A screening system for smear-negative pulmonary tuberculosis using artificial neural networks

CrossMark

João B. de O. Souza Filho [a,b,*], José Manoel de Seixas [a,c], Rafael Galliez [d],
Basilio de Bragança Pereira [d], Fernanda C. de Q Mello [d],
Alcione Miranda dos Santos [e], Afranio Lineu Kritski [d]

[a] Polytechnical School (POLI), Electronics and Computer Engineering Department (DEL), Avenida Athos da Silveira Ramos, 149, Technological Center, Building H, room H-219 (room 20), Federal University of Rio de Janeiro, Rio de Janeiro, Brazil
[b] Electrical Engineering Postgraduate Program (PPEEL), Federal Centre of Technological Education Celso Suckow da Fonseca, Rio de Janeiro, Brazil
[c] Electrical Engineering Postgraduate Program (PEE), Alberto Luiz Coimbra Institute (COPPE), Federal University of Rio de Janeiro, Rio de Janeiro, Brazil
[d] Academic Tuberculosis Program, Faculty of Medicine and University Complex (HUFF and IDT), Federal University of Rio de Janeiro, Rio de Janeiro, Brazil
[e] Postgraduate Program on Collective Health, Federal University of Maranhão, Maranhão, Brazil

ARTICLE INFO

SUMMARY

Objectives: Molecular tests show low sensitivity for smear-negative pulmonary tuberculosis (PTB). A screening and risk assessment system for smear-negative PTB using artificial neural networks (ANNs) based on patient signs and symptoms is proposed.
Methods: The prognostic and risk assessment models exploit a multilayer perceptron (MLP) and inspired adaptive resonance theory (iART) network. Model development considered data from 136 patients with suspected smear-negative PTB in a general hospital.
Results: MLP showed higher sensitivity (100%, 95% confidence interval (CI) 78–100%) than the other techniques, such as support vector machine (SVM) linear (86%; 95% CI 60–96%), multivariate logistic regression (MLR) (79%; 95% CI 53–93%), and classification and regression tree (CART) (71%; 95% CI 45–88%). MLR showed a slightly higher specificity (85%; 95% CI 59–96%) than MLP (80%; 95% CI 54–93%), SVM linear (75%, 95% CI 49–90%), and CART (65%; 95% CI 39–84%). In terms of the area under the receiver operating characteristic curve (AUC), the MLP model exhibited a higher value (0.918, 95% CI 0.824–1.000) than the SVM linear (0.796, 95% CI 0.651–0.970) and MLR (0.782, 95% CI 0.663–0.960) models. The significant signs and symptoms identified in risk groups are coherent with clinical practice.
Conclusions: In settings with a high prevalence of smear-negative PTB, the system can be useful for screening and also to aid clinical practice in expediting complementary tests for higher risk patients .

1. Introduction

The directly observed therapy strategy (DOTS) has improved tuberculosis (TB) cure rates in several nations, although its use remains low. Globally, smear-negative pulmonary tuberculosis (PTB) accounts for 20–50% of active TB.[1]

The World Health Organization (WHO) has recommended that those with a chronic cough, i.e., a cough lasting 2 or more weeks, should have expectorated sputum evaluated by acid-fast bacillus (AFB) smear examination or using the Xpert MTB/RIF (Xpert) assay.[1,2] The use of AFB testing and the detection of PTB based solely on a chronic cough show low accuracy in smear-negative PTB patients.[2–4]

To improve the detection of smear-negative PTB cases, the WHO 2007 algorithm recommends the use of culture, chest X-ray (CXR), and new diagnostic tests, such as the Xpert assay.[3] A recent meta-analysis compared the performance of Xpert, the microscopic observation drug susceptibility assay (MODS), and the WHO 2007 algorithm for the diagnosis of smear-negative PTB.[5] The pooled sensitivity and specificity were 67% and 98% for Xpert, 73% and 91% for MODS, and 61% and 69% for the WHO 2007 algorithm, respectively.[5] In addition, in spite of various surmountable barriers

observed in the implementation of the Xpert assay in nine countries, the detection of TB cases did not, in fact, increase.[6]

At screening of smear-negative PTB patients, without a standardized clinical work-up, misdiagnosis rates may be as high as 38% in case of assuming that any CXR abnormality would be considered to define a positive prognostic.[7] Other TB risk factors, such as symptoms suggestive of PTB, i.e., chest pain or haemoptysis, or systemic symptoms such as weight loss, night sweats, fever, chills, fatigue, and loss of appetite, would be helpful to rule in or rule out the diagnosis of TB.[8–14]

Several studies have proposed scoring systems to identify PTB using multivariate logistic regression (MLR),[15] classification trees,[16] and artificial neural networks (ANNs).[17–20] ANNs are powerful modelling tools, useful for solving complex problems, and explore existing non-linear relationships between variables extracted automatically from data. Several medical applications have benefited from the use of ANNs. This set of techniques usually outperforms classical methods, such as logistic regression,[21] when modelling clinical data.

A decision support system (DSS) as a first approach for the diagnosis of smear-negative PTB based on ANNs is proposed here.[22] This DSS serves as a TB prognostic and risk assessment tool and is based solely on patient signs and symptoms. As a prognostic it can be useful for the selection of patients for molecular tests, such as the Xpert assay, while as a risk assessment tool it can form the basis on which to expedite the use of complementary examinations for higher risk patients.

2. Study population, design, and methods

2.1. Dataset description

The dataset was based on 136 adults with symptoms and/or signs suggestive of smear-negative PTB who were referred to the teaching hospital of the Federal University of Rio de Janeiro (a general hospital) from January 1, 2010 to December 30, 2011. The recruitment and enrolment of subjects followed the approach described in a previous study.[23] In brief, subjects who had two consecutive samples of spontaneous sputum that were AFB smear-negative, or who had absence of expectorated sputum, excluding those who had already received anti-TB treatment, were included.

The patients underwent a standardized interview with questions on demographic variables and clinical history (e.g., smoking, alcohol abuse, and HIV infection). All underwent HIV testing by ELISA. Western blot of reactive ELISA cases confirmed an HIV infection in 64 patients (47%). AIDS cases were identified according to the Caracas definition of AIDS.[24] Clinical samples were sent to the mycobacteriology laboratory for culture, drug susceptibility testing, and identification to the species level using techniques described fully elsewhere.[25,26]

Patients with a positive culture for *Mycobacterium tuberculosis* in a respiratory specimen (confirmed smear-negative PTB) and those for whom some clinical improvement was observed after 6 months of solely anti-TB treatment (presumptive smear-negative PTB) were considered as PTB cases. Subjects for whom (1) AFB smears and culture for *M. tuberculosis* were negative, (2) CXR showed no change after 6 months of follow-up, or (3) the notifiable diseases surveillance database (SINAN) did not signal a positive PTB diagnosis within 2 years from the date of their enrolment in the study, were considered as non-PTB cases. A panel of experts formed of two pulmonologists reviewed all presumed smear-negative PTB cases. A third physician reviewed all cases with discordant results.

Based on a panel of three TB expert researchers, the following 12 variables were selected for the development of the ANN models: age, presence of cough, fever, haemoptysis, anorexia, weight loss, HIV, night sweats, absence of dyspnoea, smoking, association with extrapulmonary TB, and previous hospitalization. All signs and symptoms were coded as +1 and −1 to represent their presence or absence, and as 0 when the information was not available. Age was normalized to have 0 mean and to be within the range of −1 to +1.

Relationships among these variables and the diagnosis of smear-negative PTB were inferred using univariate and multivariate logistic regression (ULR and MLR) models.[27] For simplicity, this set of variables was also evaluated using a wrapper feature selection method,[28] considering a logistic regression model, according to both forward and backward sequential floating search procedures (FSFS and BSFS).[29] In this analysis, the probability of inclusion of each variable was 25% and the probability of exclusion of each variable was 30%. This process considered the use of the likelihood ratio test.[28]

2.2. Development of artificial neural networks

ANNs are biologically inspired models able to learn through examples. Composed of artificial neurons arranged in layers and interconnected by synaptic weights, the ANNs may acquire knowledge by weight adaptation during the training process.[22]

The adoption of ANNs to solve a particular problem involves two main phases: training and operation. In the first, network synaptic weights are adapted to solve a particular task, extracting knowledge from training data. Once the training phase is complete, ANNs operate only producing outputs based on the stored knowledge (operating phase).

The proposed screening system exploits a three-layer multi-layer perceptron (MLP) responsible for TB prognostics,[22] and a model inspired on adaptive resonance theory (ART),[30] here referred to as iART, to classify patients into low, medium, and high risk levels. As MLP and iART modules operate in a complementary fashion, discordant results may signal an atypical case for further clinical investigation.

The development of the MLP model considered the holdout approach, resulting in the split of the dataset into a training set and a testing set.[31] The first set was used for MLP training, i.e., for synaptic weight adaptation, whereas the second was used to estimate network prediction generalization.

Since generalization of the MLP model was shown to be sensitive to the composition of the training and test sets, mainly due to the small dataset size, an instance selection procedure was applied to define the contents of these sets.[32] The procedure adopted consisted of the production of a cluster of three groups using the iART model and random splitting of the patients belonging to each group between the two sets, in the following proportions of cases: 75% training and 25% testing. The resulting training and testing sets retained a total of 102 and 34 patients, respectively, from which 45 (44.5%) and 14 (41.4%), respectively, were positive cases. This procedure aimed to result in problem representative sets, thus conjugating robust learning with a realistic model evaluation.

The MLP module employed a three-layer architecture with 12 input nodes, corresponding to the number of dataset variables selected by the TB experts as relevant to TB diagnosis, and a single output neuron, as this model was trained to give the values −1 and +1 to negative and positive cases, respectively. The number of hidden neurons was chosen using cross-validation (CV) as 15. The hyperbolic tangent was the activation function of all neurons. The output of the resulting predictive model is given by the following equation:

$$y(\mathbf{x}) = tanh\left(\sum_{i=1}^{15} c_i tanh\left(\sum_{j=1}^{12} a_{ij}x_j + b_i\right) + d_i\right) \quad (1)$$

where the vector of input variables is denoted by $\mathbf{x}$, the j-th component of $\mathbf{x}$ is given by x_j, the synaptic weights from input and

output layers are defined by the variables a_{ij}, b_i and c_i, d_i, respectively, while the function tanh is the hyperbolic tangent.

MLP training employed the resilient backpropagation algorithm (RPROP)[33] in batch mode.[22] This first-order optimization algorithm is robust to the occurrence of small-gradient values and adjusts network synaptic weights only using the signs of the gradient components. This strategy reduces the dependence of network training on both the landscape of objective function and the initial values of the synaptic weights, usually resulting in better training.

In order to mitigate possible class-imbalance effects,[34] which may be generated by the difference in number of negative and positive cases in the dataset, the following adjusted mean-square error function was considered for synaptic weight adaptation:

$$MSE_{ad} = \frac{1}{2n_{TB+}} \sum_{i=1}^{n_{TB+}} [1 - y(\mathbf{x}i | \mathbf{x}i \in S_{TB+}^{Train})]^2$$
$$+ \frac{1}{2n_{TB-}} \sum_{j=1}^{n_{TB-}} [1 + y(\mathbf{x}j | \mathbf{x}j \in S_{TB-}^{Train})]^2 \qquad (2)$$

where the variables n_{TB+} and n_{TB-} represent the cardinality of the subsets S_{TB+}^{Train} and S_{TB-}^{Train}, which correspond to positive and negative cases in the training set, respectively. This objective function forces a balanced reduction of the prediction errors related to positive and negative cases during MLP training. In this work, this performance index also controlled the number of training iterations in order to avoid overtraining,[22] according to the early stop procedure,[22] but was estimated using the testing set in this task.

To avoid local minima,[22] 10 MLP models were produced, each one considering random synaptic weights following a zero-mean Gaussian distribution with variance equal to 0.05. The model showing the highest area under the receiver operating characteristic curve (AUC)[35] for the testing set was selected for this analysis.

The iART network is a competitive clustering algorithm that automatically produces hyperspherical clusters to enclose data.[30] A vector of centre coordinates (c_i) and a vigilance radius (r) define each cluster. Usually, for simplicity, all groups share the same vigilance radius.

In order to evaluate whether a given input data $\mathbf{x}$ belongs to an arbitrary cluster with a centre $\mathbf{c_i}$, the iART model uses the following pertinence function:

$$d_i = ||\mathbf{x} - \mathbf{c}_i||^2 \qquad (3)$$

This quantity corresponds to the distance from data to cluster centre. If this value is lower than r, the data is inside the hypersphere associated with this group, thus is classified as belonging to it. If more than one group satisfies this criterion, the sample is assigned to the group with lowest value of d^i (the closest one).

During the iART training phase, dataset cases should be presented randomly to the network. For each one, the algorithm verifies whether the current input $(\mathbf{x})$ belongs to any of the already identified groups. If negative, a new group is created having this sample as the centre. Otherwise, the centre of the winner group $(\mathbf{c}_j)$ is updated according to the following equation:

$$\mathbf{c}_j(k) = \mathbf{c}_j(k-1) + \eta(k)[\mathbf{x} - \mathbf{c}_j(k-1)] \qquad (4)$$

where the variable η is the learning rate factor. It can be shown that if the value of η is progressively reduced during training, the vector $(\mathbf{c}_j)$ asymptotically converges to its corresponding cluster centre. In this work, this factor was adjusted as follows:

$$\eta(k) = \gamma\eta(k-1) \qquad (5)$$

where the constant γ is the learning decay factor, here chosen as 0.99.

With the aim of providing a risk assessment, the vigilance radius of the iART network was tuned to result in three groups, labelled as low, medium, and high risk, according to the prevalence of TB observed in each one.[30] This model was developed using all dataset cases. The existence of significant signs and symptoms for each group was verified using Fisher's test.[36]

The development and evaluation of MLP and iART models was conducted using self-made codes in Octave environment.[37] The resulting models were also implemented in C language, only considering the operating phase (no learning), and were integrated into a friendly Web-based user interface in order to make the system available for clinical use.

Other classification techniques, such as classification and regression trees (CART),[38] MLR, and support vector machines (SVM)[39] were evaluated as alternatives to MLP. The mid-p McNemar test was used to verify the effectiveness of the MLP model as compared to this set of algorithms.[40] The McNemar test is recommended for the comparison of two models produced using hold-out, thus considers only one testing set, and compares the proportion of classification errors associated with each one.[41]

Three kernel functions were considered for SVM machines: polynomial, radius basis functions (RBF), and linear. Optimal SVM parameters were defined using cross-validation and considered the AUC as the figure of merit. These methods were produced and evaluated using the same pair of sets used in the MLP development.

Model performance was assessed using the following indices: sensitivity, specificity, accuracy, and AUC. For the first three indices, the confidence intervals (w^-, w^+) were estimated using the Wilson score interval method,[27] given by:

$$(w^-, w^+) \equiv (\hat{p} - z_{\alpha/2}\hat{s}, \hat{p} + z_{\alpha/2}\hat{s}) \qquad (6)$$

with:

$$\hat{p} = \left(p + \frac{z_{\alpha/2}}{2n}\right) \Big/ \left(1 + \frac{z_{\alpha/2}^2}{n}\right) \qquad (7)$$

$$\hat{s} = \sqrt{\frac{p(1-p)}{n} + \frac{z_{\alpha/2}^2}{4n^2}} \Big/ \left(1 + \frac{z_{\alpha/2}^2}{n}\right) \qquad (8)$$

$$p = \frac{m}{n} \qquad (9)$$

Here, the variables m and n, which define the proportion p, are dependent on the performance index taken into consideration. Such values were also evaluated considering the testing set. The critical value $Z_{\alpha/2}$ was defined to produce 95% confidence intervals (CI).

In the case of the receiver operating characteristic curve (ROC), the CIs (AUC^-, AUC^+) were produced using the modified Wald interval with continuity correction,[42] as follows:

$$(AUC^-, AUC^+) \equiv (\widehat{AUC} + 1/(2n) - z_{\alpha/2}\hat{s}, \widehat{AUC} + 1/(2n) + z_{\alpha/2}\hat{s}) \quad (10)$$

$$\hat{s} = \sqrt{\frac{\widehat{AUC}(1-\widehat{AUC})}{0.75n-1}} \qquad (11)$$

where the number of testing set events is given by n.

2.3. Ethics statement

This study was approved by the Institutional Review Board of the Hospital Universitário Clementino Fraga Filho/Faculdade de Medicina (HUCFF/FM) (n. 060/1999). Patients included in this study gave their written informed consent.

Table 1

Descriptive analysis of the signs and symptoms, as selected by TB experts, in the 136 subjects suspected to have smear-negative pulmonary tuberculosis attending an outpatient clinic at a teaching hospital[a]

		PTB Mean (95% CI)		Non-PTB Mean (95% CI)		t-test p-value	ULR (p-Value)	MLR (p-Value)
Age		37.39 (33.39–41.39)		47.91 (44.15–51.67)		$p < 0.001$	<0.001	<0.001
		PTB		Non-PTB		OR (95% CI)	ULR (p-Value)	MLR (p-Value)
		n	%	n	%			
Cough	Yes	57	97.0	65	86.0	4.82 (1.03–22.68)	0.039	0.001
	No	2	3.0	11	14.0			
Haemoptysis	Yes	15	27.3	16	22.2	1.31 (0.58–2.96)	0.510	0.170
	No	40	72.7	56	77.7			
Night sweats	Yes	27	48.0	31	41.0	1.32 (0.66–2.65)	0.424	0.333
	No	29	52.0	44	59.0			
Fever	Yes	34	61.0	32	44.0	1.98 (0.97–4.02)	0.059	0.238
	No	22	39.0	41	56.0			
>10% weight loss	Yes	20	37.0	23	32.0	1.25 (0.59–2.63)	0.527	0.523
	No	34	63.0	49	68.0			
Dyspnoea	Yes	13	24.0	27	37.0	0.54 (0.25–1.18)	0.147	0.223
	No	41	76.0	46	63.0			
Anorexia	Yes	17	32.0	23	32.0	0.99 (0.47–2.13)	0.940	0.794
	No	37	68.0	50	68.0			
Smoking	Yes	13	24.0	10	13.0	2.03 (0.81–5.06)	0.089	0.125
	No	41	76.0	64	87.0			
Extrapulmonary TB	Yes	1	1.9	2	2.7	0.72 (0.06–8.22)	0.188	0.131
	No	51	98.1	74	97.3			
Past hospitalization	Yes	6	11.0	22	30.0	0.30 (0.11–0.81)	0.035	0.028
	No	46	89.0	51	70.0			
AIDS (Caracas definition)	Yes	1	4.3	4	11.8	0.34 (0.03–3.27)	0.847	0.258
	No	22	95.7	30	88.2			

PTB, pulmonary tuberculosis; CI, confidence interval; ULR, univariate logistic regression; MLR, multivariate logistic regression; OR, odds ratio.

[a] Some variables data are missing.

3. Results

Table 1 summarizes the frequencies of the sign and symptoms (those selected by the TB experts) observed in the dataset and the p-values obtained by ULR and MLR. Large odds ratio CIs and high p-values were observed in most cases. According to the ULR results, anorexia and AIDS did not perform well as isolated predictors of TB, but when associated with other variables exhibited significantly lower p-values, thus may contribute to the diagnosis of PTB.

Due to the high p-values achieved by ULR and MLR, related in part to the small sample size, no strong level of association between signs and symptoms and TB outcome was observed. This fact offers a rationale for the use of a non-linear complex prediction model such as MLP for this problem, as it explores non-linear relationships between problem variables.

FSFS identified the following six variables for model construction: age, presence of cough, absence of dyspnoea, smoking, association with extrapulmonary TB, and previous hospitalization. BSFS selected eight variables, including those proposed by the

FSFS, plus fever and AIDS. However, MLP models based on these sets of variables exhibited lower generalization performance. Thus, all models analysed in this work considered the set of variables selected by the TB experts.

Table 2 describes the performance achieved by alternative computational intelligence methods applied as prognostic models for smear-negative PTB. The MLP model showed higher sensitivity (100%, 95% CI 78–100%) than the SVM linear model (86%, 95% CI 60–96%). The accuracy of the MLP model (88%, 95% CI 73–95%) was also higher than MLR (83%, 95% CI 67–92%). According to the mid-p McNemar test, the MLP and SVM linear models performed similarly, with some evidence of lower classification errors for MLP ($p = 0.1445$).

Figure 1 shows the ROC curves for CART, SVM employing linear kernel function (SVM linear), MLR, and MLP. The MLP model (0.918) achieved higher values for AUC than the SVM linear (0.796) and MLR (0.782) models.

Table 3 summarizes the performance of the iART model. In this case, the analysis considered all patients. Low, medium, and high risk clusters retained 56, 36, and 44 patients, with a TB prevalence

Table 2

Results of the models evaluated in terms of sensitivity, specificity, accuracy, and AUC and their associated 95% confidence intervals

	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)	AUC (95% CI)
CART	71% (45–88%)	65% (39–84%)	67% (51–81%)	0.702 (0.536–0.898)
Multivariate logistic regression	79% (53–93%)	85% (59–96%)[a]	83% (67–92%)	0.782 (0.633–0.960)
SVM polynomial	86% (60–96%)	60% (35–81%)	70% (54–83%)	0.782 (0.633–0.960)
SVM RBF	86% (60–96%)	75% (49–90%)	80% (63–90%)	0.793 (0.647–0.968)
SVM linear	86% (60–96%)	75% (49–90%)	80% (63–90%)	0.796 (0.651–0.970)
Multilayer perceptron network	100% (78–100%)[a]	80% (54–93%)	88% (73–95%)[a]	0.918 (0.824–1.000)[a]

CI, confidence interval; AUC, area under the receiver operating characteristic curve; CART, classification and regression trees; SVM, support vector machines; RBF, radius basis functions.
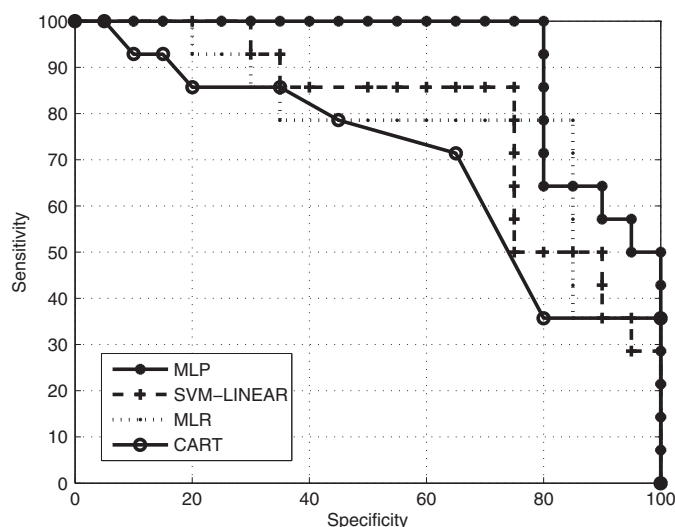
[a] Highest values.

Figure 1. ROC curves for the classification and regression tree (CART), support vector machine (SVM) linear, multivariate logistic regression (MLR), and multilayer perceptron (MLP) models.

of 33.9%, 41.7%, and 56.2%, respectively. Lower values for age were identified in the high-risk group (p = 0.035), and no difference in age was observed between the low-risk and medium-risk groups (p = 0.523), irrespective of TB outcome. In the low-risk group, a prevalence higher than 50% among positive PTB patients was observed only for cough. For the medium-risk group, as well as cough, this criterion was also satisfied by fever, weight loss, and anorexia, while for the high-risk group it was satisfied only by fever and night sweats. Among the three risk groups, Fisher's test identified significant differences for the prevalence of the following signs and symptoms: night sweats, fever, weight loss, and anorexia (all p < 0.001).[36] Except for fever, for which the medium-risk and high-risk groups shared the same prevalence according to statistical tests, loss of weight and anorexia represented significant signs and symptoms in the case of the medium-risk group. Night sweats showed a similar behaviour in the case of the high-risk group.

4. Discussion

Using only clinical standardized interviews, the ANN-based predictive model for smear-negative PTB achieved a high

sensitivity (78–100%) and moderate specificity (54–93%). As compared to other studies involving only suspected smear-negative PTB cases and models not based on ANNs, the sensitivity achieved by the present model was higher than those of Soto et al.[9] (sensitivity of 29.9% and specificity of 95.4%) and Siddiqi et al.[10] (sensitivity of 59% and specificity of 86%). The clinical score associated with radiological and laboratory tests proposed by Alavi-Naini et al. achieved a sensitivity of 94% and specificity of 74%.[11] Following a similar approach, the scores of Swai et al.[7] and Aguiar et al.[16] showed sensitivity values ranging from 38% to 71% and specificity from 58% to 76%.

The WHO algorithms, which include clinical work-up and chest CXR analysis, have shown good performance among smear-negative PTB cases. Reported sensitivity values have ranged from 58.8% to 95%, and specificity from 79.4% to 98%.[8,13,43] Thus, only using signs and symptoms, the present model performs similarly to other scores that include CXR.

Regarding screening models for PTB based on MLP, the present clinical score using 12 variables has an accuracy (88%) similar to those reported by Orhan et al.[19] (38 variables and accuracy of 93.3%), El-Solh et al.[17] (21 variables and accuracy of 92%), and Elveren and Yumuşak[18] (38 variables and accuracy of 94.9%). As expected for a screening approach, the sensitivity in the present study was higher than that reported for new diagnostic tests such as Xpert MTB/RIF (67%) in patients with suspected smear-negative PTB, but the specificity was lower (85% vs. 98%).[5]

Furthermore, the significant signs and symptoms identified in the medium- and high-risk groups produced by the iART architecture coincide with those identified as relevant for the diagnosis of PTB.[4,9,12,14] Risk group assignment may be useful in clinical practice, as additional approaches such as CXR, culture, and molecular techniques may be expedited for subjects classified as belonging to the medium- or high-risk groups.

Currently, CXR, Xpert MTB/RIF, and sputum culture are recommended for all presumed PTB patients attending hospitals. CXR has shown high sensitivity in prevalence surveys, but is restricted to hospitals and reference centres.[2,3] Wisnivesky et al. reported that the use of chronic symptoms, fever, and upper lobe abnormalities on CXR to diagnose PTB may result in sensitivity values ranging from 81% to 100%, and specificity from 19% to 84%.[14] Swai et al. analysed the accuracy of evaluating those with a chronic cough and any abnormal CXR and found sensitivity and specificity values of 38.1% and 74.5%, respectively.[7] Sputum culture, however, confirmed positive TB cases in only 38.1% of patients who had started anti-TB drugs. Swindells et al. evaluated the standard of

Table 3
Description of risk-group signs and symptoms according to PTB diagnosis

	PTB			Non-PTB		
	Low	Medium	High	Low	Medium	High
Age	39.21 (31.60–46.81)	41.87 (33.55–50.18)	33.32 (27.23–39.41)	47.27 (42.80–51.74)	52.33 (45.14–59.52)	41.95 (36.85–47.24)
Signs and symptoms	PTB			Non-PTB		
	Low	Medium	High	Low	Medium	High
Cough	89.5% (17/19)[a]	100.0% (15/15)[a]	100.0% (25/25)[a]	73.0% (27/37)	100% (21/21)	94.4% (17/18)
Haemoptysis	31.6% (6/19)[a]	33.3% (5/15)[a]	19.0% (4/21)[a]	16.2% (6/37)	15.0% (3/20)	46.7% (7/15)
Night sweats	0% (0/19)	42.9% (6/14)	91.3% (21/23)	5.4% (2/37)	71.4% (15/21)	82.4% (14/17)
Fever	10.5% (2/19)	73.3% (11/15)[a]	95.5% (21/22)[a]	8.1% (3/37)	75.0% (15/20)	87.5% (14/16)
Weight loss	10.5% (2/19)	85.7% (12/14)	28.6% (6/21)	10.8% (4/37)	70.0% (14/20)	33.3% (5/15)
Absence of dyspnoea	16.7% (3/18)[a]	26.7% (4/15)[a]	28.6% (6/21)[a]	13.5% (5/37)	47.6% (10/21)	80.0% (12/15)
Anorexia	5.6% (1/18)	100.0% (15/15)	4.8% (1/21)	2.7% (1/37)	100% (21/21)	6.7% (1/15)
Smoking	5.6% (1/18)[a]	28.6% (4/14)[a]	36.4% (8/22)[a]	10.8% (4/37)	14.3% (3/21)	18.8% (3/16)
Previous hospitalization	5.9% (1/17)[a]	21.4% (3/14)[a]	9.5% (2/21)[a]	24.3% (9/37)	33.3% (7/21)	40.0% (6/15)
HIV	0% (0/6)[a]	14.3% (1/7)[a]	0.0% (0/21)[a]	11.1% (2/18)	12.5% (1/8)	12.5% (1/8)

PTB, pulmonary tuberculosis.
[a] Groups showing no statistically significant difference in the observed prevalence values at the 5% confidence level.

care screening in 801 presumed TB subjects infected with HIV.[43] In their conclusion, the authors suggested that a more sensitive diagnostic approach is required for HIV-infected patients, since the screening procedure analysed showed a sensitivity of 54% (95% CI 40–67%) and specificity of 76% (95% CI 72–80%).

Recently, Creswell et al. described barriers to the programmatic implementation of Xpert MTB/RIF testing in nine countries.[6] Moreover, Theron et al. suggested that innovative approaches for the evaluation of new diagnostic tests should include patient outcome and strategies to improve access to health systems.[44] As proposed by Van't Hoog et al., a screening strategy based on a sensitive, but not necessarily highly specific rapid test, such as CXR or a prognostic model, could be useful to select patients for Xpert MTB/RIF and might result in a more affordable diagnosis framework.[45]

This study has some limitations. First, a small number of smear-negative PTB patients were enrolled and there was a high proportion of active TB among those attending an outpatient clinic at the hospital. Second, the models should be evaluated in terms of external generalizability, thus this approach warrants further clinical validation in other settings with different TB and HIV prevalences.

This study evaluated the performance of a simple and low-cost screening approach based on neural network models for further diagnosing smear-negative PTB. In these cases, the use of new rapid molecular tests for diagnosis is usually expensive, as multiple specimens may be required to achieve an adequate sensitivity. Outpatients attending a teaching hospital in a high TB prevalence setting formed the study population.

The proposed diagnosis system serves as a TB prognostic and risk assessment tool. As a prognostic, it may be a useful screening approach for the use of molecular tests, such as the Xpert MTB/RIF assay, while as a risk assessment tool it allows the use of complementary tests to be expedited for higher risk patients. Furthermore, the TB diagnosis chain can easily incorporate this DSS through the implementation of its Web-based interface, or a custom application for smartphones and/or tablets could easily be developed for this purpose.

The system has been shown to be adequate for screening, since it achieved high sensitivity and moderate specificity, and also identified risk factors coherent with clinical practice. However, a larger patient sample should be used for further validation. The proposed system also needs to be trialled in addition to new molecular tests and/or CXR using a standardized protocol in different regions to evaluate the clinical impact and cost-effectiveness of its incorporation into the clinical routine.

Acknowledgements

References

1. World Health Organization. Global tuberculosis report 2015. Geneva: WHO; 2015 Available at: http://apps.who.int/iris/bitstream/10665/191102/1/9789241565059_eng.pdf?ua=1http://www.who.int/tb/publications/global_report/indicators_global_and_ regional_summaries.pdf (accessed October 20, 2015)
2. World Health Organization. Systematic screening for active tuberculosis: principles and, recommendations. WHO/HTM/TB/2013.04. Geneva: WHO; 2013.
3. World Health, Organization. Improving the diagnosis and treatment of smear-negative pulmonary and extrapulmonary tuberculosis among adults and adolescents: recommendations for HIV-prevalent and resource-constrained settings. Geneva: WHO; 2007, Available at: http://whqlibdoc.who.int/hq/2007/WHO_HTM_TB_2007.379_eng.pdf (accessed November 7, 2014)
4. Castro CB, Costa PA, Ruffino-Netto A, Maciel EL, Kritski AL. Assessment of a clinical score for screening suspected pulmonary tuberculosis cases. *Rev Saúde Pública* 2011;**45**:1110–6.
5. Walusimbi S, Bwanga F, De Costa A, Haile M, Joloba M, Hoffner S. Meta-analysis to compare the accuracy of GeneXpert, MODS and the WHO 2007 algorithm for diagnosis of smear-negative pulmonary tuberculosis. *BMC Infect Dis* 2013;**13**:507. http://dx.doi.org/10.1186/1471-2334-13-507
6. Creswell J, Codlin AJ, Andre E, Micek MA, Bedru A, Carter EJ, et al. Results from early programmatic implementation of Xpert MTB/RIF testing in nine countries. *BMC Infect Dis* 2014;**14**:2. http://dx.doi.org/10.1186/1471-2334-14-2
7. Swai HF, Mugusi FM, Mbwambo JK. Sputum smear negative pulmonary tuberculosis: sensitivity and specificity of diagnostic algorithm. *BMC Res Notes* 2011;**4**:475. http://dx.doi.org/10.1186/1756-0500-4-475
8. Alamo ST, Kunutsor S, Walley J, Thoulass J, Evans M, Muchuro S, et al. Performance of the new WHO diagnostic algorithm for smear-negative pulmonary tuberculosis in HIV prevalent settings: a multisite study in Uganda. *Trop Med Int Health* 2012;**17**:884–95. http://dx.doi.org/10.1111/j.1365-3156.2012.03003.x
9. Soto A, Solari L, Gotuzzo E, Acinelli R, Vargas D, Van der Stuyft P. Performance of an algorithm based on WHO recommendations for the diagnosis of smear-negative pulmonary tuberculosis in patients without HIV infection. *Trop Med Int Health* 2011;**16**:424–30. http://dx.doi.org/10.1111/j.1365-3156.2010.02715.x
10. Siddiqi K, Walley J, Khan MA, Shah K, Safdar N. Clinical guidelines to diagnose smear-negative pulmonary tuberculosis in Pakistan, a country with low-HIV prevalence. *Trop Med Int Health* 2006;**11**:323–31.
11. Alavi-Naini R, Cuevas LE, Squire SB, Mohammadi M, Davoudikia AA. Clinical and laboratory diagnosis of the patients with sputum smear-negative pulmonary tuberculosis. *Arch Iran Med* 2012;**15**:22–6. 012151/AIM.007.
12. Getahun H, Kittikraisak W, Heilig CM, Corbett EL, Ayles H, Cain KP, et al. Development of a standardized screening rule for tuberculosis in people living with HIV in resource-constrained settings: individual participant data meta-analysis of observational studies. *PLoS Med* 2011;**8**:e1000391. http://dx.doi.org/10.1371/journal.pmed.1000391
13. Koole O, Thai S, Khun KE, Pe R, van Griensven J, Apers L, et al. Evaluation of the 2007 WHO guideline to improve the diagnosis of tuberculosis in ambulatory HIV-positive adults. *PLoS One* 2011;**6**:e18502. http://dx.doi.org/10.1371/journal.pone.0018502
14. Wisnivesky JP, Serebrisky D, Moore C, Sacks HS, Iannuzzi MC, McGinn T. Validity of clinical prediction rules for isolating inpatients with suspected tuberculosis. A systematic review. *J Gen Intern Med* 2005;**20**:947–52. http://dx.doi.org/10.1111/j.1525-1497.2005.0185.x
15. Solari L, Acuna-Villaorduna C, Agapito J, Soto A, Perez F, Samalvides F, et al. A clinical prediction rule for pulmonary tuberculosis in emergency departments. *Int J Tuberc Lung Dis* 2008;**12**:619–24.
16. Aguiar FS, Almeida LL, Ruffino-Netto A, Kritski AL, Mello FC, Werneck GL. Classification and regression tree (CART) model to predict pulmonary tuberculosis in hospitalized patients. *BMC Pulm Med* 2012;**12**:40. http://dx.doi.org/10.1186/1471-2466-12-40
17. El-Solh AA, Hsiao CB, Goodnough S, Serghani J, Grant BJ. Predicting active pulmonary tuberculosis using an artificial neural network. *Chest* 1999;**116**:968–73.
18. Elveren E, Yumuşak N. Tuberculosis disease diagnosis using artificial neural network trained with genetic algorithm. *J Med Syst* 2011;**35**:329–32.
19. Orhan ER, Temurtas F, Tanrikulu AC. Tuberculosis disease diagnosis using artificial neural networks. *J Med Syst* 2010;**34**:299–302.
20. Seixas JM, Faria J, Souza Filho JB, Vieira AF, Kritski A, Trajman A. Artificial neural network models for diagnosis support of pleural tuberculosis in adult patients. *Int J Tuberc Lung Dis* 2013;**17**:682–6. http://dx.doi.org/10.5588/ijtld.12.0829
21. Eftekhar B, Mohammad K, Ardebili HE, Ghodsi M, Ketabchi E. Comparison of artificial neural network and logistic regression models for prediction of mortality in head trauma based on initial clinical data. *BMC Med Inform Decis Mak* 2005;**5**. http://dx.doi.org/10.1186/1472-6947-5-3
22. Haykin S. Neural network and learning machines, 3rd ed., Pearson; 2008.
23. Mello FC, Bastos LGV, Soares SLM, Rezende VMC, Conde MB, Chaisson EL, et al. Predicting active smear negative pulmonary tuberculosis with classification trees and logistic regression models. *BMC Public Health* 2006;**6**:43. http://dx.doi.org/10.1186/1471-2458-6-43
24. Weniger BG, Quinhões EP, Sereno AB, de Perez MA, Krebs JW, Ismael C, et al. A simplified surveillance case definition of AIDS derived from empirical clinical data. The Clinical AIDS Study Group and the Working Group on AIDS Case Definition. *J Acquir Immune Defic Syndr* 1992;**5**:1212–23.
25. Guidelines for TB control. Brazil: Ministry of Health; 2010. Available at: http://www.saude.gov.br/tuberculose.(accessed August 24, 2013)
26. World Health Organization. Laboratory services in tuberculosis control microscopy—part II. Geneva: WHO; 1998, Available at: http://whqlibdoc.who.int/hq/1998/WHO_TB_98.258(part2).pdf (accessed June 20, 2014)
27. Agresti A. Categorical data analysis, 3rd ed., Wiley; 2012.
28. Saeys Y, Inza I, Larrañaga P. A review of feature selection in bioinformatics. *Bioinformatics* 2007;**23**:2507–17.
29. Pudil P, Novovicova J, Kittler J. Floating search methods in feature selection. *Pattern Recognition Letters* 1994;**15**:1119–25. http://dx.doi.org/10.1016/0167-8655(94)90127-9
30. Souza Filho JB, de Seixas JM. Statistical data selection for passive sonar signal neural classifiers (in Portuguese). *VII Brazilian Conference on Neural Networks.* 2005.
31. Theodoridis S, Koutroumbas K. Pattern recognition, 4th ed., Academic Press; 2009
32. Olvera-López JA, Carrasco-Ochoa JA, Martinez-Trinidad JF, Kittler J. A review of instance selection methods. *Artificial Intelligence Review* 2010;**34**:133–43.

33. Riedmiller M, Braun H. A direct adaptive method for faster backpropagation learning: the RPROP algorithm. *Proceedings of the IEEE International Conference on Neural Networks*. March 28–April 1 1993. p. 586–91.
34. He H, Garcia EA. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering* 2009;**21**:1263–84.
35. Hand DJ. Assessing the performance of classification methods. *International Statistical Review* 2012;**80**:400–14. http://dx.doi.org/10.1111/j.1751-5823.2012.00183.x
36. Sheskin DJ. Handbook of parametric and nonparametric statistical procedures, 5th ed., Chapman and Hall; 2011.
37. Eaton JW, Bateman D, Hauberg S. GNU Octave manual, 3rd ed., Network Theory Ltd; 2008.
38. Breiman L, Friedman J, Stone C, Olshen R. Classification and regression trees. Belmont CA: Wadsworth Statistics; 1984.
39. Schölkopf B, Smola AJ. Learning with kernels—support vector machines, regularization, optimization and beyond. MIT Press; 2001.
40. Fagerland MW, Lydersen S, Laake P. The McNemar test for binary matched-pairs data: mid-*p* and asymptotic are better than exact conditional. *BMC Med Res Methodol* 2013;**13**. http://dx.doi.org/10.1186/1471-2288-13-91

41. Dietterich TG. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation* 1998;**10**:1895–923.
42. Kottas M, Kuss O, Zapf A. A modified Wald interval for the area under the ROC curve (AUC) in diagnostic case–control studies. *BMC Med Res Methodol* 2014;**14**. http://dx.doi.org/10.1186/1471-2288-14-26
43. Swindells S, Komarow L, Tripathy S, Cain KP, MacGregor RR, Achkar JM, et al. Screening for pulmonary tuberculosis in HIV-infected individuals: AIDS Clinical Trials Group Protocol A5253. *Int J Tuberc Lung Dis* 2013;**17**:532–9. http://dx.doi.org/10.5588/ijtld.12.0737
44. Theron G, Peter J, Dowdy D, Langley I, Squire SB, Dheda K. Do high rates of empirical treatment undermine the potential effect of new diagnostic tests for tuberculosis in high-burden settings? *Lancet Infect Dis* 2014;**14**:527–32. http://dx.doi.org/10.1016/S1473-3099(13)70360-8
45. Van't Hoog AH, Cobelens F, Vassall A, van Kampen S, Dorman SE, Alland D, et al. Optimal triage test characteristics to improve the cost-effectiveness of the Xpert MTB/RIF assay for TB diagnosis: a decision analysis. *PLoS One* 2013;**8**: e82786. http://dx.doi.org/10.1371/journal.pone.0082786