A patch-based and multi-instance learning strategy for pneumothorax classification on chest X-rays

Yuchi Tian1, Yan Chang2,3 and Xiaodong Yang1,2,3,*

1Academy of Engineering and Technology, Fudan University, Shanghai, 200433, China
2Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou 215163, China
3Jihua Laboratory, Foshan 528000, China

*Corresponding author. Email: xiaodong.yang@sibet.ac.cn

Abstract. Pneumothorax is a lung emergency. Automated computer-aid pneumothorax diagnosis based on chest X-ray can help reduce the diagnostic time and save valuable time for the treatment. A total of 21,759 patient's frontal-view chest X-ray images from one medical center are used in this study. The dataset is divided into two categories: pneumothorax and non-pneumothorax, which are evaluated by two radiologists with over ten years of practical experience. A two-stage training for pneumothorax classification based on multi-instance learning (MIL) are proposed, first training a patch-level classifier, followed by an image-level classifier training, which is initialized with the patch pre-trained weights. The image-level classifier initialized with patch pre-trained weights achieves good classification performance with the F1-score, accuracy and recall of 0.869, 0.915 and 0.843 respectively, which are larger compared to that of the model initialized without patch pre-trained weights (0.785, 0.878 and 0.783). The two-stage training strategy can improve the performance of pneumothorax classification and does not require too high GPU memory and long training time.

1. Introduction

A pneumothorax is an abnormal collection of air coming from the lung in the pleural space. It is a common medical emergency that can be life threatening and requires rapid treatment [1]. According to different studies, the annual incidence of pneumothorax is between 7.4 and 18 per 100,000 men and between 1.2 and 6 per 100,000 women in America [2,3]. The simple standard chest X-rays is recommended for the initial diagnosis and evaluation of pneumothorax [3,4]. As mentioned above, the pneumothorax is a lung emergency, which needs immediate treatment after diagnosis. However, the common method for pneumothorax diagnosis mainly depend on human participation, which implies low efficiency. Therefore, reliable and fully automated computer-aid pneumothorax diagnosis based on chest X-ray is highly desirable.

Recently, the application of convolutional Neural Networks (CNNs) in the medical field is becoming an increasingly important technique to assist clinical diagnosis and treatment planning. More specific uses of CNNs in medical imaging are segmentation, classification, and detection of various anatomical regions of interest [5-11], with a reliable performance throughout. Automatically detecting pneumothorax via CNN has not stayed untouched in research either. Wang X’s team applied four classic neural network models to identify pneumothorax with good results in an AUC of 0.7891 [12]. Cicero et
al discussed the training and validation of a CNN to detect pneumothorax on X-chest radiographs and reached a sensitivity, specificity, and AUC of 0.78, 0.78, and 0.861 respectively [13]. Zhe Li’s research team presented a unified model that jointly models disease identification and localization with limited localization annotation data, with a good result in the classification of pneumothorax [14]. Their researches promote the study of automatic diagnosis of pneumothorax on chest X-ray.

However, with the limitations of the specific CNN framework they used and computation capability they may have, most of these pneumothorax identification methods compressed the training image size from high resolution to low resolution for the training model. Cicero et al downsampled the training image resolution to 224 x 224 pixels; Zhe Li’s research team compressed the training data sizes to 512 x 512 pixels; Wang X’s team resized X-rays images as a 1024 x 1024 bitmap images. While the typical dimensions of pneumothorax X-ray image is usually greater than 3000 pixels in each dimension, compressing the resolution will cause significant loss in detail and decrease the detection accuracy. Furthermore, due to the limited GPU memory and computing hardware capacity, it is almost impossible to deal with high-resolution images directly without downsampling to low resolutions.

In this paper, we propose a method to improve the accuracy of pneumothorax classification by reducing the distortion of pneumothorax information while maintaining low GPU memory and computing requirements. The method is inspired by the curriculum learning [15] approach and multi-instance learning (MIL) [16] strategy. For training, we divide the full chest-X-rays into several patches, relying on whether there is a pneumothorax lesion area to create labels for each patch. We first train patch-level CNN classifiers and then use the learned features to fine-tune an image-level training classifier using MIL strategy. We demonstrate that our approach can effectively relieve the image resolution loss without increasing the burden on computing power and computing hardware, and significantly improves the performance for automatic pneumothorax classification on chest-X-rays.

2. Methods

Unlike those methods [12,13,14] of pneumothorax detection that directly learn the disease feature from a full image. Our approach consists of a two stage training, first a patch-level classifier training, followed by an image-level classifier training using the MIL scheme. An overview of our approach is shown in Figure 1.

2.1. Patch-level classifier

2.1.1. Patch-level training model. Compared to using the image in full resolution with a high distortion after compression, inputting small sized patches results in lower distortion, meaning more information is kept. In addition, it reduces noise and improves the overall data quality. Therefore, we first apply the patch-based training model to learn pneumothorax disease features instead of the entire image training model. The image is divided into a set of patches, each patch contained entirely within the image. The patch including the pneumothorax is assigned to a positive label, while the patch without the pneumothorax is assigned to a negative label. For the patch-level training model, we use the Inception-ResNet-v2 architecture [17], which is a convolutional neural network combining the Inception architecture with residual connections and achieves excellent performance on the test set of the ImageNet classification (CLS) challenge [18]. As shown in Figure 1(a), inputting the patches into the patch-level classifier, Inception-ResNet-v2 bottom layers extract the features for each patch, followed by a global average pooling layer and the sigmoid classification layer to generate a classification result.

2.1.2. Loss function. Due to the serious imbalance of the data ratio — the proportion of negative samples is much larger than that of positive samples, it is challenging to learn positive instances well. In order to balance the loss weigh of each class, each batch contains at least one positive sample, and the average loss of positive and negative samples are calculated respectively in each batch. Finally, the half of their sum is calculated as the final loss.
For example, the Cross Entropy Loss with class balancing strategy for Patch-based classifier is defined as
\[
L_{\text{patch-level}} = -\frac{1}{2} \left( \sum_{j=0}^{J} \log(p_j) + \sum_{k=0}^{K} \log(1 - p_k) \right),
\]
where \( J \) and \( K \) are the total number of positive patches and negative patches in a batch of training data, respectively; \( p \) indicates a patch probability of pneumothorax; where \( j \) and \( k \) are the index of positive and negative sample respectively.

Figure 1. Schematic of the approach consisting of two training stages.

First a patch-level classifier is trained, and the final patch-trained weights of Inception-Resnet-V2 bottom layers in classifier are transferred into Inception-Resnet-V2 bottom layers of image-level classifier for transfer learning, while the layers of top in image-level model are trained from scratch. The image-level prediction for pneumothorax is calculated according to different MIL strategies. The \( Q \) in (b) indicate the instances/patches of the full image.

2.2. Image-level classifier

2.2.1. Image-level training model. In order to prepare for image-level training, we make some changes to the Inception-ResNet-v2 architecture. We keep the Inception-ResNet-v2 bottom section fixed, removing global pooling layers and final sigmoid classification layer, adding four layers at the end of Inception-ResNet-v2 bottom section. As shown in Figure 1(b), inputting the full image into the framework, then outputting a set of abstracted feature maps from Inception-ResNet-v2 bottom section, processed by local average pooling across feature maps with the sliding-window scheme at each channel. After that, the 1x1 convolutions are applied for dimensionality reduction, decreasing the number of feature map to one. An element-wise sigmoid activation function is applied to the feature map, followed by a flatten operation, finally a one-dimensional tensor of predicted pneumothorax probabilities of instances/patches is obtained. The image-level label prediction for pneumothorax is calculated by these scores according to the MIL strategy, which will be described in more detail below.

With a model of this form, an image-level training based on the MIL strategy can be implemented and the model can be fine-tuned using patch-trained weights.
2.2.2. MIL for image-level pneumothorax classification. We treat the image-level training as a multiple instance learning (MIL) problem [19] — each learning image contains a bag of instances/patches instead of a single feature vector.

Given an image \(I\), a set of feature maps of multiple channels is obtained after the Inception-ResNet-v2 bottom layers. Applying local average pooling across feature maps with the sliding-window scheme at each channel, followed by a 1x1 convolutions, an element-wise sigmoid activation function, and flatten operations. Finally a one-dimensional tensor \(p = \{p_1, p_2, p_3, \ldots, p_n\}\) is obtained, where \(p\) is the predicted pneumothorax probability for each instance/patch within the \(I\), corresponding to the flattened patches \(Q = \{Q_1, Q_2, \ldots, Q_n\}\) as shown in Figure 1(b), where \(n\) is the number of patches.

We explore two kinds of MIL strategies for image-level training: probability max-based MIL and probabilities joint-based MIL.

- **Probability max-based multiple instance learning**
  It is assumed that if there is at least one patch in the image belongs to positive instance, the image-level label will be regarded as positive, otherwise it is negative when all instances are negative instances. In other words, the probability of \(I\) is same as the one with the highest probability among all instances. In this assumption, the predicted pneumothorax probability of \(I\) is supposed to be formulized as
  \[
  P(y = 1|I) = \max(p_1, p_2, p_3, \ldots, p_n),
  \]
  where \(P(y|I)\) is the image-level probability, \(p\) is the probability of the patch/instance, \(i\) is the index of a patch probability of pneumothorax, and \(n\) is the total number of patches/instances in the \(I\).

- **Probabilities joint-based multiple instance learning**
  The downside of probability max-based MIL is that it only considers the patch of the max pneumothorax probability and does not exploit information from other patches. The probabilities joint-based MIL takes more patches into consideration for training, which calculates the predicted pneumothorax probability of \(I\) by all patches within the \(I\). We know that there must be at least one patch classified as positive to make the \(I\) a positive label. Therefore, the image-level probability of \(I\) being pneumothorax can be defined as
  \[
  P(y = 1|I) = 1 - \prod_{i=1}^{n}(1 - p_i),
  \]
  where \(P(y|I)\) is the image-level probability, \(p\) is the probability of the patch/instance, \(i\) is the index of a patch probability, and \(n\) is the total number of patches/instances in the \(I\).

2.2.3. Loss function. The Cross Entropy Loss with same class balancing strategy as patch-based classifier is utilized for image-level classifier stage. Regardless of the MIL method used for image-level training, the cross entropy-based cost function for image-level classifier is formulized as follows
  \[
  \mathcal{L}_{\text{image-level}} = -\frac{1}{\mathcal{J}} \sum_{j=0}^{\mathcal{J}} \log(P(y = 1|I_j)) + \frac{1}{\mathcal{K}} \sum_{k=0}^{\mathcal{K}} - \log(P(y = 0|I_k)),
  \]
  where \(P(y = 1|I)\) is the image-level probability of pneumothorax, \(P(y = 0|I)\) indicates the image-level probability of non-pneumothorax; where \(\mathcal{J}\) and \(\mathcal{K}\) are the total number of positive and negative samples in a batch of training data, respectively; where \(j\) and \(k\) are the indexes of positive and negative samples respectively.

3. Experiment

3.1. Dataset
The dataset consists of 22,905 frontal-view chest X-ray images in DICOM format (22,905 patients), which is obtained from the Second Affiliated Hospital Zhejiang University School of Medicine from January 1, 2009 to December 31, 2015. Based on the results of the radiological reports, the dataset is divided into two categories: pneumothorax and non-pneumothorax, which are evaluated by two radiologists with over ten years of practical experience. They diagnose these cases independently, finally the 21,759 images with consistent result are kept for this experiment.
Table 1. The dataset of the experiment for the patch-level classifier and image-level classifier.

| Data   | Training set | Validation set | Test set |
|--------|--------------|----------------|----------|
| Image-level | p:1438       | p:359          | p:235    |
|         | n:14242      | n:3561         | n:1924   |
| Patch-level | p:2583       | p:646          | p:470    |
|         | n:85452      | n:21366        | n:11544  |

We split the dataset into a training set (72%), a validation set (18%) and test sets (10%) for image-level training. The total numbers of images in the training, validation and test sets are: 15680, 3920 and 2159, respectively. For patch-level training, we divide each images into patches using the method described in section 3.2. As a result, the total number of patches in the train, validation and test sets are 88035, 22012, and 12014 respectively (the patches in these sets are divided from the image-level sets correspondingly).

3.2. Data preprocessing
The original image is divided into six patches of the same size for patch-based training. For example, given an image with size $h \times w$, the each patch size will be $(h/3, w/2)$. For non-pneumothorax image, all six patches of it are marked as negative labels. For pneumothorax image, only the patch including the pneumothorax is marked with a positive label, while the patch without the pneumothorax is rejected. The task of patch label assignment is done by two radiologists. The average original images resolution is roughly 3000 x 2500, which is reduced to 512 x 512 after resizing for image-level training. The average resolution of the patches is roughly 1000 x 1250, which is reduced to 512 x 512 as well for patch-based training. The pixel values in each channel are scaled between -1 and 1. There is no data augmentation techniques applied.

3.3. Training

3.3.1. Patch-level training stage. We use Inception-ResNet-v2 model pre-trained on ImageNet to train a patch-based classifier, and only the last classification layer is trained from scratch. It can make the model converge faster than training completely from scratch. The batch-size is set as 8, and an SGD optimizer with learning rate 0.0001 is used; The loss function with class balancing strategy mentioned in Section 2.1 is applied. Each convolution in a block is proceeded by batch normalization [20] followed by ReLU activation. The patch-level training is stopped when the highest F1 score on the validation set is reached.

3.3.2. Image-level training stage. The image-level training model described in Section 2.2 is used as the model. In order to get the predicted pneumothorax probabilities of six patches within the whole image for MIL, it is expected to obtain the tensor in the form of 3x2x1 before the element-wise sigmoid activation function is applied. Because the output of the Inception-ResNet-v2 bottom layers is a set of feature maps with shape 14x14x1536, the output is processed by 3x3 average pooling with a 1x1 stride, followed by 4 x 6 average pooling with 4x6 stride and 1x1 convolutions with a 1x1 stride. Finally a 3x2x1 tensor of predictions of six patches can be obtained after element-wise sigmoid activation. The final patch-trained weights of Inception-Resnet-V2 bottom layers in classifier are transferred into Inception-Resnet-V2 bottom layers of image-level classifier for transfer learning, while the layers of top in image-level model are trained from scratch. To expect the bottom layers be more preserved than the top layers. First training only the top layers for 6 epochs, and then freeze the top layers and just train the bottom layers until achieving the best pneumothorax classification performance on validation set. The network hyper-parameter settings are the same as patch-based classifier training except for the learning rate, which is modified to 0.001.

The experiment code is written in Keras [21], and implemented on an Nvidia GeForce GTX 1070 GPU with 8GB of GDDR5 memory.
3.4. Result
We evaluate the model performance of the classification using the F1-measure score [22, 23], accuracy and recall, and the F1-measure score is used as the indicator to guide when to stop the model training for both patch-level and image-level training. In addition, the precision-recall (PR) curves [29] of different experiments for image-level training are also given.

Table 2 shows that the F1-score of patch-level classifier on the test set is 0.85 after training with 31 epochs reaching the best F1 score on validation set, and the accuracy and recall at this epoch are 0.88 and 0.791 respectively.

Table 2. Test F1 score, accuracy and recall of the patch classifiers using the Inception-Res-V2. The number of epochs indicates the amount of epochs required to reach the best validation F1 score.

| Patch-level model | F1-score | Accuracy | Recall | Epochs |
|-------------------|----------|----------|--------|--------|
| Test set          | 0.850    | 0.88     | 0.791  | 31     |

We conduct three sets of comparative experiments on MIL strategies, pre-trained weights and image resolution for image-level training. Table 3 and 4 demonstrates the performance of different experiments in the image-level classifier. We first try the two MIL strategies for image-level training: Probability Max-based MIL & Probabilities joint-based MIL. The probabilities joint-based MIL achieves a better result than probability max-based MIL on the test set, with F1-scores of 0.869 and 0.824, accuracy of 0.915 and 0.893, recall of 0.843 and 0.804, respectively. Next we examine the influence of pretrained weights. The F1-score, accuracy and recall of the model initialized with patch-trained weights are 0.869, 0.915 and 0.843, which are larger compared to that of the model initialized with ImageNet pre-trained weights (0.785, 0.878 and 0.783). Finally, we compare the performance of three different image resolutions. The model initialized by patched-trained weights with an input size of 512x512 reaches better results of F1-score, accuracy and recall compared to results of 0.79, 0.867, and 0.753 with size of 256x256. The model initialized by ImageNet pre-trained weights with input size of 512x512 obtains the results of F1-score, accuracy and recall are 0.785, 0.878 and 0.783 respectively, which performs similarly to the model using patch pre-trained weights with input size of 256x256, on F1-score and accuracy except for recall, while uses more training time than it, with 39,913s and 33,709s respectively, in addition, the GPU memory allows the model with size 256x256 to set training batch size as 16, which is twice that of model with size 512x512.

Table 3. Test performance of the different experiments in image-level classifier. The number of epochs indicates the amount of epochs required to reach the best validation F1 score.

| MIL          | Pre-trained | Image-size | F1-score  | Accuracy | Recall | Epochs | Batch-size | Time       |
|--------------|-------------|------------|-----------|----------|--------|--------|------------|------------|
| Ps Max-based | Patch-trained | 512x512 | 0.824     | 0.893    | 0.804  | 21     | 8          | 40,290s    |
| Ps joint-based | Patch-trained | 512x512 | 0.869     | 0.915    | 0.843  | 19     | 8          | 36,981s    |
| Ps joint-based | ImageNet      | 512x512 | 0.785     | 0.878    | 0.783  | 20     | 8          | 39,913s    |
| Ps joint-based | Patch-trained | 512x512 | 0.869     | 0.915    | 0.843  | 19     | 8          | 36,981s    |
| Ps joint-based | Patch-trained | 256x256 | 0.790     | 0.867    | 0.753  | 28     | 16         | 33,709s    |
| Ps joint-based | ImageNet      | 512x512 | 0.785     | 0.878    | 0.783  | 20     | 8          | 39,913s    |
Table 4. The precision-recall (PR) curves of different experiments: a. experiment of MIL strategies, b. experiment of pre-trained weights, c. experiment of image resolutions.

4. Discussion

Because the typical chest X-ray image with high resolution needs to be compressed to low resolutions for model training, it can cause significant loss in detail and decrease the detection accuracy. In this paper, we proposed a developed way to reduce the image distortion, which improves the performance of pneumothorax classification. We believe that the efficiency of the classifier training can be improved by inputting multiple smaller sized patches with low distortion.

The basic idea of our approach is to first train a patch-level classifier, meaning the elementary training unit is patch-based instead of the whole image, and then use the final weights of patch-level classifier to initialize the image-level classifier with a MIL strategy. In order to prove the effectiveness of our method, we have set up three sets of comparative experiments on MIL strategies, pre-trained weights, and image resolution. The experiment result as illustrated in Table 3 and 4. In the experiment of different MIL strategies for image-level training, the MIL of probabilities joint-based is more efficient than the probability max-based MIL on the performance of pneumothorax classification. The reason for this result may be that the probabilities joint-based MIL takes more patches information into consideration to train the image-level model, while the probability max-based MIL only exploits information from one patch with the max probability. In the experiment of pre-trained weights, our proposed method performs better than model without patch-level training. A possible explanation for this is the decreased image distortion rate and increased pneumothorax detail. The average original image size is about 3000x2500, but if it is resized to 512x512, its resolution is roughly reduced 24-fold. However, with patch-based training, dividing the whole image into 6 patches of the same size of 1000×1250, each patch will be compressed by only 5-fold after compressing each patch to 512×512. This means that more information can be retained. Finally, we examine the effect of training with images of different sizes on the performance of pneumothorax classification. The result in Table 3 and 4 illustrate that the resolution has a great impact on the classification performance. With higher image resolutions, there will be a better classification performance. It is worth noting that the model initialized by patched-trained weights with an input size of 256x256 obtains a similar performance result to the model using ImageNet pre-trained weights with 512x512 input size obtaining, but uses less training time and bigger batch size, and their PR curves basically overlap as Table 4 shown. The result demonstrates that, with using patch-trained weights, even training with low-resolution images can still achieve similar pneumothorax classification performance as training with higher resolution image without patch-trained initialization, and use less time to train and occupy smaller GPU memory per sample.

Put together, these experiment results are consistent with our hypothesis that patch-based training can reduce the image distortion rate and keep more information about pneumothorax, which improves the efficiency of pneumothorax classification. Besides, we suggest that the two-stage training with probabilities joint-base MIL strategy can further effectively improve the performance of pneumothorax classification.
There are some examples available in literature using a similar method. For example, Hou et al. used two-stage training for the whole slide tissue image classification, first training a patch-based CNN and then predicting the image-level labels by an image-level decision fusion model which is trained on histograms of patch-level predictions [24]. By training a nodule-level CNN model firstly, followed by MIL model, Shen et al. built a network for patient-level lung cancer prediction [25]. These approaches both used probability max-based MIL for image-level training stage, which only considers the patch of max probability. Lotter et al. applied the pre-trained classifier of patch-level as feature extractors in a sliding-window fashion to build an image-level model for mammogram classification [26]. Li Shen utilized two stage training to first train a model to recognize localized patches and then convert it into a whole image classifier for breast cancer diagnosis based on mammograms [27]. For the above two researches, no MIL strategy is used.

Theoretically, in the patch-level training of two-stage training, splitting the whole image into more and smaller sized patches for training can reduce the information loss rate more, but we suggest that the specific number and size of patches split should depends on specific problems such as experimental objective and the characteristics of the disease. For example, if it is an object localization problem, more and smaller patches can not only decrease information distortion, but also improve the localization resolution. However, more patches split means more heavy and finer work of ROI annotations, and also requires more computation during training. In addition, for the disease diagnosis which needs to consider the associated information between local information. The spatial information of disease will be lost if the image is divided into too many patches. In our case, it is a classification task. We divide the pneumothorax into six patches based on the characteristics of pneumothorax on chest X-ray and chest anatomy. Pneumothorax is a lung disease where air is collected in the pleural cavity. In general, it is diagnosed by examining the border between the air and the compressed lung from a chest X-ray film [28]. In chest radiography diagnosis, two horizontal lines are used to indicate the lesion location on the chest X-ray. These are in the front of the second and the forth rib, which divide the lung into three fields: upper field, middle field, and lower field. A pneumothorax may appear in either one of the lungs or in both. In this case, the two lungs can be separated with a longitudinal division. Consequently, according to anatomical feature of the lung and the pathological feature of a pneumothorax in X-ray images, a full image can be divided into six parts without significant effect on the characteristics of the pneumothorax in chest X-ray image. Even though there are some radiographs (full images) that are off center or at an angle, there is no influence on identification of pneumothorax after splitting the radiograph up into patches, due to the characteristics of the pneumothorax in medical imaging are still kept in all patches.

5. Conclusions
In this paper, we propose a two-stage training for pneumothorax classification by reducing the image distortion. Instead of directly feeding the training model with the entire chest-X-ray image in low resolution, our approach first trains a patch-level classifier, which can reduce the image distortion, followed by an image-level classifier training using a scanning window MIL scheme. Experimental results demonstrate with our algorithm, more medical details are kept and the performance of pneumothorax classification is improved without high GPU memory occupied and long training time cost. Furthermore, the effect of image input size and two different MIL schemes, i.e., probabilities max-based MIL and probabilities joint-based MIL, are also investigated in our study. Our work is meaningful in clinical application, which to help reduce the pneumothorax diagnostic time and save valuable time for the treatment. In future work, we plan to extend the current work by localizing the pneumothorax lesion area, and achieve both pneumothorax identification and lesion area localization automatically.

Acknowledgements
The authors acknowledged the funding from Jihua Laboratory under No. X190131TD190 and Technology in Fudan University for lab facilities support.
References

[1] Light RW. Pleural Diseases. 6th ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2013.

[2] Noppen M. Spontaneous pneumothorax: epidemiology, pathophysiology and cause. Eur Respir Rev. 2010;19:217-219.

[3] Sahn SA1, Heffner JE. Spontaneous pneumothorax. N Engl J Med. 2000;342:868-874.

[4] Bintcliffe OJ, Hallifax RJ, Edey A, et al. Spontaneous pneumothorax: time to rethink management? Lancet Respir Med. 2015;3:578-588.

[5] Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017; 542:115–118.

[6] Bar, Y.; Diamant, I.; Wolf, L.; Greenspan, H. Deep learning with non-medical training used for chest pathology identification. In Proceedings of the Medical Imaging 2015: Computer-Aided Diagnosis, Orlando, FL, USA, 3–7 November 2015; Volume 9414.

[7] Choi, H., Jin, K.H. Fast and robust segmentation of the striatum using deep convolutional neural networks. J. Neurosci. Methods. 2016; 274:146–153.

[8] Kumar, D., Wong, A., Claudi, D.A. Lung nodule classification using deep features in CT images. In Proceedings of the 2015 12th Conference on Robot VisionComputer and Robot Vision (CRV), Halifax, NS, Canada, 3–5 June 2015; 133–138.

[9] Samala, R.K., Chan, H.P., Hadjiiski, L.M., Cha, K., Helvie, M.A. Deep-learning convolution neural network for computer-aided detection of microcalcifications in digital breast tomosynthesis. In Proceedings of the Medical Imaging 2016: Computer-Aided Diagnosis, San Diego, CA, USA, 27 February–3 March 2016; Volume 9785.

[10] Lakhani P, Sundaram B. Deep learning at chest radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. Radiology.2017; 84:574-582.

[11] Rajpurkar P, Irvin J, Zhu K et al. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. https://arxiv.org/pdf/1711.05225.pdf. Accessed November 20, 2017.

[12] Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Disease. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017;3462 – 3471.

[13] Cicero M, Bilbily A, Colak E, Dowdell T, Gray B, Perampaladas K, Barfett J. Training and Validating a Deep Convolutional Neural Network for Computer-Aided Detection and Classification of Abnormalities on Frontal Chest Radiographs. Invest Radiol. 2017;52:281-287.

[14] Zhe Li, Chong Wang, Mei Han, Yuan Xue, Wei Wei, Li-Jia Li, Li Fei-Fei. Thoracic Disease Identification and Localization with Limited Supervision. http://cn.arxiv.org/pdf/1711.06373. Accessed November 20, 2017.

[15] Bengio, Y., Louradour, J., Collobert, R., et al.: Curriculum learning. In: ICML(2009)

[16] O. Maron and T. Lozano-Perez, “A framework for multiple-instance learning,” in Advances in Neural Information Processing Systems,1998; 10:570–576.,

[17] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning[C]//Thirty-first AAAI conference on artificial intelligence. 2017.

[18] O. Russakovsky et al., “ImageNet Large Scale Visual Recognition Challenge,” Int J Comput Vis, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[19] Babenko B. Multiple instance learning: algorithms and applications[J]. View Article PubMed/NCBI Google Scholar, 2008: 1-19.

[20] Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv:1502.03167 (2015)

[21] Ketkar N. Introduction to keras[M]/Deep learning with Python. Apress, Berkeley, CA, 2017: 97-111.
[22] Pastor-Pellicer J, Zamora-Martínez F, España-Boquera S, et al. F-measure as the error function to train neural networks[C]//International Work-Conference on Artificial Neural Networks. Springer, Berlin, Heidelberg, 2013: 376-384.

[23] Y. Sasaki. The truth of the F-measure, Sasaki, Version: 26th October, 2007

[24] L. Hou, D. Samaras, T. M. Kurc, Y. Gao, J. E. Davis, and J. H. Saltz. Patch-based convolutional neural network for whole slide tissue image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2424–2433, 2016.

[25] Shen, W., Zhou, M., Yang, F., et al.: Learning from experts: Developing transferable deep features for patient-level lung cancer prediction. In: MICCAI (2016)

[26] Lotter W, Sorensen G, Cox D. A multi-scale CNN and curriculum learning strategy for mammogram classification[M]//Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer, Cham, 2017: 169-177.

[27] Shen L. End-to-end training for whole image breast cancer diagnosis using an all convolutional design[J]. arXiv preprint arXiv:1711.05775, 2017.

[28] O’connor A R, Morgan W E. Radiological review of pneumothorax[J]. Bmj, 2005, 330(7506): 1493-1497.

[29] Takaya S , Marc R , Guy B . The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets[J]. Plos One, 2015, 10(3):e0118432.-.