A Novel Approach for Tuberculosis Screening Based on Deep Convolutional Neural Networks

Sangheum Hwang*[a1], Hyo-Eun Kim*[a], Jihoon Jeong[b], Hee-Jin Kim[c]
[a]Lunit Inc., Seoul, Korea;
[b]Department of Mobile Convergence Technology, Kyung Hee Cyber University, Seoul, Korea;
[c]The Korean Institute of Tuberculosis, Seoul, Korea

ABSTRACT

Tuberculosis (TB) is one of the major global health threats especially in developing countries. Although newly diagnosed TB patients can be recovered with high cure rate, many curable TB patients in the developing countries are obliged to die because of delayed diagnosis, partly by the lack of radiography and radiologists. Therefore, developing a computer-aided diagnosis (CAD) system for TB screening can contribute to early diagnosis of TB, which results in prevention of deaths from TB. Currently, most CAD algorithms adopt carefully designed morphological features distinguishing different lesion types to improve screening performances. However, such engineered features cannot be guaranteed to be the best descriptors for TB screening. Deep learning has become a majority in machine learning society. Especially in computer vision fields, it has been verified that deep convolutional neural networks (CNN) is a very promising algorithm for various visual tasks. Since deep CNN enables end-to-end training from feature extraction to classification, it does not require objective-specific manual feature engineering. In this work, we designed CAD system based on deep CNN for automatic TB screening. Based on large-scale chest X-rays (CXRs), we achieved viable TB screening performance of 0.96, 0.93 and 0.88 in terms of AUC for three real field datasets, respectively, by exploiting the effect of transfer learning.

Keywords: Tuberculosis screening, computer-aided diagnosis, convolutional neural network, transfer learning

1. INTRODUCTION

Since 1990 there have been substantial achievements in detection and treatment of tuberculosis (TB): mortality from TB has fallen by 47%. Despite of the improvements, TB still ranks alongside HIV as a leading cause of death worldwide. In 2014, 1.5 million and 1.2 million people were died due to TB and HIV, respectively [1]. TB patients can be recovered with over-90% cure rate if they reach timely treatments under early diagnosis. However, due to the lack of radiography and radiologists which are crucial factors for early diagnosis and treatments, the death toll is still significantly high especially in developing countries. Therefore, developing an accurate and reliable computer-aided diagnosis (CAD) system can largely contribute to reduce mortality caused by TB worldwide.

CAD systems for TB have been studied on a conventional framework of computer vision so far. Specifically, previous approaches commonly have a serial procedure consisting of several processing steps such as image preprocessing, boundary segmentation, feature extraction and classification. For example, histogram analysis of the image (e.g., histogram equalization) is usually applied to enhance the contrast at the boundaries as preprocessing, and then region-of-interests (ROIs) that are informative regions for diagnosis (i.e. lung in case of TB detection) are segmented. To extract features from them, carefully designed shape and texture descriptors are considered, and finally popular classification algorithms are trained to classify normal and abnormal images. For recent and detailed survey, please refer to [2].

The most important step in the conventional approaches is to extract informative and discriminative features from images. These features have been designed based on domain-specific knowledge. However, such manually designed features are limited to describe a number of variations existing within abnormal images. Recently, deep CNN has shown

* These two authors contributed equally to this work, and are listed alphabetically
[1] shwang@lunit.io; phone 82 10 8897-8113; lunit.io

Medical Imaging 2016: Computer-Aided Diagnosis, edited by Georgia D. Tourassi, Samuel G. Armato III,
Proc. of SPIE Vol. 9785, 97852W · © 2016 SPIE · CCC code: 1605-7422/16/$18 · doi: 10.1117/12.2216198

promising performances in various computer vision tasks including object classification, detection, and segmentation. It does not require any domain knowledge if we have a large-scale dataset since it extracts and learns meaningful features to discriminate target classes during training. These data-driven features are more effective in terms of discriminability.

In this work, we show the effectiveness of deep CNN for TB screening. To the best of our knowledge, this is the first CNN-based TB screening system without lesion-specific manual feature engineering. For this, we collected relatively large-scale CXRs (approximately 10k CXRs) and their corresponding true labels (i.e. TB or normal). To overcome difficulties in training deep neural networks, we employed a transfer learning strategy. It is verified that transferring low-level filters from pre-trained models based on large-scale general images is very effective for training. From cross-validated results and cross-dataset evaluations, it is shown that deep CNN gives a promising result without any domain knowledge and complicated image processing techniques.

2. TB SCREENING WITH DEEP CNN

In this section, deep CNN architecture for TB screening and the concept of transfer learning are introduced. Although there are no valid pre-trained models in medical applications for transfer learning, it is found that an initialization for filter weights of low-level convolutional layers from the models pre-trained on general images can be effective for training CNN with CXRs.

2.1 Deep convolutional neural networks (CNN)

CNN is a feed-forward neural network where the individual nodes are arranged in tiled to model the visual receptive field. Each set of parameters to be trained in convolutional layers (called as a convolutional filter) extracts meaningful visual concepts from original input images, while the set of parameters to be trained in fully-connected layers classifies the extracted visual features into target classes (e.g., TB or normal). The convolutional layers hierarchically abstract visual concepts from the raw input images such that lower convolutional layers extract low-level features such as colors and/or shapes, while higher ones extract high-level visual concepts such as sub-parts of target objects [7].

Table 1. Deep convolutional neural network architecture in this study

Layer #*	Type (activation)	Input Shape**	Filter Size*** – stride
C1	Convolution (ReLU)	(1, 500, 500)	(96, 1, 11, 11) – 4
M1	Max Pool		(3, 3) – 2
C2	Convolution (ReLU)	(96, 61, 61)	(256, 96, 5, 5) – 1
M2	Max Pool		(3, 3) – 2
C3	Convolution (ReLU)	(256, 30, 30)	(384, 256, 3, 3) – 1
C4	Convolution (ReLU)	(384, 30, 30)	(384, 384, 3, 3) – 1
C5	Convolution (ReLU)	(384, 30, 30)	(256, 384, 3, 3) – 1
M5	Max Pool		(3, 3) – 2
C6	Convolution (ReLU)	(256, 15, 15)	(256, 256, 3, 3) – 1
M6	Max Pool		(3, 3) – 2
F7	Fully Connected (ReLU)	(256*7*7)	(2048, 12544)
D7	Dropout		
F8	Fully Connected (ReLU)	(2048)	(2048, 2048)
D8	Dropout		
F9	Fully Connected (Softmax)	(2048)	(2, 2048)

*Layer # = C (Convolution), M (Max Pool), F (Fully Connected), D (Dropout)
**Input Shape = Convolution (# of channels, image height, image width), Fully Connected (# of flattened nodes)
***Filter Size = Convolution (# of filters, # of channels, kernel height, kernel width), Max Pool (kernel height, kernel width), Fully Connected (# of output nodes, # of input nodes)

Since CNN extracts the most discriminative features according to target objective (e.g., TB classification) from given data by itself, it does not require hard manual feature engineering relying on domain-specific knowledge. In previous TB screening algorithms, for example, morphological characteristics of various lesion cases should be preliminarily defined appropriately [3].

For this study, we designed our deep CNN based on well-known CNN architecture called *alexnet* which is trained for general image recognition [4]. We added one extra convolutional layer for feature extraction since the resolution of the input CXR is relatively high compared to general object recognition tasks, i.e. the input size in this study is 520×520 while [4] used 256×256 sized images. Furthermore, we reduced the number of hidden nodes in fully-connected layers (i.e. classifier) because of fewer training images and classes. Table 1 summarizes our deep CNN architecture. Note that the input size of the first convolutional layer C1 is 500×500 since we randomly cropped 520×520 images to 500×500 for dataset augmentation.

2.2 Transfer Learning (TL)

In practice, it is unusual to have sufficient amount of data for training deep networks. To overcome this issue, it is common to use pre-trained CNN models based on a large-scale dataset such as an ImageNet dataset which contains 1.2 million images with 1000 classes [8]. The pre-trained models are used as initial values of network weights. Based on such well-defined initial weights, deep CNN can be trained properly even with the relatively small-scale datasets. This scheme is called transfer learning. Unfortunately, this is not directly applicable to medical images since the characteristics of medical images are quite different from those of general images.

Given the training and validation datasets, we trained the CNN with the architecture described in Table 1 with randomly initialized weights in all layers. We observed that the number of given images is not sufficient for training such deep networks (see Figure 2). In other words, it is difficult to obtain well-trained parameters (network weights) in terms of target objective function (e.g., cross-entropy loss modeling classification errors) in such a high-dimensional parameter space (33,317,827-dimensional parameter space in this architecture).

As shown in Figure 2 (see cost curves labeled to "Training loss w/o TL" and "Validation accuracy w/o TL"), training loss decreases slowly with unsatisfactorily converged performance of validation accuracy. Note that the experiments were repeated three times to reduce the effect of random split, and the corresponding datasets are called Dataset 1, 2, and 3, respectively. Since lower convolution layers normally extract low-level features such as edges and/or curves from the input images, we reused pre-trained convolutional parameters (filters from the first and second convolution layers in deep CNN trained based on the ImageNet classification dataset [4]) for the initial values of parameters in lower convolutional layers (C1 and C2) of our CNN. Figure 1 visualizes the filters of the first convolutional layer C1 transferred from the pre-trained model. Note that channel size of original filters is three since the input images used for training in their architecture have RGB color components. To transfer those filters into our CNN to be trained using single-channel CXRs, we simply sum filter values along channel dimension. From Figure 1, it can be observed that these are useful in terms of detecting low-level information from input images such as edges and orientations.

Figure 2 shows that transfer learning with low-level filters is effective to train our CNN. Optimization proceeds from much better initial points with the transferred low-level filters, and consequently the network easily finds a good local optimum compared to the network with random initialization. Training procedures and experimental results will be discussed in more detail in Section 3.

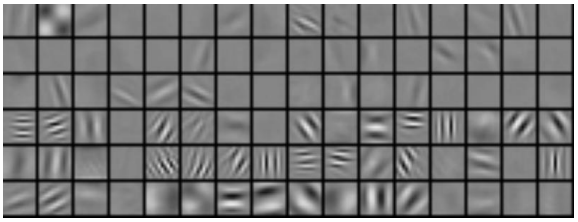
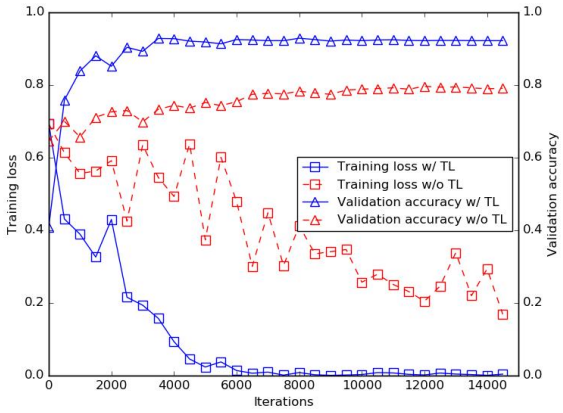
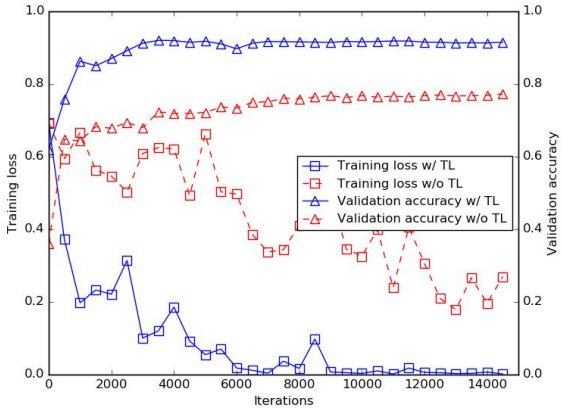


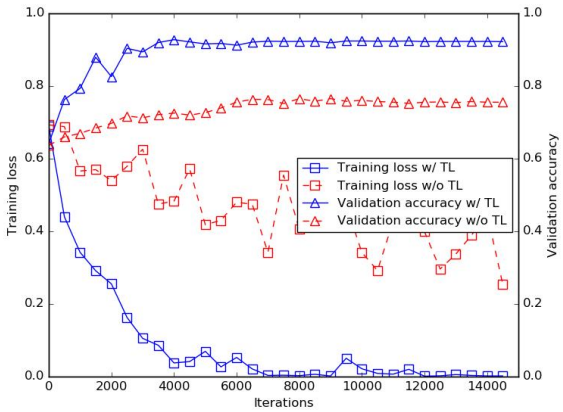
Figure 1. Visualization of initial filters in the first convolution layer C1. The total number of filters in C1 is 96 and each filter has 11×11 size.



(a) Training curves for Dataset 1



(b) Training curves for Dataset 2



(c) Training curves for Dataset 3

Figure 2. Training loss (cross-entropy loss) and validation accuracy curves for three different datasets. The blue-sold lines and red-dashed lines represent curves with and without transfer learning, respectively. The triangle and square marker represent validation accuracy and training loss, respectively.

3. EXPERIMENTAL RESULTS

3.1 Datasets and experimental setup

We used three CXR datasets, namely KIT, MC, and Shenzhen sets as described in Table 2 in this study. All the CXRs used in this work were de-identified by the corresponding image providers.

Table 2. Descriptions of datasets used in this study.

Dataset Name	Description
KIT	10,848 DICOM data, consisting of 7,020 normal and 3,828 abnormal (TB) cases, from the Korean Institute of Tuberculosis (KIT) under Korean National Tuberculosis Association (KNTA), South Korea.
MC	138 PNG data, consisting of 80 normal and 58 abnormal (TB) cases, from National Library of Medicine, National Institutes of Health, Bethesda, MD, USA [3, 5].
Shenzhen	662 PNG data, consisting of 326 normal and 336 abnormal (TB) cases, from Shenzhen No. 3 People's Hospital, Guangdong Medical College, Shenzhen, China [3, 5].

To verify the screening performance of deep CNN, KIT set (10,848 CXRs) was randomly divided into training (70%), validation (15%) and test (15%) sets. The training set is used for training deep CNN, while the validation set is used for checking the validity of trained deep CNN, and finally the screening performance is measured using test set. This random split was repeated 3 times to evaluate the performances (i.e. 3-fold cross validation) in an unbiased way. Other two datasets, MC and Shenzhen sets, were utilized to show the cross-dataset performances of deep CNN trained with KIT set.

We used 520×520 resized CXRs for efficiency of training, and designed CNN with an architecture described in Table 1. During training, resized CXRs are randomly cropped to 500×500 and also randomly flipped horizontally (i.e. mirroring). Such data augmentation techniques improve the generalization capability of CNN by providing subtle variations to input images. Note that any additional data augmentation that may distort the content of images was not used.

All weights in each layer (except for the first and second convolutional layers C1 and C2; their weights were transferred from pre-trained model [4]) were initialized from a zero-mean Gaussian distribution with standard deviation 0.01, and initial biases were set to 0. We considered an initial learning rate 0.01 and it was decreased by a factor of 2 for every 30 epochs. For transfer learning of low-level filters, we set an initial learning rate to 0.001 for the corresponding layers, i.e. C1 and C2. The network was trained via stochastic gradient descent with momentum 0.9 and the minibatch size was set to 64. The weight decay parameter was determined by a grid search through the comparison of validation accuracy. All the experiments described in this work were performed based on *caffe*, a public deep learning framework [6].

3.2 Transfer learning effect and performances for the KIT set

Figure 2 shows the effects of transfer learning of low-level filters. As shown in this figure, training cost converges much faster with higher validation accuracy (i.e. better local minima) due to the effect of transfer learning. Such transfer learning of lower pre-trained filters is effective since both CXRs and general images certainly share the low-level visual concepts (e.g., edges, curves) although they are semantically different in high-level concepts (e.g., parts of objects). It should be noted that this finding is important especially in medical imaging domain since the number of available images for training is usually limited.

Figure 3 shows ROC curves with area-under-curve (AUC) of three repeated experiments. As shown in this figure, all the three experiments show comparable ROC curves with high AUC (0.964 in average).

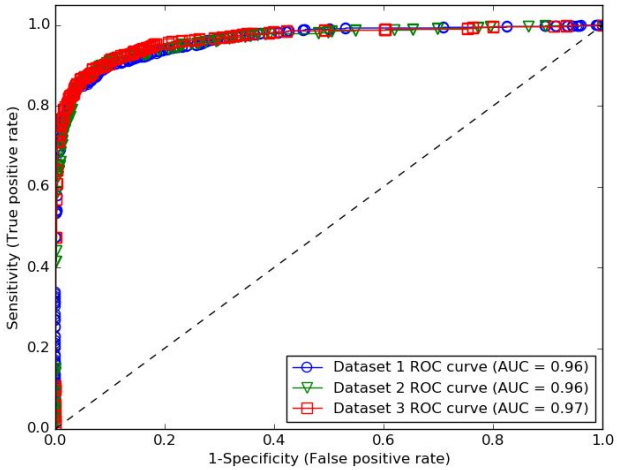


Figure 3. ROC curves with AUC values for three different datasets. Circle, triangle and square represent ROC curves from each dataset. Each AUC value is shown in the legend.

Screening performances between deep CNN with and without transfer learning of low-level filters are summarized in Table 3. As expected, deep CNN with transfer learning shows better performance in terms of AUC, accuracy and average precision for each class.

Table 3. Performance comparison between deep CNN with and without transfer learning

	w/o Transfer Learning				w/ Transfer Learning			
	AUC	Accuracy	AP(pos)	AP(neg)	AUC	Accuracy	AP(pos)	AP(neg)
Dataset 1	0.824	0.773	0.750	0.884	0.963	0.902	0.951	0.973
Dataset 2	0.828	0.788	0.759	0.879	0.963	0.905	0.950	0.974
Dataset 3	0.796	0.758	0.721	0.868	0.967	0.903	0.957	0.976
Average	0.816	0.773	0.743	0.877	0.964	0.903	0.953	0.974

3.3 Cross-dataset performance analysis

To verify the cross-dataset performances of the CNN trained with KIT set, additional experiments were performed using MC and Shenzhen sets. Table 3 summarizes the performance of the proposed deep CNN as well as the performance of the previous work on the same test sets [3]. Note that this is not for the performance comparison, but for estimating the screening ability of our deep CNN. Net *k* denotes the trained deep CNN using Dataset *k*. We averaged the class probabilities from Net 1, 2, and 3 to obtain ensemble results. The ROC curves are plotted in Figure 4.

Table 4. Screening performances for MC and Shenzhen sets

	Montgomery County (MC)				Shenzhen Hospital (Shenzhen)			
	AUC	Accuracy	AP(pos)	AP(neg)	AUC	Accuracy	AP(pos)	AP(neg)
Net 1	0.877	0.652	0.864	0.895	0.916	0.831	0.931	0.898
Net 2	0.864	0.826	0.870	0.869	0.918	0.834	0.933	0.901
Net 3	0.845	0.601	0.860	0.846	0.919	0.830	0.935	0.904
Ensemble	0.884	0.674	0.890	0.890	0.926	0.837	0.940	0.910
TMI'14 [3]	0.869	0.783	-	-	0.900	0.841	-	-

The screening performances for MC and Shenzhen sets are slightly lower than those obtained from KIT set since there exist different modalities between these datasets such as various nationalities, different X-ray equipment for image acquisition, etc. We observed that output probability distributions from MC and Shenzhen sets are shifted due to those different modalities. The best accuracies for MC and Shenzhen sets are 0.877 at probability threshold 0.8 and 0.847 at probability threshold 0.6, respectively. These issues can be dealt with if the deep CNN is trained using more datasets that cover a wide-range of such modalities.

Table 4 shows that the trained network classifies Shenzhen set better than MC set. We found that it is caused by intrinsic image characteristics of CXRs. More than 90% of training images (i.e. KIT set) are CXRs from Digital Radiography, and MC and Shenzhen sets are from Computed Radiography and Digital Radiography, respectively. Despite of such different nature between datasets, our CNN shows considerable and robust screening performances.

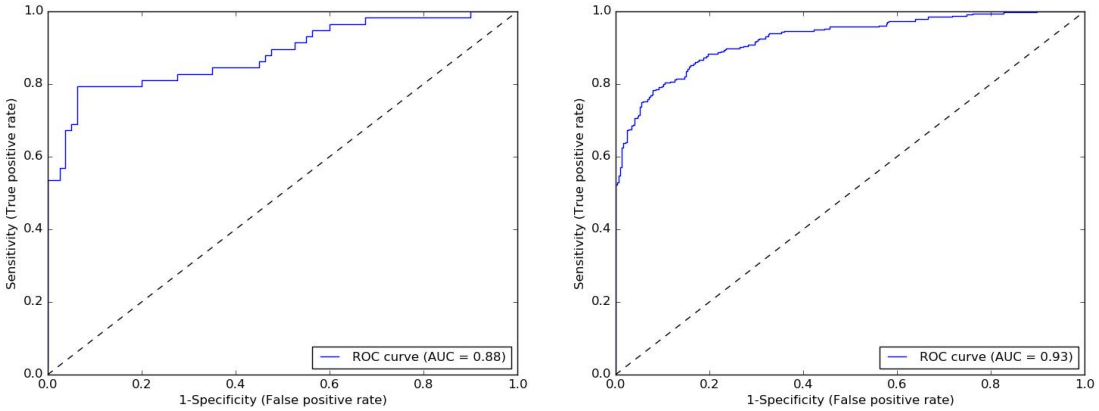


Figure 4. ROC curves for the MC set (left) and Shenzhen set (right). Each AUC value is shown in the legend.

4. CONCLUSION

In this work, we propose an automatic TB screening system based on deep CNN. This is the first deep CNN-based TB screening system trained on CXRs. Since CNN extracts the most discriminative features according to target objective from given data by itself, the proposed system does not require manually designed features for TB screening. Also, it is shown that transfer learning from lower convolutional layers of pre-trained networks resolves the difficulties in handling high-resolution medical images and training huge parameters with limited number of images. Computational experiments are conducted using three real field datasets, KIT, MC and Shenzhen sets, and the results show that the proposed system has high screening performance in terms of various performance metrics.

ACKNOWLEDGEMENTS

This work was supported in part by The Korean Institute of Tuberculosis (KIT) under Korean National Tuberculosis Association (KNTA), and in part by Korea Digital Hospital Export Agency (KOHEA). The authors would like to thank KIT under KNTA for providing CXRs for the KIT set, and the authors of [3,5] for kindly providing CXRs for the MC and Shenzhen sets.

REFERENCES

[1] World Health Organization (WHO), "Global tuberculosis report," 2015, http://www.who.int/tb/publications/global_report/en/

[2] S. Jaeger, A. Karargyris, S. Candemir, J. Siegelman, L. Folio, S. Antani, and G. Thoma, "Automatic screening for tuberculosis in chest radiographs: a survey," Proc. AME Quantitative Imaging in Medicine and Surgery (2013).

[3] S. Jaeger, A. Karargyris, S. Candemir, L. Folio, J. Siegelman, F. Callaghan, Zhiyun Xue, K. Palaniappan, R. K. Singh, S. Antani, G. Thoma, Yi-Xiang Wang, Pu-Xuan Lu, and C. J. McDonald, "Automatic tuberculosis screening using chest radiographs," IEEE Trans. Medical Imaging 33(2), 233-245 (2014).

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Proc. Advances in Neural Information Processing Systems (2012).

[5] S. Candemir, S. Jaeger, K. Palaniappan, J. P. Musco, R. K. Singh, Zhiyun Xue, A. Karargyris, S. Antani, G. Thoma, and C. J. McDonald, "Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration," IEEE Trans. Medical Imaging 33(2), 577-590 (2014).

[6] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," Proc. ACM International Conference on Multimedia (2014).

[7] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," Proc. European Conference on Computer Vision (2014).

[8] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei- Fei., "Imagenet: A large-scale hierarchical image database," Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 248–255 (2009).