An Explainable Deep-Learning for Detection and Classification of Pulmonary Tuberculosis

Doyun Kim, PhD; YiRang Shin, MS, Massachusetts General Hospital; Sehyo Yune, MD, MPH, MBA; Myeongchan Kim, MD; Kyoungdoo Song, MD; Joo Won Chung, MD; Jin Kyem Kim, MD; Synho Do, PhD

Introduction
Deep learning for chest radiography classification remains limited for mass adoption in clinical practice due to small and specific datasets as well as the inability to explain decisions made by deep learning algorithms. In this study, we utilize models pre-trained on different publicly open dataset of Chest x-ray14 and CheXpert for triaging and screening pulmonary tuberculosis (TB). We aim to mimic the cognitive function of doctors to learn different radiographic manifestation of TB through transfer learning. Also, a prediction basis is retrieved from an atlas created in training process for each significant feature of TB to understand the reasons for the predictions made by each model.

Hypothesis
Through extracting features of lung manifestations with transfer learning, we expect to overcome the restrictions of small and specific datasets in medical domain for screening TB.

Methods
In this study, different DCNNs (DenseNet121) trained on Montgomery and Shenzhen pulmonary TB open datasets (National Library of Medicine, National Institutes of Health, Bethesda, MD, USA) are used to classify whether a radiograph have manifestations of TB or not. The pre-trained weights are obtained from training of 100,000 labeled CXR images of a largest publicly available dataset of Chest x-ray14 and CheXpert, labels selected by radiologists that are thought to be related to TB observations. In order to understand the basis of the algorithms’ decision, we feed all the training images through the trained network and collect significant feature maps of all convolutional layers to create a feature atlas. Significant features are selected by maximum activation values. During classification of a new image, a Gradient-weighted Class Activation Map (Grad-CAM) is generated for the image and the feature with the lowest L2 distance with the Grad-CAM is retrieved from the atlas to justify the prediction of the network.

![Fig.1 System overview. An illustration of the explainable TB detection and classification. The system includes DCNNs pre-trained on Chest x-ray 14 and CheXpert, and a framework of atlas creation and prediction basis retrieval](image-url)
Results
Using transfer-learning with chest x-ray datasets, the Area under the ROC curve (AUC) improved considerably, where Chest x-ray14 pre-trained model yields the best performance with 0.949 AUC. To show that the proposed models introduce more confident localization of TB manifestations, a region with top 20% of the Grad-CAM is generated and evaluated against ground-truth (GT) segmentation annotated by radiologist. The Intersection over union ratio (IoU) is considered, where Chest X-ray14 pre-trained model significantly outperforms ImageNet pre-trained model. Finally, the prediction basis is evaluated by re-labeling each chest radiograph into TB manifestation categories (upper consolidation, middle consolidation, lower consolidation, cavitation, nodule/mass, pleural effusion, hilar/mediastinal lymphadenopathy) set by 2 radiologists. The percentage of the positive instances correctly classified as positive from the prediction basis with the test image label as ground truth is computed.

<table>
<thead>
<tr>
<th>Pre-trained Model</th>
<th>ImageNet</th>
<th>Chest X-ray 14</th>
<th>CheXpert</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUC</td>
<td>0.896</td>
<td>0.949</td>
<td>0.911</td>
</tr>
<tr>
<td>IoU</td>
<td>0.107</td>
<td>0.258</td>
<td>0.181</td>
</tr>
<tr>
<td>Prediction Basis Evaluation</td>
<td>53.3%</td>
<td>64.0%</td>
<td>43.0%</td>
</tr>
</tbody>
</table>

Table 1. AUCs of ROC curves, pathology localization accuracy comparison, and prediction basis evaluation for tuberculosis classification in different DCNN pre-trained model and

Fig 2. Example of localization result and prediction basis. Example of correct ground truth segmentation (in red), region of top 20% of Grad-CAM (in yellow) plotted over the original image. The prediction basis shows retrieved atlas images from different pre-trained model.

Conclusion
We present different transfer-learning approaches for CNN-based TB classification with satisfactory results obtained with networks pre-trained on Chest x-ray14 datasets. As a solution to the ‘black box’ problem, an atlas of distinguishing features was used to provide ‘visual explanations’ of the system’s decision for clinical assistance.
Statement of Impact
We have utilized transfer learning to extract features of TB manifestations and rationalized the model predictions to improve clinician’s trust in ML algorithms and facilitate the adoption of ML application in healthcare.

Keywords
deep learning, 3D convolutional neural network, tuberculosis, computer assisted diagnosis