Deep Learning for Automated Classification of Abnormal Chest Radiograph Associated with Tuberculosis (DAC4TB) in the U.S. Hospital-scale CXR Database

Krit Pongpirul, MD, MPH, PhD[1,2,3*], Seelwan Sathitratanacheewin, MD[1,2], Panasun Sunanta, BSc[2]
Kittipat Kampa, PhD[2]

[1] Department of Preventive and Social Medicine, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand
[2] Thai Health AI Foundation, Bangkok, Thailand.
[3] Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

Thai Health AI Foundation

Chulalongkorn University
จุฬาลงกรณ์มหาวิทยาลัย
Pillar of the Kingdom

Background

Tuberculosis (TB) is a major health problem in many regions of the world. As part of the World Health Organization (WHO) systematic screening strategy to ensure early and correct diagnosis for all people with tuberculosis, Chest X-ray (CXR) is one of the primary tools for triaging and screening for TB because of its high sensitivity (87% to 98%, depending on how the CXR is interpreted). However, significant intra- and inter-observer variations in the reading of CXR can lead to overdiagnosis or underdiagnosis of tuberculosis. Deep convolutional neural network (DCCN) has emerged as an attractive technique TB surveillance and detection. Recently, the National Institute of Health (NIH) released the ChestX-ray8 dataset with 108,948 anonymized chest images and their associated data which compiled from 32,717 patients. Although these data allow researchers to develop an algorithm for classifying lung abnormalities, recent studies have not targeted TB detection. Our objectives are to develop a deep learning model for automated classification of chest radiography and to estimate the prevalence of CXR consistent with TB in NIH ChestX-ray8 database.

Materials and Methods

Two de-identified HIPAA-compliant datasets, the NIH ChestX-ray8 database and the National Library of Medicine (NLM) Shenzhen No.3 Hospital X-ray set were included in this study. In the Shenzhen Hospital X-ray set, CXR have been collected in JPEG format as part of the routine care at Shenzhen No.3 Hospital in Shenzhen, Guangdong providence, China. There are 336 normal and 326 abnormal x-rays showing various manifestations of tuberculosis. The ChestX-ray8 comprises CXR images acquired as a part of routine care at NIH Clinical Center, Bethesda, Maryland, USA. It comprises 112,120 frontal view X-ray images of 32,717 unique patients with the text mined eight disease image labels (each image can have more than one label), from their corresponding radiological reports using natural language processing. All personal identifiable information were checked and removed from the dataset by the NIH Clinical Center.

Firstly, the Shenzhen Hospital X-ray set which consisted of 336 chest radiographs with manifestations of TB and 326 normal radiographs were used to develop DCCN. The dataset was split into training (75%), validation (15%), and intramural test (10%) sets. Based on TensorFlow framework, Inception-v3, a novel pre-trained DCCN, was augmented with several techniques to classify each image as having TB characteristics or as healthy. Receiver operating characteristic (ROC) curves and areas under the curve (AUC) were used to assess model performance and to define the optimal cut point for tuberculosis detection.

Next, 112,120 CXR (60,362 normal and 51,760 abnormal images with one of fourteenth common thoracic abnormalities including atelectasis, cardiomegaly, consolidation, edema, effusion, emphysema, fibrosis, hernia, infiltration, mass, nodule, pleural thickening, pneumonia and pneumothorax) from ChestX-ray8 dataset were used to construct an extramural test set to examine the generalizability of DCCN model to classify normal and other chest radiographs in addition to test set from intramural database(Shenzhen No.3 Hospital). Lastly, prevalence of tuberculosis associated radiographs in the U.S. radiograph dataset was estimated by using the final DCCN model.

Results

Total=112,782 (662+112,120)	Human Reading-Abnormal	Human Reading-Normal
NLM Senzhen No.3 Hospital	336 (Consistent with TB)	326
NIH Clinical Center	51,760 (14 common pathologies, no TB)	60,362

Table1: Datasets for DAC4TB development

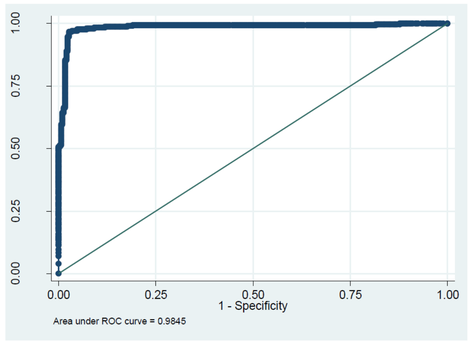

Figure 1: ROC curve of DAC4TB in train dataset using Shenzhen No.3 Hospital dataset

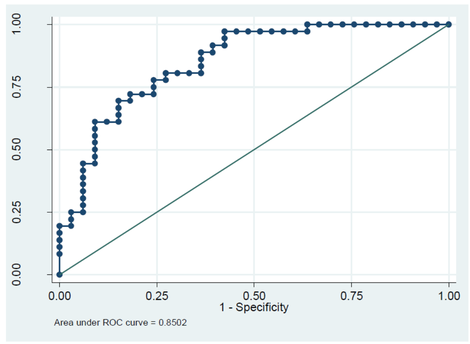

Figure 2: ROC curve of DAC4TB in test data set using Shenzhen No.3 Hospital dataset

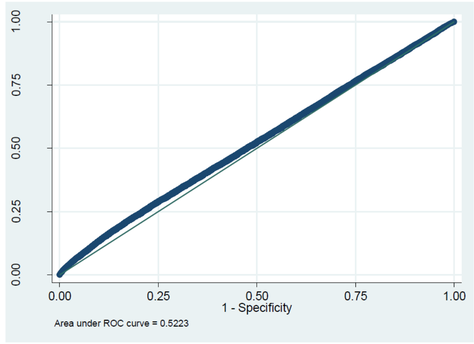

Figure 3: ROC curve of DAC4TB in NIH Chest Xray14 dataset

Table 2: DAC4TB Prediction

CXR Abnormalities	Suggested Tuberculosis (normal score>0.90)	Not Suggested Tuberculosis (normal score>0.90)
Atelectasis	2,762	8,797
Cardiomegaly	761	2,015
Consolidation	1,005	3,662
Edema	375	1,928
Effusion	3,676	9,641
Emphysema	745	1,771
Fibrosis	635	1,051
Hernia	85	142
Infiltration	4,950	14,944
Mass	1,870	3,912
Nodule	2,272	4,059
Pleural Thickening	1,129	2,256
Pneumonia	377	1,054
Pneumothorax	1,501	3,801
Normal (No Finding)	14,632	45,729
Total = 141,537 (112,120)	36,775(29,223)	104,762(82,897)

Conclusions

Findings suggest different training data set are likely to engage in variation of diagnostic accuracy of deep learning in different populations, highlighting the utility of context specific approach in the use of AI for automated CXR classification.

References

1. World Health Organization. (2016). Chest radiography in tuberculosis detection: summary of current WHO recommendations and guidance on programmatic approaches. World Health Organization. http://www.who.int/iris/handle/10665/252424

Corresponding Author

Assist. Prof. Krit Pongpirul, MD, MPH, PhD.
Department of Preventive and Social Medicine, Faculty of Medicine, Chulalongkorn University
Email: doctorkrit@gmail.com Phone:+668-6605-5088

October 3-7 • San Francisco, CA • www.idweek.org
IDWeek 2018™
Advancing Science, Improving Care