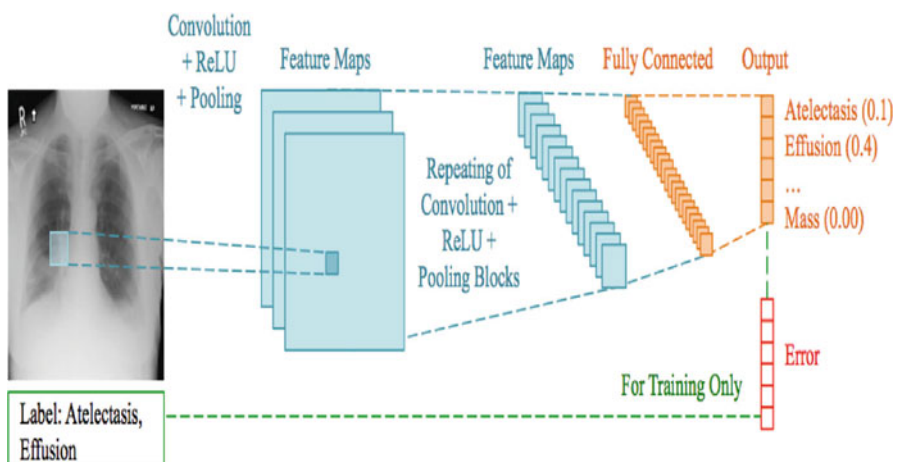


Fig. 3 Deep CNN model for multiple disease classification

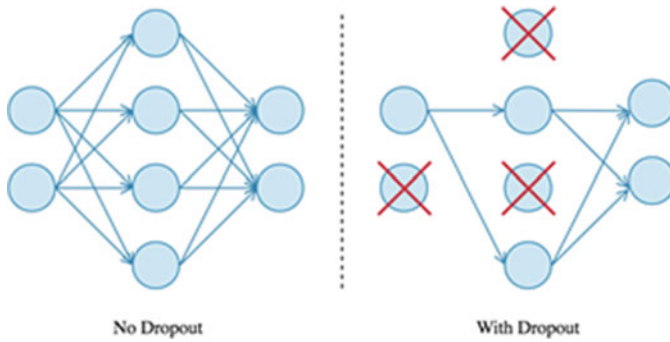


Fig. 4 Dropout of nodes

time. At each iteration, a neuron is temporarily “dropped” or disabled with probability as shown in Fig. 4. Dropout can be applied to the input or hidden layer nodes but not the output nodes. Adding dropout is the most popular regularization technique for a deep neural network’s accuracy boost.

2.3 Training Types

Transfer learning is a training strategy of new deep models, where training is initiated with a pretrained model, instead of from scratch. The pretrained model is used, when we have limited availability of training data. The final layers are fine-tuned with the same available training data. The pretrained models are trained with many databases, which contain all types of images. Advantages of transfer learning are performance enhancement in the new deep model, reduction in training time for the new model, and less training data which may be enough for the new task domain.

3 Multiple Disease Classification

The deep neural networks (DNNs) provide good results, only if we feed enough data to them [5]. CAD for thoracic diseases is very challenging due to limited number of chest radiograph database in digital form.

3.1 Evaluation Metrics Used

The evaluation of all existing methods in multiple disease classification frameworks is done using a metric called area under the curve (AUC). It measures the area under

the receiver operating characteristic (ROC) curve. The performance of the classifier on individual pathology is evaluated using per-class AUC, and for multiple diseases, the average value of AUC is taken. Hence, a higher average AUC indicates a better classifier. Both AUCs are always in the range of 0–1. Most of the time, AUCs are expressed in percentages.

3.2 Popular Dataset

The first ever publicly available chest X-ray dataset is Open I [6]. It has 7470 X-ray images in digital form. This dataset came in 2016 with automated image annotation, but none of the diseases were detected. Major achievement happened next year, when a new chest X-ray database was made public by NIH (National Institute of Health). It had a total of eight thoracic disease labels. These labels are created using natural language processing (NLP) process [5]. They also proposed a thoracic disease classification framework with four CNN pertained models AlexNet, GoogLeNet, VGG-16, and Resnet-50 for the initialization of a network. Out of these, Resnet-50 shows better performance with an average AUC of 73.8%.

However, they used a binary relevance approach for classification even though most of the images contain multiple pathology as shown in Fig. 5. Later, NIH

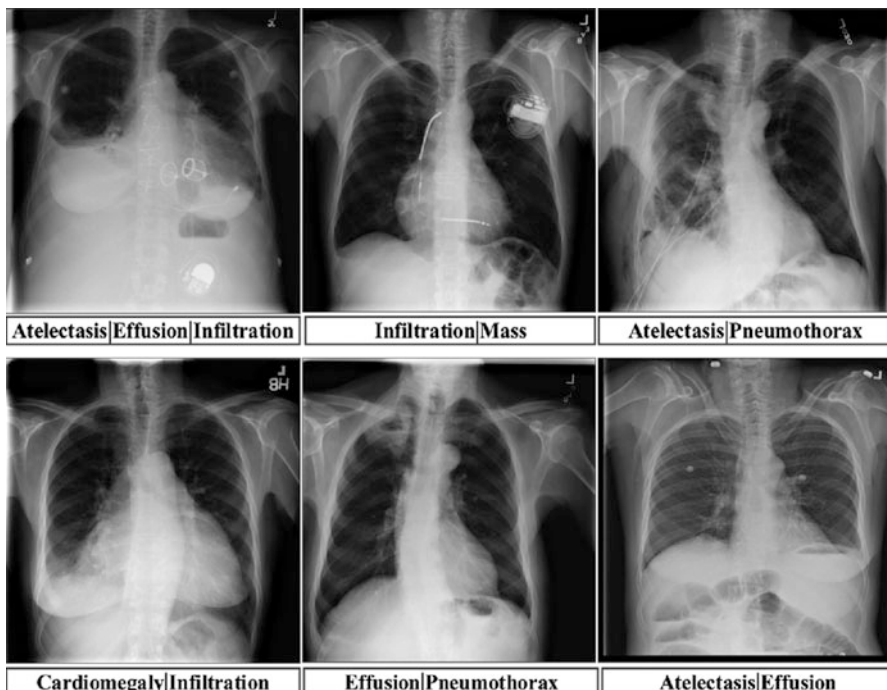


Fig. 5 Multilabel images of chest X-ray14 dataset

expanded this database from 8 to 14 thoracic diseases and named it has chest X-ray14 dataset [5]. More recently, different state-of-the-art models were proposed by researchers to address chest X-ray14 multiple disease classification problems, which are discussed next.

3.3 *State-of-the-Art Models*

Yalo et al. [7] proposed a DenseNet model with an average AUC score of 80.1%. They tried to discover inter-relation among all 14 pathologies. Their model is trained from scratch without any transfer learning, and their designed LSTM (long short-term memory) model ignored label dependencies also.

Rajpurkar et al. [8] proposed a method called as CheXNet. It is pretrained with DenseNet-121 with batch normalization and achieved an average AUC score of 84.2%. The heatmaps are generated with the help of CAM (class activation map). In CheXNet method, AUC score on pathology pneumonia equaled the prediction of radiologists.

Kumar et al. [9] developed cascaded DenseNet for thoracic disease classification task. The performance classification of thoracic disease, cardiomegaly, was improved with per class AUC score of 91.33% than previous methods. However, the above methods used the entire image for the detection of diseases even though some pathology areas are very small in the images. Usage of entire image in training process resulted in more computational complexity and noise.

Guan et al. [10] developed AG-CNN, a deep learning framework, and obtained good score of 87.1%. Here, they introduced a fusion branch in their method to fuse both global and local information of X-ray image. But here, some misclassification of pathology are observed due to CAM-generated heatmaps.

Wang et al. [11] introduced a model on chest X-ray14 database and called as ChestNet. They introduced Grad-CAM-based attention block to regulate CNN-based feature extractor. The CNN pretrained ResNet-152 is used as a main network with a support of attention block. Their proposed model outperformed previous models with an average AUC result of 89.6%. Figure 6 shows Grad-CAM [12]-generated heatmaps from ChestNet for thoracic disease localization.

Segmentation-based deep fusion network [13] does thoracic pathology detection in a unique way. Here, two CNNs are used to act on entire X-ray and also on a cropped lung region of X-ray. Both features are later fused together for classification work. The average AUC of 81.5% was achieved with very high computation complexity and more computation time.

Chen et al.'s [14] DualCheXNet architecture was introduced. It works with dual asymmetric feature learning from both ResNet and DenseNet architectures. It does embed complementary feature learning using both CNNs. This model achieved a good average AUC score of 85.6%. Figure 7 shows the ROC (receiver operating characteristics) curve plot of their model. The ROC curve of all 14 pathologies of the standard chest X-ray14 dataset is shown.

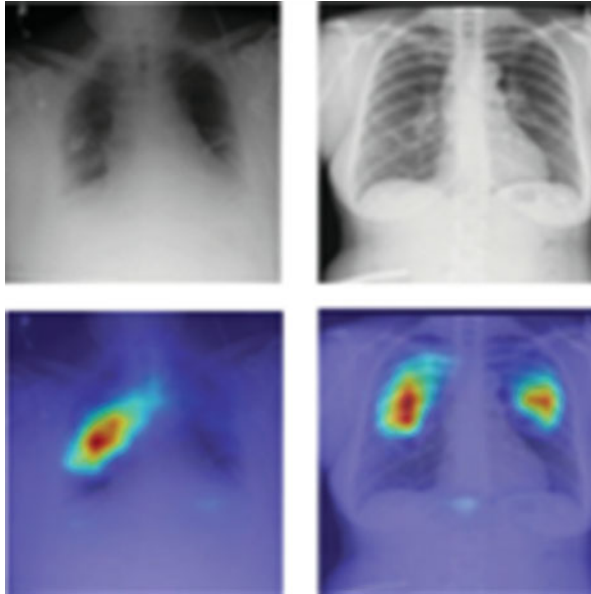


Fig. 6 Grad-CAM generates heatmap for a localization of atelectasis (left) and nodule

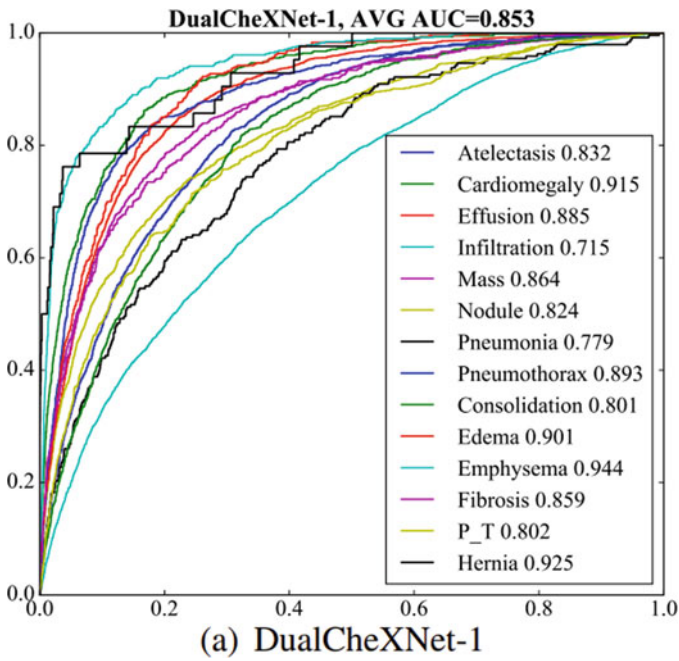


Fig. 7 The ROC curve plot of DualCheXNet

The area under the ROC curve of each pathology is calculated. Later average of all 14 AUC values is used to find the average AUC value of the model. The disadvantage of the above method is its failure in detecting textures of lung portion in an image. Table 1 gives a brief summary of all state-of-the-art deep models using NIH dataset.

4 Discussion and Future Trends

This paper discussed various CNN-based learning algorithms for the classification of multiple diseases belonging to thoracic cavity. It is proved that algorithms' performance reached radiologists' level predictions in most of the pathologies. This could be the main direction toward improved health care for patients of developing countries. However, the limitation of all these deep learning methods is that they are data-hungry. Still, only few chest X-ray dataset are available in digital form. The present dataset has very imbalanced disease class distribution. This imbalance affects model performance on minority class pathology.

Coming to multiple disease classification problems, the future direction is accurate detection of Covid-19 and also degeneration of pathology in the human body, for example, change of Covid-19 to pneumonia, etc. The other difficulty is that thoracic disease patterns in X-ray are very diverse in nature, with many overlaps. In most of the models, localization of pathology is done in a weakly supervised fashion. A lot of work needs to be done on the accurate localization of pathology on chest X-rays.

The selection of appropriate data sampling techniques and loss function also play a critical role in success of diagnosis. There are three sampling techniques. They are under-sampling training data, penalizing the misclassification, and oversampling. Under-sampling the training data results in wasting part of the training dataset, which has already been relatively small in chest X-ray [14]. Penalizing the misclassification is aimed to impose an additional cost on the model for making classification mistakes or misclassifications of the minority class during training. Weighted cross-entropy loss functions or standard binary cross-entropy loss functions are used to do this penalization in [11]. Here the difficulty is manually fixing weights for different misclassifications. A moderate improvement is noticed in their model [11] to overcome data imbalance issues. A type of oversampling technique called adaptive augmentation with unweighted cross-entropy loss function is used in [16]. A good improvement is observed here to overcome the class imbalance issue of the chest X-ray dataset.

Table 1 The overview of all the models

	Description	Avg AUC	Advantage
	Methodology		
Wang et al. [5]	Pretrained with 4 CNN + binary relevance approach	74.2%	First model
Yao et al. [7]	DenseNet and LSTM	80.2%	Correlation among pathology labels is discovered
Rajpurkar et al. [8]	Pretrained with DenseNet-121 with batch normalization and CAM-generated heatmaps	83.5%	Score of pneumonia (76.8%) exceeded predictions of radiologist
Kumar et al. [9]	DenseNet with cascade learning	78.5%	Score of cardiomegaly (91.33%); more than any model
Guan et al. [10]	DenseNet-based pretrained CNN with attention guidance	87.1%	Improvement in small nodule detection
Wang et al. [11]	Grad-CAM-supported ResNet	88.4%	High average AUC
Liu et al. [13]	Segmentation-based deep fusion network	81.5%	The 95.84% lung region identified
Chen et al. [14]	ResNet and DenseNet used in complementary fashion	86.7%	Unique type of feature learning
Hongyu et al. [15]	Triple attention learning-based DenseNet	82.6%	Three-way attention learning
Wang et al. [16]	Adaptive sampling-based DenseNet	82%	Adaptive augmentation to overcome class imbalance issue successfully

5 Conclusion

This survey discussed various artificial intelligence-based works on chest X-ray radiography, along with a brief background. Transfer learning with pretrained CNN models is used more instead of learning from scratch. At the end, fine-tuning was done in end-to-end manner. Out of several pretrained CNN models, DenseNet showed promising results as a feature extractor than ResNet. In most of the models, sigmoid activation functions with a specific number of neurons are kept at the last layer for the purpose of classification. Evaluation metrics such as AUCs are used for the performance analysis of classifiers. Under localization of abnormal regions, Grad-Cam-generated heatmaps gave better accuracy than CAM-generated heatmaps. In future, these AI-assisted diagnoses will provide improved health care to society and decreased workload to medical community.

References

1. K. Yasaka, O. Abe, Deep learning and artificial intelligence in radiology: Current applications and future directions. *PLoS Med.* **15**(11), e1002707 (2018) <https://doi.org/10.1371>
2. G. Chartrand, P.M. Cheng, E. Vorontsov, M. Drozdal, S. Turcotte, C.J. Pal, S. Kadoury, A. Tang, Deep learning: A primer for radiologists. *Radiographics* **37**(7), 2113–2131 (2017)
3. A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **1**, 1097–1105 (2012)
4. I. Allauzi, M.B. Ahmed, A novel approach for multi-label chest X-ray classification of common thorax diseases. *IEEE Access* **7**, 64279–64288 (2019)
5. X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, R.M. Summers, 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, 2017, pp. 2097–2106
6. Open-i: An open access biomedical search engine
7. L. Yao, J. Prosky, E. Poblenz, B. Covington, K. Lyman, Weakly supervised medical diagnosis and localization from multiple resolutions (2018), arXiv preprint arXiv: 1803.07703
8. P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, A. Y. Ng, CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning (2017), arXiv:1711.05225 [cs, stat]. *arXiv: 1711.05225*
9. P. Kumar, M. Grewal, M.M. Srivastava, Boosted cascaded convnets for multilabel classification of thoracic diseases in chest radiographs (2017), arXiv:1711.08760. [Online]
10. Q. Guan, Y. Huang, Z. Zhong, Z. Zheng, L. Zheng, Y. Yang, Diagnose like a radiologist: Attention guided convolutional neural network for thorax disease classification (2018), *arXiv:1801.09927*
11. H. Wang, H. Jia, L. Lu, Y. Xia, Thorax-net: An attention regularized deep neural network for classification of thoracic diseases on chest radiography. *IEEE J. Biomed. Health Inform.* **24**(2), 475–485 (2019)
12. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: Visual explanations from deep networks via gradient-based localization, in Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 618–626
13. H. Liu, L. Wang, Y. Nan, F. Jin, Q. Wang, J. Pu, SDFN: Segmentation-based deep fusion network for thoracic disease classification in chest X-ray images. *Comput. Med. Imaging Graph.* **75**, 66–73 (2019)

14. B. Chen, J. Li, X. Guo, G. Lu, DualCheXNet: Dual asymmetric feature learning for thoracic disease classification in chest X-rays. *Biomed. Signal Process. Control* **53**, 101554 (2019)
15. H. Wang, S. Wang, Z. Qin, Y. Zhang, R. Li, Xia, Triple attention learning for classification of 14 thoracic diseases using chest radiography. *Med. Image Anal.* **67**, 101846 (2020)
16. H. Wang, Y.Y. Yang, Y. Pan, P. Han, Z.X. Li, H.G. Huang, S.Z. Zhu, Detecting thoracic diseases via representation learning with adaptive sampling. *Neurocomputing* **406**, 354–360 (2020)