Study on the TB and non-TB diagnosis using two-step deep learning-based binary classifier

S.H. Yoo, a H. Geng, a T.L. Chiu, a S.K. Yu, a D.C. Cho, b J. Heo, b M.S. Choi, b I.H. Choi, b C.V. Cung, c N.V. Nhung c and B.J. Min d, 1

a Medical Physics and Research Department, Hong Kong Sanatorium & Hospital, 2 Village Road, Happy Valley, Hong Kong, China
b Division of Artificial Intelligent Clinical Research, Radisen, 128, Gongduk B/D, Saechang-ro, Mapo-gu, Seoul, Korea
c National Tuberculosis Programme, Vietnam National Lung Hospital, Hanoi, Vietnam
d Department of Radiation Oncology, Chungbuk National University Hospital, 776, 1 Sunhwan-ro, Seowon-gu, Cheongju, Korea

E-mail: bjmin@cbnuh.or.kr

ABSTRACT: A deep learning-based binary classifier was proposed to diagnose tuberculosis (TB) and non-TB disease using a chest X-ray radiograph. The proposed classifier comprised two-step binary decision trees, each trained by a deep learning model with convolution neural network (CNN) based on the PyTorch frame. Normal and abnormal images of chest X-ray was classified in the first step. The abnormal images were predicted to be classified into TB and non-TB disease by the second step of the process. The accuracies of first and second step were 98% and 80% respectively. Moreover, re-training could improve the stability of prediction accuracy for images in different data groups.

KEYWORDS: Medical-image reconstruction methods and algorithms, computer-aided diagnosis; X-ray radiography and digital radiography (DR)

© 2020 The Author(s). Published by IOP Publishing Ltd on behalf of Sissa Medialab. Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

https://doi.org/10.1088/1748-0221/15/10/P10011
1 Introduction

Tuberculosis (TB) is classified as the fifth leading cause of death worldwide, with about 10 million new cases and 1.5 million deaths per year [1]. Being one of the world’s biggest threats and being rather easy to cure, the World Health Organization recommends systematic and broad use of screening to extirpate the disease. Posteroanterior chest radiography, in spite its low specificity and difficult interpretation, is one of the preferred tuberculosis screening methods. Unfortunately, since TB is primarily a disease of poor countries, the clinical officers trained to interpret these chest X-Rays are often lacking [2, 3].

Several skin tests based on immune response are available for determining whether an individual has been exposed to TB. However, skin tests are not an indicator of active disease, and are also affected by vaccinations. A definitive test for diagnosing TB is the identification of the bacteria in a clinical sputum or pus sample, which is the current gold standard [4, 5]. However, it may take several months to identify this slow-growing organism in the laboratory. Another technique is sputum smear microscopy, in which bacteria in sputum samples are observed under a microscope.

Recent advances in automated diagnostic tests are based on DNA analysis [4]. The test is fast and accurate, but it is expensive making it unsuitable for population screening. As a result, taking posterior-anterior (PA) chest X-rays (CXR) continues to remain as an inexpensive and mandatory part of every evaluation for TB [6, 7]. However, there can be non-TB disease images in real examination for CXRs and these non-TB images need to be classified to another category compared to TB or normal images.
Many of the computer aided diagnosis (CAD) papers dealing with abnormalities in chest radiographs do so without focusing on any other specific disease (non-TB). Most of CAD systems specializing in TB detection have been reported, such as [8]–[14]. Further, TB screening is not just limited to normal/abnormal classification, but also to spot the location in CXRs where the abnormality lies [15]. In [16], Hogeweg et al. use a combination of pixel classifiers and active shape models for clavicle segmentation. Note that the clavicle region is a notoriously difficult region for TB detection because the clavicles can obscure manifestations of TB in the apex of the lung. Freedman et al. showed in a recent study that an automatic suppression of ribs and clavicles in CXRs could significantly increase a radiologist’s performance for nodule detection [17]. Arzhaeva et al. use dissimilarity-based classification to cope with CXRs for which the abnormality is known but the precise location of the disease is unknown [18]. They report classification rates comparable to rates achieved with region classification on CXRs with known disease locations. More information on existing TB screening systems can be found in our recent survey [19].

In real examination of lung disease, TB and Non-TB images are mixed, thus, classification into normal, TB and Non-TB categories is necessary for lung disease detection. If we use the algorithm trained by only TB/Normal data, the accuracy can be lower in examination which include Non-TB images [20, 21]. In addition to TB, several studies on the detection of the various kinds of Non-TB disease have been conducted [22]–[24]. Also, abnormal image detection including Non-TB was investigated by using a multi-CNN model [25]. ChestX-ray14 dataset is from the National Institutes of Health (NIH) Clinical Center [23]. Due to its size, the ChestX-ray14 consisting of 112,120 frontal CXR images from 30,805 unique patients attracted considerable attention in the deep learning community. Triggered by the work of Wang et al. [23] using convolution neural networks (CNNs) from the computer vision domain, several research groups have begun to address the application of CNNs for CXR classification. In the work of Yao et al. [26], they presented a combination of a CNN and a recurrent neural network to exploit label dependencies. As a CNN backbone, they used a DenseNet [27] model which was adapted and trained entirely on X-ray data. Li et al. [28] presented a framework for pathology classification and localization using CNNs. More recently, Rajpurkar et al. [29] proposed transfer-learning with fine tuning, using a DenseNet-121 [27], which raised the AUC results on ChestX-ray14 for multi-label classification even higher.

In this study, we suggest two-step of process for detection of normal, TB and non-TB disease using binary classifier for each step. The accuracies of algorithm for each step were investigated and the results showed that this process can be a good candidate for the clinical purpose.

2 Method

2.1 Workflow

A two-step process was suggested in order to categorize examined chest X-ray images (CXR) into three clinical states (Normal, TB and Non-TB) through binary classification for each step as shown in figure 1. All CXR is checked by radiographer when it is taken before being predicted by a CAD algorithm to check their imaging quality. In our suggested workflow, there are two automated X-ray imaging systems (AXIRs: AXIR1 and AXIR2 which are for the classification of Abnormal/Normal
and TB/Non-TB, respectively). When a good quality CXR is taken by AXIR1 and predicted as normal state, there is no more action taken for that patient. However, the image which is classified as abnormal state is predicted by next step (AXIR2) to determine whether it is TB or Non-TB. Finally, medical doctor reviews the classified images and makes a decision for further investigation of clinical assessment.

In each step, the prediction algorithm is trained by with an optimized data group for its intended purpose. AXIR1 is trained with a combination of normal and abnormal data. In this case, the abnormal data include TB and Non-TB data. In the case of the Non-TB data, ChestX-ray14 dataset (NIH) was used for the training. AXIR2 is trained by the TB and Non-TB cases taken from the ChestX-ray14 dataset. Each of the classifiers in the step system is designed for binary classification.

![Workflow](image)

**Figure 1.** Workflow of two-step process for determining of Normal, TB and Non-TB lung diseases.

### 2.2 Patient data and augmentation

In an effort to provide sufficient training data for the research community, allowing benchmark tests, the U.S. National Library of Medicine has made two datasets of posterio-anterior (PA) chest radiographs available: the MC set and the Shenzhen set. Both datasets contain normal and abnormal chest X-rays with manifestations of TB and include associated radiologist readings. The categorization of TB is based on the final pathological diagnosis. The Shenzhen set was used in this study for TB dataset. The open-source Shenzen data comprise 326 TB and 226 normal cases (https://lhncbc.nlm.nih.gov/publication/pub9931). The average image size for Shenzen data is $3000 \times 3000$.

Also, the NIH (National Institutes of Health, U.S.) Clinical Center recently released over 100,000 anonymized chest x-ray images and their corresponding data to the scientific community. The database from NIH is available online at [https://nihcc.app.box.com/v/ChestXray-NIHCC/folder/36938765345](https://nihcc.app.box.com/v/ChestXray-NIHCC/folder/36938765345). The size of each image is $1024 \times 1024$. This dataset is categorized into 14-lung disease sub-patterns which are based on pathological diagnosis. However, this dataset does not show TB or Non-TB classification with pathological confirmation.
In addition to above open dataset, we used East Asian Hospital dataset (which is based on cooperation with Radisen) for TB and Non-TB disease. This dataset is based on pathological diagnosis. In the case of the non-TB data, all images were categorized into NIH-14 based sub-patterns. The images were taken by a Vicomed system and viewed with a Radisen Detector (17” × 17”). The average image size is 2484 × 3012. The image acquisition conditions were as follows: voltage = 105 kVp, current = 125 mA, charge = 10 mAs, time = 80 ms, source-to-image distance = 130–150 cm. For the correct collection of the data, two radiologists review dataset and only data which is agreed by two radiologists were used in this study. From those data set (open source and East Asian Hospital dataset), trained and tested images were randomly selected, excluding children’s images.

In the case of the AXIR1, among the 1,170 patients in the total dataset (table 1), 170 patients (14.5%) were randomly selected for testing. Among these 170 patients, 85 were abnormal and 85 were normal. The remaining 1,000 patients were split in a 50% : 50% ratio to make abnormal (500 patients) and normal (500 patients) cases for training. Among the 585 abnormal cases, 442 were from NIH data and 143 were from East Asian hospital data. The normal data images also came from the NIH and East Asian countries’ data.

| Abn/Nor | Data group                              | Number of images | Sub Total                      |
|---------|-----------------------------------------|-----------------|--------------------------------|
| Abn     | NIH data (TB+Non-TB)                    | 442             | 585 (Train 500, Test 85)       |
|         | East Asian hospital data (TB+Non-TB)    | 143             | (Train with augmentation 1500)  |
| Nor     | (NIH data + East Asian hospital data)   |                 | 585 (Train 500, Test 85)       |
|         |                                         |                 | (Train with augmentation 1500)  |

1TB and Non-TB are not classified with pathological diagnosis.
2TB and Non-TB are classified with pathological diagnosis.

In the case of the AXIR2, among the 984 patients in the total dataset (table 2), 164 patients (16.7%) were randomly selected for testing. Among these 164 patients, 82 were TB cases and 82 were non-TB (other) cases. The remaining 820 patients were split in a 50% : 50% ratio to make TB (410 patients) and non-TB cases (410 patients) for training. Among the 492 TB patients, 372 images were from East Asian data, and the remaining 120 TB cases were from Shenzhen data. All 492 non-TB lung disease images were taken from East Asian hospital data.

We applied several data augmentation algorithms to improve the training and classification accuracy of the CNN model and achieved remarkable validation accuracy. 1500 images including augmented data are used to train the model. The settings applied for the image augmentation are shown below in table 3.

The rotation range denotes the range in which the images were randomly rotated during training, i.e., 10 degrees. Width shift denotes the horizontal translation of the images by 0.2 percent, and height shift is the vertical translation of the images by 0.2 percent. And finally, the images were flipped horizontally.
Table 2. Data structure for AXIR2.

| TB/Non-TB | Data group                     | Number of images | Sub Total                                |
|-----------|--------------------------------|------------------|------------------------------------------|
| TB        | TB (East Asian hospital data)  | 372              | 492 (Train 410, Test 82)                 |
|           | TB (Shenzhen data)             | 120              | (Train with augmentation 1500)           |
| Non-TB    | Non-TB (East Asian hospital data) |                 | 492 (Train 410, Test 82)                 |
|           |                                |                  | (Train with augmentation 1500)           |

Table 3. Setting for the image augmentation.

| Method          | Setting |
|-----------------|---------|
| Rotation angle  | 10      |
| Width shift     | 0.2     |
| Height shift    | 0.2     |
| Horizontal flip | True    |

We performed stability test for AXIR1 using other group data (Shenzhen data). First of all, 100 TB and 100 normal data were used for the test. And the same test was repeated, after re-training the model with 226 TB and 100 normal data.

2.3 Annotation and classification

We further categorized abnormal patterns of the NIH data as TB or non-TB. TB patterns were then subcategorized into seven active TB pattern classes (consolidation, cavitation, pleural effusion, miliary, interstitial, tree in bud, and lymphadenopathy) and an additional inactive TB pattern class. These TB pattern classes are based on those used by the Centers for Disease Control and Prevention (CDC) of the U.S.A. [30]. Also, we subcategorized the Shenzhen TB data into sub-classes (according to the ChestX-ray14 dataset (NIH data)), which are infiltration, consolidation, pleural effusion, pneumonia, fibrosis, atelectasis, nodule, pneumothorax, pleural thickening, mass, hernia, cardiomagaly, edema, and emphysema. These additional categorizations are based on the radiological readings for the purpose of the detailed analysis for sub-patterns of FN and FP. The resulting annotated data were reviewed by two radiologists independently. In the case of the TB data, most data are related with infiltration, consolidation, plural effusion, fibrosis, and nodule. These patterns are strongly related with TB patterns.

We also used East Asia countries’ data (TB and Non-TB disease) from our cooperating hospitals to diversify the trained dataset. Non-TB disease dataset for Easter Asia Hospital is categorized based on NIH-14 disease pattern related with final confirmed diagnosis. In the case of the TB data (Eastern Asian Hospital), we annotated more detailed sub-patterns according to radiological sub-TB patterns. However, these sub-TB patterns are categorized with only radiological readings.
2.4 Deep learning algorithm and training/testing technique

Herein, we used 2D CNN algorithm with PyTorch frame for the training and testing for our two-step CAD process. Figure 2 shows the overall architecture of the proposed CNN model which is based on the pre-trained ResNet18 [31, 32], using ImageNet dataset. ResNet is one of the most popular CNN architecture, which provides easier gradient flow for more efficient training, and was the winner of the 2015 ImageNet competition. The core idea of ResNet is introducing a so-called identity shortcut connection that skips one or more layers. This would help the network to provide a direct path to the very early layers in the network, making the gradient updates for those layers much easier. ResNet18, which is similar in performance to the ResNet50 structure, was employed to reduce the training and verification time [33]. The schematic diagram of ResNet18 model for lung disease detection is illustrated in figure 2.

We analyzed the performance of the prediction results with accuracy, sensitivity, specificity, precision and area under curve (AUC) as following:

\[
\text{Accuracy} = \frac{TP+TN}{TP+FN+FP+TN}, \quad \text{Sensitivity} = \frac{TP}{TP+FN}, \quad \text{Specificity} = \frac{TN}{TN+FP}, \quad \text{Precision} = \frac{TP}{TP+FP} \quad \text{and} \quad \text{AUC} = \text{The area under a curve between two points can be found by doing a definite integral between the two points, where TP = True Positive, FP = False Positive, TN = True Negative and FN = False Negative.}
\]

![Figure 2. The architecture of ResNet18 model [31].](image)

3 Results

3.1 Accuracy for AXIR1 and AXIR2

We performed the prediction and the results are shown in table 4 and 5. The Accuracy of the AXIR1 and AXIR2 are 0.98 and 0.80 respectively.

Also, we investigated the classification capability of AXIR2 for normality using 100 normal test data. The normal data can be classified into TB and others with the ratio of 1 : 3 by AXIR2.
Table 4. Performance of classification for AXIR1.

| Accuracy | Sensitivity | Specificity | Precision | AUC  |
|----------|-------------|-------------|-----------|------|
| 0.98     | 0.99        | 0.97        | 0.97      | 0.98 |

Table 5. Performance of classification for AXIR2.

| Accuracy | Sensitivity | Specificity | Precision | AUC  |
|----------|-------------|-------------|-----------|------|
| 0.80     | 0.72        | 0.89        | 0.87      | 0.80 |

In addition, we analyzed detailed distribution of false negative (FN) and false positive (FP) classification errors in AXIR2. The decision tree made 9 FP and 23 FN predictions, as shown in table 6. Atelectasis and plural thickening (67% of all FPs) and infiltration, consolidation and fibrosis (NIH classification) (83% of all FNs) are the categories most likely to be involved in such errors. These three sub-patterns can be categorized as consolidation sub-pattern in CDC sub-TB classification.

Table 6. Detailed classification for FP and FN in AXIR2.

| Prediction Results | NIH classification | Number of images | TB pattern in [30]* | Total number of images |
|--------------------|--------------------|------------------|---------------------|-----------------------|
| FP (from East Asian Hospital data) | Infiltration | 2 | Non-TB | 9 |
|                     | consolidation | 1 | | |
|                     | Atelectasis     | 3 | | |
|                     | Plural thickening | 3 | | |
| FN (from Shenzhen data and East Asian Hospital data) | Infiltration | 5 | Consolidation | 23 |
|                     | Consolidation   | 8 | Consolidation | |
|                     | Plural effusion | 2 | Plural effusion | |
|                     | Fibrosis        | 6 | Consolidation | |
|                     | Nodule          | 2 | Milliary | |

* Sub-patterns are classified with only radiological readings.

3.2 Stability test for other group data

We evaluated the stability of the studied model for Shenzhen data set (100 TB and 100 Normal data) applying the same trained algorithm. In this case, the accuracy and sensitivity become much lower than those of the NIH-14 and East Asian test data. Thus, we re-trained the model using a Shenzhen data (236 TB cases and 100 Normal cases, which are different from the test set).

For the Shenzhen test dataset before re-training, the accuracy of AXIR1 was low as 0.50 and was improved up to 0.82 after re-training (table 7). For the NIH test dataset, the accuracy after re-training show not so significantly difference (2%) from that before re-training as shown in table 8.

The effect of number of Shenzhen TB images added on the accuracy of the re-trained model was shown in figure 3. The accuracy are increased with number of images re-trained and saturated.
Table 7. Performance of classification for re-trained algorithm for Shenzen test data.

|                        | Accuracy for AXIR1 |
|------------------------|--------------------|
| before re-training     | 0.50               |
| After re-training      | 0.82               |

Table 8. Performance of classification for re-trained algorithm for NIH-14 test data.

|                        | Accuracy for AXIR1 |
|------------------------|--------------------|
| before re-training     | 0.98               |
| After re-training      | 0.96               |

Figure 3. The accuracy of AXIR1 classifier.

after adding 200 images or more into the re-training dataset. For this re-training, augmentation method was not applied.

4 Discussions

The accuracy of AXIR1 is 0.98. This prediction results show that our trained algorithm can be used clinically for the screening of abnormality and normality. The sensitivity is slightly higher than specificity by 2%, but this is not so significant (table 4). However, the accuracy of AXIR2 (0.80) is lower than that of AXIR1 (0.98) as shown in table 5. The sensitivity and specificity of AXIR2 show 0.72 and 0.89, respectively. For analysis of the distribution of sub-patterns of false positive (FP) and false negative (FN), we used 14 classifications from NIH data (Infiltration, Consolidation, Plural
Table 9. Sub-Classification of TB and non-TB disease.

| Normal          | No finding          |
|-----------------|---------------------|
| TB              |                     |
| Active          | Consolidation       |
|                 | Cavitation          |
|                 | Plural effusion     |
|                 | Milliary            |
|                 | Interstitial        |
|                 | Tree in Bud         |
|                 | Lymphadenopathy     |
|                 | Inactive            |
| Abnormal        |                     |
| Non-TB (Other)  | Infiltration        |
|                 | Consolidation       |
|                 | Plural effusion     |
|                 | Pneumonia           |
|                 | Fibrosis            |
|                 | Atelectasis         |
|                 | Nodule              |
|                 | Pneumothorax        |
|                 | Plural thickening   |
|                 | Mass                |
|                 | Hernia              |
|                 | Cardiomegaly        |
|                 | Edema               |
|                 | Emphysema           |

effusion, Pneumonia, Fibrosis, Atelectasis, Nodule, Pneumothorax, Plural Thickening, Edema and Emphysema) and sub-TB patterns according to ref. [30] (see table 9).

Table 6 shows the detailed distribution among these classes for 9 FP and 23 FN predictions of AXIR2. Atelectasis and plural thickening show a higher portion of FP (67%) and infiltration, consolidation and fibrosis are higher portion for FN (83%). In atelectasis and plural thickening, the region can be similar to plural effusion in lower lobe in both side of lung. Thus, the algorithm’s prediction results can show larger portion of FP due to atelectasis and plural thickening specifically. Clinically, plural effusion is strongly associated with TB, but atelectasis or plural thickening is not. Also, infiltration, consolidation and fibrosis, as defined by the NIH, are strongly associated with TB; however, some of these radiological sub-patterns can be predicted as non-TB. When these sub-patterns are annotated as TB, the position of occurrence was frequently in the upper apex region. The most of cases of FP show that the locations of occurrence are middle or lower lobe area. Thus, applying position-based feature filtering in the algorithm might improve the accuracy. For this
analysis of sub-pattern, we just categorized data with only radiological readings. However, the prediction results for TB and Non-TB, all training data structure is based on pathological diagnosis.

The accuracy depends on not only the optimization of model, also properties of the trained dataset. We performed stability test of the accuracy between data groups. First of all, we trained our algorithm by using NIH dataset and East Asian dataset which is same dataset of test. Then we compared our prediction results with other data group (Shenzhen dataset) for the test. The results show that accuracy become lower significantly for different test dataset as shown in table 7. The accuracy can be improved again, as we re-trained the model using Shenzhen dataset additively (see table 8). Also, the number of images that is needed to re-train the algorithm of 200 images can be enough for the improvement as shown in figure 3. These results implies that it is important to conduct maintenance in which CAD algorithm need to be re-trained by using appropriate dataset (same data group or similar pattern of annotation) in order to use it as a clinical purpose in local hospital.

There are many data augmentation strategies for image data. We added more training data to our deep learning model using easy-to-implement methods such as horizontal flip, rotations, and shifts. Image processing techniques using stochastic region sharpening, elastic transforms, randomly erasing patches, and many more to augment data can be considered to improve the performance of the resulting model. Therefore, further studies (enlargement, contrast adjustment, and baseline correction) are needed on advanced augmentation techniques for building better models and creating a system that does not require gathering a lot of training data to get a reliable statistical model.

According to our pre-study, the effect of the image size from 512 to 2048 does not affect the performance significantly. However, in some cases, mixed image sizes could affect the performance. Therefore, the different image sizes were normalized to $1024 \times 1024$, which is the smallest image size of the obtained images, to avoid image size effect on the performance.

5 Conclusion

In this study, we proposed two-step process with binary filter to classify Normal, TB and Non-TB lung disease. 2D CNN algorithm (ResNet18 model) with PyTorch frame was used for each step with optimized trained data. In first step (AXIR1), Abnormal/Normal data are used for training to perform screening of normal image from patient data. In second step, TB and Non-TB data are used for training to classify them into TB and Non-TB lung diseases. To analyze the tendency of the accuracy according to sub-patterns, we annotated our dataset with TB/Non-TB items and NIH-14 based disease items separately. The accuracy of 0.98 and 0.80 can be achieved for AXIR1 and AXIR2 respectively and stability can be satisfied with re-training by using a same data group with test/training data group. The reason for low accuracy for AXIR2 compared to AXIR1 is resulted from the large portion of FP with atelectasis and plural thickening. Also, large portion of FN with infiltration, consolidation and fibrosis can result in the low accuracy of AXIR2.

Acknowledgments

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2020R1A2C4001910 and 2020M2D9A1094075)
References

[1] L. Anderson et al., *Global tuberculosis report 2015*, World Health Organization (2015) and online at https://apps.who.int/iris/handle/10665/191102.

[2] A.H. Hoog et al., *High sensitivity of chest radiograph reading by clinical officers in a tuberculosis prevalence survey*, *Int. J. Tuberc. Lung Dis.* 15 (2011) 1308.

[3] J. Melendez et al., *An automated tuberculosis screening strategy combining X-ray-based computer-aided detection and clinical information*, *Sci. Rep.* 6 (2016) 25265.

[4] A. Baddeley et al., *Global tuberculosis report 2014*, World Health Organization (2014) and online at https://apps.who.int/iris/handle/10665/137094.

[5] World Health Organization and Stop TB Partnership, *The global plan to stop TB 2011–2015: transforming the fight towards elimination of tuberculosis*, World Health Organization (2010) and online at https://apps.who.int/iris/handle/10665/44437.

[6] C.C. Leung, *Reexamining the role of radiography in tuberculosis case finding*, *Int. J. Tuberc. Lung Dis.* 15 (2011) 1279.

[7] L. Folio, *Chest Imaging: An Algorithmic Approach to Learning*, Springer (2012).

[8] B. van Ginneken, S. Katsuragawa, B. ter Haar Romeny, K. Doi and M. Viergever, *Automatic detection of abnormalities in chest radiographs using local texture analysis*, *IEEE Trans. Med. Imaging* 21 (2002) 139.

[9] L. Hogeweg, C. Mol, P.A. de Jong, R. Dawson, H. Ayles and B. van Ginneken, *Fusion of local and global detection systems to detect tuberculosis in chest radiographs*, in *Medical Image Computing and Computer-Assisted Intervention — MICCAI 2010*, Lecture Notes in Computer Science 6363, Springer (2010), pp. 650–657.

[10] R. Shen, I. Cheng and A. Basu, *A hybrid knowledge-guided detection technique for screening of infectious pulmonary tuberculosis from chest radiographs*, *IEEE Trans. Biomed. Eng.* 57 (2010) 2646.

[11] T. Xu, I. Cheng and M. Mandal, *Automated cavity detection of infectious pulmonary tuberculosis in chest radiographs*, in proceedings of the 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Boston, MA, U.S.A., 30 August–3 September 2011, pp. 5178–5181.

[12] C.K. Ajmal Shan and D.L. Binoy, *An Optimal Way for Tuberculosis Detection*, *Int. Res. J. Eng. Technol.* 3 (2016) 391.

[13] C.C. Pawar and S.R. Ganorkar, *Tuberculosis screening using digital image processing techniques*, *Int. Res. J. Eng. Technol.* 3 (2016) 623.

[14] Z.Z. Qin et al., *Using artificial intelligence to read chest radiographs for tuberculosis detection: A multi-site evaluation of the diagnostic accuracy of three deep learning systems*, *Sci. Rep.* 9 (2019) 15000.

[15] B. van Ginneken and B.M. ter Haar Romeny, *Automatic segmentation of lung fields in chest radiographs*, *Med. Phys.* 27 (2000) 2445.

[16] L. Hogeweg, C.I. Sánchez, P.A. de Jong, P. Maduskar and B. van Ginneken, *Clavicle segmentation in chest radiographs*, *Med. Image Anal.* 16 (2012) 1490.

[17] M.T. Freedman, S.-C.B. Lo, J.C. Seibel and C.M. Bromley, *Lung nodules: Improved detection with software that suppresses the rib and clavicle on chest radiographs*, *Radiology* 260 (2011) 265.
[18] Y. Arzhaeva, D. Tax and B. van Ginneken, Dissimilarity-based classification in the absence of local ground truth: Application to the diagnostic interpretation of chest radiographs, Pattern Recognit. 42 (2009) 1768.

[19] S. Jaeger et al., Automatic screening for tuberculosis in chest radiographs: a survey, Quant. Imaging. Med. Surg. 3 (2013) 89.

[20] S. Antani, Automated Detection of Lung Diseases in Chest X-Rays, Ph.D. Thesis, U.S. National Library of Medicine, LHNCBC, Bethesda MD U.S.A. (2015).

[21] Y.-X. Tang, Y.-B. Tang, M. Han, J. Xiao and R.M. Summers, Abnormal chest X-ray identification with generative adversarial one-class classifier, in proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, 8–11 April 2019, pp. 1358–1361.

[22] I.M. Baltruschat, H. Nickisch, M. Grass, T. Knopp and A. Saalbach, Comparison of deep learning approaches for multi-label chest X-ray classification, Sci. Rep. 9 (2019) 6381.

[23] 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 proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, U.S.A., 21–26 July 2017, pp. 3462–3471.

[24] A. Elkins, F.F. Freitas and V. Sanz, Developing an App to interpret Chest X-rays to support the diagnosis of respiratory pathology with Artificial Intelligence, arXiv:1906.11282.

[25] P.N. Kieu, H.S. Tran, T.H. Le, T. Le and T.T. Nguyen, Applying multi-CNNs model for detecting abnormal problem on chest x-ray images, in proceedings of the 2018 10th International Conference on Knowledge and Systems Engineering (KSE), Ho Chi Minh City, Vietnam, 1–3 November 2018, pp. 300–305.

[26] L. Yao, E. Poblenz, D. Dagunts, B. Covington, D. Bernard and K. Lyman, Learning to diagnose from scratch by exploiting dependencies among labels, arXiv:1710.10501.

[27] G. Huang, Z. Liu, L.V.D. Maaten and K.Q. Weinberger, Densely connected convolutional networks, in proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, U.S.A., 21–26 July 2017, pp. 2261–2269.

[28] Z. Li et al., Thoracic disease identification and localization with limited supervision, in proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, U.S.A., 18–23 June 2018, pp. 8290–8299.

[29] P. Rajpurkar et al., Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning, arXiv:1711.05225.

[30] A.C. Nachiappan et al., Pulmonary tuberculosis: Role of radiology in diagnosis and management, Radiographics 37 (2017) 52.

[31] K. He, X. Zhang, S. Ren and J. Sun, Deep Residual Learning for Image Recognition, in proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, U.S.A., 27–30 June 2016, pp. 770–778 [arXiv:1512.03385].

[32] H.J. Yoon, Y.J. Jeong, H. Kang, J.E. Jeong and D.-Y. Kang, Medical image analysis using artificial intelligence, Prog. Med. Phys. 30 (2019) 49.

[33] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani and G.J. Soufi, Deep-COVID: Predicting COVID-19 from chest x-ray images using deep transfer learning, Med. Image Anal. 65 (2020) 101794.