CrossMark

ORIGINAL ARTICLE

Combination of texture and shape features to detect pulmonary abnormalities in digital chest X-rays

**Alexandros Karargyris[1] · Jenifer Siegelman[2,3] · Dimitris Tzortzis[4] ·
Stefan Jaeger[1] · Sema Candemir[1] · Zhiyun Xue[1] · K. C. Santosh[1] ·
Szilárd Vajda[1] · Sameer Antani[1] · Les Folio[5] · George R. Thoma[1]**

Abstract

Purpose To improve detection of pulmonary and pleural abnormalities caused by pneumonia or tuberculosis (TB) in digital chest X-rays (CXRs).

Methods A method was developed and tested by combining shape and texture features to classify CXRs into two categories: TB and non-TB cases. Based on observation that radiologist interpretation is typically comparative: between left and right lung fields, the algorithm uses shape features to describe the overall geometrical characteristics of the lung fields and texture features to represent image characteristics inside them.

Results Our algorithm was evaluated on two different datasets containing tuberculosis and pneumonia cases.

Conclusions Using our proposed algorithm, we were able to increase the overall performance, measured as area under the (ROC) curve (AUC) by 2.4 % over our previous work.

Keywords Tuberculosis · Screen · Software · Remote · Telemedicine

✉ Alexandros Karargyris
akarargyris@gmail.com

1. Communications Engineering Branch, Lister Hill National Center for Biomedical Communications, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA

2. Division of Emergency Radiology, Department of Radiology, Brigham and Women's Hospital, Boston, MA, USA

3. Center for Evidence Based Imaging, Harvard Medical School, Boston, MA, USA

4. Ugeianet Diagnostic Center, General Hospital of Athens KAT, Athens, Greece

5. Radiology Department, Clinical Center, National Institutes of Health, Bethesda, MD, USA

Introduction

According to the World Health Organization (WHO), tuberculosis (TB) is one of the world's deadliest communicable diseases [1]. In 2013 alone, roughly 9 million people acquired TB and 1.5 million died from the disease, 360,000 of whom were HIV positive [1]. Despite considerable progress in diagnosis and treatment to eradicate the disease, TB remains a major health threat due to the following factors:

- Opportunistic coinfection with HIV-positive populations
- Emergence of multidrug resistance strains

Screening for TB is critical to fight disease spread and help with early treatment. There are several established guidelines for determining active TB infection (Centers for Disease Control [2]). These can be summarized as follows:

- The TB skin test (also called the Mantoux tuberculin skin test) is a common test to detect whether someone has been exposed to TB. It is a highly sensitive test and does not confirm active disease.
- A more reliable method is the sputum test where cultured sputum samples are microscopically analyzed. Though definitive in its determination, it is slow and depends on the ability of the patient to produce sputum—which may be a challenge when coinfected with other disease, or in certain age groups.
- Finally the most reliable method is blood testing (also called interferon-gamma release assays or IGRAs). However, due to high cost, this method is used as a last resort when all other methods fail. Further, high cost makes it difficult for population screening applications.
- Chest x-ray analysis has long been accepted as definitive indicator of disease. It is relatively inexpensive, fast, and also a good indicator severity [3].

🙋 Springer

Digital CXRs enable a radiologist interpretation, while the patient is still on-site. This offers a unique opportunity for the radiologist to immediately assess the situation and refer the patient for further consultation and limit disease exposure. However, in many under-resourced locations, having a radiologist on-site is not an option, and teleradiology is limited in remote areas. Commonly, an operator/technician manages care of all tasks, including acquisition of the image, and patient instructions. To address such limitation the National Library of Medicine (NLM) is developing an automatic CXR screening system. As an initial test, the system has been deployed in rural western Kenya due to high HIV and TB prevalence in the region. The deployment is in collaboration with the Indiana University School of Medicine, AMPATH (Academic Model Providing Access to Healthcare), the largest AIDS treatment program in the third world and Moi University in Eldoret, Kenya.

Building a robust automated screening software is a difficult task since TB detection is still an open research problem. The reason lies on the likely variable image quality of CXR and the diversity in size, shape, and texture of image patterns indicative of pulmonary abnormality. While texture information is one critical resource for image classification, it alone is not sufficient to screen for the presence of pulmonary anomalies, in particular those presented by TB infection, and overall geometrical shape characteristics of lung give context, normalizing to the patient anatomy significantly aiding in classification performance.

The paper is organized as follows: in "Materials and methods" section, we first introduce the reader to our proposed screening system, by briefly describing the research so far, along with the system components and underlying algorithms. We present and discuss our research motivation. Furthermore, we describe the main focus of this paper, which is the improvement of our system by incorporating shape features along with the texture descriptors. In "Experiments" section, we provide extensive results on system evaluation. Finally, in "Discussion" section, we discuss new ideas and challenges on our research. The reader is advised to review the following survey papers [4] and [5] on related research.

Materials and methods

Previous work

We have developed a screening software system based on a multistage framework (Fig. 1). The first stage is described in detail in [6–8] and [9], and it consists of methods that extract lung anatomy. Specifically, in [6] we discussed a method to automatically segment anatomic structures in CXRs, and Log-Gabor filter banks were used to segment lung fields and ribs by adjusting orientation and frequency.
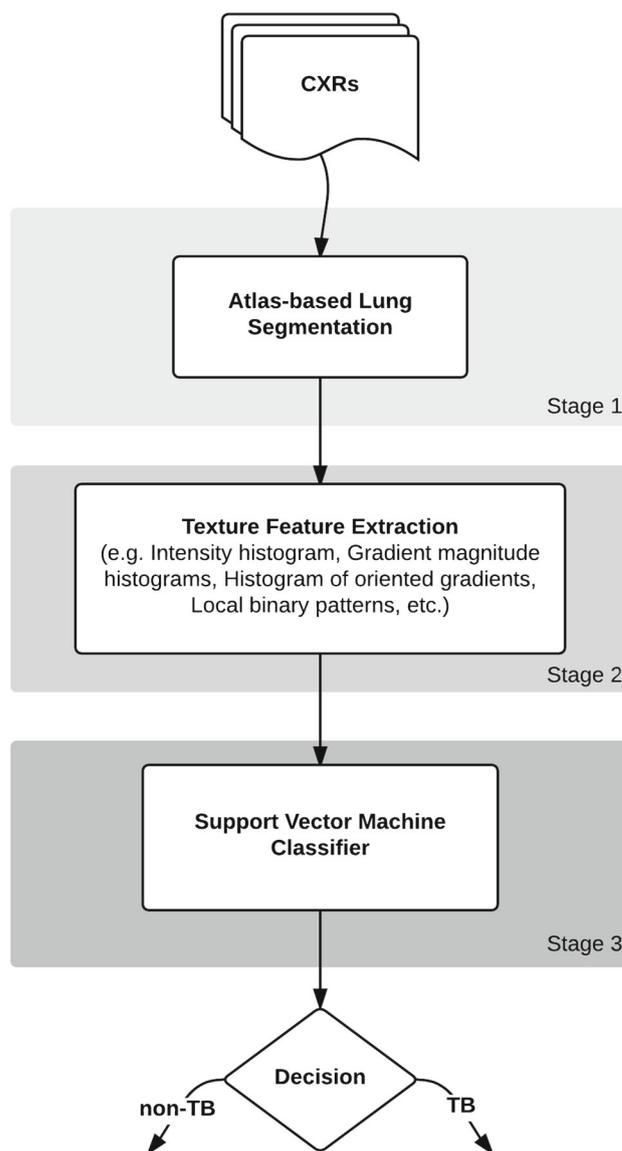

Fig. 1 Overview of our current screening system with its processing stages

In [7] we presented an improved method for segmenting lung fields. This method consists of three steps: (i) a content-based image retrieval approach for identifying training images (with binary masks) most similar to the patient CXR using a partial Radon transform [10] and Bhattacharyya shape similarity measure [11], (ii) creating the initial patient-specific anatomic model of lung shape using SIFT-flow for deformable registration of training masks to the patient CXR, and (iii) extracting refined lung boundaries using a graph cut optimization approach with a customized energy function. The next stages in the system are feature extraction and classification. In [8] we used texture features to describe the extracted lungs from the first stage. A histogram that shows the distribution of the different descriptor values is
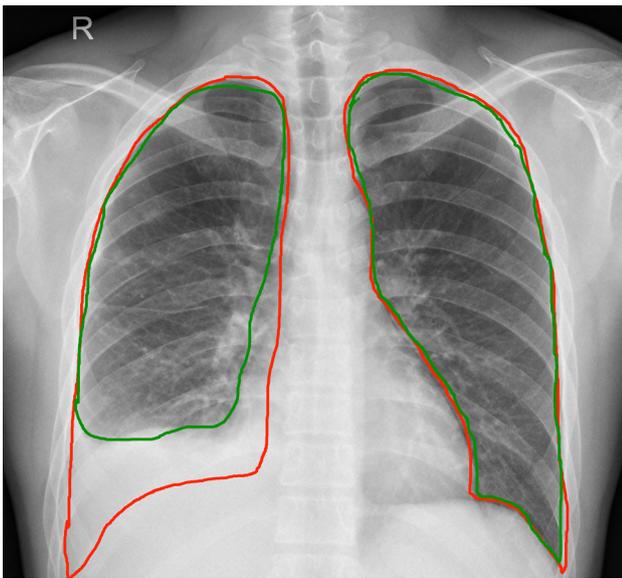
Fig. 2 Two different segmentation approaches: **a** air cavity segmentation shown in *green* and **b** expected lung anatomy segmentation shown in *red*

measured across the lung field. Each histogram bin is a feature, and features from all descriptors are concatenated to form a 192-dimensional feature vector that is the input to the classifier (i.e., support vector machine). We obtained an AUC of 88.5 % in identifying a CXR as TB or non-TB on the Shenzhen dataset (see "Appendix"). Figure 1 illustrates each stage of our system.

Motivation

Our goal is to build a system that can perform population screening in resource-constrained regions in which use of CXR for diagnosis purposes is prominent. This requires us to build a robust CXR screening system with high sensitivity (∼100 %) to avoid missing any TB cases out in the field as this can have devastating consequences to the local community.

The study proposed in this paper was conducted to determine whether accurate segmentation of anatomic lung boundaries is needed for disease detection, even when they are not clearly visible—possibly due to manifestation of disease or other radiological occlusions. Figure 2 presents this notion where a TB manifestation (pleural effusion) causes the CXR impression to not exhibit a clear anatomic lung boundary.

More specifically, in the previous paragraph we discussed that the current implementation of our segmentation module extracts only the air cavity within the lung in a CXR (green outline in Fig. 2). It does not consider the expected anatomy of lungs (i.e., surrounding lung area), which could be useful for disease detection. That is because our current image features are extracted within the air cavity segmented part and

not the whole anatomic lung boundary. Note, however, that for healthy lungs the two outlines, i.e., the air cavity and the anatomic lung boundary, can be considered as overlapping.

Assumption 1 We hypothesized that by adjusting our segmentation module to follow the anatomy of lungs (red outline in Fig. 2) instead of air cavities, overall classification performance would increase.

When collaborating with radiologists, we witnessed that during examination, they consider geometrical features (i.e., lung shape) and they perform a comparative reading between left/right lung fields.

Assumption 2 We hypothesized that geometric differences between lung fields (given normal anatomic asymmetry such as cardiac silhouette) as depicted in CXRs can be useful for TB identification by an algorithm as it is for the human reader.

In an effort to improve the overall classification accuracy of our system (Fig. 1), we wanted to expand the CXR atlas model set of the segmentation module [7] (stage 1 in Fig. 1) because until now we used only CXR models from non-TB cases.

Assumption 3 We hypothesized that adding CXR models of TB cases would improve the overall classification performance.

To address "Assumption 1," we invited human observers to provide their input and expertise into helping us identify an appropriate segmentation approach. That is whether to use segmentation based on air cavity or lung anatomy. Since we do not have a clear medical definition of lung segmentation, we sought medical expertise from the radiologists.

We selected 48 CXRs exhibiting TB manifestations from the Shenzhen dataset (see "Appendix"), and we asked two experts with long experience in radiology reading to mark their perceived boundaries of the lung fields manually. Their manual segmentation would be used as guidance for our segmentation module. In this way, we would also acquire new CXR models that would be used for "Assumption 3." For clarification reasons, the question toward the two radiologists was formulated as follows:

> For the provided CXR dataset, please mark the boundaries of left and right lung fields excluding the area of the diaphragm, heart, and aorta.

Our collaborating radiologists used Firefly software [12] to mark the lung field boundaries.[1]

We calculated the inter-observer agreement between the two radiologists by measuring the overlap of their markings.

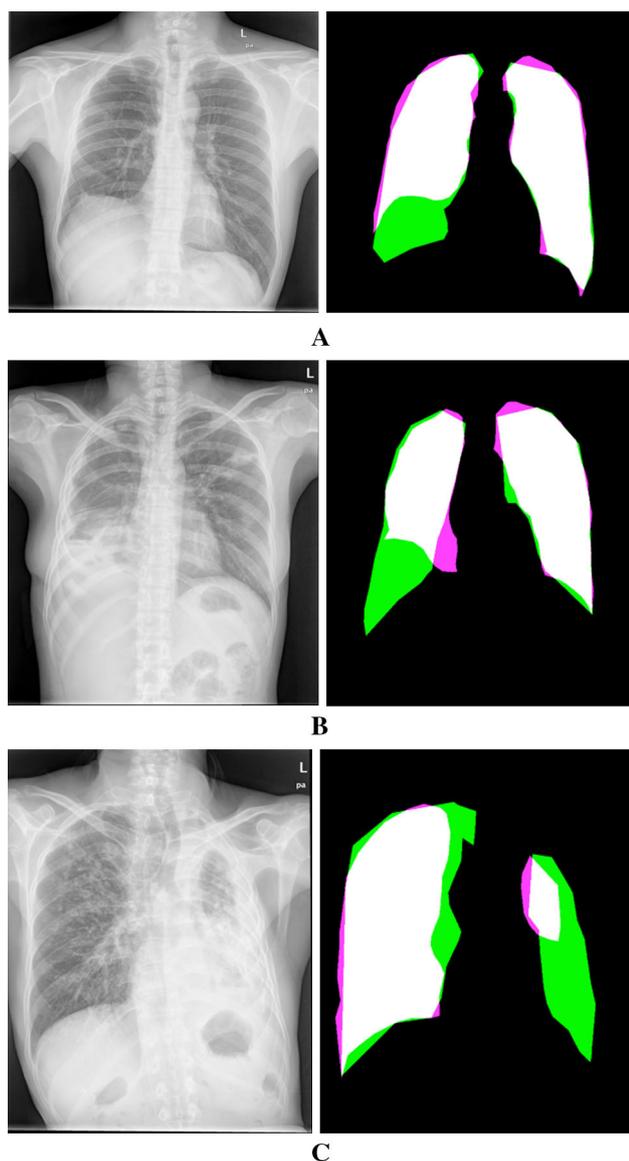[1] Firefly was developed by University of Missouri, and it is an online annotating toolbox.

Fig. 3 *Left*: abnormal cases, *right*: radiologists' markings (*purple*: reviewer A, *green*: reviewer B, *white*: overlapping area between reviewers). **a** Right lung field: 69.0 %, left lung field: 91.9 %. **b** Right lung field: 59.8 %, left lung field: 89.9 %. **c** Right lung field: 80.7 %, left lung field: 22.2 %

We noticed that although the radiologists had an excellent inter-observer agreement on normal lung fields ($(\mu, \sigma) = (86\,\%, 13.6)$), they scored significantly lower on abnormal lung fields ($(\mu, \sigma) = (73\,\%, 18.1)$). Figure 3 shows sample cases of abnormal lung fields along with radiologists' annotations.

This high disagreement in radiologists' markings in abnormal lung fields prompted us to investigate further.

Each radiologist was interviewed and asked to describe how they produced their markings. Their responses were clear and well defined: reviewer A simply mentioned that they only included air cavities in their markings because they wanted to have these markings as defined as possible. Reviewer B mentioned that they included air cavities as well as areas covered in fluid in their marking because: (a) the abnormal lung field is still present even with fluids covering its cavity (i.e., summation shadow); (b) they followed a comparative approach to "mirror" the normal lung field onto the abnormal one. Finally they both agreed that marking the boundaries of abnormal lung fields was a very tedious and subjective task because they were not visible.

We derived two interesting conclusions from these discussions. First, reviewer A represented one approach, and that was to segment only the air cavity part of the lung field. Reviewer B represented a different approach, and that was to segment the lung field following its expected anatomy. These two approaches counter each other and as such they inevitably do not provide a clear definition of lung segmentation, which in turn plays an important role in designing an automated segmentation module (i.e., "Assumption 1"). The second point came from reviewer B who examined the anatomy of both lung fields before marking the boundary. It shows that there is an underlying comparative reviewing when human experts examine a CXR. This approach is directly related to comparative diagnosis, which is a very common process during CXR reading [13]. Interestingly enough this point comes as a support to the original "Assumption 2" in our motivation.

Experiments

From the evaluation above, we were able to conclude that during reading process, radiologists do not focus on anatomic lung boundaries. It seems that they have accumulated an intuitive interpretation experience that disregards strict lung anatomies and focuses on identifying diseases. However, this ambiguity in a clear definition of lung segmentation is an obstacle in developing a final version of the lung segmentation method.

To address this, we performed a series of experiments. Specifically, we performed four experiments and we used the error of the classification module (i.e., Decision Stage in Fig. 1) as the overall evaluation to our system. In the first experiment, we tried to incorporate the segmentation behavior of the two experts. More specifically, we modified parameters in the segmentation module such that its output matched the annotated lung boundaries, and we evaluated the performance of the classification module. In the second experiment, we adapted the algorithm to include the notion of comparative diagnosis (i.e., "Assumption 2"). For this we extracted shape features to describe lung field differences in addition to our existing texture features. In the third experiment, we tested whether adding binary masks of abnormal CXRs would improve classification accuracy (i.e., "Assump-

tion 3"). Finally in the fourth experiment, we split the shape and texture features and we classified them separately using two classifiers because we wanted to identify whether a parallel versus a serial configuration of classifiers would improve the overall performance.

First experiment: effect of segmentation module on overall classification error

As described in [7], our segmentation methodology uses non-rigid registration. By adjusting the number of iterations, we can effectively modify the elasticity of the registration and match our segmentation results to the experts' annotations. By assigning a large number of iterations (>60), our segmentation (*Air Cavity Segmentation*) results matched reviewer A (see green outline in Fig. 2), while for a small number of iterations (>3) our segmentation (*Lung Anatomy Segmentation*) achieved similar results with reviewer B (see red outline in Fig. 2). We evaluated the two setups on Shenzhen dataset (see "Appendix") using our current texture feature descriptors and classification module. Table 1 shows results.

Examining Table 1 *Air Cavity Segmentation* seems to have an overall better performance in ROC area than *Lung Anatomy Segmentation* by a small margin. However, when reviewing recall and *F*-measure, *Lung Anatomy Segmentation* performed better in TB classification. That means that if we wanted to focus more on TB detection (i.e., increase sensitivity for TB cases), then *Lung Anatomy Segmentation* is a better candidate.

Second experiment: effect of combining shape features and texture features on overall classification error

In the second experiment, we aimed at incorporating comparative analysis process as reviewer B suggested. More specifically, the notion of anatomy to help mark lung boundaries brought up by review B was translated to geometry characteristics. So we extracted shape features from each segmented lung field and then we calculated their differences to mimic the expert's comparative analysis process. We used standard shape features [14]: size, orientation, eccentricity, extent, centroid coordinates, bounding box coordinates. While these geometry characteristics are basic, they are powerful enough to capture expert's intuition when comparing left and right lung fields. For this round of experiments, we used the previous segmentation setups (Round 1) and we extracted texture and shape features from the lung masks. Table 2 synopsizes this round experiments.

This time *Air Cavity Segmentation* performed significantly better than *Lung Anatomy Segmentation* in all metrics. What is more interesting is comparing Table 2 with Table 1: *Air Cavity Segmentation* of the second experiment was overall better for both setups of the first experiment. When it

Table 1 Evaluation of our system using the two segmentations

	Class	True positive rate	False positive rate	Precision	Recall	F-measure	ROC area
Setup A—Air cavity segmentation	Normal	0.896	0.234	0.791	0.896	0.84	0.91
	TB	0.766	0.104	0.882	0.766	0.82	0.91
Setup B—Lung anatomy segmentation	Normal	0.885	0.208	0.809	0.885	0.845	0.907
	TB	0.792	0.115	0.874	0.792	0.831	0.907

Table 2 Classification results using combination of shape and texture features

	Class	True positive rate	False positive rate	Precision	Recall	F-measure	ROC area
Setup C—Air cavity segmentation	Normal	0.921	0.199	0.822	0.921	0.868	0.934
	TB	0.801	0.079	0.91	0.801	0.852	0.934
Setup D—Lung anatomy segmentation	Normal	0.9	0.231	0.794	0.9	0.844	0.908
	TB	0.769	0.1	0.886	0.769	0.823	0.908

comes to comparing individual metrics between *Air Cavity Segmentation* of the second experiment and *Lung Anatomy Segmentation* of the first experiment *Air Cavity Segmentation* behaved almost the same in the case of TB recall (0.801 vs. 0.792, respectively), while it behaved much better in all other metrics (e.g., precision). This is an important finding because it shows that either the texture information from the whole lung anatomy (*Lung Anatomy Segmentation*) or the combined shape and texture information from the air cavities (*Air Cavity Segmentation*) leads to the same detection completeness of TB cases. However, combined shape and texture information offers a more accurate detection of TB cases.

Third experiment: evaluation of binary masks from abnormal CXRs

The previous rounds of experiments showed that *Air Cavity Segmentation* is a better segmentation approach than *Lung Anatomy Segmentation* when combined with shape and texture features. In this round of experiments, we tested whether adding binary masks of abnormal CXRs would improve the overall segmentation process. So we included the binary masks annotated by reviewer A into our atlas and we re-evaluated the accuracy of our segmentation on the JSRT dataset (see "Appendix"). The reason we used the JSRT dataset is because it contained manual annotations of lung fields. We did not measure any significant increase when adding these new masks.

Fourth experiment: performance comparison of two-stage classifier and one-stage classifier

At this point, we have to mention that our screening software will be configured to operate at 100 % sensitivity in detecting TB cases because it is going to be deployed in a real environment where no TB case should be missed. With the previous experiments, we confirmed that shape and texture features using *Air Cavity Segmentation* give the best results. In this last experiment, we evaluated two classifier configurations while maintaining based on their specificity scores while maintaining 100 % sensitivity. More specifically instead of having one classifier (Fig. 4 top) to process the whole input feature set, we used two classifiers in a sequential setup (Fig. 4 bottom). First, the shape features are fed into "Classifier A," and if the output is "Abnormal," then the process stops here: The input image is labeled as TB case. If the output is "Normal," then the process continues to "Classifier B" into which just the texture features are fed. For single setup specificity was at 65 %, while for cascade setup specificity was at 84 %. Therefore, with this two-stage classifier configuration, we achieved 19 % accuracy increase compared to one classifier while maintaining zero true negatives (i.e., TB cases classified as normal).
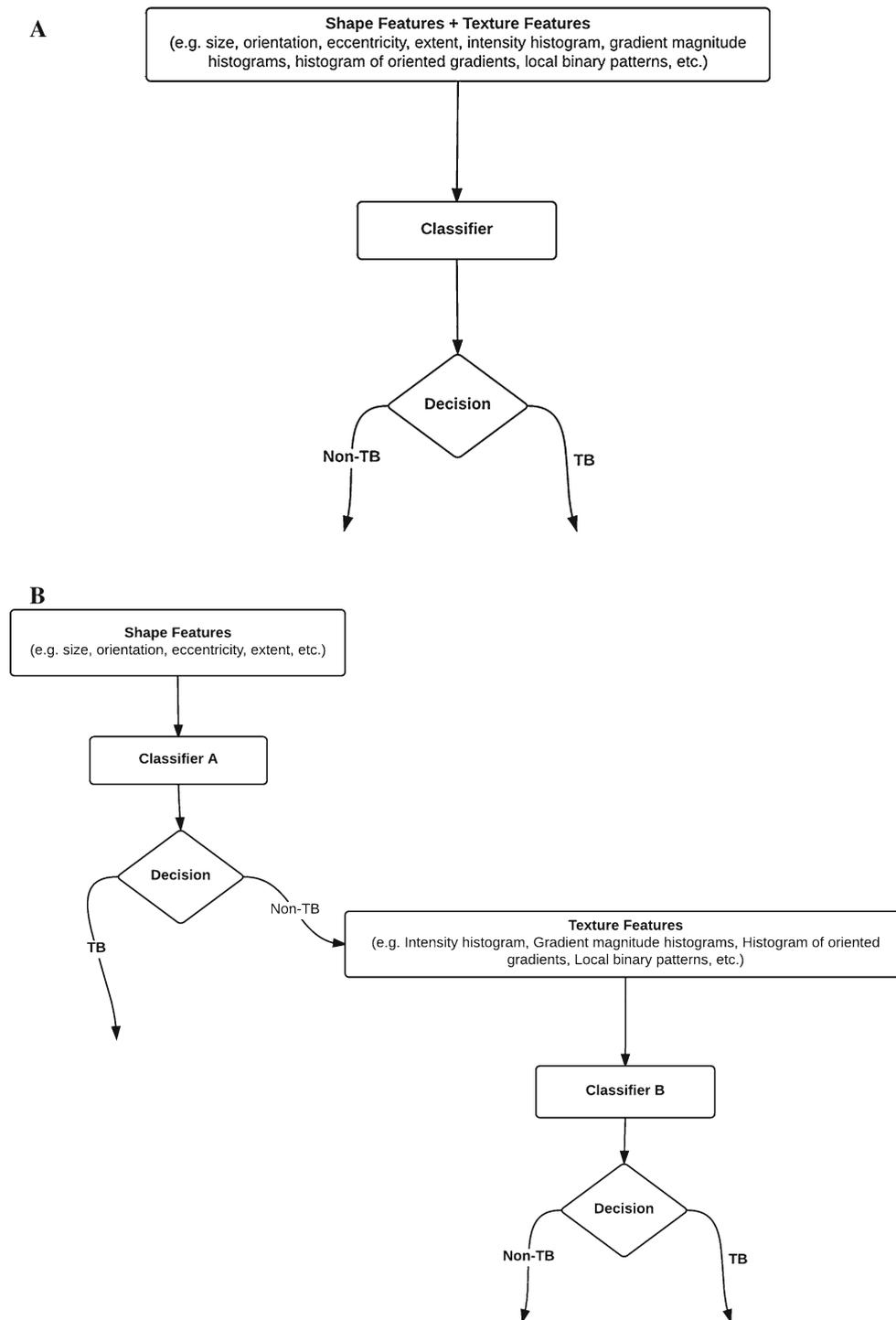
Fig. 4 *Top*: single classifier configuration. *Bottom*: two-stage classifier configuration

Discussion

This paper describes an elaborate evaluation of a current state-of-the-art segmentation algorithm applied to CXRs with TB findings [7,8]. To the best of our knowledge, this is the first attempt to assess various segmentation approaches using abnormal CXR cases. More specifically, through this study, we acquired experience and knowledge, which can act as useful lessons to the rest of the research community. The key points in this paper are synopsized in the following paragraphs.

First, preliminary study with two experts indicated that there is *no reliable agreement* in identifying lung boundaries in abnormal cases. In contrast, there is a stronger agreement for normal cases. This has to do with the fact that normal cases have clearly defined lung boundaries, while abnormal manifestations like pleural effusion, silhouette and miliary pattern affect how these boundaries are depicted in CXRs making it hard to detect them. Without a clear definition of lung segmentation, we had to identify the best automated segmentation approach to increase TB detection.

Second, there is an *underlying comparative process* when a human expert performs manual lung segmentation. It is derived from the fact that left and right lung fields are symmetric around the vertical-axis. This comparative notion was utilized in our segmentation module by changing it to perform a more rigid registration and therefore has the output masks be symmetric. We tested this modified segmentation module on the TB classification and recall *increased by 2.6 %*, proving that a comparative process improves texture classification (see Table 1, row 4).

Third, with the help of the aforementioned comparative process, we incorporated *shape features to describe differences in the geometry of lung fields* in addition to their texture information. When combining shape and texture features, the area under the curve (ROC) in detecting TB cases *increased by 2.4 %* (see Table 2, row 2) compared to using only texture features.

Fourth, *splitting shape and texture features significantly improves accuracy*. Also, because shape classifier (Classifier A, bottom, Fig. 4) acts as a pre-filtering module, it leads to *faster CXR processing* which is a major priority in low-resourced settings like in Kenya.

Finally, adding abnormal masks to our atlas does not help with segmentation performance regardless of our motivation's assumption.

Compliance with ethical standards

Conflict of interest   Alexandros Karargyris, Jenifer Siegelman, Dimitris Tzortzis, Stefan Jaeger, Sema Candemir, Zhiyun Xue, KC Santosh, Szilárd Vajda, Sameer Antani, Les Folio and George R. Thoma declare that they have no conflict of interest.

Appendix

There are two datasets that are used in this paper:

1. Shenzhen dataset [11]: It was acquired from Shenzhen Hospital in China. It has a good variety of TB cases. They were captured over a month period as part of the daily routine at Shenzhen Hospital, using a Philips DR Digital Diagnost system. The set contains 340 normal CXRs and 275 abnormal CXRs with TB along with radiologist readings. The dataset is publicly available here: http://archive.nlm.nih.gov/repos/chestImages.php.

2. JSRT dataset [15]: This is a popular publicly available dataset from the Japanese Society of Radiological Technology (JSRT). This dataset contains 154 nodule and 93 non-nodule CXRs along with manual annotations of the lung fields.

References

1. http://www.who.int/tb/publications/global_report/gtbr14_executive_summary.pdf?ua=1. Viewed in Dec 2014
2. Centers for Disease Control and Prevention—Testing for TB Infection. http://www.cdc.gov/tb/topic/testing/default.htm. Viewed in Dec 2014
3. Van Cleeff M, Kivihya-Ndugga L, Meme H, Odhiambo J, Klatser P (2005) The role and performance of chest X-ray for the diagnosis of tuberculosis: a cost-effectiveness analysis in Nairobi, Kenya. BMC Infect Dis 5:111. doi:10.1186/1471-2334-5-111
4. Jaeger S, Karargyris A, Candemir S, Siegelman J, Folio L, Antani S, Thoma G (2013) Automatic screening for tuberculosis in chest radio-graphs: a survey. Quant Imaging Med Surg 3(2):89–99
5. Ginneken B, Hogeweg L, Prokop M (2009) Computer-aided diagnosis in chest radiography: beyond nodules. Eur J Radiol 72(2):226–230
6. Karargyris A, Antani S, Thoma G (2011) Segmenting anatomy in chest X-rays for tuberculosis screening. In: Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE, pp 7779–7782, Aug 30, 2011–Sept 3, 2011. doi:10.1109/IEMBS.2011.6091917
7. Candemir S, Jaeger S, Palaniappan K, Musco JP, Singh RK, Xue Z, Karargyris A, Antani S, Thoma G, McDonald CJ (2014) Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration. IEEE Trans Med Imaging 33(2):577–590
8. Jaeger S, Karargyris A, Candemir S, Folio L, Siegelman J, Callaghan F, Xue Z, Palaniappan K, Singh RK, Antani S, Thoma G, Wang Y-X, Pu-Xuan L, McDonald CJ (2014) Automatic tuberculosis screening using chest radiographs. IEEE Trans Med Imaging 33(2):233–245
9. Santosh KC, Candemir S, Jaeger S, Karargyris A, Antani S, Folio L, Thoma G (2015) Automatically detecting rotation in chest radiographs using principal rib-orientation measure for quality control. Int J Pattern Recognit Artif Intell. doi:10.1142/S0218001415570013
10. Radon Transform. http://mathworld.wolfram.com/RadonTransform.html. Viewed in Jan 2015
11. Bhattacharyya distance. http://en.wikipedia.org/wiki/Bhattacharyya_distance. Viewed in Jan 2015
12. Beard D (2009) Firefly—web-based interactive tool for the visualization and validation of image processing algorithms. M.S. thesis, Univ. Missouri, Columbia
13. Folio LR (2012) Chest imaging: an algorithmic approach to learning. Springer, Berlin
14. The MathWorks, Inc., Matlab—measure properties of image regions. http://www.mathworks.com/help/images/ref/regionprops.html
15. http://www.jsrt.or.jp/jsrt-db/eng.php. Viewed in Dec 2014