TX-CNN: DETECTING TUBERCULOSIS IN CHEST X-RAY IMAGES USING CONVOLUTIONAL NEURAL NETWORK

Chang Liu⋆*, Yu Cao*⋆*, Marlon Alcantara*⋆*, Benyuan Liu*⋆*, Maria Brunette*⋆*, Jesus Peinado*†*, Walter Curioso*††

⋆ University of Massachusetts, Lowell, MA, USA
† Partners in Health Perú
†† University of Washington, Seattle, WA, USA

ABSTRACT

In Low and Middle-Income Countries (LMICs), efforts to eliminate the Tuberculosis (TB) epidemic are challenged by the persistent social inequalities in health, the limited number of local healthcare professionals, and the weak healthcare infrastructure found in resource-poor settings. The modern development of computer techniques has accelerated the TB diagnosis process. In this paper, we propose a novel method using Convolutional Neural Network(CNN) to deal with unbalanced, less-category X-ray images. Our method improves the accuracy for classifying multiple TB manifestations by a large margin. We explore the effectiveness and efficiency of shuffle sampling with cross-validation in training the network and find its outstanding effect in medical images classification. We achieve an 85.68% classification accuracy in a large TB image dataset, surpassing any state-of-art classification accuracy in this area. Our methods and results show a promising path for more accurate and faster TB diagnosis in LMICs healthcare facilities.

***Index Terms*—** convolutional neural network, image classification, deep learning, tuberculosis diagnosis, computer-aided diagnosis

1. INTRODUCTION

Tuberculosis(TB) is a global epidemic that causes the death of 1.8 million people worldwide, annually. Recent data from the World Health Organization (WHO, 2016)[1] indicates that there are more than 9 million new cases found all over the world, among which over 80% are found in South-East Asia, Western Pacific, and Africa. TB is curable and preventable, but in resource-poor and marginalized communities with weak healthcare infrastructure, it is difficult to diagnose because of the high prevalence and constrained resources for better diagnosis and effective treatment follow-up. The global TB report from WHO declares that it is necessary to eliminate TB epidemic using innovative methods such as mHealth to improve the diagnosis process and promoting patient adherence to the medical treatment.

The modern development of computer techniques has accelerated the TB diagnosis among the resource-poor areas, however, there is still a wide gap between the technological advancement and the clinical practices. The gap is mainly caused by two major barriers: 1) lacking large-scale, real-world, well-annotated and public-available X-ray image database. Even though some databases like ImageCLEF, JSRT Digital Image Database[2] and ANODE Grand Challenge Database contain TB images, they are mostly focused on one or two TB manifestations(e.g, pulmonary nodule); 2) lacking high-performance computing system for accurate diagnosis by analyzing the chest X-ray images. The use of computer-aided chest radiography for TB screening and diagnosis[3, 4, 5] has been limited due to the modest sensitivity and specificity, and high inter- and intra-observer differences in reporting shapes of tuberculosis on radiographic images. How to use precise computer algorithm for automatic TB screening and detection still remains to be a challenging problem for researchers.

Efforts described here are part of a mobile health (mHealth) integrative project aimed at reducing patient wait time to be diagnosed with TB by implementing a socio-technical solution to optimize the diagnosis process in a high-burden TB area in Lima, the capital of Perú. In this paper, we propose a novel deep learning method with CNN and transfer learning for classifying TB manifestations in chest X-ray images. Our algorithm and training protocol show outstanding accuracy and are proven to be practical and stable for various CNN architectures(e.g., AlexNet[6], GoogLeNet[7]). Experimental results show a wide potential for medical images analysis and TB diagnosis.

2. BACKGROUND AND RELATED WORK

Most of the current research in computer-aided image analysis for TB screening from X-ray image are focused on two aspects. The first aspect is the computer-aided screening and scoring algorithms using chest radiographic features for the TB diagnosis[3, 4, 5, 8, 9]. Researchers develop various types of visual features and classification algorithms to score and screen different types of TB manifestations. Related papers employ the texture features(e.g., Local binary pattern(LBP)), Daubechies wavelets or geometry features (e.g., circularity, Hessian shape features). For classification algorithm, they range from simple threshold-based approach or k-nearest neighbors(k-NN) algorithm to decision tree and support vector machine(SVM). Another aspect is to explore the organ and pathology level for X-ray image categorization[10]. Most of the research focus on studying of local patch representation of the image content(e.g., visual bag of words(BoW) approach). After extracting these simple features from the dense samples, more non-linear kernel-based classifier is applied to discriminate between the healthy and pathological cases.

In recent years, deep learning[11, 12] shows promising advances for feature learning and representation. It aims to learn multiple levels of representation and abstraction from data such as image, video, audio, and text. Astonishing progress has been made in computer vision, speech recognition, multimedia analysis and drug design. In medical image analysis[13, 14], researchers try various models to learn powerful features for tumor detection[15, 16], cancer diagnosis[17, 18] and illness prevention. Due to the limited training image and specific image characteristics in the medical image, even though some progress has been made in the various fields, it remains to be a challenging problem for accurate and efficient diagnosis using deep learning. In this paper, we explore the feasibility of applying deep learning in TB diagnosis and detection using chest X-ray images. To the best of our knowledge, it is the first research trial and shows outstanding performance comparing with any other methods.

3. PROPOSED APPROACH

The ultimate goal of our research is to design and deploy a reliable and efficient system to classify various TB manifestations. Several efforts have been explored: 1) we study and evaluate various CNN architectures and training parameters for TB X-ray image dataset; 2) we apply the transfer learning techniques in chest X-ray images. We finetune the pretrained CNN model from natural image dataset(ImageNet[19]) to our medical X-ray image dataset to detect TB and find its effectiveness and efficiency than training from scratch. Several other research efforts(e.g., shuffle sampling, cross-validation) and tasks are explored to verify its correctness. More details are discussed in the following subsections.

3.1. Convolutional Neural Network

Convolutional Neural Network(CNN) is widely used in multiple computer vision tasks, such as image classification, object detection and visual question answering. While many efforts have been given to general tasks, few of them focus on medical images. Here we study two convolutional neural network architectures(AlexNet[6] and GoogLeNet[7]) with different model parameters. Comparing with general image applications[6, 7], deep learning models for medical images tend to be smaller[17], as the region-of-interest(ROI) are usually small. We adopt this schema and use various CNN models and smaller size of kernels to find the best fit for the TB chest X-ray images.

Fig.1 shows a basic structure for TB classification using LeNet[20]. A CNN model is usually consist of convolutional layers, pooling layers, and fully-connected layers. Each layer is connected to the previous layer via kernels that have predefined, fixed-size receptive field. The weights within each layer are shared to reduce complexity and computation. CNN model learns the parameters from a large-scale dataset to represent the global and local features in the image. Every model architecture has various types of layers and activation functions to exhibit strong feature representation ability than human-engineered features. More details of the network structure are discussed in [6, 7, 11].

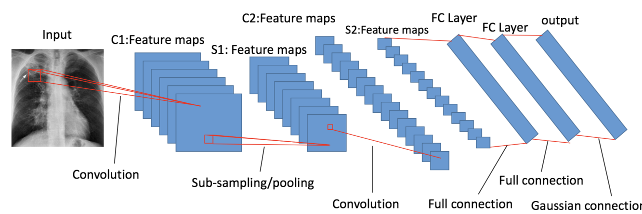

Fig. 1. CNN architecture(LeNet[20]) for TB classification.

3.2. Transfer Learning

Transfer learning aims to store the learned knowledge from one domain and apply it to another different but related domain. When training from scratch, it usually takes a lot of time because model parameters are all initialized with random Gaussian distribution and convergence are achieved after at least 30 epochs with a batch size of 50 images. Another challenge is that in the medical domain it's usually very hard to obtain large-scale, well-annotated images. Lacking of medical data usually makes it very hard to learn precise models for accurate predictions.

Recent studies[16, 18, 21] show that using a pretrained model from ImageNet dataset, then finetune with a more specific dataset yield outstanding classification and detection results. The reason behind this training protocol is that CNN gains general representation capability from pretraining in natural images. After finetuning, the model adjusts the parameter for representing the unique features in the specific images, while retaining the abilities to represent general image. We inherently adopt this training strategy, combine with shuffle sampling and cross-validation, and creatively apply it to the chest X-ray images for classifying TB manifestations. Our experiment shows the outstanding performance for TB diagnosis.

Category(Name of TB Manifestation)	Total Image
Miliary Disease(MI)	25
Cavitation(CA)	1182
Lympahadenopathy(LI)	202
Ghon Focus(GH)	27
Alveolar Infiltrates(AI)	2252
Other(OT)	560

Table 1. Data distribution in TB dataset.

4. EXPERIMENTS

We conducted several experiments on a private TB X-ray image dataset from Perú and followed the standard evaluation protocol, used multiple training models and got the average as the final accuracy.

4.1. Dataset Details

The dataset is from our Peruvian partners at "Socios en Salud", Partners In Health in Lima, Perú. This dataset contains 4701 images, 453 of them are labeled as normal(which means the patients don't have the TB) and 4248 are labeled as abnormal that contain various TB manifestations. Among the abnormal TB images, there are 6 categories, which indicate 6 different types of TB manifestations: miliary pattern, cavitation, lymphadenopathy, ghon focus, alveolar infiltrates and others. Table 1 illustrates the characteristics of the data. This less-category and unbalanced distribution casts a unique technical problem for classifying the image. In the next sections, we will show our architecture design and optimization that improves the performance by a large margin.

4.2. Architecture Details

We revise the AlexNet and GoogLeNet architecture for image classification. On the top of AlexNet/GoogLeNet, we place a softmax layer to get the score for each category. Each score ranges from 0 to 1 and represents the probability of classifying the manifestation correctly. When training AlexNet, we use the standard 7-layer network structure[6], which is consist of five convolutional layers and two fully-connected layers. Dropout and ReLU are deployed to address overfitting and convergence issue. For GoogLeNet, we use the 22-layer network structure[7] that employ the Inception model with dropout and ReLU. We implement our architecture using Caffe[22]. Our model is trained using Stochastic Gradient Descent(SGD), with a lot of combinations of different parameters. By experiment, we find that for AlexNet, a base learning rate of 0.01, a momentum of 0.9 and weight decay of 0.0005, would yield the best result within reasonable time. For GoogLeNet, we use a base learning rate of 0.001, a momentum of 0.9 and weight decay of 0.0002. Note that learning rate is not fixed and will decrease exponentially as iteration grows. Since the chest X-ray images are still not abundant for training large complex model, we use the pretrained model from ImageNet that are public available in Model Zoo among the Caffe[22] community and then finetune on our chest X-ray image database.

4.3. Shuffle Sampling

For very unbalanced dataset, it's very hard to learn a general classifier for all categories. How to augment data is crucial to learn an expressive classifier to classify all TB manifestations. Inspired by the previous work from [23], we propose a shuffle sampling technique to augment data. Our method is done before the training of CNN models, so it doesn't affect the training time substantially. According to our experiments, the training time for using shuffle sampling increases about 2 hours for our 5K images with AlexNet, while the accuracy boost from 53.02%[21] to 85.68%.

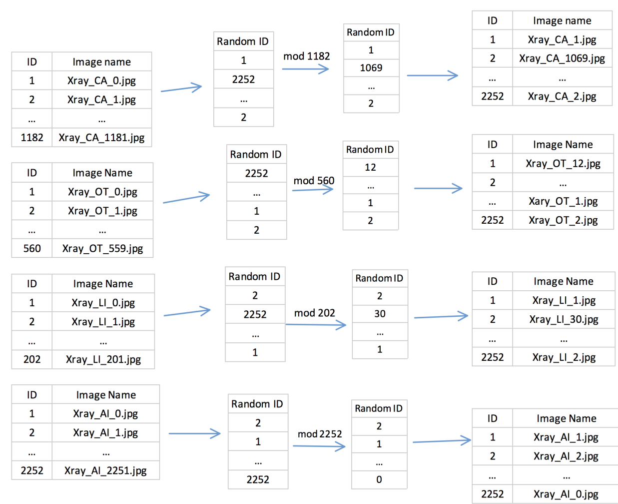

Fig. 2. Shuffle sampling for unbalanced data.

As Fig.2 shows, we first select the largest number of category as the baseline number of instances, marked as N. Then for each category, we generate N *unique* integers, each integer represents the index for one image. For each category, we calculate the mod value $I_i\%N_c$ as the final index, I_i is the index from the unique sequence of each category, N_{ci} is the number of maximum images for each category. Here N is 2252, and I_i is from the random array. After these shuffle samplings, we expand our dataset into $N * c$, where N is the maximum number of images for all the categories, and c is the number of categories.

4.4. Results

We study several cases of classifying the manifestations and evaluate their performance. For each model, we use non-shuffle and shuffle sampling settings for four common manifestations, these manifestations have more training images than others. We first pretrain on AlexNet using ImageNet

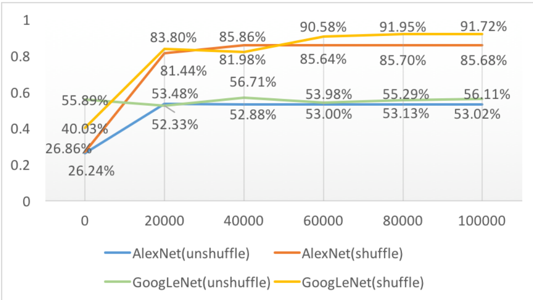

Fig. 3. Non-shuffle vs. shuffle classification accuracy.

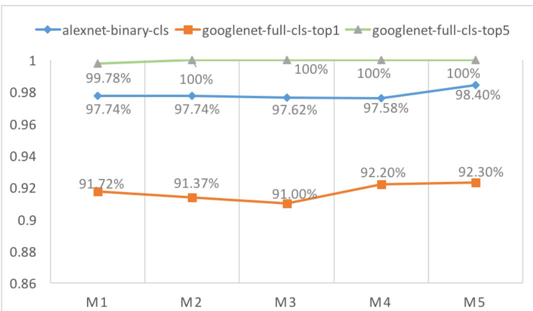

Fig. 4. Classification accuracy with cross validation. Use AlexNet for binary classification and GoogLeNet for full classification

dataset, then finetune using the same network structure with chest X-ray images. Fig.3 shows that shuffle one is more accurate than non-shuffle one in all iterations and remains stable in general cases.

The final classification accuracy using AlexNet is about 85.68%, a significant improvement from non-shuffle sampling's 53.02% in[21]. Note that since this strategy may generate repeated images for testing, influencing the final evaluation of accuracy, we use the trained-well model to retest all the images in the training data with the unique id to verify its correctness, we calculate the precision, recall, f1-score and miss rate. Table 2 shows the stability of our method to classify the images on the original training set.

We also conduct an experiment on abnormal detection using AlexNet, which is to determine whether a chest X-ray image contains TB or not. This is a binary classification problem, we use cross-validation repeatedly. By dividing the whole dataset into 5 equal folders, we used 4 folders as the training set and the remaining 1 folder as the test set. We also explore the deployment of GoogLeNet for all six manifestations with shuffle sampling and cross-validation to enhance the model prediction. Results are shown in Fig.4 and Fig.5.

Finally we evaluate the computation time. We use a Telsa K80 GPU for training and testing, it roughly takes 2 to 3 days nonstop to train the AlexNet or GoogLeNet models. After we have trained the model, we use pycaffe, scikit-learn and

Set	Class name	Precision	Recall	F1-score	Miss rate
Train	CA	0.90	0.94	0.92	0.06
	OT	0.87	0.99	0.93	0.01
	LI	0.91	1.00	0.95	0
	AI	0.98	0.91	0.94	0.09
Test	CA	0.87	0.93	0.90	0.07
	OT	0.90	1.00	0.95	0
	LI	0.91	0.98	0.94	0.02
	AI	0.97	0.90	0.93	0.10

Table 2. Evaluation on training and testing set

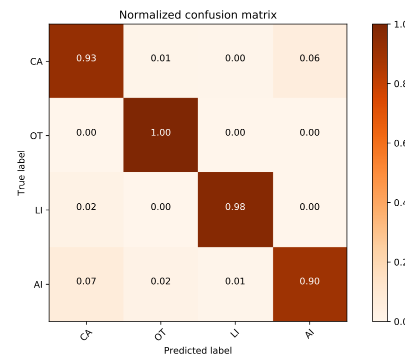

Fig. 5. Confusion matrix on test set

pandas-ml to evaluate model performance, it takes less than 1 minute to test one image when using GPU.

5. ACKNOWLEDGEMENT

This project is supported by NIH (NO: 1R01EB021900), NSF (NO.1547428, 1541434, 1440737, 1229213, and 1156639).

6. CONCLUSIONS AND FUTURE WORK

We design a novel method to apply CNN models to detect and classify TB manifestations in X-ray images. Work presented here is the first research trial using CNN for TB detection in a large TB dataset. Based on the research result and the specific technical problems in this large unbalanced, less-category dataset, we use a set of optimization solutions to further improve the accuracy. Our method shows the stability and universality in various CNN architectures. The next step is to collaborate with health science and engineering researchers to annotate the regions of the chest images for more accurate classification and localization. We will use more region-level information for preprocessing and study the algorithms to further improve the accuracy. We will also deploy a user-centered, mobile device-based computing system to expedite the TB diagnosis process and conduct the field-testing in TB clinics in a high-burden TB area in Lima, the capital of Perú.

7. REFERENCES

[1] W. H. Organization *et al.*, "Global tuberculosis report 2016," *World Health Organization*, 2016.

[2] J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi *et al.*, "Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules," *American Journal of Roentgenology*, vol. 174, no. 1, pp. 71–74, 2000.

[3] S. Jaeger, A. Karargyris, S. Antani, and G. Thoma, "Detecting tuberculosis in radiographs using combined lung masks," in *Engineering in Medicine and Biology Society*. IEEE, 2012.

[4] T. Xu, I. Cheng, R. Long, and M. Mandal, "Novel coarse-to-fine dual scale technique for tuberculosis cavity detection in chest radiographs," *EURASIP Journal on Image and Video Processing*, vol. 2013, no. 1, p. 3, 2013.

[5] Y.-L. Song and Y. Yang, "Localization algorithm and implementation for focal of pulmonary tuberculosis chest image," in *Machine Vision and Human-Machine Interface*. IEEE, 2010.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[7] C. Szegedy, W. Liu, Y. Jia, Sermanet *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.

[8] P. Maduskar, L. Hogeweg, R. Philipsen *et al.*, "Improved texture analysis for automatic detection of tuberculosis (tb) on chest radiographs with bone suppression images," in *SPIE Medical Imaging*. International Society for Optics and Photonics, 2013.

[9] O. M. Rijal, H. Ebrahimian, and N. M. Noor, "Determining features for discriminating ptb and normal lungs using phase congruency model," in *Biomedical and Health Informatics*. IEEE, 2012.

[10] U. Avni, H. Greenspan, E. Konen, M. Sharon, and J. Goldberger, "X-ray categorization and retrieval on the organ and pathology level, using patch-based visual words," *IEEE Transactions on Medical Imaging*, vol. 30, no. 3, pp. 733–746, 2011.

[11] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT Press, 2016.

[12] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *science*, vol. 313, no. 5786, pp. 504–507, 2006.

[13] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio *et al.*, "A survey on deep learning in medical image analysis," *arXiv preprint arXiv:1702.05747*, 2017.

[14] S. K. Zhou, H. Greenspan, and D. Shen, "Deep learning for medical image analysis," *Academic Press*, 2017.

[15] W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, and D. Shen, "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation," *NeuroImage*, vol. 108, pp. 214–224, 2015.

[16] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu *et al.*, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.

[17] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in *International Conference on Medical Image Computing and Computer-assisted Intervention*. Springer, 2013, pp. 411–418.

[18] H.-C. Shin, H. R. Roth, M. Gao, L. Lu *et al.*, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.

[19] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Computer Vision and Pattern Recognition*. IEEE, 2009, pp. 248–255.

[20] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[21] Y. Cao, C. Liu, B. Liu, M. J. Brunette, N. Zhang *et al.*, "Improving tuberculosis diagnostics using deep learning and mobile health technologies among resource-poor and marginalized communities," in *Connected Health: Applications, Systems and Engineering Technologies (CHASE)*. IEEE, 2016.

[22] Y. Jia, E. Shelhamer, J. Donahue *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014, pp. 675–678.

[23] L. Shen, Z. Lin, and Q. Huang, "Relay backpropagation for effective learning of deep convolutional neural networks," *arXiv preprint arXiv:1512.05830*, 2015.