INT J TUBERC LUNG DIS 17(12):1613–1620

http://dx.doi.org/10.5588/ijtld.13.0325

# Detection of tuberculosis using digital chest radiography: automated reading vs. interpretation by clinical officers

**P. Maduskar,* M. Muyoyeta,† H. Ayles,†‡ L. Hogeweg,* L. Peters-Bax,§ B. van Ginneken***

***Diagnostic Image Analysis Group, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands; †Zambia AIDS-Related Tuberculosis Project (ZAMBART), University of Zambia School of Medicine, Lusaka, Zambia; ‡Department of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London, UK; §Department of Radiology, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands**

**SUMMARY**

**SETTING: A busy urban health centre in Lusaka, Zambia.**

**OBJECTIVE: To compare the accuracy of automated reading (CAD4TB) with the interpretation of digital chest radiograph (CXR) by clinical officers for the detection of tuberculosis (TB).**

**DESIGN: A retrospective analysis was performed on 161 subjects enrolled in a TB specimen bank study. CXRs were analysed using CAD4TB, which computed an image abnormality score (0–100). Four clinical officers scored the CXRs for abnormalities consistent with TB. We compared the automated readings and the readings by clinical officers against the bacteriological and radiological results used as reference. We report here the area under the receiver operating characteristic curve (AUC) and kappa (κ) statistics.**

**RESULTS: Of 161 enrolled subjects, 97 had bacteriologically confirmed TB and 120 had abnormal CXR. The AUCs for CAD4TB and the clinical officers were respectively 0.73 and 0.65–0.75 in comparison with the bacteriological reference, and 0.91 and 0.89–0.94 in comparison with the radiological reference. *P* values indicated no significant differences, except for one clinical officer who performed significantly worse than CAD4TB ($P < 0.05$) using the bacteriological reference. κ values for CAD4TB and clinical officers with radiological reference were respectively 0.61 and 0.49–0.67.**

**CONCLUSION: CXR assessment using CAD4TB and by clinical officers is comparable. CAD4TB has potential as a point-of-care test and for the automated identification of subjects who require further examinations.**

**KEY WORDS:** digital chest radiograph; computer-aided diagnosis; tuberculosis diagnostics; medical image analysis; automated screening

TUBERCULOSIS (TB) remains a global health problem, with almost 9 million new cases and 1.4 million deaths reported in 2011.[1] As TB is infectious but curable, its early detection is crucial in reducing the disease burden. Various strategies have been recommended by the World Health Organization (WHO) and other organisations for active case finding and prevalence surveys, whereby chest radiography is used as a screening tool and those subjects with an abnormal chest radiograph (CXR) undergo further, more expensive and time-consuming examinations.[2–8] One of the reasons for the increased interest in CXR is the improvement in image quality obtained with digital radiography. In populations with a high prevalence of HIV/AIDs (human immunodeficiency virus/acquired immune-deficiency syndrome), existing TB diagnostics such as sputum smear microscopy are less reliable,[9] and using CXR in the diagnostic algorithm can improve TB detection.[10–13] In various studies, CXR is used as one of the diagnostic tools for TB detection, primarily in resource-limited settings, where confirmatory diagnostic tests such as sputum culture are unavailable.[14,15] In resource-constrained countries with a high TB incidence, the limited availability of skilled clinical officers to read CXRs is a major obstacle to their use as an adjunct tool for TB diagnosis. Computer-aided detection (CAD) systems may be exploited to accelerate active case finding and antituberculosis treatment in such settings.[16] CAD can also be a valuable tool for clinics where bacteriological tests are recommended based on CXR findings.

In the present study, we evaluated and compared the performance of a CAD system with clinical officers who read CXRs on a daily basis in their regular practice in a primary health care centre in Lusaka, Zambia. This is the first study to evaluate the performance of a CAD system for TB detection in comparison with the performance of clinical officers.

## MATERIAL AND METHODS

### *Study setting*

The study was conducted at a busy urban health centre in Lusaka, Zambia, among TB suspects presenting

Correspondence to: Pragnya Maduskar, Diagnostic Image Analysis Group, Department of Radiology, Radboud University Nijmegen Medical Center 766, Postbus 91016500 HB Nijmegen, The Netherlands. Tel: (+31) 24 366 8112. e-mail: pragnya.maduskar@radboudumc.nl

*Article submitted 6 May 2013. Final version accepted 31 July 2013.*

with symptoms. The health centre serves a population of 146 000 and notifies 2000 TB cases per year; the HIV coinfection rate is approximately 70%.[17]

### *Study procedures*

Data used in this analysis were collected as part of a WHO TB Specimen Bank study in 2010–2011, where sputum, serum and urine samples were preserved to devise and test new diagnostic tools for the detection of TB.[18] Inclusion criteria were patients aged >18 years with a persistent cough of at least 3 weeks, who were willing to provide clinical specimens for storage and to undergo HIV testing. Patients who had received anti-tuberculosis treatment in the last 2 months were excluded.

Patients were asked to submit two sputum samples before the start of any treatment; the samples were examined at the Chest Diseases Laboratory, Lusaka (also known as the Zambian National TB Reference Laboratory). Sputum smears were examined and graded for acid-fast bacilli (AFB) using fluorescence microscopy. Sputum specimens were cultured in liquid media using BACTEC™ MGIT™ (Mycobacteria Growth Indicator Tube; BD Sparks, MD, USA) and in solid Löwenstein-Jensen media. Clinical information was collected using standardised case report forms, as described elsewhere.[18] Posterior-anterior digital CXRs (1600–1800 pixels image width, isotropic pixel size 0.25 mm, Odelca-DR; Delft Imaging Systems, Veenendaal, The Netherlands) were acquired at the site and read by a clinical officer (the field officer) who classified the CXR as 'normal' or 'abnormal' based on absence or presence of abnormalities consistent with TB. The CXR reading by the field officer along with the above bacteriological tests were used to decide the treatment regimen.

### *Ethics statement*

Ethics approval for the study was obtained from the University of Zambia Ethics Committee as part of the WHO Specimen Bank study. Formal written consent to use the data for research purposes was obtained from all study participants.

### *Study design*

In this retrospective study, 161 subjects were enrolled from the WHO TB Specimen Bank study, for whom we were able to link the CXRs with clinical information on symptoms, sputum smear microscopy, HIV status, CD4 count, solid and liquid AFB culture, and CXR reading by the field officer. The study population demographics are shown in Table 1. Smear microscopy and culture results were used as the bacteriological reference. Active TB was defined as ⩾1 sputum smear-positive and/or ⩾1 culture-positive results having ⩾1+ growth of *Mycobacterium tuberculosis* on solid media or growth on liquid media identified as *M. tuberculosis*. All mycobacterial isolates defined as growth containing AFB on Ziehl-Neelsen staining were further identified using the GenoType® Mycobacterium CM assay (Hain Lifescience, Nehren, Germany). The radiological reference was retrospectively determined by an experienced chest radiologist at Radboud University Nijmegen Medical Centre, who labelled the CXRs as normal or abnormal for abnormalities indicative of active TB.

**Table 1** Demographics, symptoms and clinical findings of patients who participated in the WHO TB Specimen Bank study in Lusaka, Zambia[18]

| | *n* (%) |
|---|---|
| Total | 161 |
| Sex | |
| Female | 42 (22.1) |
| Male | 119 (73.9) |
| Age, years, mean ± SD | 35.8 ± 9.6 |
| Persistent cough | 161 (100) |
| HIV-positive | 110 (68.3) |
| CD4 count (HIV-positive only) | |
| <200 | 63 |
| 200–500 | 30 |
| >500 | 14 |
| Unknown | 3 |
| Weight loss | 105 (65.2) |
| Fever | 137 (85.0) |
| Chest pain | 129 (80.1) |
| Night sweats | 121 (75.2) |
| Haemoptysis | 26 (16.2) |
| Contact with active case | 67 (41.6) |
| Bacteriological findings | |
| Smear-positive, culture-positive | 69 (42.8) |
| Smear-negative, culture-positive | 28 (17.4) |
| Culture-negative | 64 (39.8) |

WHO = World Health Organization; TB = tuberculosis; SD = standard deviation; HIV = human immunodeficiency virus.

### *Radiological assessment by clinical officers*

The CXRs were independently assessed by four clinical officers working at the same urban health centre in Lusaka. All clinical officers had received a 3-year diploma in medicine in Zambia and read CXRs regularly at the clinic. The medical programme included brief training on reading CXRs and 2 months' practicum where ward duties included CXR reading. For this study, the clinical officers were given 30-min training by a researcher in reading protocol and the use of a DICOM (digital imaging and communications in medicine) image viewing software tool for scoring CXRs. The clinical officers were blinded to the CAD scores and all clinical information. They were instructed to assess each CXR on a continuous scale of 0 to 100, where abnormality scores of >50 indicated an abnormal image. The scores given by the clinical officers reflect their confidence in interpreting abnormalities consistent with active TB.

### *Automated reading*

A software system developed for automated detection of pulmonary TB (CAD4TB, version 1.08, Diagnostic

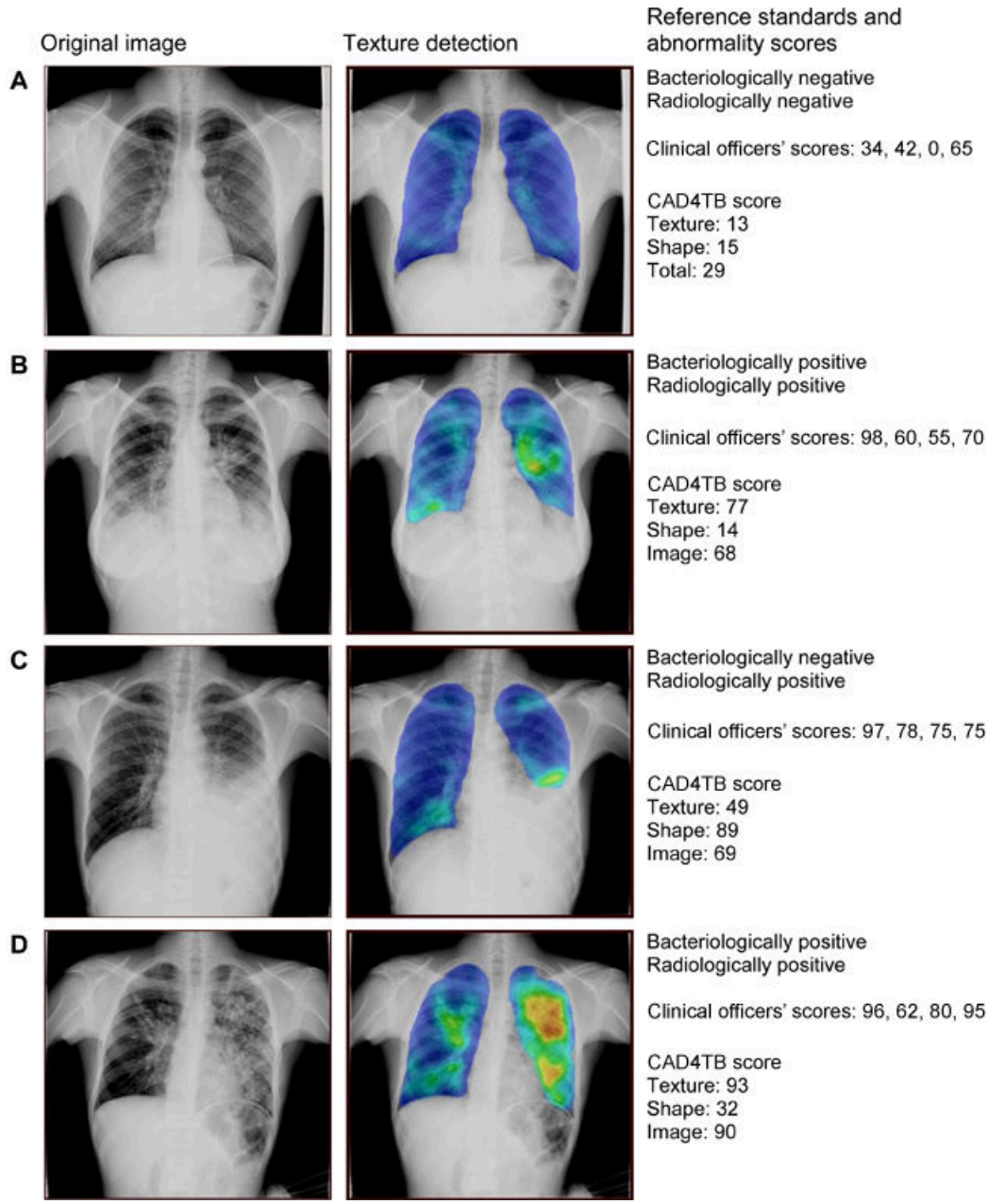


**Figure 1** CAD4TB output. **A.** Normal CXR with low image abnormality score. **B.** Abnormal CXR with high texture and image abnormality score. **C.** Abnormal CXR with low texture abnormality score; abnormalities detected by the shape detection system resulting in a high image abnormality score. **D.** Highly abnormal CXR. This image can be viewed online in colour at http://www.ingentaconnect.com/content/iuatld/ijtld/2013/00000017/00000012/art00020

Image Analysis Group, Nijmegen, The Netherlands), as described by Hogeweg et al.,[19] was also used to analyse the CXRs. Examples of output images with CAD abnormality scores are shown in Figure 1. The software is based on supervised machine learning methodology, whereby the software system is trained with labelled samples (examples) of various classes to produce an inference function that is used to label an unknown sample.[20] The software was trained with labelled normal and abnormal CXRs to predict the probability of an unseen CXR being abnormal. The training set consisted of 945 consecutive digital CXRs (514 abnormal, 431 normal) acquired from two high TB prevalence sites in sub-Saharan Africa, Lusaka, Zambia, and Cape Town, South Africa. The software combines the output of two detection systems, namely textural abnormality detection and shape abnormality detection.

Both detection systems require the automated segmentation of un-obscured lung fields as an initial step.[21] The training procedure for textural abnormality detection included extracting descriptive features from normal and abnormal circular patches in lung fields to train a *k*-nearest neighbour (*k*-NN) classifier[20] to differentiate between a normal and an abnormal location in the image. The descriptive features were based on moments of intensity distribution of

Gaussian derivative filtered images at each patch location and its relative position in the lung fields.[22] Patches in the lung fields of the new CXR were classified and assigned a probability score of being abnormal. These patch probabilistic labels were then aggregated into a textural abnormality score for the CXR.

Automated lung field segmentation may be inaccurate if the CXR has large abnormalities, particularly when these are in the pleural space (Figure 1C).[23] As abnormalities outside the segmented lung fields would go undetected by the texture detection system, the second system analyses the shape of the extracted lung fields, which could indicate the presence of abnormalities. Hence, a shape model as proposed in Hogeweg et al. was built using lung shapes of normal CXRs in the training data set.[19] Using this shape model, a shape abnormality score of between 0 and 100 was computed. A high shape abnormality score reflected an abnormal image. Texture and shape abnormality scores were used as image descriptive features to train a *k*-NN classifier, which was then used to estimate the combined abnormality score of a new CXR image.

### *Statistical data analysis*

The results were evaluated against the bacteriological and radiological reference results. The area under the receiver operating characteristic (ROC) curve (AUC)[24] was constructed using both references for an average clinical officer, for the four clinical officers individually and for CAD4TB. A previously published study of breast cancer detection in mammography reported higher performance when the score was averaged among several readers than in an individual reader.[25] We compared the performance of CAD4TB with a simulated average clinical officer. To construct an ROC curve for the average clinical officer, the abnormality scores of the clinical officers were averaged to obtain a mean abnormality score per image. The significance of differences in performance was measured using bootstrapping.[26,27] Using this procedure, a new data set is constructed by sampling subjects with replacement from the test data set 5000 times. For each resampling, ROC curves were constructed for all the six reader categories (i.e., CAD4TB, the four clinical officers and the average clinical officer readings). The difference in mean sensitivity ($\Delta S$) was calculated pair-wise between CAD4TB and the five other reader categories. *P* values were defined as the fraction of $\Delta S$ values that were negative or zero.[26] Differences in performance were considered significant if $P < 0.05$. A statistical analysis tool developed in-house based on the statistical software package R, v2.15.0 (R Computing, Vienna, Austria) was used to calculate *P* values and construct ROC curves.

We compared the sensitivity and specificity of the clinical officers with that of the CAD4TB by labelling the CXRs as normal or abnormal using the threshold clinical officer's abnormality score (>50 was abnormal). To calculate these values for CAD4TB, we thresholded the CAD4TB abnormality score to obtain the same specificity as the field officer, using the bacteriological reference. This threshold allowed us to analyse the performance of CAD4TB if it were used as a reader in the field in place of the field officer.

Intra-observer agreement kappa (κ) and 95% confidence intervals (CIs) were estimated using Cohen's κ statistics for agreement between the radiological reference and all reader categories. κ values were measured using the R statistical software package (v 2.15.0). We assessed the sensitivity of the four clinical officers and CAD4TB at the specificity of the field officer.

**Table 2** Area under the ROC curve, *P* values and other performance measures for all the reader categories compared to the bacteriological and radiological reference

| Test | AUC (95%CI) | Difference in AUC (95%CI) | *P* value (CAD4TB reader) | Sensitivity | Specificity | Sensitivity (95%CI) |
|---|---|---|---|---|---|---|
| Bacteriological reference | | | | | | At 0.41 specificity |
| CAD4TB | 0.73 (0.64–0.80) | NA | NA | 0.88 | 0.41 | 0.86 (0.75–0.94) |
| Clinical officer 1 | 0.69 (0.61–0.77) | 0.034 (−0.043–0.115) | 0.20 | 0.80 | 0.53 | 0.83 (0.75–0.90) |
| Clinical officer 2 | 0.65 (0.56–0.73) | 0.077 (0.003–0.152) | 0.02* | 0.76 | 0.41 | 0.77 (0.67–0.87) |
| Clinical officer 3 | 0.75 (0.68–0.82) | −0.030 (−0.095–0.034) | 0.82 | 0.79 | 0.55 | 0.85 (0.75–0.93) |
| Clinical officer 4 | 0.69 (0.60–0.77) | 0.038 (−0.024–0.103) | 0.12 | 0.90 | 0.22 | 0.82 (0.73–0.90) |
| Average clinical officer | 0.73 (0.65–0.81) | −0.008 (−0.069–0.052) | 0.69 | 0.80 | 0.48 | 0.83 (0.75–0.91) |
| Field officer | NA | NA | NA | 0.81 | 0.41 | 0.79 (0.66–0.89) |
| Radiological reference | | | | | | At 0.66 specificity |
| CAD4TB | 0.91 (0.86–0.95) | NA | NA | 0.92 | 0.68 | 0.91 (0.81–0.98) |
| Clinical officer 1 | 0.90 (0.84–0.96) | 0.007 (−0.057–0.075) | 0.42 | 0.86 | 0.88 | 0.94 (0.87–0.98) |
| Clinical officer 2 | 0.89 (0.83–0.93) | 0.025 (−0.033–0.083) | 0.20 | 0.85 | 0.76 | 0.87 (0.80–0.94) |
| Clinical officer 3 | 0.92 (0.87–0.96) | −0.008 (−0.061–0.042) | 0.62 | 0.83 | 0.85 | 0.94 (0.87–0.98) |
| Clinical officer 4 | 0.91 (0.86–0.96) | −0.001 (−0.060–0.056) | 0.53 | 0.96 | 0.46 | 0.93 (0.88–0.97) |
| Average clinical officer | 0.94 (0.89–0.97) | −0.025 (−0.073–0.020) | 0.86 | 0.87 | 0.83 | 0.95 (0.86–0.99) |
| Field officer | NA | NA | NA | 0.86 | 0.66 | 0.80 (0.59–0.92) |

*Significant ($P < 0.05$).
ROC = receiver operating characteristic; AUC = area under the ROC curve; CI = confidence interval; NA = not applicable.

**Table 3** Agreement between all reader categories and the radiological reference*

| Reader category | Reference: abnormal; reader: abnormal | Reference: normal; reader: abnormal | Reference: abnormal; reader: normal | Reference: normal; reader: normal | κ (95%CI) |
|---|---|---|---|---|---|
| CAD4TB | 110 | 10 | 13 | 28 | 0.61 (0.44–0.74) |
| Clinical officer 1 | 103 | 17 | 5 | 36 | 0.67 (0.54–0.78) |
| Clinical officer 2 | 102 | 18 | 10 | 31 | 0.57 (0.41–0.70) |
| Clinical officer 3 | 100 | 20 | 6 | 35 | 0.62 (0.47–0.73) |
| Clinical officer 4 | 115 | 5 | 22 | 19 | 0.49 (0.32–0.64) |
| Average clinical officer | 104 | 16 | 7 | 34 | 0.65 (0.51–0.79) |
| Field officer | 103 | 17 | 14 | 27 | 0.50 (0.36–0.66) |

*'Reference' = radiological reference; 'reader' = each reader category specified in the first column.
CI = confidence interval.

## RESULTS

According to the bacteriological reference, the data set contained 97 active TB cases and 64 subjects without TB. The average reading time per clinical officer per case was 37 ± 27 s. CAD4TB computed an abnormality score per case in ~100 s using standard PC hardware.

If we look at the sensitivity and specificity of the human readers, the performance of CAD4TB was comparable using both references, as indicated in Table 2 (Columns 5 and 6). The clinical officers and CAD4TB had moderate (κ = 0.49–0.67) and substantial (κ = 0.61) inter-reader agreement,[28] respectively, with the radiological reference (Table 3, column 6).

ROC curves using the bacteriological reference (Figure 2) for all the readers were comparable, with similar AUC values (0.65–0.75, Table 2), and the *P* values indicated no significant differences between any clinical officer and CAD4TB, except for one clinical officer (AUC = 0.65), who performed significantly worse than CAD4TB (AUC = 0.73, $P = 0.02$). The field officer achieved a sensitivity of 0.81 at a specificity of 0.41. At this specificity, the CAD4TB had higher sensitivity than the field officer and all the clinical officers (Table 2, column 7).

We also analysed the performance of all of the readers against the radiological reference results (Figure 3). The data set contained 120 abnormal CXRs as read by the experienced chest radiologist in The Netherlands. All readers obtained higher AUC values (0.89–0.94) than those calculated using the bacteriological reference, and there were no significant differences in performance between CAD4TB and the clinical officers (Table 2).

## DISCUSSION

Previous studies have shown substantial variations in the performance of human readers in interpreting

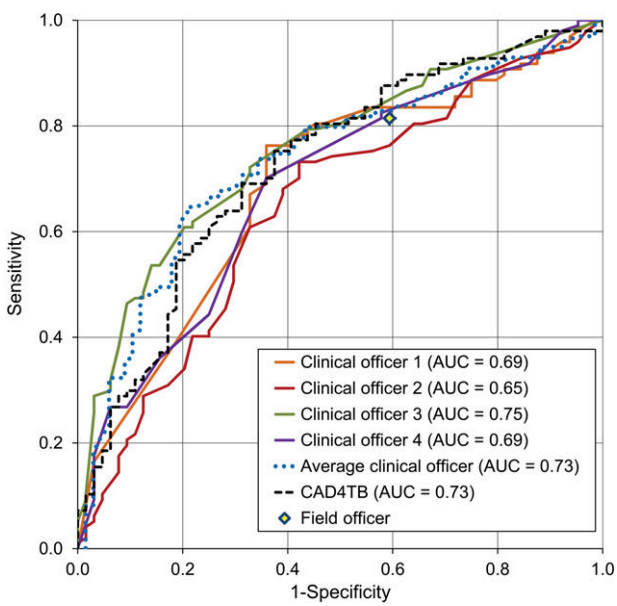


**Figure 2** ROC curves using bacteriological reference. AUC = area under the ROC curve; ROC = receiver operating characteristic. This image can be viewed online in colour at http://www.ingentaconnect.com/content/iuatld/ijtld/2013/00000017/00000012/art00020

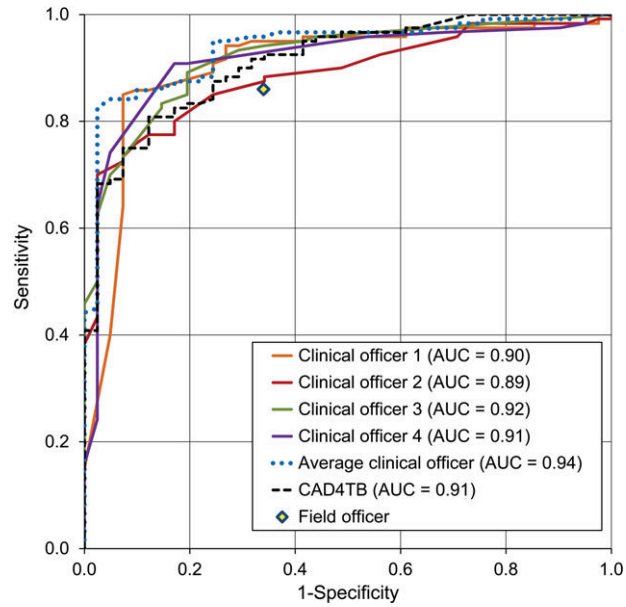


**Figure 3** ROC curves using radiological reference. AUC = area under the ROC curve; ROC = receiver operating characteristic. This image can be viewed online in colour at http://www.ingentaconnect.com/content/iuatld/ijtld/2013/00000017/00000012/art00020

**Table 4** CXR performance for detection of TB

| Author, reference, year | Setting | Sample size *n* | Reference standard | κ | Sensitivity % | Specificity % | AUC | HIV cases only | Type of readers | CXR abnormality type | Analogue/ digital |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Story, 2012[5] | High-risk group screening | 47 510 | Bacteriological | NA | 82 | 99 | NA | No | Radiographers | TB | Digital |
| van't Hoog, 2012[7] | Prevalence survey | 20 566 | Bacteriological | NA | 94 | 73 | 0.83 | No | Clinical officers | Any | Analogue |
| van't Hoog, 2011[29] | Prevalence survey | 1 143 | Bacteriological | 0.78*<br>0.50–0.62† | 82*<br>95† | 76*<br>73† | NA | No | Experts, clinical officers | Any | Analogue |
| Dawson, 2010[30] | ART service | 203 | Bacteriological | 0.61 | 68 | 53 | NA | Yes | Physicians | TB | Analogue |
| Lewis, 2009[31] | Miners screening | 1 955 | Bacteriological | NA | 26 | 99 | NA | No | Physicians | TB | NA |
| Shah, 2009[14] | Urban voluntary counselling and testing clinic | 438 | Bacteriological | 0.53 | 59 | 83 | NA | Yes | Radiologists | TB | Analogue |
| Day, 2006[11] | Miners | 899 | Bacteriological | NA | 66<br>73 | 86<br>79 | NA | Yes | Doctor | TB<br>Any | Analogue |
| den Boon, 2006[6] | Prevalence survey | 1 170 | Bacteriological | NA | 90<br>97 | 83<br>67 | NA | No | Pulmonologist | TB<br>Any | Analogue |
| Zellweger, 2006[32] | Immigrant screening | 377 | Radiological | 0.64<br>0.55 | NA | NA | NA | No | Physicians | TB<br>Any | Analogue |
| Balabanova, 2005[33] | General clinic | 50 | Radiological | 0.39<br>0.38‡<br>0.45§<br>0.39¶ | NA | NA | 0.88‡<br>0.81§<br>0.81¶ | No | TB specialists, radiologists, respiratory specialists | TB | Digital |
| Van Cleef, 2005[34] | Chest clinic | 998 | Bacteriological | 0.75<br>NA | 91<br>92 | 67<br>63 | NA | No | Radiologists | TB<br>Any | Analogue |

* Experts.
† Clinical officers.
‡ TB specialists.
§ Radiologists.
¶ Respiratory specialists.
CXR = chest X-ray; TB = tuberculosis; AUC = area under the ROC curve; ROC = receiver operating characteristic; HIV = human immunodeficiency virus; NA = not available; ART = antiretroviral therapy.

abnormalities consistent with TB on CXRs. Sensitivity and specificity vary respectively from 0.26 to 0.95 and from 0.56 to 0.99 (Table 4, Figure 4). Most of the studies report only sensitivity, specificity and κ values, because in clinical practice human readers interpret an image in a binary manner as either 'normal' or 'abnormal'. Only a few studies have used the ROC paradigm to determine the accuracy of a diagnostic TB algorithm.[7,33,35,36] Binary image classification leads to differences in sensitivity and specificity between readers, which was also evident in our study (Table 2). This source of variation could potentially be eliminated with automated reading, as CAD4TB provides a continuous abnormality score on a scale of 0 to 100. In practical applications of CAD4TB, an optimal threshold could be determined based on its role in the TB diagnostic algorithm. In a screening setting, a threshold with very high sensitivity should be chosen to select subjects who require definitive examinations, i.e., sputum culture or Xpert® MTB/RIF testing (Cepheid, Sunnyvale, CA, USA). This threshold may also depend upon laboratory capacity and the costs incurred in resource-constrained settings, where the number of available Xpert cartridges may be limited.

The performance of the readers and the software was similar, but AUC values were relatively low (0.65–0.75) against the bacteriological reference (Table 2). This might be attributed to the subject inclusion criteria of the WHO TB Specimen Bank study, where only subjects with TB symptoms were enrolled (Table 1).[18] According to the expert reading, 64.1% of bacteriologically negative subjects had an abnormal CXR with obvious TB-related abnormalities that were scored as highly suspicious by the clinical officers and CAD4TB. As a result, the use of a radiological reference resulted in considerably higher AUC values (0.89–0.94). Furthermore, the study population had a high prevalence of HIV (68%), which is known to lead to atypical abnormalities on CXR.[37] The performance we obtained with CAD4TB is comparable to what is reported in the literature on human reading of CXR in HIV-positive individuals (Figure 4).[11,14,30]

CAD4TB obtained a very high AUC value (0.91) and substantial agreement (κ = 0.61) with the radiological reference. These results indicate that CAD4TB performs well in detecting radiological abnormalities, and could be a useful tool in resource-constrained settings for the diagnosis of smear-negative TB. For CAD4TB to be used independently as a diagnostic tool

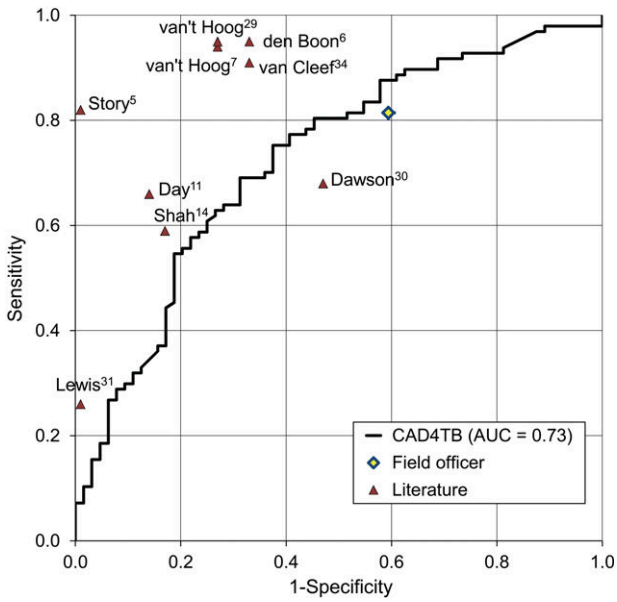


**Figure 4** The curve shows the CAD4TB ROC curve as reported in this study using the bacteriological reference. For comparison, sensitivity/specificity pairs reported in the literature are plotted. Numbers are reference numbers. Studies 11, 14 and 30 use data from HIV-positive individuals only. Table 4 provides further details about the literature studies. AUC = area under the ROC curve; ROC = receiver operating characteristic; HIV = human immunodeficiency virus. This image can be viewed online in colour at http://www.ingentaconnect.com/content/iuatld/ijtld/2013/00000017/00000012/art00020

in a clinic, very high sensitivity should be achieved at an acceptably high specificity. Future research includes training the software with various subtypes of abnormalities present in bacteriologically proven TB cases.

This study has some limitations. The study population was unbalanced, i.e., a large proportion of subjects with an abnormal CXR or HIV-positive status were residing in this high TB prevalence community. Our findings need to be validated in larger cohorts that include more healthy subjects, and the data should ideally be collected from a multicentre study with subjects from varied ethnicities. This would provide stronger support for the use of the CAD4TB for point-of-care testing or in prevalence surveys.

In conclusion, automated reading using CAD4TB has similar performance to that of clinical officers in assessing abnormalities indicative of TB. CAD4TB has the potential to be used as a point-of-care decision tool, assist human readers or identify subjects who should undergo further testing.

## *Acknowledgements*

The authors thank all the clinical officers: B Mwiche, M R Ndhlovu, L Banda and N Lubamba, at the urban health centre in Lusaka, for their reading of CXRs, and the staff at ZAMBART (Zambia AIDS Related Tuberculosis) for help in setting up the reader study and for providing CXRs and the clinical information.

This study was supported by the European and Developing Countries Clinical Trials Partnership, the Evaluation of Multiple Novel and Emerging Technologies for TB Diagnosis in Smear-negative and HIV-infected Persons in High-burden Countries (TB-NEAT study).

Conflict of interest: none declared.

## *References*

1 World Health Organization. Global tuberculosis report, 2012. WHO/HTM/TB/2012.6. Geneva, Switzerland: WHO, 2012.
2 World Health Organization. Systematic screening for active tuberculosis: principles and recommendations. WHO/HTM/TB/2013.04. Geneva, Switzerland: WHO, 2013.
3 World Health Organization. Tuberculosis prevalence surveys: a handbook. WHO/HTM/TB/2010.17. Geneva, Switzerland: WHO, 2011.
4 Golub J E, Mohan C I, Comstock G W, Chaisson R E. Active case finding of tuberculosis: historical perspective and future prospects. Int J Tuberc Lung Dis 2005; 9: 1183–1203.
5 Story A, Aldridge R W, Abubakar I, et al. Active case finding for pulmonary tuberculosis using mobile digital chest radiography: an observational study. Int J Tuberc Lung Dis 2012; 16: 1461–1467.
6 den Boon S, White N W, van Lill S W P, et al. An evaluation of symptom and chest radiographic screening in tuberculosis prevalence surveys. Int J Tuberc Lung Dis 2006; 10: 876–882.
7 van't Hoog A H, Meme H K, Laserson K F, et al. Screening strategies for tuberculosis prevalence surveys: the value of chest radiography and symptoms. PLOS ONE 2012; 7: e38691.
8 Hoa N B, Cobelens F G J, Sy D N, Nhung N V, Borgdorff M W, Tiemersma E W. Yield of interview screening and chest X-ray abnormalities in a tuberculosis prevalence survey. Int J Tuberc Lung Dis 2012; 16: 762–767.
9 Getahun H, Harrington M, O'Brien R, Nunn P. Diagnosis of smear-negative pulmonary tuberculosis in people with HIV infection or AIDS in resource-constrained settings: informing urgent policy changes. Lancet 2007; 369: 2042–2049.
10 Harries A D, Hargreaves N J, Kwanjana J H, Salaniponi F M. Clinical diagnosis of smear-negative pulmonary tuberculosis: an audit of diagnostic practice in hospitals in Malawi. Int J Tuberc Lung Dis 2001; 5: 1143–1147.
11 Day J H, Charalambous S, Fielding K L, Hayes R J, Churchyard G J, Grant A D. Screening for tuberculosis prior to isoniazid preventive therapy among HIV-infected gold miners in South Africa. Int J Tuberc Lung Dis 2006; 10: 523–529.
12 Shah N S, Anh M H, Thuy T T, et al. Population-based chest X-ray screening for pulmonary tuberculosis in people living with HIV/AIDS, An Giang, Vietnam. Int J Tuberc Lung Dis 2008; 12: 404–410.
13 Reid M J A, Shah N S. Approaches to tuberculosis screening and diagnosis in people with HIV in resource-limited settings. Lancet Infect Dis 2009; 9: 173–184.
14 Shah S, Demissie M, Lambert L, et al. Intensified tuberculosis case finding among HIV-infected persons from a voluntary counseling and testing center in Addis Ababa, Ethiopia. J Acquir Immune Defic Syndr 2009; 50: 537–545.
15 Bakari M, Arbeit R D, Mtei L, et al. Basis for treatment of tuberculosis among HIV-infected patients in Tanzania: the role of chest X-ray and sputum culture. BMC Infect Dis 2008; 8: 32.
16 van Ginneken B, Schaefer-Prokop C M, Prokop M. Computer-aided diagnosis: how to move from the laboratory to the clinic. Radiology 2011; 261: 719–732.
17 Zachary D, Schaap A, Muyoyeta M, Mulenga D, Brown J, Ayles H. Changes in tuberculosis notifications and treatment delay in Zambia when introducing a digital X-ray service. Public Health Action 2012; 2: 56–60.
18 Nathanson C-M, Cuevas L E, Cunningham J, et al. The TDR Tuberculosis Specimen Bank: a resource for diagnostic test developers. Int J Tuberc Lung Dis 2010; 14: 1461–1467.
19 Hogeweg L, Mol C, de Jong P A, Dawson R, Ayles H, van Ginneken B. Fusion of local and global detection systems to detect

tuberculosis in chest radiographs. Med Image Comput Comput Assist Interv 2010; 13 (Pt 3): 650–657.
20 Duda R O, Hart P E, Stork D G. Pattern classification. 2nd ed. New York, NY, USA: John Wiley and Sons, 2001.
21 van Ginneken B, Stegmann M B, Loog M. Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database. Med Image Anal 2006; 10: 19–40.
22 Arzhaeva Y, Tax D M J, van Ginneken B. Dissimilarity-based classification in the absence of local ground truth: application to the diagnostic interpretation of chest radiographs. Pattern Recognit 2009; 42: 1768–1776.
23 Seghers D, Loeckx D, Maes F, Vandermeulen D, Suetens P. Minimal shape and intensity cost path segmentation. IEEE Trans Med Imaging 2007; 26: 1115–1129.
24 Metz C E. ROC methodology in radiologic imaging. Invest Radiol 1986; 21: 720–733.
25 Karssemeijer N, Otten J D M, Rijken H, Holland R. Computer aided detection of masses in mammograms as decision support. Br J Radiol 2006; 79: S123–S126.
26 Samuelson F W, Petrick N. Comparing image detection algorithms using resampling. In: 3rd IEEE International Symposium on Biomedical Imaging: from nano to macro, Arlington, VA, USA, 2006. Pitscataway, NJ, USA: IEEE, 2006: 1312–1315.
27 Samuelson F W, Petrick N, Paquerault S. Advantages and examples of resampling for CAD evaluation. In: 4th IEEE International Symposium on Biomedical Imaging: from nano to macro, Arlington, VA, USA, 2007. Pitscataway, NJ, USA: IEEE, 2007: 492–495.
28 Landis J R, Koch G G. The measurement of observer agreement for categorical data. Biometrics 1977; 33: 159–174.
29 van't Hoog A H, Meme H K, van Deutekom H, et al. High sensitivity of chest radiograph reading by clinical officers in a tuberculosis prevalence survey. Int J Tuberc Lung Dis 2011; 15: 1308–1314.
30 Dawson R, Masuka P, Edwards D J, et al. Chest radiograph reading and recording system: evaluation for tuberculosis screening in patients with advanced HIV. Int J Tuberc Lung Dis 2010; 14: 52–58.
31 Lewis J J, Charalambous S, Day J H, et al. HIV infection does not affect active case finding of tuberculosis in South African gold miners. Am J Respir Crit Care Med 2009; 180: 1271–1278.
32 Zellweger J-P, Heinzer R, Touray M, Vidondo B, Altpeter E. Intra-observer and overall agreement in the radiological assessment of tuberculosis. Int J Tuberc Lung Dis 2006; 10: 1123–1126.
33 Balabanova Y, Coker R, Fedorin I, et al. Variability in interpretation of chest radiographs among Russian clinicians and implications for screening programmes: observational study. BMJ 2005; 331: 379–382.
34 Van Cleeff M R A, Kivihya-Ndugga L E, Meme H, Odhiambo J A, Klatser P R. The role and performance of chest X-ray for the diagnosis of tuberculosis: a cost-effectiveness analysis in Nairobi, Kenya. BMC Infect Dis 2005; 5: 111.
35 Soto A, Solari L, Díaz J, Mantilla A, Francine F, van der Stuyft P. Validation of a clinical-radiographic score to assess the probability of pulmonary tuberculosis in suspect patients with negative sputum smears. PLOS ONE 2011; 6: e18486.
36 Lam P K, LoBue P A, Catanzaro A. Clinical diagnosis of tuberculosis by specialists and non-specialists. Int J Tuberc Lung Dis 2009; 13: 659–661.
37 Kisembo H N, Boon S D, Davis J L, et al. Chest radiographic findings of pulmonary tuberculosis in severely immunocompromised patients with the human immunodeficiency virus. Br J Radiol 2012; 85: e130–e139.

## RÉSUMÉ

**CONTEXTE : Un centre de santé urbain très actif à Lusaka, Zambie.**

**OBJECTIF : Comparer la précision de la lecture automatique (CAD4TB) avec l'interprétation des agents cliniques concernant les radiographies digitales du thorax (CXR) pour la détection de la tuberculose (TB).**

**SCHÉMA : On a mené une analyse rétrospective sur 161 sujets recrutés dans une étude d'une banque d'échantillons de TB. Les radiographies ont été analysées au moyen de CAD4TB qui computérise un score d'anomalie d'image (0–100). Quatre agents cliniques ont donné un score au CXR concernant les anomalies compatibles avec la TB. Nous avons comparé les lectures automatiques et celles des agents cliniques avec une référence radiologique et bactériologique. Nous exposons la zone sous la courbe des caractéristiques opératoires du receveur (AUC) ainsi que les statistiques kappa (κ) entre les lecteurs et le CAD4TB.**

**RÉSULTATS : Sur 161 sujets recrutés, la TB a été confirmée par l'examen bactériologique dans 97 cas et les CXR ont été anormaux chez 120. L'AUC pour CAD4TB et pour les agents cliniques a été respectivement de 0,73 et de 0,65–0,75 par rapport à la référence bactériologique et de 0,91 et 0,89–0,94 par rapport à la référence radiologique. Les valeurs *P* n'ont pas indiqué de différences significatives, à l'exception d'un agent clinique dont les performances ont été significativement plus mauvaises que celles de la lecture automatique ($P < 0{,}005$) par rapport à la référence bactériologique. Les valeurs κ entre les agents cliniques et la référence radiologique ont été de 0,49–0,67.**

**CONCLUSION : L'évaluation des séries de CXR est comparable entre les CAD4TB et les agents cliniques. Le CAD4TB a des potentialités comme test sur les lieux de soins et pour l'identification automatique des sujets qui exigent des examens complémentaires.**

## RESUMEN

**MARCO DE REFERENCIA: Un centro de salud urbano de gran actividad de Lusaka, en Zambia.**

**OBJETIVO: Comparar la precisión de la lectura automática de las radiografías numéricas de tórax (CXR) por una sistema CAD4TB de detección asistida por computadora con la interpretación médica de las mismas en la detección de la tuberculosis (TB) pulmonar.**

**MÉTODO: Se llevó a cabo un análisis retrospectivo de 161 pacientes inscritos en un estudio de muestras recogidas con fines de creación de un banco. Se analizaron las radiografías mediante el sistema CAD4TB, el cual computaba una escala de puntuación de las imagines anormales (de 0 a 100). Cuatro médicos calificaron las anomalías radiográficas indicativas de TB. Luego se compararon ambas lecturas con una referencia bacteriológica y radiológica. En el estudio se presenta la curva de eficacia diagnóstica (AUC) y el coeficiente κ de concordancia entre los lectores y el sistema CAD4TB.**

**RESULTADOS: De los 161 pacientes inscritos, 97 obtuvieron confirmación bacteriológica del diagnóstico y 120 presentaron imágenes anormales en la CXR. Al considerar la referencia bacteriológica, el AUC del sistema CAD4TB fue 0,73 y el AUC de la lectura por los médicos osciló entre 0,65 y 0,75; con respecto a la referencia radiográfica, el AUC fue de 0,91 y de 0,89 a 0,94 respectivamente. Los valores de *P* demostraron que no existían diferencias significativas entre ambos sistemas de interpretación, con la excepción de uno de los lectores, cuyo desempeño fue considerablemente inferior al de la lectura automática ($P < 0{,}05$) al usar la referencia bacteriológica. Los coeficientes κ de concordancia entre la lectura de los médicos y la referencia radiológica oscilaron entre 0,49 y 0,67.**

**CONCLUSIÓN: La evaluación de las CXR numéricas mediante el sistema CAD4TB y la lectura realizada por los médicos es comparable. El sistema asistido por computadora ofrece posibilidades de uso en el lugar de atención y puede detectar de manera automática a las personas que precisan mayores investigaciones.**