AI Chest 4 All

P. Thammarach, S. Khaengthanyakan, S. Vongsurakrai, P. Phienphanich, P. Pooprasert, A. Yaemsuk, P. Vanichvarodom, N. Munpolsri, S. Khwayotha, M. Lertkowit, S. Tungsagunwattana, C. Vijitsanguan, S. Lertrojanapunya, W. Noisiri, I. Chiawiriyabunya, N. Aphikulvanich, and C. Tantibundhit*

Abstract— **AIChest4All is the name of the model used to label and screening diseases in our area of focus, Thailand, including heart disease, lung cancer, and tuberculosis. This is aimed to aid radiologist in Thailand especially in rural areas, where there is immense staff shortages. Deep learning is used in our methodology to classify the chest X-ray images from datasets namely, NIH set, which is separated into 14 observations, and the Montgomery and Shenzhen set, which contains chest X-ray images of patients with tuberculosis, further supplemented by the dataset from Udonthani Cancer hospital and the National Chest Institute of Thailand. The images are classified into six categories: no finding, suspected active tuberculosis, suspected lung malignancy, abnormal heart and great vessels, Intrathoracic abnormal findings, and Extrathroacic abnormal findings. A total of 201,527 images were used. Results from testing showed that the accuracy values of the categories heart disease, lung cancer, and tuberculosis were 94.11%, 93.28%, and 92.32%, respectively with sensitivity values of 90.07%, 81.02%, and 82.33%, respectively and the specificity values were 94.65%, 94.04%, and 93.54%, respectively. In conclusion, the results acquired have sufficient accuracy, sensitivity, and specificity values to be used. Currently, AIChest4All is being used to help several of Thailand's government funded hospitals, free of charge.**

Clinical relevance— **AIChest4All is aimed to aid radiologist in Thailand especially in rural areas, where there is immense staff shortages. It is being used to help several of Thailand's goverment funded hospitals, free of charege to screening heart disease, lung cancer, and tubeculosis with 94.11%, 93.28%, and 92.32% accuracy.**

I. INTRODUCTION

Heart disease is the leading cause of death worldwide. It is responsible for 17.9 million lives globally, accounting for 31% of total deaths worldwide. In Thailand, 430,000 people are diagnosed with heart diseases each year, and 21,000 Thai people are predicted to die from it. Additionally, lung cancer is the 2nd most type of cancer for men and women. Annually there are 2.1 million newly registered cases of lung cancer, and 1.7 million deaths per year. In our focused region, there are 170,000 new cases of lung cancer every year, and 110,000 deaths per year. Tuberculosis is a bacterial infection that is the leading cause of death for infectious diseases, taking 1.3 million lives each year, with 93,000 new cases yearly in our region with a mortality rate of 13%. Intrathoracic diseases are diseases related to the thorax, such as an abnormal mediastinum, while extrathoracic abnormalities are abnormalities around the chest and thorax area.

One method of detecting and diagnosing diseases in a cost-effective and efficient manner is via using a chest X-ray. X-ray is a radioactive electromagnetic wave that produces an image on a film when it comes into contact with body tissues and bones. Normally, chest X-ray images will be analysed and diagnosed by radiologists for accurate findings. However, there are shortages in radiologists in various developing and developed countries, which can lead to untimely diagnosis resulting in unnecessary death rates.

There are three main approaches to automating chest X-ray interpretations: image processing, machine learning, and deep learning. Image processing is a method used to perform operations on an image in order to enhance the image or retrieve useful information using several methods. However, image processing still lacks the ability to learn features from data, which leads to the invention of machine learning. Machine learning involves using artificial intelligence to improve its accuracy by training. Generally, machine learning algorithms tend to have better accuracy and are more efficient than image processing methods, with the main advance being the ability to extract features (handcrafted).

However, machine learning still lacks the ability to extract features automatically, and some key features of images may be missed. As a result, when using machine learning, time, money, and a specialized technician is needed to supervise the machine. This leads to the development of deep learning with the ability to automate feature extraction with high precision. Deep learning structures differentiate algorithms into layers that are in order, forming a neural network that can automatically evaluate and make intelligent decisions such as diagnosing diseases in chest X-rays.

In image processing, one of the first methods to detect diseases was introduced by automatically constructing a procedure based on a sample figure by choosing a procedure frame and constructing a detailed procedure based on the frame chosen. This was then developed to a directional contrast filter to compare and detect abnormal patterns in X-ray images, with results of 71.30% accuracy, 76.70% sensitivity, and 69.20% specificity [1]. Al-Tarawneh looked into enhancing image quality and efficiently accurate method

*C. Tantibundhit (e-mail: tchartur@engr.tu.ac.th), P. Thammarach, S. Khaengthanyakan, P. Phienphanich, A. Yaemsuk, and P. Vanichvarodom are with Center of Excellence in Intelligence Informatics, Speech, and Language Technology, and Service Innovation (CILS), Faculty of Engineering. S. Vongsurakrai is with CILS and Shrewsbury International School Bangkok, Thailand. P. Pooprasert is with CILS and Faculty of Medicine, Cardiff University School of Medicine, United Kingdom. N. Munpolsri, S. Khwayotha, M. Lertkowit, and I. Chiawiriyabunya are with Udonthani Cancer Hospital, Thailand. S. Tungsagunwattana, C. Vijitsanguan, S. Lertrojanapunya, and W. Noisiri are with Central Chest Institute of Thailand. N. Aphikulvanich is with Department of Medical Services, Thailand.

978-1-7281-1990-8/20/$31.00 ©2020 IEEE

to extract features from the image [2]. He proposed a method that enhances images using Gabor filter, with up to a 38.03% enhancement of images, then the images are segmented using a marker controlled watershed approach. The extracted feature is then analyzed and given a diagnosis. Chaudhary and Singh utilized MATLAB for a thresholding approach. MATLAB was used to increase accuracy and quality of the thresholding approach, as well as being able to support several filters for enhancing images such as the Gabor filter [3].

Lingayat and Tarambale developed a computer-based system that extracted maximum features from segmented, suspicious areas from X-ray images [4]. The system's ability to maximize features increases accuracy of the diagnosis, and the system can also differentiate between malign and benign tumors. Machine learning algorithms were used to help radiologists diagnose diseases. Ballard and Sklansky proposed a method to scan and consolidate images into different resolutions [5]. After image enhancement, the images were analyzed by a hierarchic tumor recognition process, with 83% of images scanned having 100% in detecting and evaluating nodules. However, this ladder structured decision tree can only one shape of curve, and storage required exponential increases with the number of parameters in that curve. Brown et al. used a modular system architecture to incorporate the model, image processing routines, and the inference engine (for decision making and control) and blackboard (for communication) to automatically identify lung boundary and abnormal features in the lung [6]. The proposed method gained a specificity of 95% and a sensitivity of 88%. However, detection can be inaccurate with images with a low contrast edge. Murphy et al. looked into an algorithm using local features of shape index and curvedness to detect structures, which possibly have nodules [7]. Two consecutive k-NN classifiers were then applied to the structure to reduce the number of false positives. This method had a sensitivity of 80%, and over 90% of true nodules were correctly identified.

Sivakumar and Chandrasekar developed an efficient lung nodule detection scheme by performing nodule segmentation through fuzzy clustering models, then classifying the images using a support vector machine (SVM) [8]. There are different kernel types that support the SVM, with the best result being the Radial Basis Function (RBF) kernel, with 80.36% accuracy, 76.47% specificity and 82.05% sensitivity. Lastly, Santosh and Antani proposed a method by analyzing lung region symmetry by using multi scale shape features and edge plus texture features to exploit representation of lung regions and take internal content structure [9]. Then, three classifiers: Bayesian network, random forest, and multilayer perception were employed to classify images. Results showed that the best classifier was the Bayesian network, with an accuracy of 91% and area under ROC curve of 0.96.

Deep learning was then introduced to the scientific community, Bar et al. used a Decaf pretrained CNN and baseline descriptors GIST and BoVW with a feature extractor to detect pathologies, with maximum AUC of 0.94 [10]. Lakhani and Sundaram combined with 2 CNNs, AlexNet and GoogLeNet to classify for tuberculosis in chest X-ray images [11]. This method had AUC of 0.99, sensitivity of 97.3%, and specificity of 100%. However, when the 2 CNNs disagree, a proper radiologist has to evaluate the picture himself. Rajpurkar et al. proposed a method by training Densenet, a 121 layer CNN to evaluate and detect pneumonia. Overall, it had a better record than the average radiologists, increasing its reliability [12]. Wang et al. then proposed a multi purpose, end to end trainable multi task CNN-RNN framework to classify images using image features and extracted text embeddings [13]. Furthermore, the proposed method had an average AUC of over 0.9 and also capable of reporting on the images classified. Rajpurkar et al. proposed a method to detect pneumonia from chest X-ray images by using a 121-layer CNN, which is trainable, called CheXNet [14]. Results showed that CheXNet performs better than the average radiologist. Lastly, Irvin at al. presented a dataset of 224,316 chest X-ray images from 65,240 patients, called CheXpert, as well as designing a labeler that automatically detects 14 observations based on the results, with 3 main steps: mention extraction, mention classification, and mention aggregation, and results shows that CheXpert performs better than two thirds of radiologists, with the AUCs of each observation ranging from 0.85 to 0.97 [15].

Our objective in this paper is to develop an automated X-ray screening algorithm that uses deep learning to identify and classify chest X-ray images into 6 different categories, including: no finding, lung cancer, pulmonary tuberculosis, heart disease, intrathoracic abnormal findings (abnormal mediastinum, abnormal ling disease, and abnormal pleura and pleural cavity), and extrathoracic abnormal findings (abnormal bony structures and other non-specific findings of chest wall and upper abdomen).

Even though our work is similar to CheXpert [15] and many others [1], [14], we specifically highlighted diseases that are critical in Thailand, unlike CheXpert that classifies X-ray images into 14 observations, therefore it will be more beneficial to areas where these diseases are widespread, and places that are lacking in radiologists. Overall, there are several diseases that are critical in rural areas of Thailand, such as lung cancer, tuberculosis, and heart diseases. Chest X-ray is a useful method to help diagnose these diseases but there is a shortage in the number of radiologists who can interpret them. In the past, we have used several methods to detect and classify different diseases from X-ray images, such as image processing and machine learning techniques, as well as deep learning. We propose a method to use deep learning to detect and classify diseases based on the severity of the diseases in our region, in order to help prevent deaths and to aid radiologists in reading each image.

The remainder of this paper is organized as follows: Section II describes the proposed dataset and labeling. Section III and IV shows the methodologies and experimental results, respectively. Discussion and conclusion are given in Section V and future work in VI.

TABLE I: THE DATASET USED.

Category	Montgomery	Shenzhen	NIH	CCIT	UDCH
No finding	80	326	60,361	24,017	17,152
Suspected active TB	58	336	0	1,315	8,475
Suspected lung malignancy	0	0	11,207	2,612	2,661
Abnormal heart and great vessels	0	0	4,952	1,448	9,026
Intrathoracict abnormal findings	0	0	43,201	7,970	18,430
Extrathoracict abnormal findings	0	0		9,444	6,774
Total	138	662	110,994	41,074	48,659

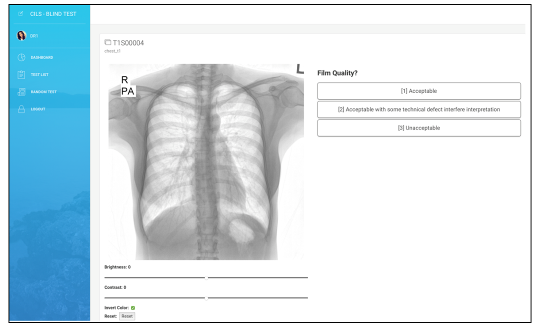

Fig. 1: Example page of the blind test system.

II. DATA

A. Dataset

The chest X-ray images used in this paper are datasets from the National Institutes of Health (NIH), and the Shenzhen and Montgomery. The images in the dataset are from the National Institute of Health and were classified into 14 findings and are grouped into 4 main categories, consisting of no finding, suspected lung malignancy, abnormal heart and great vessels, and other abnormal findings, whereas the images in the Shenzhen and Montgomery dataset were classified into two sets: no finding and suspected active tuberculosis. Another set of images were from patients in our focus area, collected from Udonthani Cancer Hospital (UDCH) and Chest Institute of Thailand (CCIT), and these images were classified into 6 main categories: no finding, suspected active tuberculosis, suspected lung malignancy, abnormal heart and great vessels, Intrathoracic abnormal findings and extrathoracic abnormal findings. The total number of information used is shown in the Table I, where the NIH dataset was used for training under the categories intrathoracic abnormal findings and extrathoracic abnormal findings.

B. Data Labeling

The datasets of images of patients in Thailand were labelled by professional radiologists via blind test system. The value of the intereater reliability acquired, tested by Krippendorff's alpha [16], 0.86, showed that every radiologist shares similar results in the blind test. In this system, each image was randomly given to 3 of 7 radiologists. The radiologists were asked 4 questions about the image: 1) Is the film quality acceptable or unacceptable? as shown in Fig. 1. 2) Is the image normal or abnormal? 3) Are abnormalities consistent with these diseases (suspected active TB, suspected lung malignancy, abnormal heart and great vessels and other abnormal findings)?, in which there can be multiple answers. 4) Are there any other abnormal findings (abnormal mediastinum, abnormal lung disease, abnormal pleura and pleural cavity, normal bony structures, and other non-specific findings of chest wall and upper abdomen)? The answer can also contain multiple abnormal findings. The images used to create a model has to have at least two out of three radiologists confirming that the quality of the film is acceptable, and all other questions have to be answered identically. Overall, a total of 101,783 images were inserted in the blind test system, and 89,733 images are valid for to be inserted into the model.

III. METHODS

A. DenseNet

The model of deep learning used in our methodology is the 121-layer Densely Connected Convolutional Network (DenseNet-121), which connects each layer by using the feed-forward fashion feature, and can also adjust the number of layers to increase depth. Furthermore, DenseNet-121 can solve the vanishing-gradient problem, and uses minimal memory and calculations, yet achieving high efficiency and quality [17].

B. Data preprocessing

Chest X-ray images have to be pre-processed before being put into the model. By using a technique called Contrast Limited Adaptive Histogram Equalization (CLAHE), an improvement from Histogram Equalization (HE), the image was converted into grayscale, and was evaluated. If the score is higher than the average score for the images, the excess value will be spread into a histogram that affects every pixel on the image, increasing the quality of the images, using this as the first input for the pre-trained model. The original image that was used in the CLAHE was inverted and used as the second input for the pre-trained model, and both images were resized to have a size of 224 $\times$ 224 pixels, and it is ready to be used in the model.

C. Model implementation

DenseNet-121 [17] was used as a pre-trained model, using two initialized weights on ImageNet [18], removing out the entire last layer, using it for feature extraction. The two initialized weights were then connected with Global Average Pooling 2D, and connected with fully connected artificial neural network (ANN), containing two hidden layers as a classifiers. The results the model are probabilities that the chest X-ray images will be diseased. For model setup, learning rate of 10^{-3} and Adam optimizer were used, with the highest number of epochs possible, one hundred, and an early stopping code was used for validation. The total value

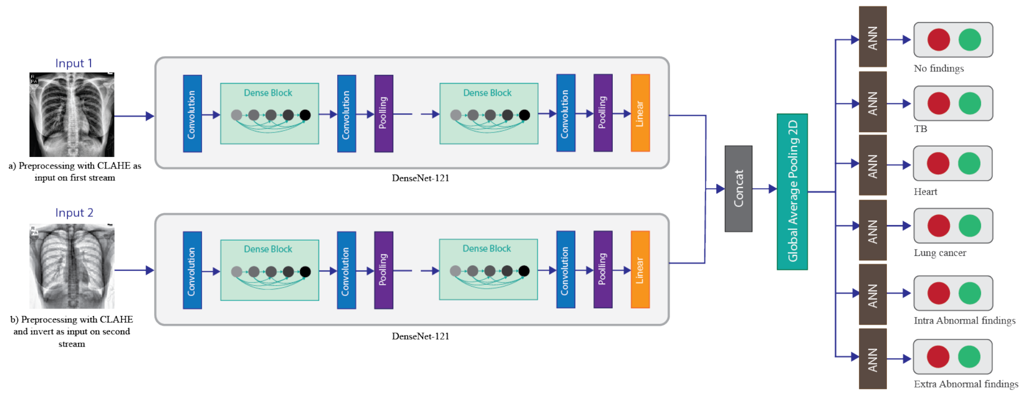

Fig. 2: AIChest4All model architecture.

of training loss and validating loss were combined, giving a ratio bigger than 0.8, and the code was stopped on the next fifteen epoch, batch size thirty-two and dropout 0.5. Ten percent of images in the test set for each disease were used for training, by using the 5-fold cross validation. The remaining 90% of the images were divided into 5 equal folds, 4 of them used for training and the last fold was used as a validation set. Each group of images takes turns to be the validation set, thus validating a total of five times, in which in each fold, we spreaded the images so that each fold has the same or similar number of images for each disease. Figure 2. shows model architecture of AIChest4All.

D. *Train on public dataset*

The training method was made up of two steps, the first the model was trained the NIH [19], Shenzhen and Montgomery [20] set by using initial weights of ImageNet in order to train the model to extract features from the chest X-ray images. The model was trained by using the 5-fold cross validation technique, and the model having the best result was selected to be the initial weights of the model in the second step.

E. *Freeze and train very last layers*

In the second step, the dataset from UDCH and the CCIT was used so the model can have a greater understanding of patients in our focused area. The feature extraction technique in DenseNet-121 [17] was frozen, as the class “other abnormal findings” falls in a wide, varying range of categories, which makes it hard to develop methods to increase accuracy of the model. Therefore, the class was separated into two smaller classes, “intrathoracic abnormal findings” and “extra thoracic abnormal findings”, and then the model was trained again using the 5-fold cross validation technique. The ROC and AUC values from the validation set from each round of training were separated into the 6 categories as shown in Fig. 3. The model with the highest AUC was selected to be used in the test set.

IV. EXPERIMENTAL RESULTS

For each of the 6 classes stated earlier, the model with the highest AUC based on 5-fold cross validation technique was used to investigate its efficiency on the test set from UDCH and CCIT. The category ‘no finding’ consisted of 4,116 images, ‘pulmonary TB’ had 979 images, ‘lung cancer’ had 527 images, ‘heart disease’ had 1,047 images, ‘intrathoracic abnormal findings’ had 2,640 images, and ‘extrathoracic abnormal findings’ had 1,621 images. Accuracy, sensitivity, and specificity values for 6 findings are summarized in Table II.

V. DISCUSSION AND CONCLUSION

From the model used in our methodology to classify X-ray images, an accuracy of over 90% was achieved when classifying tuberculosis, lung cancer, and heart diseases. The disease with the lowest accuracy was intrathoracic abnormal findings, which achieved an accuracy of 79.96%, as there can be several abnormalities that fit in this category in a single image, thus making detection more difficult. The category extrathoracic findings also followed the same trend, with a slightly lower accuracy, due to the same reason that there are several abnormalities that fit in that category. From the confusion matrix, the false negative numbers were low, resulting in low chances that the model will not detect tuberculosis, lung cancer, and heart diseases, all the diseases that are in our focus.

The model proposed in this paper, called “AIChest4All”, is now being used in Thailand’s government funded hospitals, particularly in rural areas, in order to support hospitals that lack in radiologists, as well as deducting the time used to diagnose each X-ray image. This can help patients to be treated quicker, reducing risks from the disease.

VI. FUTURE WORK

One way to improve our model is to change pre-processing techniques on the X-ray images, currently CLAHE was used on all three RGB layers, which gives the same results, rendering the other CLAHE useless. To improve on this, different preprocessing techniques are to be used on different layers. Moreover, a feedback system collected opinions (agreement/disagreement) from radiologists using AIChest4All is also implemented. These opinions together with the collected X-ray images across hospitals in Thailand can be used further in the data labeling system then these images will be used to improve the accuracy of AIChest4All accordingly.

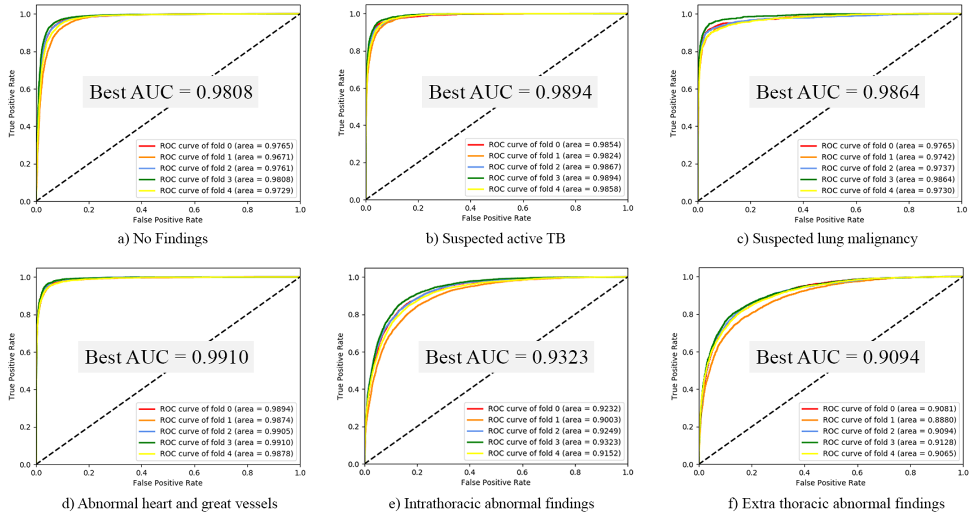

Fig. 3: ROC and AUC for each class for 5-fold cross validation.

TABLE II: Results from the test set.

Category	Acc	Sen	Spec
No finding	86.82%	86.83%	86.82%
Suspected active TB	92.32%	82.33%	93.54%
Suspected lung malignancy	93.28%	81.02%	94.04%
Abnormal heart and great vessels	94.11%	90.07%	94.65%
Intrathoracict abnormal findings	79.96%	76.40%	81.44%
Extrathoracict abnormal findings	80.12%	76.13%	81.00%

References

[1] H. Suzuki, N. Inaoka, H. Takabatake, M. Mori, H. Natori, and A. Suzuki, "Experimental system for detecting lung nodules by chest x-ray image processing," in *Biomedical Image Process. II*, vol. 1450, pp. 99–107, International Society for Optics and Photonics, 1991.

[2] M. S. Al-Tarawneh, "Lung cancer detection using image processing techniques," *Leonardo Electronic Journal of Practices and Technologies*, vol. 11, no. 21, pp. 147–58, 2012.

[3] A. Chaudhary and S. S. Singh, "Lung cancer detection on CT images by using image processing," in *2012 Int. Conf. Comput. Sci.*, pp. 142–146, IEEE, 2012.

[4] N. S. Lingayat and M. R. Tarambale, "A computer based feature extraction of lung nodule in chest x-ray image," *International Journal of Bioscience, Biochemistry and Bioinformatics*, vol. 3, no. 6, p. 624, 2013.

[5] D. H. Ballard and J. Sklansky, "A ladder-structured decision tree for recognizing tumors in chest radiographs," *IEEE Trans. Comput.*, no. 5, pp. 503–513, 1976.

[6] M. S. Brown, L. S. Wilson, B. D. Doust, R. W. Gill, and C. Sun, "Knowledge-based method for segmentation and analysis of lung boundaries in chest X-ray images," *Computerized medical imaging and graphics*, vol. 22, no. 6, pp. 463–477, 1998.

[7] K. Murphy, B. van Ginneken, A. M. Schilham, B. De Hoop, H. Gietema, and M. Prokop, "A large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbour classification," *Medical image analysis*, vol. 13, no. 5, pp. 757–770, 2009.

[8] S. Sivakumar and C. Chandrasekar, "Lung nodule detection using fuzzy clustering and support vector machines," *International Journal of Engineering and Technology*, vol. 5, no. 1, pp. 179–185, 2013.

[9] K. Santosh and S. Antani, "Automated chest X-ray screening: Can lung region symmetry help detect pulmonary abnormalities?," *IEEE Trans. Med. Imag*, vol. 37, no. 5, pp. 1168–1177, 2017.

[10] Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen, and H. Greenspan, "Chest pathology detection using deep learning with non-medical training," in *2015 IEEE 12th Int. Symp. Biomedical Imaging (ISBI)*, pp. 294–297, IEEE, 2015.

[11] P. Lakhani and B. Sundaram, "Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks," *Radiology*, vol. 284, no. 2, pp. 574–582, 2017.

[12] P. Rajpurkar, J. Irvin, R. L. Ball, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. P. Langlotz, *et al.*, "Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists," *PLoS medicine*, vol. 15, no. 11, p. e1002686, 2018.

[13] X. Wang, Y. Peng, L. Lu, Z. Lu, and R. M. Summers, "Tienet: Text-image embedding network for common thorax disease classification and reporting in chest x-rays," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, pp. 9049–9058, 2018.

[14] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, *et al.*, "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," *arXiv preprint arXiv:1711.05225*, 2017.

[15] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, H. Marklund, B. Haghgoo, R. Ball, K. Shpanskaya, *et al.*, "Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison," in *Proc. AAAI Conf. Artificial Intell.*, vol. 33, pp. 590–597, 2019.

[16] K. Krippendorff, *Content Analysis: An Introduction to Its Methodology*. Sage publications, 2018.

[17] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, pp. 4700–4708, 2017.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Adv. Neural Inform. Process. Syst.*, pp. 1097–1105, 2012.

[19] 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 *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, pp. 2097–2106, 2017.

[20] S. Jaeger, S. Candemir, S. Antani, Y.-X. J. Wáng, P.-X. Lu, and G. Thoma, "Two public chest X-ray datasets for computer-aided screening of pulmonary diseases," *Quantitative imaging in medicine and surgery*, vol. 4, no. 6, p. 475, 2014.