INT J TUBERC LUNG DIS 23(7):805–810

http://dx.doi.org/10.5588/ijtld.18.0004

Automated chest X-ray reading for tuberculosis in the Philippines to improve case detection: a cohort study

R. H. H. M. Philipsen,* C. I. Sánchez,* J. Melendez,* W. J. Lew,†‡ B. van Ginneken*

*Diagnostic Image Analysis Group, Radboud University Medical Center, Nijmegen, The Netherlands; †World Health Organization Representative Office in Mongolia, Ulaanbaatar, Mongolia

SUMMARY

BACKGROUND: DetecTB (Diagnostic Enhanced Tools for Extra Cases of TB), an intensified tuberculosis (TB) case-finding programme targeting prisons and high-risk communities was implemented on Palawan Island, the Philippines.

OBJECTIVE: To evaluate the performance of TB detection based on computerised chest radiography (CXR) readings.

DESIGN: Data from 14 094 subjects were analysed from September 2012 to June 2014. All CXRs were read by a physician and by software. Individuals with TB symptoms or CXR abnormalities according to the physician underwent Xpert® MTB/RIF testing, the remaining persons were considered TB-negative (screening reference). A subset of 200 CXRs was read by an independent human reader (radiological reference). This reader also re-read a subset of the most abnormal cases as identified using the software but read as normal by the physician (discordant cases).

RESULTS: A total of 10 755 individuals were included in the analysis, 2534 of whom had a positively assessed CXR; 298 cases were Xpert-positive. Using the screening reference, the area under the receiver operating characteristic curve for software readings was 0.93 (95%CI 0.92–0.94), with a sensitivity of 0.98 (95%CI 0.97–0.99) and a specificity of 0.69 (95%CI 0.40–0.98). Based on the radiological reference, the physician performed slightly worse than the software (sensitivity, 0.82, 95%CI 0.74–0.89 and specificity, 0.87, 95%CI 0.81–0.96 vs. sensitivity, 0.83, 95%CI 0.71–0.93 and specificity, 0.87, 95%CI 0.75–0.95), although this was not statistically significant. Of the 291 discordant cases, 70% were assessed as positive, resulting in a 22% increase in TB detection when extrapolated to the full cohort.

CONCLUSION: The performance of automated CXR reading is comparable to that of the attending physicians in DetecTB, and its use as a second reader could increase TB detection.

KEY WORDS: TB; computer-aided detection; chest radiography; computerised image analysis

TUBERCULOSIS (TB) IS ONE of the leading causes of death worldwide and, as of 2015, ranked above HIV/AIDS (human immunodeficiency virus/acquired immune-deficiency syndrome).[1,2] The Millennium Development Goals target deadlines for 2015 have been achieved, and the TB incidence has declined, with the prevalence and mortality falling slowly. However, as the TB burden remains high, the End TB Strategy was initiated for a period of 20 years, until 2035. Key components of this new strategy are early diagnosis of TB and the development of new screening tools and diagnostic algorithms.[1]

In the present study, we evaluate the performance of CAD4TB 5 (Thirona, Nijmegen, The Netherlands), an automated chest X-ray (CXR) reading software from September 2012 to June 2014. This tool could facilitate rapid and early diagnosis of TB. Specifically, we analyse its application in the DetecTB (Diagnostic Enhanced Tools for Extra Cases of TB) Project. This project was conceptualised and developed by the World Health Organization (WHO) Country Office in the Philippines. The main objectives of DetecTB include increasing TB case detection and making possible same-day TB diagnosis in high-risk groups. To achieve these goals, systematic screening in accordance with the WHO's 2D TB screening strategy was implemented.[4] This strategy includes symptom screening combined with CXR, followed by confirmatory Xpert® MTB/RIF (Cepheid, Sunnyvale, CA, USA) testing if either screening result was positive.

To ensure efficient and successful TB detection, CXR and symptom pre-screening should 1) be rapid to reduce the logistic load, 2) be cheap to ensure low costs, 3) have high sensitivity to prevent missing positive cases, and 4) preferably have high specificity

‡ Deceased 27 March 2017.

Correspondence to: Rick H H M Philipsen, Department of Radiology and Nuclear Medicine, Radboud University Medical Center, Internal Postal Code 766, 6500 HB Nijmegen, The Netherlands. e-mail: rickphilipsen@thirona.eu

Article submitted 24 January 2018. Final version accepted 4 December 2018.

to reduce the burden and costs associated with follow-up Xpert tests. CXR interpretation requires skilled personnel, who are often scarce, so that radiographers or X-ray technicians with limited experience and knowledge are used instead. Training human readers is expensive, readers are subject to fatigue,[5] and readings are subject to inter- and intra-reader variability.[6–8]

To meet the requirements of pre-screening, and to overcome the challenges associated with the use of human readers, automated reading can be used in conjunction. Automated reading provides a fast and objective interpretation of CXRs for the presence of TB-related abnormalities. As results are available within 1 min after acquiring the CXR, follow-up actions can be undertaken rapidly. As the software produces a continuous output score for each CXR, the operating point can be chosen to optimally balance performance and costs according to the setting used.[9] For example, a moderate threshold can be chosen when using the software as a stand-alone reader, while a higher threshold could be set when the software is used for second readings, ensuring that abnormalities are not missed.

In this study, CXRs in the DetecTB project were retrospectively read using the CAD4TB 5 software. The performance of the human reader was compared with that of the software. We hypothesise that the sensitivity and specificity of the software are comparable to that of the attending physician. We also investigate whether the software could be used to improve TB detection rates when used as a second reader.

MATERIALS AND METHODS

Study setting

The study was performed on Palawan Island, the Philippines, among high TB risk groups. These include people living in slums, poor or marginalised areas, indigenous people and prisoners. The screening took place in a mobile clinic, which contained a digital radiography unit (Sitec Medical, Gyeonggi-do, Korea) and several GeneXpert IV machines. The mobile clinic was operated by two teams running on a biweekly basis. Each team comprised a physician, a nurse, a medical technologist and an X-ray technician. The mobile clinic travelled from one area to another to reach the targeted populations and to offer access to TB diagnosis.

As this study was part of a routine assessment of the ongoing project and data were extracted from existing records, ethical clearance was not required according to local regulations.

All participants aged ≥15 years underwent symptom screening and a digital posterior-anterior (PA) CXR (3072×3072 pixels, 0.1743^2 mm pixel spacing) in the mobile clinic. The CXR was read by one of the attending physicians. All physicians had previously completed a 2-week course of intensified training in digital CXR reading. Each CXR was read on a 5-category scale: 1, active TB; 2, TB activity undetermined; 3, inactive TB; 4, other disease; 5, normal. Handwritten transcripts were used to keep track of all the participants' study identifiers, names, and interview and CXR results. If the CXR was read as Category 1 or 2, or the participant had positive symptoms, a sputum sample was requested for Xpert testing. A positive Xpert result was considered confirmatory of TB diagnosis, after which treatment was initiated. Electronic records of the Xpert test results were kept. Transcripts were digitised in order to link the clinical information, the Xpert results and the CXRs. Figure 1 shows the number of cases for which all information could be linked.

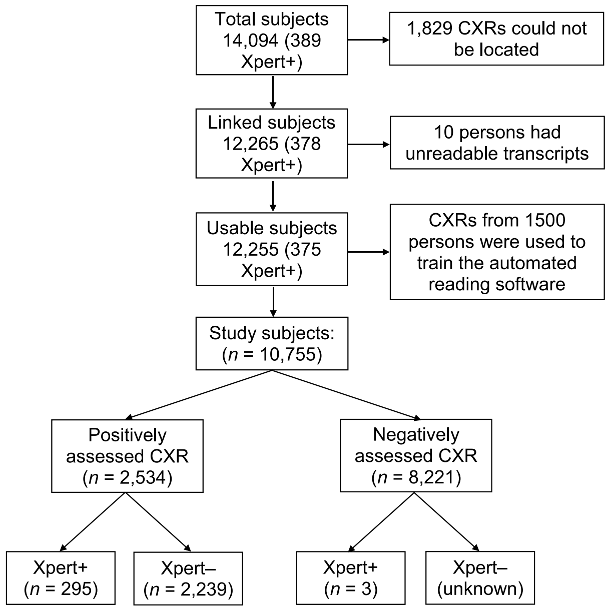


Figure 1 Simplified flowchart of data distribution.

Automated CXR reading

Automated CXR reading was performed using the CAD4TB 5 software, which automatically processes digital PA CXRs for the presence of abnormalities and scores them on a scale from 0 to 100, where a higher score indicates a higher likelihood for the CXR to be abnormal. The software uses supervised machine learning techniques.[10] First, the CXR is normalised to make it scanner-independent.[11] Second, the lungs are segmented using a technique described by van Ginneken et al.[12] A number of texture features are computed for each pixel in the lungs. A classifier is then applied to assign a score to the pixels within the lung fields. Finally, these scores are added into a single value using a quantile rule.[13]

The system is trained with labelled regions from both normal and abnormal images. A total of 1500 CXRs from the original data set were used: 629

random, physician-read CXRs from Categories 1–4 and 871 random CXRs from Category 5. Abnormal regions in the CXRs were outlined and were used as input to train the software in texture analysis. All 1500 training images were excluded for analysis, leaving a final test set of 10 755 images.

Study design and reference standards

Three different analyses were conducted: in the first two, the performance of the software and that of the physician were compared using different reference standards; in the last, we investigated a scenario where the software was used as an additional reader.

In the first analysis, the reference standard used (the screening reference) was the Xpert test result of subjects with TB symptoms and/or a positive CXR (Category 1 or 2). All other subjects (who did not undergo Xpert testing) were assumed to be negative for TB.

To compare the performance of the software with that of the study physician, a second analysis was conducted. The reference standard (the radiological reference) was set by an independent human expert who blindly read the CXRs from 100 randomly selected cases that were originally read by the physician as abnormal for TB (Category 1 and 2), and 100 randomly selected CXRs with no TB abnormalities (Category 3, 4 and 5). The 200 CXRs were read as either suspicious for TB or highly unlikely TB.

In the last analysis, CXRs read as normal by the attending physician were considered suspicious for TB if the computer reading score exceeded the predefined threshold of 60. We believe that individuals with a score above this threshold should undergo additional Xpert testing. To investigate if this approach would increase the TB detection rate, the independent human expert blindly read a subset of all CXRs read as normal by the physician, stratified by CAD4TB score using the same categories as in the second analysis.

To compare the performance of the software and the physician, the sensitivity and the specificity were estimated and a receiver operator characteristic (ROC) curve analysis was performed. Bootstrapping methods were used to calculate 95% confidence intervals (CIs) for sensitivity, specificity and area under the ROC curves.[14,15] McNemar's test was used to determine significant differences.

RESULTS

Of all 14 094 subjects included in the programme from September 2012 to June 2014, 389 (2.7%) bacteriologically tested positive on Xpert. Of these 389 Xpert-positive individuals, 378 could be matched to the clinical information and CXR. Another 10 individuals had unreadable clinical records, leading to a final data set of 12 255 cases with 375 Xpert-positive cases (Figure 1). Excluding the CXRs used for training, the remaining data set comprised 10 755 CXRs, including 298 Xpert-positive cases (2.8%). Of these 10 755 CXRs, 2534 (23.6%) were read as abnormal for TB by the attending physician. Table 1 lists the exact distribution of cases across the five CXR categories as read by the physician. The results for the software are shown in Table 2. This table includes the numbers of cases that were read as either positive or negative by the physician, and shows the number of Xpert-positives.

Table 1 Case distribution and number of Xpert-positive cases among CXR categories as read by the physician

CXR category	Total *n* (%)	Xpert-positive *n*
All, *n*	10 755	298
TB active	731 (6.8)	167
TB activity undetermined	1 803 (16.8)	128
TB inactive	294 (2.7)	2
Other disease	256 (2.4)	0
Normal	7 671 (71.3)	1

CXR = chest X-ray; TB = tuberculosis.

The performance of the software when used as standalone reader (analysis 1) based on the screening reference are shown in the ROC curve in Figure 2. Using a threshold of 60, the software had an area under the ROC curve (AUC) of 0.93 (95%CI 0.91–0.94), a sensitivity of 0.98 and a specificity of 0.69. The corresponding 2×2 contingency table is shown in Table 3.

To set the radiological reference for the second analysis, the independent observer read 100 randomly selected cases read as negative by the physician and 100 randomly selected cases read as positive by the physician. From these 200 cases, the independent observer read 108 cases as abnormal for TB active, and the remaining 92 were read as TB-negative. Using this reference, the ROC curve of CAD4TB and the operating point of the physician are shown in Figure

Table 2 Distribution of CAD4TB scores, physician readings and Xpert results

CAD4TB	Total *n*	Physician–* *n*	Physician+† *n*	Xpert-positive *n/N* (%)‡
0–30	4165	3995	170	2/170 (1)
31–40	1670	1516	154	2/154 (1)
41–50	1364	1164	200	1/200 (1)
51–60	935	702	233	6/233 (3)
61–70	955	553	402	17/402 (4)
71–80	700	200	500	53/500 (11)
81–90	265	35	230	54/230 (23)
91–95	241	21	220	44/220 (20)
>95	460	35	425	119/425 (28)

* Number of cases read as negative for tuberculosis (Category 3, 4 and 5) by the physician.
† Number of cases read as positive (Category 1 and 2) by the physician.
‡ Number of Xpert-positive cases compared to the total number of Xpert tests performed.
– = negative; + = positive.

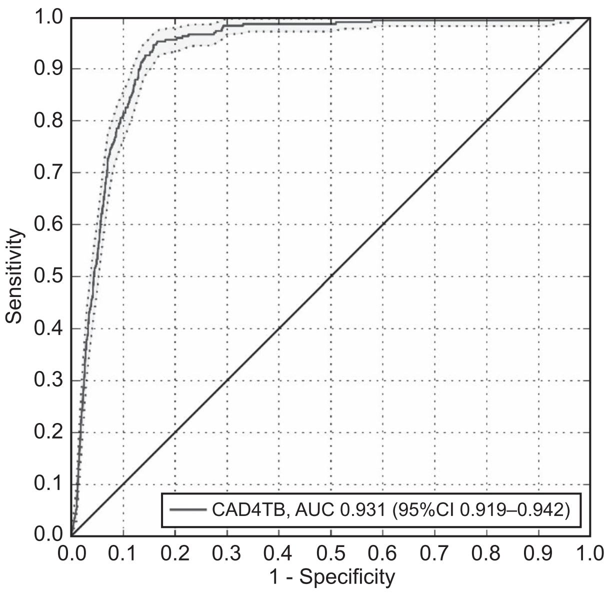


Figure 2 ROC curve for the automated software. The dotted lines depict the confidence interval. AUC = area under the ROC curve.

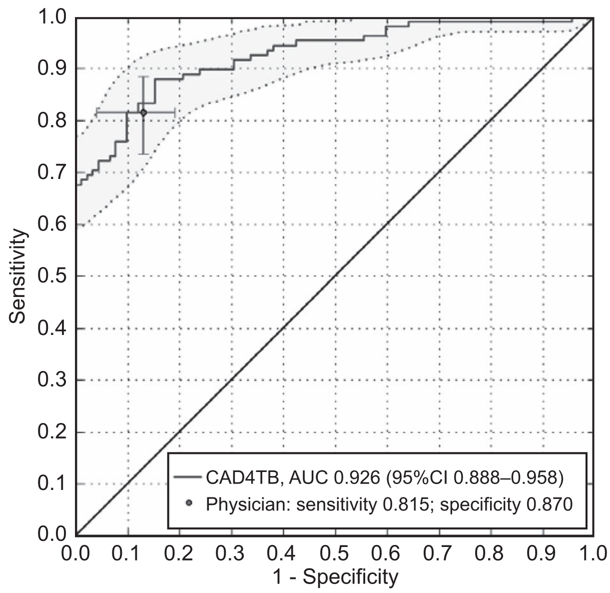


Figure 3 Automated vs. human reading using the radiological reference. Sensitivity and specificity of the physician was 0.82 (95%CI 0.74–0.89) and 0.87 (95%CI 0.81–0.96). The dotted lines depict the confidence interval. AUC = area under the ROC curve.

3. The AUC value of the software is 0.93 (95%CI 0.89–0.96); the physician had a sensitivity and specificity of respectively 0.82 (95%CI 0.74–0.89) and 0.87 (95%CI 0.81–0.96). At the same specificity, the software had a slightly higher sensitivity than the physician (0.83, 95%CI 0.72–0.94; P = 0.739), although the difference was not statistically significant. Table 4 gives the contingency table comparing physician and software readings using a threshold of 60. At this threshold, the software had a sensitivity and a specificity of respectively 0.74 and 0.92.

The distribution of CAD4TB scores and the number of Xpert-positive cases in the last analysis is shown in Table 2. This table shows that 2621 of 10 755 software-read CXRs (24.8%) were scored above 60 (Table 1), which is comparable to the number of cases read as abnormal by the physician: 2534 of 10 755 (23.6%). Of these 2621 CXRs, 844 cases (32%) were read by the physician as TB-negative (Table 2, column 3) and 287 (11%) cases were Xpert-positive, i.e., 11% of all cases with a score above 60 were Xpert-positive, amounting to 96% of all Xpert-positive cases in the data set. The results of all the cases read by the independent observer and presumed negative by the physician are shown in Table 5. This shows that the independent reader considered 70% (204/291) of those presumably normal images, but with a score above 60, as abnormal. Given that 844 persons received a score above 60 on the software but were read as negative by the physician, a misclassification rate of 70% would mean that an additional 591 individuals should undergo Xpert testing. Assuming a similar TB rate of 11% among cases with a score above 60 (see above, 287/2621 were Xpert-positive), this would result in the detection of around 65 additional TB cases, in addition to the original 298 cases, which is a substantial increase of 22% in the TB detection rate.

Figure 4 gives some sample CXRs, including the software output, the physician's reading, the independent reading, the Xpert results, and the CAD4TB scores and heat map indicating abnormal areas found by the software.

DISCUSSION

This study compared the performance of automated CXR reading with reading by human readers in different analyses. Using the screening reference, software readings had a sensitivity of 0.98 at a

Table 3 Contingency table for automated readings using the screening reference (threshold = 60)

	Automated reading	
Screening reference	Positive *n*	Negative *n*
Positive	293	5
Negative	3263	7194

Table 4 Contingency table for physician and automated readings using the radiological reference (threshold = 60)

	Physician		Automated reading	
Radiological reference	Positive *n*	Negative *n*	Positive *n*	Negative *n*
Positive	88	20	80	28
Negative	12	80	7	85

Table 5 Independent CXR readings of all 691 cases in Analysis 3*

	Independent reader	
CAD4TB bin	Positive *n*	Negative *n*
Normal		
0–30	5	95
31–40	6	94
41–50	17	83
51–60	21	79
Abnormal		
61–70	45	55
71–80	81	19
81–90	32	3
91–95	21	0
>95	25	10

* A threshold of 60 was used for the software. Only discordant cases (read as negative for TB by the physician) were used. For the six bins with lowest CAD4TB scores (0–30 until 71–80), a random sample of 100 cases was read, for the other three bins with the highest CAD4TB scores (81–90 to >95), all available CXRs were read. CXR = chest X-ray; TB = tuberculosis.

specificity of 0.69. Although comparisons with the physician's readings might be relevant, this was not done due to the following two limitations. First, the physician may have taken into account clinical information, as the physician had seen the participant before reading the CXR. Second, all asymptomatic cases with a presumptive normal CXR were negative for TB in this reference standard, i.e., the physician was, by definition, always assumed to be correct in these cases. The fact that the attending physician missed only 3/298 Xpert-positive cases indicates that the proportion of persons with only TB symptoms and a CXR read as normal was very small (1%).

The performance of the physician is slightly poorer than the automated reading in analysis 2: at the specificity level of the physician (0.87), the software had at a sensitivity of 0.82 (95%CI 0.74–0.0.88) vs. 0.83 (95%CI 0.70–0.92), respectively, although the difference was not significant (P = 0.739). The performance of humans and the software in reading CXRs was therefore comparable. While the physicians were asked not to take into account any clinical information, the results (only three TB-positive cases based on symptoms only) indicate that they might have done so.

On comparing both ROC curves, a difference between the physician's reading and software score was observed at high specificity values: the first ROC curve (Figure 2) is less steep at the start, which suggests that some of the most abnormal CXRs according to the software were in fact Xpert-negative.

To achieve the highest TB detection rate, active case finding programmes such as DetecTB could use a combination of human and automated CXR reading. When any one of the two CXR readings is positive, the subject should undergo follow-up Xpert testing. For analysis 3, a threshold of 60 on the CAD4TB score seems appropriate, as this would lead to the selection of a similar proportions of screened subjects and physician's readings, around a fourth, for additional testing. The results of this analysis (Table 5) show that this approach has potential: 70% of the CXRs with a high CAD4TB score that were read as normal by the physician, were re-read as abnormal by the independent reader. Extrapolating these results to the entire data set, and assuming a TB rate of 11% as was found in the original data, the use of the CAD4TB software as a second reader would result in the detection of around 65 additional TB cases. In addition to the original 298 detected cases, this represents a substantial increase of 22% in TB detection rate.

To date, this is the first study in the literature to examine the use of automated CXR reading for active case finding in an Asian population. All other studies have been conducted in African[16–18] or European

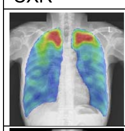
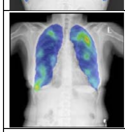
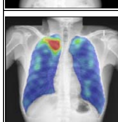
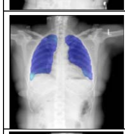
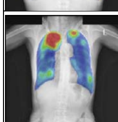
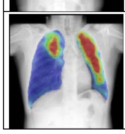
CXR	Details	CXR	Details
	Xpert: + Physician: + CAD4TB: 96 Reader 2: +		Xpert: not performed Physician: – CAD4TB: 16 Reader 2: –
	Xpert: + Physician: + CAD4TB: 49 Reader 2: +		Xpert: not performed Physician: – CAD4TB: 88 Reader 2: –
	Xpert: not performed Physician: – CAD4TB: 22 Reader 2: +		Xpert: not performed Physician: – CAD4TB: 93 Reader 2: +
	Xpert: + Physician: – CAD4TB: 89 Reader 2: +		Xpert: not performed Physician: – CAD4TB: 90 Reader 2: +

Figure 4 Sample CXRs with reading and bacteriological information. The two examples in the last row are cases read as 'Inactive TB' by the physician, the left case was symptom-positive and thus has Xpert results. CXR = chest X-ray; + = positive; – = negative; TB = tuberculosis.

settings,[19] and were mostly used for passive case finding or prevalence surveys.[20,21] In agreement with other study findings, our results indicate that the CAD4TB software achieves similar performance as human field readers.

The study has several limitations: in the DetecTB programme, only persons with symptoms or TB-positive CXRs underwent Xpert testing. This renders it difficult to compare the performance of attending physician to that of the software, as doing so would lead to a potential bias for the attending physician and gives an unconfirmed diagnosis of no TB for the screening reference. Xpert should ideally be available for all individuals in a research study setting. As a result, the number of additional cases detected by complementing human reading by software reading can only be estimated. Additional research is needed to optimise and evaluate the use of software and human reading, taking costs and cost-effectiveness factors into account.[9]

CONCLUSION

The performance of the CAD4TB software in reading CXRs is similar to that of the attending physician. Our findings indicate that the TB detection rate could be substantially increased if CAD4TB were integrated as an additional reader in selecting individuals for additional follow-up during active case finding.

Our findings indicate that a substantial number of presumptive TB cases were missed by the physician, although this was not confirmed by Xpert testing. By using a second software-based CXR reading technique, additional TB-positive cases that were initially missed can be detected at limited additional costs.

Acknowledgements

The authors thank L Infante for his help in digitising handwritten transcripts of clinical information and physicians' readings.

Disclaimer: The contents of this manuscript are the sole responsibility of the authors, and the funding body had no role in the decision to publish the manuscript.

Conflict of interest: RP and JM are employed by Thirona (Nijmegen, The Netherlands) as developers of the CAD4TB software.

References

1 World Health Organization. Global tuberculosis report, 2016. WHO/HTM/TB/2016.13. Geneva, Switzerland: WHO, 2016.
2 World Health Organization. Global tuberculosis report, 2015. WHO/HTM/TB/2015.22. Geneva, Switzerland: WHO, 2015.
3 World Health Organization. WHO Representative Office, Philippines. http://www.wpro.who.int/philippines/areas/communicable_diseases/tb/story_detectb_project_palawan/en/ Accessed June 2019.
4 World Health Organization. Systematic screening for active tuberculosis: principles and recommendations. WHO/HTM/TB/2013.04. Geneva, Switzerland: WHO, 2013.
5 Vertinsky T, Forster B. Prevalence of eye strain among radiologists: influence on viewing variables on symptoms. AJR Am J Roentgenol 2005; 184: 681–686.
6 Sakurada S, Hang N T, Ishizuka N, et al. Inter-rater agreement in the assessment of abnormal chest X-ray findings for tuberculosis between two Asian countries. BMC Infect Dis 2012; 12: 31.
7 Graham S, Das G K, Hidvegi R J, et al. Chest radiograph abnormalities associated with tuberculosis: reproducibility and yield of active cases. Int J Tuberc Lung Dis 2002; 6: 137–142.
8 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.
9 Philipsen R H H M, Sánchez C I, Maduskar P, et al. Automated chest radiography as a triage for Xpert testing in resource-constrained settings: a prospective study of diagnostic accuracy and costs. Sci Rep 2015; 5: 12215.
10 Duda R, Hart E, Stork D G. Pattern classification. 2nd ed. Hoboken, NJ, USA: John Wiley & Sons, 2000.
11 Philipsen R H H M, Maduskar P, Hogeweg L, Melendez J, Sánchez C I, van Ginneken B. Localized energy-based normalization of medical images: Application to chest radiography. IEEE Trans Med Imaging 2015; 34: 1965–1975.
12 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.
13 Loog M, van Ginneken B. Static posterior probability fusion for signal detection: applications in the detection of interstitial diseases in chest radiographs. In: Proceedings of the 17th International Conference on Pattern Recognition, 23–26 August 2004, Cambridge, UK: pp 644–647.
14 Metz C E, Herman B A, Roe C A. Statistical comparison of two ROC curve estimates obtained from partially-paired datasets. Med Decis Making 1998; 18: 110–121.
15 Metz C E, Herman B A, Shen J H. Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data. Stat Med 1998; 17: 1033–1053.
16 Muyoyeta M, Maduskar P, Moyo M, et al. The sensitivity and specificity of using a computer-aided diagnosis program for automatically scoring chest X-rays of presumptive TB patients compared with Xpert MTB/RIF in Lusaka Zambia. PLOS ONE 2014; 9: e93757.
17 Maduskar P, Muyoyeta M, Ayles H, Hogeweg L, Peters-Bax L, van Ginneken B. Detection of tuberculosis with digital chest radiography: automatic reading versus interpretation by clinical officers. Int J Tuberc Lung Dis 2013; 17: 1613–1620.
18 Breuninger M, van Ginneken B, Philipsen R H H M, et al. Diagnostic accuracy of computer-aided detection of pulmonary tuberculosis in chest radiographs: a validation study from sub-Saharan Africa. PLOS ONE 2014; 9: e106381.
19 Melendez J, Hogeweg L, Sánchez C I, et al. Accuracy of an automated system for tuberculosis detection on chest radiographs in high-risk screening. Int J Tuberc Lung Dis 2018; 22: 567–571.
20 Rahman T, Codlin A, Rahman M, et al. An evaluation of automated chest radiography reading software for tuberculosis screening among public- and private-sector patients. Eur Respir J 2017; 49: 1602159.
21 Melendez J, Philipsen R H H M, Chanda-Kapata P, Sunkuntu V, Kapata N, van Ginneken B. Automatic versus human chest radiograph reading in the Zambia national tuberculosis prevalence survey. Int J Tuberc Lung Dis 2017; 21: 880–886.

RÉSUMÉ

CONTEXTE : DetecTB (Diagnostic Enhanced Tools for Extra Cases of TB), un programme intensifié de recherche des cas de tuberculose (TB) ciblant les prisons et les communautés à haut risque dans les îles Palawan, Philippines.

OBJECTIF : Evaluer la performance, en matière de détection de la TB, de la lecture de radiographies pulmonaires (CXR) numérisées.

SCHÉMA : Les données de 14 094 sujets enrôlés dans le programme DetecTB ont été analysées. Toutes les CXR ont été lues par un médecin et par un logiciel. Les sujets ayant des symptômes de TB ou une CXR anormale, selon le médecin, ont eu un test Xpert® MTB/RIF sinon ils ont été considérés comme négatifs pour la TB (référence du dépistage). Un sous-groupe de 200 CXR a été lu par une personne indépendante (référence radiologique). Ce lecteur a également relu le sous-groupe des cas les plus anormaux identifiés par le logiciel mais considérés comme normaux par le médecin (cas discordants).

RÉSULTATS : Sur 10 755 sujets qui ont été utilisés pour l'analyse, dont 2534 ont eu une CXR évaluée comme positive. Au total, 298 cas ont été positifs pour le Xpert. En utilisant la référence du dépistage, le logiciel a abouti à une zone sous la courbe de la fonction d'efficacité du récepteur de 0,93 (IC95% 0,92–0,94), avec une sensibilité de 0,98 (IC95% 0,97–0,99) et une spécificité de 0,69 (IC95% 0,40–0,98). En utilisant la référence radiologique, le médecin a eu une performance légèrement, mais pas significativement, moins bonne que le logiciel : sensibilité de 0,82 (IC95% 0,74–0,89) et spécificité de 0,87 (IC95% 0,81–0,96) contre une sensibilité de 0,83 (IC95% 0,71–0,93) et une spécificité de 0,87 (IC95% 0,75–0,95). En ce qui concerne les 291 cas discordants, 70% ont été lus comme positifs par le lecteur indépendant, résultant en l'extrapolant à la cohorte entière, en une augmentation de 22% de la détection de TB.

CONCLUSION : La lecture automatisée des CXR a une performance comparable à celle des médecins dans le programme DetecTB, et son utilisation comme deuxième lecteur augmenterait le taux de détection de la TB.

RESUMEN

MARCO DE REFERENCIA: El programa DetecTB (Diagnostic Enhanced Tools for Extra Cases of TB) de búsqueda intensificada de casos de tuberculosis (TB), dirigido a los entornos carcelarios y las comunidades de alto riesgo en las Islas Palawan de Filipinas.

OBJETIVO: Evaluar la eficacia diagnóstica de la lectura computarizada de la radiografía de tórax (CXR).

MÉTODO: Se analizaron los datos de 14 094 personas inscritas al programa DetecTB. Todas las CXR fueron interpretadas por un médico y por el programa informático. A las personas con síntomas indicativos de TB o CXR anormales según el médico, se practicó la prueba Xpert® MTB/RIF y las demás se consideraron con un resultado negativo para TB (referencia del tamizaje). Además, un lector independiente interpretó un subgrupo de 200 CXR (referencia radiológica). Este lector también realizó una segunda lectura de un subgrupo de los casos más anormales según la interpretación del programa, pero que el médico calificó como normales (casos discordantes).

RESULTADOS: De las 10 755 personas incluidas en el análisis, la CXR de 2534 casos se consideró positiva. En total, 298 casos obtuvieron un resultado positivo en la prueba Xpert. Con respecto a la referencia del tamizaje, el programa informático alcanzó un área bajo la curva de eficacia diagnóstica de 0,93 (IC95% 0,92–0,94), con una sensibilidad de 0,98 (IC95% 0,97–0,99) y una especificidad de 0,69 (IC95% 0,40–0,98). Con respecto a la referencia radiológica, el desempeño del médico fue ligeramente inferior al del programa informático, pero la diferencia no alcanzó significación estadística; su sensibilidad fue 0,82 (IC95% 0,74–0,89) y su especificidad de 0,87 (IC95% 0,81–0,96) contra 0,83 (IC95% 0,71–0,93) y 0,87 (IC95% 0,75–0,95) respectivamente. En los 291 casos discordantes, el lector independiente interpretó un 70% como positivo, lo cual, extrapolado a la cohorte completa significó un aumento del 22% de detección de la TB.

CONCLUSIÓN: La lectura automática de la CXR ofrece un rendimiento comparable al de la interpretación realizada por el médico tratante en el marco del programa DetecTB y si se utiliza como una segunda lectura, aumentaría la tasa de detección de la TB.