Local-Global Classifier Fusion for Screening Chest Radiographs

Conference Paper · February 2017

DOI: 10.1117/12.2252459

CITATIONS
17

READS
159

7 authors, including:

Meng Ding
National Institutes of Health
23 PUBLICATIONS **302** CITATIONS
SEE PROFILE

Sameer K Antani
National Library of Medicine
335 PUBLICATIONS **5,884** CITATIONS
SEE PROFILE

Stefan Jaeger
National Institutes of Health
85 PUBLICATIONS **2,899** CITATIONS
SEE PROFILE

Zhiyun Xue
U.S. Department of Health and Human Services
76 PUBLICATIONS **1,484** CITATIONS
SEE PROFILE

Local-Global Classifier Fusion for Screening Chest Radiographs

Meng Ding[1], Sameer Antani[1], Stefan Jaeger[1], Zhiyun Xue[1], Sema Candemir[1], Marc Kohli[2], and George Thoma[1]

[1]Lister Hill National Center for Biomedical Communications, National Library of Medicine, National Institutes of Health, Bethesda, MD, 20894, USA

[2]Department of Radiology and Biomedical Imaging, University of California San Francisco, San Francisco, CA, 94143, USA

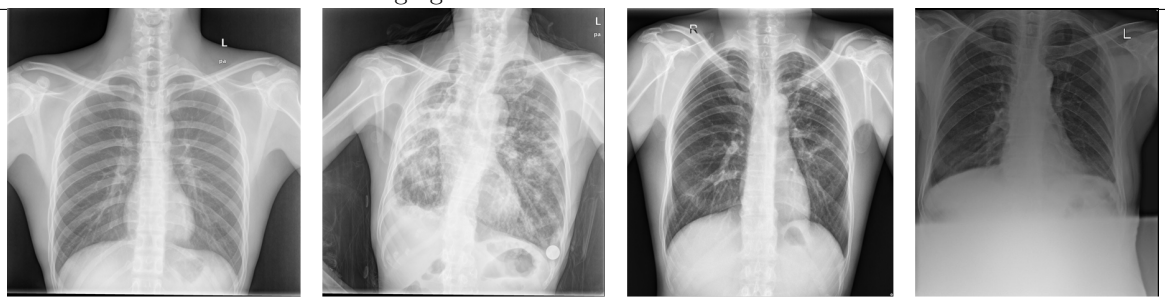
ABSTRACT

Tuberculosis (TB) is a severe comorbidity of HIV and chest x-ray (CXR) analysis is a necessary step in screening for the infective disease. Automatic analysis of digital CXR images for detecting pulmonary abnormalities is critical for population screening, especially in medical resource constrained developing regions. In this article, we describe steps that improve previously reported performance of NLM's CXR screening algorithms and help advance the state of the art in the field. We propose a local-global classifier fusion method where two complementary classification systems are combined. The local classifier focuses on subtle and partial presentation of the disease leveraging information in radiology reports that roughly indicates locations of the abnormalities. In addition, the global classifier models the dominant spatial structure in the gestalt image using GIST descriptor for the semantic differentiation. Finally, the two complementary classifiers are combined using linear fusion, where the weight of each decision is calculated by the confidence probabilities from the two classifiers. We evaluated our method on three datasets in terms of the area under the Receiver Operating Characteristic (ROC) curve, sensitivity, specificity and accuracy. The evaluation demonstrates the superiority of our proposed local-global fusion method over any single classifier.

Keywords: Chest Radiographs, Classifier Fusion, Pulmonary Abnormality Screening, Local-global, Tuberculosis (TB)

1. INTRODUCTION

Chest radiography is an efficient yet inexpensive examination for screening the pulmonary abnormality in population screening. It can be used to effectively detect various pulmonary diseases. Tuberculosis (TB), in particular,



(a) (b) (c) (d)

Figure 1. Examples of CXR images. (a) Normal, (b) Abnormal, (c) and (d) Abnormal with partial subtle manifestation of abnormalities.

is a severe comorbidity of HIV and its detection is critical for appropriate treatment. In order to address the imbalance between the large HIV affected population in resource constrained regions and the burden on the few available radiologists, researchers at the U.S. National Library of Medicine (NLM) have developed automated chest x-ray image analysis algorithms to detect and screen for pulmonary diseases with a special focus on detecting radiographic manifestations consistent with TB exposure.

There has been much interest in automated analysis of CXR images. In[1–3] , for example, various texture and shape features were extracted to train a binary classifier for TB screening with or without lung segmentation. Recently, deep learning model trained with non-medical images was exploited for chest pathology identification in[4] . While these methods are prominent to detect TB exposure with obvious and gross radiographic manifestations (e.g., consolidation, pleural effusion and blunt costophrenic angles), they could be prone to ignore the subtle and partial radiographic manifestations distributed locally (e.g., nodules and perihilar infiltrates), as shown in Fig. 1 (c) and (d). To the best of our knowledge, few works have been reported that are effective for local abnormalities in chest radiographs except[5] . However, the detection in[5] is operated in pixel level, which requires intensive annotation that is not available for middle / large size dataset.

Inspired by research in face recognition, where images are split into a number of sub-regions[6] , we propose an image sub-region based local abnormality detection, which is further enhanced by a GIST[7] descriptor based global classifier through a classifier fusion strategy. Our novel local-global classifier fusion method can detect abnormalities using the global representation of the whole image while retaining the influence of localized subtle manifestations. The proposed method has been evaluated on three datasets in terms of the area under the ROC curve (AUC), accuracy, sensitivity and specificity. The experimental results confirm the superiority of the proposed fusion method over any single classifier. Also, the results are encouraging and promising compared with state-of-the-art algorithms.

2. METHOD

Our proposed method consists of two complementary classifiers applied at different scales, i.e., a local and a global classifier, that are combined using linear fusion to improve overall classification performance. The framework with the training and testing steps is illustrated in Fig. 2.

2.1 Local Classifier

2.1.1 Feature Extraction in Sub-regions

We use a local classifier to enhance the impact of local subtle abnormalities. In this work, the lung area is first segmented by a method described in[8] . It is straightforward to obtain the bounding box of the lung area as the region of interest (ROI). The ROI is resized to 512×512 and is partitioned into six non-overlapping sub-regions as shown in Fig. 2. The reason for separating the lung area into six zones is to approximate the anatomical locations mentioned in radiology readings (upper zone: above the aortic arch, middle zone: between the aortic arch and lower pulmonary veins, lower zone: between lower pulmonary veins and the diaphragm). The proximate partitioning exploits the additional information from radiology readings and facilitates the training of sub-classifiers mentioned in Sec. 2.1.2. Different from[1] where several texture and shape features are extracted, in this work, to reduce the dimensionality of feature vector and the computational complexity, only three features, i.e., Frangi[9] , Histogram of Oriented Gradients (HoG)[10] and Local Binary Patterns (LBP)[11] are calculated in each sub-region and cascaded to compose the feature vector. Thus, we have six local feature vectors in each image.

2.1.2 Sub-Classifiers Voting Scheme

From the radiologists' reading we generate sub-labels for the sub-regions in addition to the (normal / abnormal) label for the whole image. Using the collection of feature vectors $\mathbf{X}_j$,where $j \in \{1, \cdots, 6\}$ from all training data and their corresponding sub-label vector $\mathbf{Y}_j$, we can train a Support Vector Machine (SVM) classifier for each sub-region, resulting in six sub-classifiers. The SVM in its original form is a supervised non-probabilistic classifier. Here, we use the distance between samples and the hyperplane as classification confidence and calculate a probability through a sigmoid function[12] . With this probability, we develop a voting scheme to fuse the decisions from these six classifiers. Each classifier's vote is considered valid if its probability is larger than an empirically determined threshold λ. The threshold eliminates relatively low confidence predictions leading to a more robust classification. Then, we collect all the valid votes and apply a strategy as '*one abnormality takes all*', that is if there exists abnormality in any valid votes, the result is abnormal and the probability is the mean of all confidences with valid abnormality; otherwise it is normal and the probability is the mean of valid confidences.

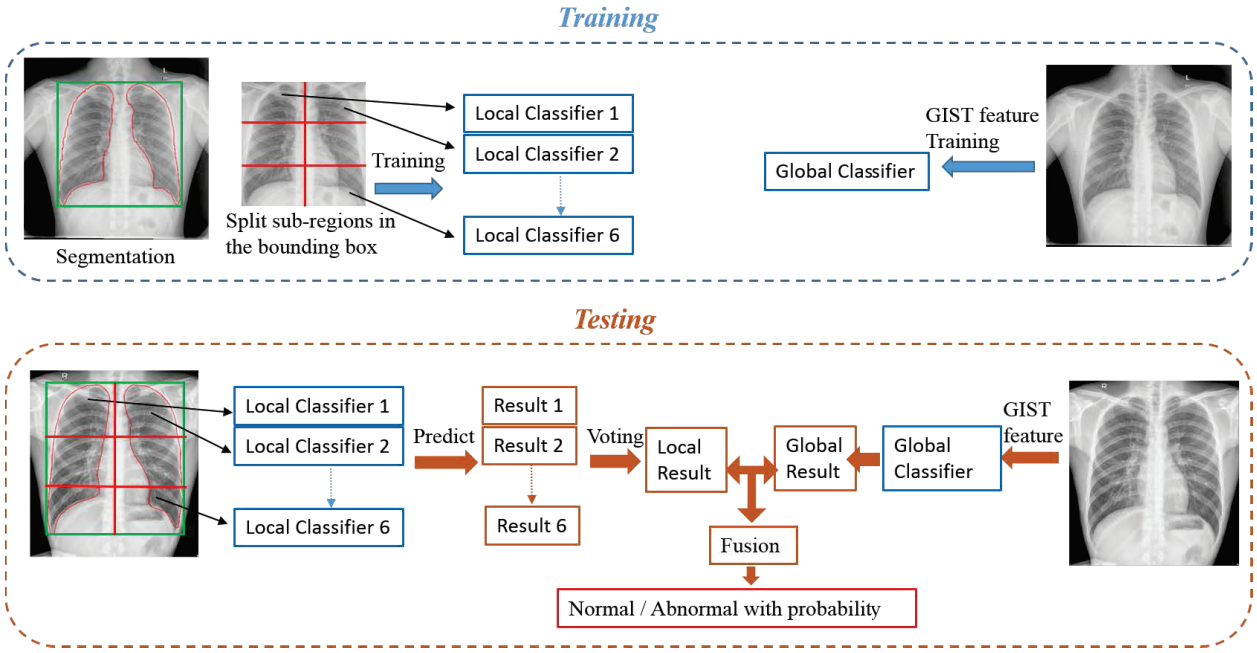


Figure 2. The framework of the proposed method which includes a local and a global classifier training and their fusion when testing.

The procedure with pseudo code is shown in Algorithm 1. This voting scheme can effectively reduce the false negative rate (*Type II Error*).

2.2 Global Classifier

To model the dominant spatial structure (shape, etc.), we use the GIST descriptor[7] to train a global classifier as a complement to our local detection scheme. The GIST descriptor, initially proposed for the scene recognition, has been successfully applied in CXR screening[2,4] . It captures the dominant spatial structure of a scene by concatenating a set of perceptual dimensions (naturalness, openness, roughness, expansion, ruggedness). In this paper, the low dimensional GIST representation of the gestalt CXR image is used to semantically discriminate normal/abnormal cases. We first extract the GIST feature vectors from resized CXR images (512×512). Then, we train a model termed global classifier using SVM with the radial basis function (RBF) kernel. Similarly, the global classifier results in a category label and a probability that represents the classification's confidence during the testing. Next, we will explain the fusion of our local and global classifier.

2.3 Local-Global Classifiers Fusion

With the labels and probabilities from the two complimentary classifiers, we develop a local-global classifier fusion method. Different with the 'one abnormality takes all' in Sec. 2.1.2, we combine the decisions by calculating

Algorithm 1 The procedure of Sub-Classifiers Voting

Partitioning the test image by 6 sub-regions.
for each sub-region **do**
    Predicting its *label* and $confidence$ using the corresponding classifiers;
    if $confidence > \lambda$ **then**
        predicted label is valid;
    else
        prediction is invalid;
    end if
end for
if *abnormality* exists in any valid prediction **then**
    $voting\ result = abnormal$;
    $propobility = mean(confidence(valid\ label == abnormal))$;
else
    $voting\ result = normal$;
    $propobility = mean(confidence(valid\ label == normal))$;
end if
return *voting result* and *propobility*;

weights for both classifiers in terms of their confidence,

$$Fused\ Decision = sign\left(\frac{p_1}{p_1+p_2} \times label_1 + \frac{p_2}{p_1+p_2} \times label_2\right), \tag{1}$$

where $(p_1, label_1)$ and $(p_2, label_2)$ are the probability and predicted label ($Normal\ is\ -1$ and $Abnormal\ is\ +1$) from the local and global classifiers, respectively. The larger the probability of the classifier, the greater the influence it has on the final decision. The $sign\ function$ makes the $Fused\ Decision$ as $Normal$ (-1) or $Abnormal$ $(+1)$, and the final probability is the larger value of p_1 and p_2.

3. EXPERIMENTAL RESULTS

3.1 Data

We evaluated our method on three datasets which are all posterior-anterior (PA) CXR images taken in standard clinical conditions. **Shenzhen, China Dataset:** This publicly available dataset[1] consists of 662 CXR images totally, including 336 cases with TB and 326 normal ones. Also, the detailed radiologists' reading that roughly

annotate the locations of abnormality are available as our ground-truth. **New Delhi, India Dataset**,[2] contains a training set (52 normal and 52 TB cases) and an independent test set (26 normal and 26 TB ones). Since the manifestations of TB in this dataset are obvious and distributed in the whole lung area, we set all the sub-labels as abnormal in the TB cases. **Kenya Dataset**, is collected in western Kenya recently by the National Library of Medicine, National Institutes of Health, USA in collaboration with Indiana University School of Medicine and Academic Model Providing Access To Healthcare (AMPATH), an NGO in Kenya providing one of the largest HIV treatment programs in sub-Saharan Africa. The dataset is composed of 729 normal and 290 abnormal cases with detailed radiologists' reading. Due to the imbalanced number of normal and abnormal cases, we randomly selected 290 normal samples combined with the 290 abnormal CXRs in our experiments.

3.2 Quantitative Results

We evaluated the methods in terms of AUC, accuracy, sensitivity and specificity. The ROC curves for the three datasets are shown in Fig. 3. Each ROC curve illustrates different possible operating points depending on the probability threshold for the classifier. The vertical axis indicates the sensitivity of the classification, while the horizontal axis indicates the false positive rate, which is $(1 - specificity)$. From Fig. 3, we can observe that training a single classifier using the three features (Frangi, HoG and LBP) extracted in the segmented lung field (marked as 'Baseline') has relatively low AUC; our local classifier boosted by six sub-classifiers obtains better result. Further, the classifier fusion enhances both the local and global classifiers and achieves the best AUC as 0.949, 0.982 and 0.76 in China, India and Kenya datasets, respectively. We summarize the accuracy, sensitivity and the corresponding specificity for the three datasets in Table 1. The quantitative results confirm the superiority of our fusion method. Also, we compared the proposed method with the state-of-the-arts methods in Table 2, which indicates that our fusion method is competitive and promising.

Table 1. Comparison of classifiers in this paper for the China, India and Kenya datasets (Accuracy, Sensitivity, Specificity). The 'Baseline' method is training a single classifier with three texture features, i.e., Frangi, HoG and LBP extracted from the segmented lung area.

Datasets	Baseline	Local Classifier	Global Classifier	Classifiers Fusion
China	(0.83, 0.84, 0.81)	(0.86, 0.85, 0.87)	(0.82, 0.81, 0.83)	(0.90, 0.91, 0.89)
India	(0.80, 0.77, 0.83)	(0.88, 0.88, 0.88)	(0.94, 0.96, 0.92)	(0.96, 0.96, 0.96)
Kenya	(0.67, 0.63, 0.71)	(0.70, 0.67, 0.72)	(0.70, 0.72, 0.68)	(0.73, 0.79, 0.67)

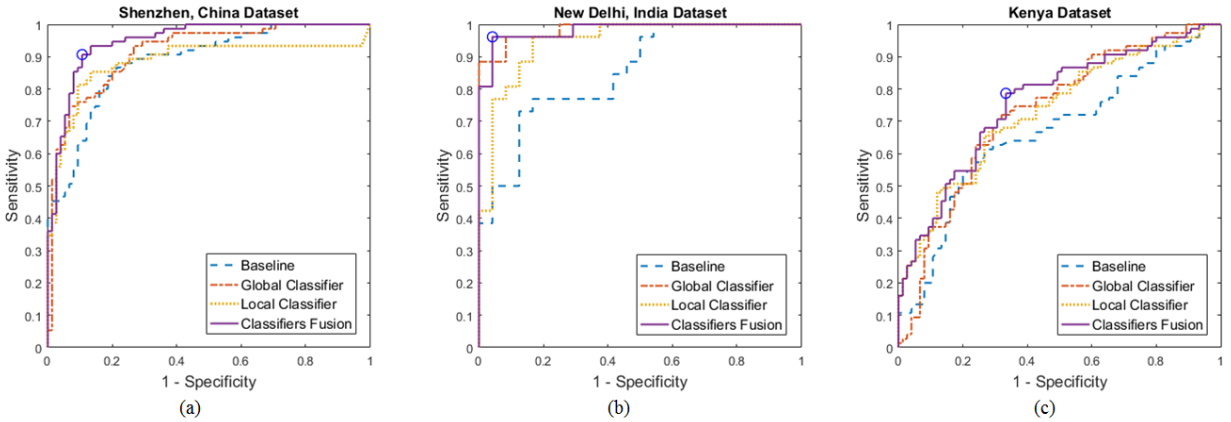


Figure 3. The ROC curves that are obtained using (a) China, (b) India and (c) Kenya datasets, respectively.

Table 2. Comparison between our classifier fusion and state-of-the-arts methods for two public datasets (AUC, Accuracy).

	Jaeger et al.[1]	Santosh et al.[3]	Chauhan et al.[2]	Classifiers Fusion
China	(0.9, 84%)	(0.93, 86.36%)	—	(0.949, 90%)
India	—	—	(0.957, 94.2%)	(0.9824, 96%)

4. CONCLUSION

We have developed a local-global classifier fusion method to improve the performance of the pulmonary abnormality screening in chest radiographs, with a particular focus on findings compatible with TB. To capture the local subtle manifestation of abnormality in the CXRs we develop a local classifier which is boosted by multiple sub-region based classifiers through a voting scheme. A global classifier is also trained using the GIST descriptor which can model the gestalt spatial structure of the images. The two complementary classifiers are combined using a linear fusion technique. We evaluated our proposed method on three datasets and the experimental results have demonstrated its improvement over any single model. Also, the results are promising when compared with state-of-the-arts methods.

ACKNOWLEDGMENTS

This research was supported by the Intramural Research Program of the National Institutes of Health (NIH), National Library of Medicine (NLM), and Lister Hill National Center for Biomedical Communications (LHNCBC).

REFERENCES

[1] Jaeger, S., Karargyris, A., Candemir, S., Folio, L., Siegelman, J., Callaghan, F., Xue, Z., Palaniappan, K., Singh, R. K., Antani, S., Thoma, G., Wang, Y. X., Lu, P. X., and McDonald, C. J., "Automatic tuberculosis screening using chest radiographs," *IEEE Transactions on Medical Imaging* **33**, 233–245 (Feb 2014).

[2] Chauhan, A., Chauhan, D., and Rout, C., "Role of gist and phog features in computer-aided diagnosis of tuberculosis without segmentation," *PLoS ONE* **9**, 1–12 (Nov 2014).

[3] Santosh, K. C., Vajda, S., Antani, S., and Thoma, G. R., "Edge map analysis in chest x-rays for automatic pulmonary abnormality screening," *International Journal of Computer Assisted Radiology and Surgery* , 1–10 (2016).

[4] Bar, Y., Diamant, I., Wolf, L., and Greenspan, H., "Deep learning with non-medical training used for chest pathology identification," in [*Proc. SPIE, Medical Imaging 2015: Computer-Aided Diagnosis*], Tourassi, L. M. H. G. D., ed., **9414** (2015).

[5] Hogeweg, L., Mol, C., de Jong, P. A., Dawson, R., Ayles, H., and van Ginneken, B., "Fusion of local and global detection systems to detect tuberculosis in chest radiographs," in [*Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010: 13th International Conference, Beijing, China*], (Sep 2010).

[6] Tan, K. and Chen, S., "Adaptively weighted sub-pattern PCA for face recognition," *Neurocomputing* **64**, 505 – 511 (2005). Trends in Neurocomputing: 12th European Symposium on Artificial Neural Networks 2004.

[7] Oliva, A. and Torralba, A., "Modeling the shape of the scene: A holistic representation of the spatial envelope," *International Journal of Computer Vision* **42**(3), 145–175 (2001).

[8] Candemir, S., Jaeger, S., Palaniappan, K., Musco, J. P., Singh, R. K., Xue, Z., Karargyris, A., Antani, S., Thoma, G., and McDonald, C. J., "Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration," *IEEE Transactions on Medical Imaging* **33**, 577–590 (Feb 2014).

[9] Frangi, A. F., Niessen, W. J., Vincken, K. L., and Viergever, M. A., "Multiscale vessel enhancement filtering," in [*Medical Image Computing and Computer-Assisted Interventation — MICCAI'98: First International Conference Cambridge, MA, USA, October 11–13, 1998 Proceedings*], Wells, W. M., Colchester, A., and Delp, S., eds., 130–137, Springer Berlin Heidelberg, Berlin, Heidelberg (1998).

[10] Dalal, N. and Triggs, B., "Histograms of oriented gradients for human detection," in [*2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*], **1**, 886–893 vol. 1 (June 2005).

[11] Ojala, T., Pietikainen, M., and Harwood, D., "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition* **29**(1), 51 – 59 (1996).

[12] Chang, C.-C. and Lin, C.-J., "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology* **2**, 27:1–27:27 (2011).