Automatic Lung Cancer Prediction from Chest X-ray Images Using the Deep Learning Approach

Worawate Ausawalaithong
Kamnoetvidya Science Academy
Rayong, Thailand
a.worawate@gmail.com

Arjaree Thirach
Kamnoetvidya Science Academy
Rayong, Thailand
arjaree.t@kvis.ac.th

Sanparith Marukatat
National Electronics and Computer Technology Center
Patumthani, Thailand
sanparith.marukatat@nectec.or.th

Theerawit Wilaiprasitporn
School of Information Science and Technology
Vidyasirimedhi Institute of Science and Technology
Rayong, Thailand
theerawit.w@vistec.ac.th

***Abstract*—Since, cancer is curable when diagnosed at an early stage, lung cancer screening plays an important role in preventive care. Although both low dose computed tomography (LDCT) and computed tomography (CT) scans provide greater medical information than normal chest x-rays, access to these technologies in rural areas is very limited. There is a recent trend toward using computer-aided diagnosis (CADx) to assist in the screening and diagnosis of cancer from biomedical images. In this study, the 121-layer convolutional neural network, also known as DenseNet-121 by G. Huang et. al., along with the transfer learning scheme is explored as a means of classifying lung cancer using chest x-ray images. The model was trained on a lung nodule dataset before training on the lung cancer dataset to alleviate the problem of using a small dataset. The proposed model yields 74.43±6.01% of mean accuracy, 74.96±9.85% of mean specificity, and 74.68±15.33% of mean sensitivity. The proposed model also provides a heatmap for identifying the location of the lung nodule. These findings are promising for further development of chest x-ray-based lung cancer diagnosis using the deep learning approach. Moreover, they solve the problem of a small dataset.**

I. Introduction

As reported by WHO, cancer caused approximately 8.8 million deaths in 2015 [1]. Almost 20% or 1.69 million of these deaths were due to *lung cancer* [1]. Cancer screening plays an important role in preventive care because it is most treatable when caught in the early stages. This study shows that the appearance of malignant lung nodules more commonly demonstrate a spiculated contour, lobulation, and inhomogeneous attenuation [2].

Presently, low dose computed tomography (LDCT) plays an important role in lung cancer screening [3]. LDCT screening has reduced lung cancer deaths and is recommended for high-risk demographic characteristics [3]. Results from LDCT screening may be further evaluated with standard dose computed tomography (CT) [4]. However, there are many barriers to implementing LDCT screening, such as providers' anxiety concerning the access to LDCT equipment and the potential financial burden on rural populations [3]. Moreover, rural populations have limited access to both primary care physicians and specialists [3]. On the other hand, chest x-rays are readily available in rural areas. Nonetheless, chest x-rays produce lower quality images compared to LDCT or CT scans and, therefore, a lower quality diagnosis is generally expected. This study explores the use of chest x-rays with a computer-aided diagnosis (CADx) system to improve lung cancer diagnostic performance.

The convolutional neural network (CNN) is proven to be very effective in image recognition and classification tasks. The development of CNNs starts from, LeNet [5], AlexNet [6], ZFNet [7], VGG [8], Inception [9] [10], ResNet [11], Inception-ResNet [12], Xception [13], DenseNet [14], and NASNet [15]. There are many studies on the use of deep CNNs to detect abnormalities in chest x-rays. For instance, M. T. Islam et al., [16] use several CNNs to detect abnormalities in chest x-rays. There is also a study by X. Wang et al., on the use of CNNs to detect thoracic pathologies from chest x-ray images. Their study also provides a large dataset as is the case in this study [17]. Among current research, some studies on the application of Densely Connected Convolutional Networks (DenseNet) [14] to detect thoracic pathologies such as ChexNet [18] and the Attention Guided Convolutional Neural Network (AG-CNN) [19]. Both studies train the neural network on a very large chest x-ray image dataset.

Lung cancer prediction with CNN faces the problem of small sample size. Indeed, CNN contains a large number of parameters for adjustment on a large image dataset. In practice, researchers often pre-train CNNs on *ImageNet*, a standard image dataset containing more than one million images. The trained CNNs are then adjusted on a specific target image dataset. Unfortunately, the available lung cancer image dataset is too small for this transfer learning to be effective, even with a data augmentation trick. To alleviate this problem, the idea of applying transfer learning was proposed several times to gradually improve the performance of the model. In this work, transfer learning is applied twice. Firstly to transfer the model from a general image into the chest x-ray domain, and secondly to transfer the model for lung cancer. Such multi-transfer learning can solve the problem of a small sample size

978-1-5386-5724-9/17/$31.00

TABLE I
TRAINING-VALIDATION-TEST SPLIT FOR EACH DATASET

	Training	Validation	Test
JSRT Dataset	80	10	10
ChestX-ray14 Dataset			
Positive cases	4992	1024	266
Negative cases	103907	1024	266
Total	108899	2048	532

and achieves a better task result compared to the traditional transfer learning technique.

Furthermore, this study also shows that it is possible to modify the model to compute a heatmap showing the predicted position of lung cancer in the chest x-ray image.

II. METHODOLOGY

In this section, the methodology used in this study is discussed. Firstly, through describing the datasets used, and subsequently the data preparation process. An explanation is provided of the model architecture and the loss function, culminating in the visualization process.

A. Datasets

1) JSRT Dataset [20]: This public dataset from JSRT (Japanese Society of Radiological Technology) consists of 247 frontal chest x-ray images; 154 of which have lung nodules (100 malignant cases, 54 benign cases) and 93 are images without lung nodules. All images have a pixel size of 2048 2048. Since this dataset is very small, 10-fold cross-validation was performed to increase performance reliability. Malignant cases were considered as positive, while both benign and non-nodule cases were considered as negative. Data were randomly split into the training, validation, and test set, with the ratio shown in Table I.

2) ChestX-ray14 Dataset [17]: This public dataset released by Wang et al. [17] contains 112,120 frontal chest x-ray images, currently the largest of its kind. Each image is individually tagged with up to 14 different thoracic pathology labels. All images have a pixel size of 1024×1024 pixels. However, this dataset does not contain any lung cancer images. The fourteen thoracic pathology labels consist of Atelectasis, Consolidation, Infiltration, Pneumothorax, Edema, Emphysema, Fibrosis, Effusion, Pneumonia, Pleural Thickening, Cardiomegaly, Nodule, Mass, and Hernia. It was decided to use this dataset to compensate for there being only 100 cases of lung cancer data, firstly by training for the recognition of nodules, considering nodule cases as positive and all remaining cases (non-nodule) as negative. Data were randomly split into the training, validation, and test set, without having any overlap of patients between data sets, as shown in Table I.

B. Data preparation

Data preparation was applied to all images, consisting of the following four steps:

- Step 1: increasing the contrast of all images using histogram equalization. This allows the intensity of images to be normalized from different datasets.
- Step 2: removing the noise from all images using median filtering with a window size of 3×3.
- Step 3: resizing images to 224×224 pixels to match the input of the model used in this study.
- Step 4: normalizing image color based on the mean and standard deviation of the ImageNet training set [21].

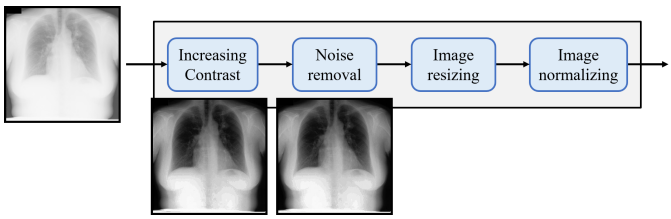

Fig. 1. Illustration of the data preparation process on an example image from the JSRT dataset.

An example of the image after processing in each step is shown in Fig. 1.

C. CNN architecture and transfer learning

Among the convolutional neural networks architectures such as Inception [9] [10], ResNet [11], and DenseNet [14], the model used in this research is the 121-layer Densely Connected Convolutional Network (DenseNet-121). This CNN has dense block architecture that improves the flow of data along the network, and also solves the vanishing gradient problem found in very deep neural networks. This CNN also performs very well on many public datasets including ImageNet [14]. The original model processes an input image pixel size of 224×224 and outputs posterior probabilities for 1,000 object categories. In this work, the last fully connected layer of the CNN was replaced by a single sigmoid node in order to output the probability of having the specified pathology.

Transfer learning was considered due to the small dataset problem. However, as the lung cancer dataset is truly small, transfer learning was applied more than usual. Specifically, in this study, it was applied twice. The first transfer learning allowed chest x-ray images to be classified as "with nodule" or "without nodule". The second allowed chest x-ray images to be classified as "with malignant nodule" and "without malignant nodule" (including chest x-rays with benign nodules and those without). Through the training process, the model became more specific to the lung cancer task. The overall training process and the model used in each step are described below and in Fig. 2, overleaf:

- Base Model: DenseNet-121 with the initial weight from the pre-trained model on the ImageNet training set having the last fully connected layer modified to one class of sigmoid activation function.
- Retrained Model A: Retrained Base Model with the least validation loss from the ChestX-ray14 validation set. The Base Model was retrained using the ChestX-ray14 training set, as shown in Table I, where positive cases were chest x-ray images with nodules and negative

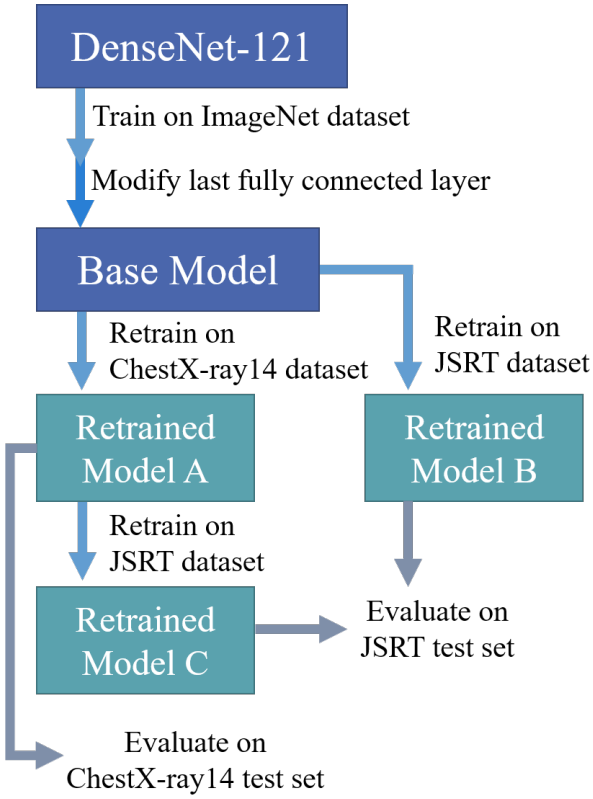

Fig. 2. Training process and models used, including Base Model, Retrained Model A, B, or C, along with the data for training each model.

cases represented chest x-ray images without nodules. The training images were randomly flipped horizontally.

- Retrained Model B: Retrained Base Model with the least validation loss from the JSRT validation set. The Base Model was retrained using the JSRT training set, as shown in Table I, along with 10-fold cross-validation. The training images were randomly rotated by an angle within 30 degrees and randomly flipped horizontally. Positive cases were chest x-ray images with malignant nodules, while negative cases represented chest x-ray images without malignant nodules.
- Retrained Model C: Retrained Model A with the least validation loss from the JSRT validation set. Since the positive cases in the JSRT dataset were all nodules, the Retrained Model A, (which already had the ability to classify "nodule" from "non-nodule", a task similar to the lung cancer classification task) was retrained to identify malignant vs. non-malignant nodules. The JSRT training set was used to retrain the model. All the JSRT training images used to retrain the model were randomly rotated within 30 degrees and randomly flipped horizontally.

The basic idea for the training of Model C is as follows: Firstly, DenseNet-121 was trained on the ImageNet dataset to gain basic knowledge about images in general. Next, the model was retrained on the ChestX-ray14 dataset to adjust this knowledge toward chest x-ray images with additional information about lung nodules. As the size of the ChestX-ray14 dataset is much larger than that of the JSRT dataset, transferring knowledge to the chest x-ray domain first allows better adaptation of the CNN trained on a general image dataset. Model A can be considered as a chest x-ray specialist. Additionally, as lung nodules are very similar to lung cancer, Model A is ready to be adjusted using the JSRT dataset. Experimentally, Model C has been found to show the best performance compared to single-step transfer learning (i.e. Model B). This underlines the effectiveness of the proposed strategy.

D. Loss and Optimizer

The two transfer learning methods involve imbalanced binary classification datasets. To counteract this problem, the following weighted binary classification loss was used:

$$\begin{aligned} L(X,y) &= -\omega_+ \cdot y \log p(Y=1|X) \\ &\quad -\omega_- \cdot (1-y) \log p(Y=0|X), \end{aligned} \tag{1}$$

where X is the image and y is the real label of the image which is labeled 0 for a negative case, or 1 for a positive case, $p(Y=i|X)$ is the predicted probability of having label i. Whereas ω is the weight applied on the loss to make the training process more efficient, ω_+ is the number of negative cases overall and ω_- is the number of positive cases overall.

The Adam (adaptive moment estimation) optimizer [22] was used in this work with the standard parameter setting, i.e. $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The initial learning rate was 0.001, decreasing by a factor of 10 when validation loss plateaued. The batch size was 32.

E. Class Activation Mappings (CAMs)

Class Activation Mappings (CAMs) [23] can be derived from Model C to show the most salient location on the image used by the model to identify the output class. This is done using the following equation:

$$M_c = \sum_k w_{c,k} f_k. \tag{2}$$

From equation (2), M_c is the map showing the most salient features on the image used by the model to classify classes. Whereas $w_{c,k}$ is the weight in the last fully connected layer of the feature map k leading to class c and f_k is the kth feature map.

III. Results and Discussion

The performance of the model in this study was evaluated using accuracy, specificity, and sensitivity. The accuracy shows the degree to which the model correctly identified both positive and negative cases. The specificity shows the degree to which the model correctly identified negative cases and sensitivity shows the degree to which the model correctly identified positive cases. By having high accuracy, specificity and sensitivity, it can be implied that the model has low error potential. The performance of Model A on classifying lung nodules was evaluated for accuracy, specificity and sensitivity using a test

TABLE II
NAME OF THE RETRAINED MODEL, DATASET USED, PURPOSE, ACCURACY, SPECIFICITY, AND SENSITIVITY OF RETRAINED MODELS A, B, AND C ON TASKS

Retrained Model	Dataset used	Purpose	Accuracy	Specificity	Sensitivity
A	ChestX-ray14	to recognize lung nodules	84.02%	85.34%	82.71%
B	JSRT	to recognize lung cancer	$65.51 \pm 7.67\%$	$80.95 \pm 20.59\%$	$45.67 \pm 21.36\%$
C	ChestX-ray14 and JSRT	to recognize lung cancer	$74.43 \pm 6.01\%$	$74.96 \pm 9.85\%$	$74.68 \pm 15.33\%$

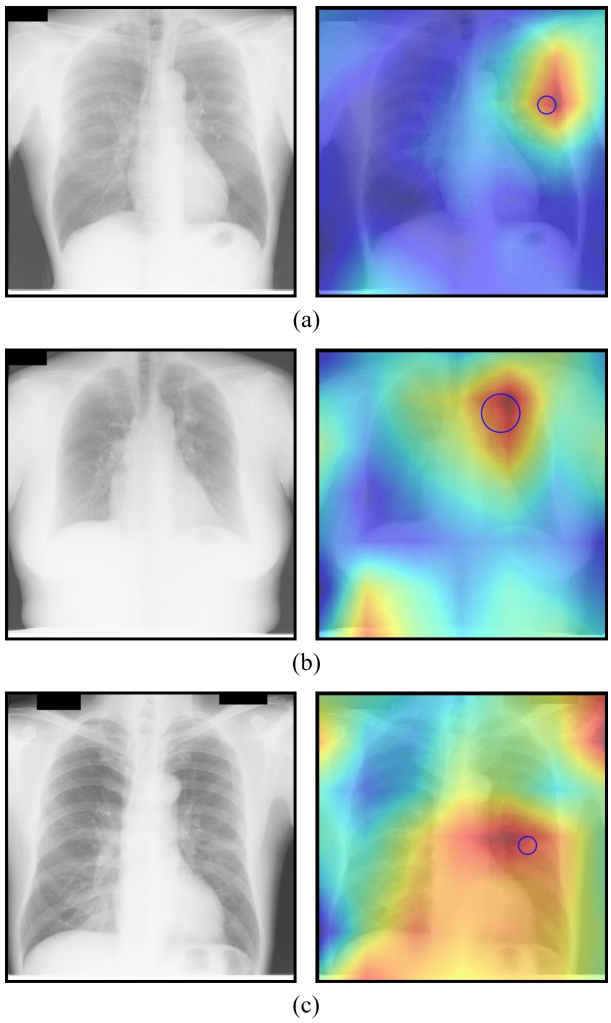

Fig. 3. Images of malignant cases from the JSRT dataset, before and after illustrating Class Activation Maps using the Retrained Model C.

set of the ChestX-ray14 Dataset as shown in Table II. The threshold of Model A is 0.55 as it gives highest value between specificity and sensitivity. The performance of Models B and C on classifying lung cancer were likewise evaluated applying average and standard deviations of accuracy, specificity and sensitivity using a test set of the JSRT dataset for 10-fold cross-validation as shown in Table II. The threshold of both models is 0.5.

As shown in Table II, after being pre-trained on the ImageNet dataset, Model A, which was trained to recognize lung nodules, performed efficiently on the test set of the ChestX-ray14 dataset. Next, Model B showed higher specificity but poorer accuracy and sensitivity than Model C. In addition, Model C also had lower standard deviation in all evaluation metrics. The findings show that retraining the model several times for specific tasks gives better results in almost all metrics.

In Fig.3, the left-hand images are original while the right-hand images are processed using CAMs to show the salient position, with the blue circle corresponding to the actual location of lung cancer, identified by information received using the JSRT images. Model C has the ability to show accurate CAMs in most of the correctly predicted images such as (a) and (b), but some images show too large an area over the actual location or show an inaccurate area. The reason for this is that the model still overfits the training set slightly due to the size of the dataset. On the other hand, CAMs illustrated by Model B do not show the accurate position of lung cancer and often show that the model used too large an image area to classify lung cancer.

IV. CONCLUSION

To conclude, the performance of the densely connected convolutional network in detecting lung cancer from chest x-ray images was explored. Due to the dataset being too small to train the convolutional neural network, a strategy was proposed to train the deep convolutional neural network using a very small dataset. The strategy involved training the model several times to learn about the final task step-by-step, which in this case, starts from the use of general images in the ImageNet dataset, followed by identifying nodules from chest x-rays in the ChestX-ray14 dataset, and finally identifying lung cancer from nodules in the JSRT dataset. The proposed training strategy performed better than the normal transfer learning method, resulting in higher mean accuracy and mean sensitivity but poorer mean specificity, as well as lower standard deviation. The region used to classify images, was illustrated using CAMs in Model C, showing quite accurate locations of lung cancer, although some were not due to the overfit problem. However, CAMs from the Retrained Model B were much poorer. Therefore, it can be concluded that the proposed method can solve the problem of a small dataset.

For future work, data augmentation could be additionally performed by randomly adding Gaussian Noise and cropping the image. The idea of an Attention-Guided Convolutional Neural Network (AG-CNN) [19] can also be used to crop the nodule part from the full image and use its real size to identify malignancy, since the appearance of malignant nodules is different to benign. Additionally, more features such

as family history and smoking can be applied in training along with images for more accuracy. The ensemble of several CNNs can further improve results. Furthermore, cooperation with doctors and radiologists can help to identify the performance of predicted results.

REFERENCES

[1] W. H. Organization, "Cancer," 2018. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs297/en/

[2] C. Zwirewich, S. Vedal, R. R Miller, and N. L Mller, "Solitary pulmonary nodule: high-resolution ct and radiologic-pathologic correlation," vol. 179, pp. 469–76, 06 1991.

[3] W. D. Jenkins, A. K. Matthews, A. Bailey, W. E. Zahnd, K. S. Watson, G. Mueller-Luckey, Y. Molina, D. Crumly, and J. Patera, "Rural areas are disproportionately impacted by smoking and lung cancer," *Preventive Medicine Reports*, vol. 10, pp. 200 – 203, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2211335518300494

[4] T. Kubo, Y. Ohno, D. Takenaka, M. Nishino, S. Gautam, K. Sugimura, H. U. Kauczor, and H. Hatabu, "Standard-dose vs. low-dose ct protocols in the evaluation of localized lung lesions: Capability for lesion characterizationilead study," *European Journal of Radiology Open*, vol. 3, pp. 67 – 73, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2352047716300077

[5] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov 1998.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," pp. 1097–1105, 2012.

[7] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 818–833.

[8] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," vol. abs/1409.1556, 2014. [Online]. Available: http://arxiv.org/abs/1409.1556

[9] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 00, June 2015, pp. 1–9. [Online]. Available: doi.ieeecomputersociety.org/10.1109/CVPR.2015.7298594

[10] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 2818–2826.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[12] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *AAAI*, 2017.

[13] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1800–1807, 2017.

[14] G. Huang, Z. Liu, L. v. d. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017, pp. 2261–2269.

[15] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," vol. abs/1707.07012, 2017. [Online]. Available: http://arxiv.org/abs/1707.07012

[16] M. T. Islam, M. A. Aowal, A. T. Minhaz, and K. Ashraf, "Abnormality detection and localization in chest x-rays using deep convolutional neural networks," vol. abs/1705.09850, 2017. [Online]. Available: http://arxiv.org/abs/1705.09850

[17] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[18] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," vol. abs/1711.05225, 2017. [Online]. Available: http://arxiv.org/abs/1711.05225

[19] Q. Guan, Y. Huang, Z. Zhong, Z. Zheng, L. Zheng, and Y. Yang, "Diagnose like a radiologist: Attention guided convolutional neural network for thorax disease classification," vol. abs/1801.09927, 2018. [Online]. Available: http://arxiv.org/abs/1801.09927

[20] J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K.-I. Komatsu, M. Matsui, H. Fujita, Y. Kodera, and K. Doi, "Development of a digital image database for chest radiographs with and without a lung nodule: Receiver operating characteristic analysis of radiologists' detection of pulmonary nodules," vol. 174, pp. 71–4, 02 2000.

[21] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. S. Bernstein, A. C. Berg, and F. Li, "Imagenet large scale visual recognition challenge," vol. abs/1409.0575, 2014. [Online]. Available: http://arxiv.org/abs/1409.0575

[22] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," vol. abs/1412.6980, 2014. [Online]. Available: http://arxiv.org/abs/1412.6980

[23] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 2921–2929.