

An Automated Screening System for Tuberculosis

Ricardo Santiago-Mozos, *Member, IEEE*, Fernando Pérez-Cruz, *Senior Member, IEEE*,
Michael G. Madden and Antonio Artés-Rodríguez *Senior Member, IEEE*,

Abstract—Automated screening systems are commonly used to detect some agent in a sample and take a global decision about the subject (e.g. ill/healthy) based on these detections. We propose a Bayesian methodology for taking decisions in (sequential) screening systems that considers the false alarm rate of the detector. Our approach assesses the quality of its decisions and provides lower bounds on the achievable performance of the screening system from the training data. In addition, we develop a complete screening system for sputum smears in tuberculosis diagnosis, and show, using a real-world database, the advantages of the proposed framework when compared to the commonly used *count detections and threshold* approach.

Index Terms—Tuberculosis, Automated screening, Decision making, Sequential analysis, Bayesian.

R. Santiago-Mozos is with the Department of Signal Theory and Communication, Universidad Rey Juan Carlos, Madrid, Spain (e-mail: rsmozos@gmail.com).

F. Pérez-Cruz is with the Department of Signal Theory and Communication, Universidad Carlos III de Madrid, Madrid, Spain. He is also a Machine Learning Researcher at Amazon (e-mail: fernando@tsc.uc3m.es).

A. Artés-Rodríguez is with the Department of Signal Theory and Communication, Universidad Carlos III de Madrid, Madrid, Spain (e-mail: antonio@tsc.uc3m.es).

M. G. Madden is with the College of Engineering and Informatics, National University of Ireland, Galway, Ireland (e-mail: michael.madden@nuigalway.ie).

I. INTRODUCTION

Tuberculosis is a contagious illness caused by the *Mycobacterium Tuberculosis*, better known as Koch bacillus. In 2011, there were an estimated 9 million incident cases of Tuberculosis (TB), 0.99 million deaths from TB among HIV-negative people and an additional 0.43 million TB deaths among HIV-positive people (classified as HIV deaths in the International Statistical Classification of Diseases) [1].

To diagnose TB, the following methods are used: thorax radiography [2]; biological culture [2], [3]; the Mantoux (tuberculin sensitivity/skin) test [4]; interferon- γ tests [5]; amplified nucleic acids-based tests [6] (which allow lab-on-chip platforms [7], [8]); and sputum smear microscopy [8], [9]. Of these tests, the two most commonly used to check whether a subject is contagious are biological culture, and sputum smear microscopy [10]. Biological culture is generally accepted as the *gold standard*, but the test can take up to four weeks [11]. Sputum smear microscopy is typically used to make a decision based on a quick examination. However, smear microscopy has a low sensitivity and is expensive, as an expert microbiologist is needed to diagnose the sputum. Therefore, automatic systems to aid the diagnosis have been proposed [12]–[17] for both auramine-stained and

Ziehl-Neelsen-stained sputum smears (fluorescence and conventional microscopy, respectively). These systems are focused on the classification of the TB bacillus by applying segmentation techniques to the images and report high sensitivities when classifying segmented objects as bacillus (e.g. 94% [13] or 95% [14]), which clearly renders them useful to assist the expert by finding objects that are likely to be bacilli, which accelerates the time needed to process slides. But, despite having good accuracy in identifying the bacillus in small object databases, their false alarm rate (1-specificity) (e.g. 1% [13] or 2% [14]) is too high to analyze subject images where the proportion of bacilli is low.

The examples of automated systems considered above have reported to need 3 detections or more [13] or 2 or more [14] to declare a TB-positive smear. Roughly, taking into account their false alarm rate, these systems declare a false bacillus per 100 objects examined. If bacilli are scarce, TB-smear might be declared positive after only analyzing non-bacilli objects. Bacilli scarcity is usually the case as World Health Organization recommends that at least 100 fields (non overlapped images from the sample) have to be examined and 1 bacillus in 100 images is enough to declare a positive sputum smear [18], [19]. The problem of the false alarm rate has been pointed out in [20], who devised a sequential test to manage the false alarm rate by augmenting the number of images until a performance criterion is met, which continued our previous approaches to this problem [21].

As no classifier is likely to have specificity equal to one, two important questions arise: can we trust these screening systems for making decisions about smears? what is the quality of these decisions based on the bacilli classifier performance (specificity and sensitivity)? The main contribution of this paper is filling the

gap that exists between classifying bacilli and subjects, i.e., between classifying objects as bacilli and classifying smears as TB-positive. We propose an automatic screening system for TB and evaluate its performance in a real-life database. We address both the bacilli classifier (detector) and the subject classifier. Our system sequentially analyses images from a subject until enough bacilli are detected or enough images are analyzed without detecting any bacilli according to the decision quality requirements. The number of images to be examined and the maximum attainable performance of the subject classifier are related to the bacilli classifier performance in the training set, and each subject (smear) diagnosis is given a quality measure. The bacilli detector architecture is constrained for real-time amenable operation. The subject classifier is automatically adjusted given the training set and it has only three parameters: the thresholds for stopping the analysis for each kind of decision (TB-positive, TB-negative) and the a priori probability of being TB-negative, whose setting is straightforward. Its design is addressed from a Bayesian point of view and compared with the commonly used count detections and threshold approach.

II. MATERIALS AND METHODS

A. Image Acquisition & Database

We have access to auramine-stained images of the subjects' sputum that were acquired from a 20 \times fluorescence Nikon microscope with a Prior automatic slide loader and a Retiga 2000R camera. We have around 300 1200 \times 1600 RGB images (8 bits per color) using manual focus for each one of 61 subjects, 12 of which are TB-positive (have TB bacilli in the sputum). In addition, we have access to around 90 images per subject for 12 TB-positive subjects acquired with an autofocus algorithm.

All the images are unlabeled. The type of subject (TB-negative or TB-positive) was confirmed by culture (even though culture is an imperfect tool to determine the existence of bacilli in the extreme low range, as it uses a decontamination procedure that may result in killing some bacilli [19]).

In addition, we have around 5000 labeled patches with bacilli from previous subjects. This database provides help to identify the bacilli in our automatic detection algorithm. Unfortunately, these patches are low quality, some only containing partial bacillus or several parts of them, or they are overlapping, or even controversial images of bacillus (not all doctors agree if the mass in the patch is a bacillus). After a careful preprocessing we are left with 1258 “mug shots”. We also have non-bacillus patches that contain the objects that would pass the Bacilli Preprocessor (see below) and have been collected from training TB-negative subjects.

We randomly split the TB-negative subjects into training and testing sets fixing a 30% for testing purposes. For the TB-positive subjects we randomly take 7 subjects for testing from the ones that have been manually acquired and 6 from the automatically acquired (we put a larger percentage of TB-positive subjects in the test set because we have other database for identifying bacilli and we want to have a similar number of positive and negative test subjects). Finally, 45 subjects are selected for training (34 TB-negative and 11 TB-positive) and 28 for testing (15 TB-negative and 13 TB-positive).

Solving this problem accurately is extremely difficult: first, because we have a global decision (whether the subject culture has TB bacilli) that might not be error free, but we need to seek for individual bacillus in the acquired images, which are not labeled; furthermore, a dead bacillus in a TB-negative subject (e.g., cured or with latent TB) looks as the same as a bacillus; and

an TB-positive subject might not have any bacilli in the acquired images (a little as one bacilli per 100 images is considered as enough to declare a positive sputum smear). We need to be able to extrapolate from bacilli finding to subject classification with high sensitivity and specificity.

B. Bacilli Classifier

The bacilli classifier has the cascade structure shown in Fig. 1. It has two parts: the bacilli preprocessor, which is fast and discards the patches that can not contain a bacilli (see below), saving processing time; and the main processing for bacilli, which performs object segmentation and analyzes these objects. An object is only declared as a bacillus if it reaches the pixels classifier and is declared bacillus.

1) *Bacilli Preprocessor*: The image is divided into patches as shown in Fig. 2. We use several overlapping grids to avoid the situation where a bacillus is located on the edge of a patch; ultimately, a bacillus detection is assigned to the patch of the first grid where most of its body lies. For the preprocessor we end up with a simple classifier that focuses on the green component of RGB images because the auramine tincture is basically green. It sends to the next step any patch whose minimum green-color value exceeds a threshold T_1 , and has P or more pixels exceeding a threshold T_2 , and discards the rest of patches. T_1 , T_2 and P are selected to ensure that all bacilli on the training set are not discarded.

2) *Main Processing for Bacilli*: In the patches selected by the Bacilli Preprocessor we perform object detection. We use Canny edge detection [22], using the green color for segmentation. Then, following [14] and [16], we extract a set of rotation and translation invariant features of each candidate object (see Table I) and classify the objects using them. This intermediate

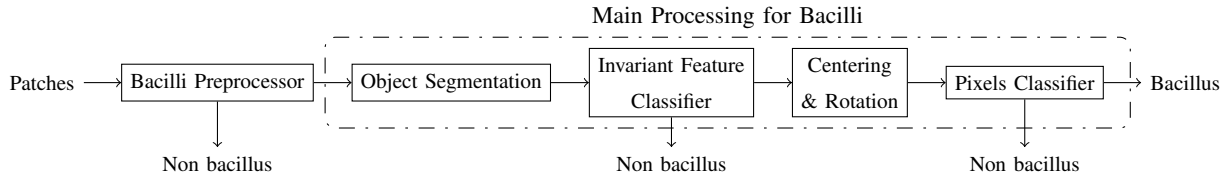


Fig. 1. Bacilli Classifier. This cascade represent the processing steps for each image patch in order to detect bacilli objects.

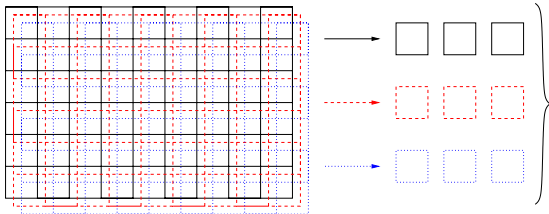


Fig. 2. Image division. Each grid outputs its patches to be analyzed but all detections are assigned to the continuous grid.

classifier is tuned for sensitivity close to 1. Finally, the

TABLE I

ROTATION AND TRANSLATION INVARIANT FEATURES.

ϕ_1	Object major axis length.
ϕ_2	Object minor axis length.
ϕ_3	$\phi_3 = \phi_1/\phi_2$.
ϕ_4	Object perimeter.
ϕ_5	Object area.
ϕ_6	$\phi_6 = \phi_4/\phi_5$.
ϕ_7-27	Object GPD invariant moments of [23].
ϕ_{28-30}	Object pixels mean (for each band).
ϕ_{31-33}	Object pixels standard deviation (for each band).
ϕ_{34-36}	Object pixels maximum value (for each band).
ϕ_{37-44}	Fourier descriptors for the boundary [24].

objects classified as *non bacillus* are discarded, and the objects classified as bacilli are rotated and centered, following the procedure used by [25]; and classified using their pixels as features. Both, the invariant feature and the pixels classifiers are implemented using a Support Vector Machine (SVM) [26], [27].

C. Subjects Classifier

The bacilli classifier provides an output $z_i \in \{0, 1\}$ for each analyzed patch that identifies whether a bacilli was detected ($z_i = 1$). We define $\mathbf{z} = \{z_1, z_2, \dots, z_K\}$ as the K outputs of the bacilli classifier for all the patches of the analyzed images of a test subject. The rate of bacilli declarations $\frac{1}{K} \sum z_i$ approximates the probability of declaring a bacillus in this test subject, which is the sum of the unknown probabilities of detection (P_D) and false alarm (P_{FA}) of the bacilli classifier in that subject.

We formalize the global decision about the smear as a binary hypothesis testing problem [28] that decides between two alternative hypotheses H_0 (TB-negative) and H_1 (TB-positive), which respectively state that an independent and identically distributed (i.i.d.) set of samples \mathbf{z} (the outputs of the bacilli classifier) has been generated according to a distribution $p(\mathbf{z}; \theta_0)$ or $p(\mathbf{z}; \theta_1)$. We model these samples using a Bernoulli distribution $p(z_i; \theta_\ell) = \theta_\ell^{z_i} (1 - \theta_\ell)^{1-z_i}$, where the success probabilities θ_0 and θ_1 are the probabilities of declaring a bacillus in a TB-negative and TB-positive smear, respectively.

We consider this binary decision when θ_0 and θ_1 are fixed but unknown, and they are described by two i.i.d. sets $\mathbf{x} = \{x_1, \dots, x_N\}$ and $\mathbf{y} = \{y_1, \dots, y_M\}$, respectively, drawn from $p(\mathbf{z}; \theta_0)$ and $p(\mathbf{z}; \theta_1)$. Here, \mathbf{x} (\mathbf{y}) are the outputs of the bacilli classifier for all the patches of the images of a prototype subject for the TB-negative (TB-positive) smear class. A prototype for each

class can be chosen as a typical subject or as a limiting subject (the one that is closest to the other class). We stress that we only know θ_0 (θ_1) through \mathbf{x} (\mathbf{y}) and therefore some uncertainty will always remain about θ_0 (θ_1) true value. We propose to address this decision problem following a Bayesian sequential approach.

In Bayesian statistics, variables and parameters are treated as random variables and our objective is to obtain the posterior distribution of the desired quantity conditioned on the observed variables (and integrating out the unknowns) [29]. Next, we detail our desired quantity, how to define the hypotheses and how to bound our confidence on the decision.

1) *Desired Quantity*: $p(H_0|\mathbf{z})$, is the probability of the test subject being TB-negative given the outputs of the bacilli classifier. We compute the posterior for hypothesis H_ℓ , $\ell = \{0, 1\}$ by applying Bayes rule:

$$p(H_\ell|\mathbf{z}) = \frac{p(\mathbf{z}|H_\ell)p(H_\ell)}{p(\mathbf{z}|H_0)p(H_0) + p(\mathbf{z}|H_1)p(H_1)}, \quad (1)$$

where $p(\mathbf{z}|H_\ell)$ are the likelihoods for each hypothesis and $p(H_\ell)$ their a priori probabilities (i.e. the probability of a generic patient being TB-negative or TB-positive).

2) *Hypotheses Definition and Likelihoods*: If θ_0 and θ_1 were known:

$$p(\mathbf{z}|H_\ell) = P(\mathbf{z}|\theta_\ell) = \prod_i \theta_\ell^{z_i} (1 - \theta_\ell)^{1-z_i}.$$

However, the only information we have for each hypothesis are the sets \mathbf{x} and \mathbf{y} . Therefore, the likelihood should be computed, by integration of the unknown random variable θ_0 , as:

$$p(\mathbf{z}|H_0) = p(\mathbf{z}|\mathbf{x}) = \int p(\mathbf{z}|\theta_0)p(\theta_0|\mathbf{x}) d\theta_0, \quad (2)$$

where $p(\theta_0|\mathbf{x})$ is the posterior of θ_0 with respect to the known set \mathbf{x} :

$$p(\theta_0|\mathbf{x}) = \frac{p(\mathbf{x}|\theta_0)p(\theta_0)}{p(\mathbf{x})} = \frac{p(\mathbf{x}|\theta_0)p(\theta_0)}{\int p(\mathbf{x}|\theta_0)p(\theta_0) d\theta_0},$$

and the prior of θ_0 , $p(\theta_0)$, is selected as a beta distribution:

$$p(\theta_0) = \text{beta}(\theta_0|a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta_0^{a-1} (1-\theta_0)^{b-1},$$

where $\Gamma(\cdot)$ is the gamma function. The beta distribution is flexible and allows a non-informative prior by setting $a = b = 1$, which results in a uniform distribution between 0 and 1. Other values of a and b can be used to incorporate different prior knowledge. The beta distribution is also a conjugate prior for the Bernoulli distribution, which makes the selection of a beta prior convenient from an analytic point of view: $p(\theta_0|\mathbf{x})$ becomes a beta distribution $p(\theta_0|\mathbf{x}) = \text{beta}(\theta_0|a_0, b_0)$, where $a_0 = n_1 + a$; $b_0 = N - n_1 + b$; N is the number of samples of \mathbf{x} and n_1 is the number of “1”s (bacilli declarations) in \mathbf{x} . Finally, for any posterior of θ_0 with the form of $\text{beta}(\theta_0|a_0, b_0)$, the likelihood for H_0 becomes:

$$p(\mathbf{z}|H_0) = \frac{\Gamma(K - k_1 + b_0)\Gamma(a_0 + b_0)\Gamma(k_1 + a_0)}{\Gamma(a_0)\Gamma(b_0)\Gamma(K + a_0 + b_0)}, \quad (3)$$

where K and k_1 are respectively the number of samples and the number of “1”s in the test sample \mathbf{z} . An equivalent expression can be computed for $p(\mathbf{z}|H_1)$.

3) *Bounds for the Achievable $p(\mathbf{z}|H_0)$* : When the number of test samples tends to infinity, the likelihood $p(\mathbf{z}|\theta_0)$ tends to a Dirac delta function $\delta(\theta_0 - \theta^*)$, where θ^* is the true value of the success probability of the Bernoulli for the test sample. Therefore (2) simplifies to

$$p(\mathbf{z}|H_0)_{as} = \int p(\mathbf{z}|\theta_0)p(\theta_0|\mathbf{x}) d\theta_0 = p(\theta_0|\mathbf{x}) \Big|_{\theta_0=\theta^*},$$

where “ as ” stands for asymptotic. An analogous reasoning gives $p(\mathbf{z}|H_1)_{as} = p(\theta_1|\mathbf{y}) \Big|_{\theta_1=\theta^*}$ and by substituting

ing these values in (1) we attain

$$p(H_0|\mathbf{z})_{as} = \frac{p(H_0)p(\theta_0|\mathbf{x})\Big|_{\theta_0=\theta^*}}{\left(p(H_0)p(\theta_0|\mathbf{x}) + p(H_1)p(\theta_1|\mathbf{y})\right)\Big|_{\substack{\theta_0=\theta^* \\ \theta_1=\theta^*}}}. \quad (4)$$

$P(H_0|\mathbf{z})_{as}$ only depends on the a priori probabilities of the hypotheses, the sets of samples which characterize the hypotheses \mathbf{x} , \mathbf{y} and the prior distributions on θ_0 and θ_1 , and is limited by the initial uncertainty in the hypotheses no matter how many test samples are used. These bounds help to set the stopping threshold probabilities (see below).

The Bayesian test may be applied sequentially. We only need to use the recursive rule for the Gamma function, i.e. $\log \Gamma(n+1) = \log n + \log \Gamma(n)$, each time a new sample from \mathbf{z} is processed.

III. RESULTS

A. Bacilli Classifier

The bacilli classifier was developed using Matlab and LIBSVM [30] package for SVM. The bacilli classifier sensitivity is obtained by 10-fold cross-validation from the training labeled bacilli. It is found to be 73.53%. On the other hand, we can obtain the specificity from the complete analysis of 15 TB-negative test subjects and it works out as 99.999% (28 “false alarms” in 6288320 patches). Operation time of the bacilli classifier is around a second per image, which is the time needed for acquiring an image (operation is pipelined: when an image is being acquired, the bacilli classifier is processing the last one). If faster processing is needed, the Invariant Feature Classifier and the Pixel Classifier can be deployed as cascade classifiers [31] and/or the complexity of the SVM boundaries can be reduced [32], with almost no penalty in accuracy. If even faster processing is required, the sensitivities of the Bacilli Preprocessor and the

Invariant Feature Classifier can be reduced, which adds a penalty in the overall bacilli classifier sensitivity.

B. Subjects Classifier

When running the bacilli classifier over all the patches of the images of the training subjects, we obtain the bacillus declarations (number of patches that contain bacillus) summarized in Tables II and III for the TB-negative and the TB-positive, where “Subject ID” is the identifier of subject; “#patches” stands for the total number of patches available; “#bacilli” for the number of patches that contain a bacillus; and “ $p_{\text{decl. bac.}}$ ” for the observed probability of declaring a bacillus.

1) *Count and Threshold Approach*: This is the most common approach in the literature [8], [9], [19]. It establishes a threshold on the number of bacilli detections. Any smear with as many or more detections is declared TB-positive. After carefully considering the number of detections in the training subjects in Tables II and III) we choose to establish the threshold in 6 bacilli. Therefore, 2 out of 34 TB-negative training subjects will be misclassified and the maximum attainable specificity of the global decision should be lowered by 2/34. Also, 5 out of 11 TB-positive training subjects will be misclassified and the maximum attainable sensitivity should be lowered by 5/11. On the other hand, we fix the number of images to analyze to 300 (412800 patches) or as many as have been acquired when 300 images are not available. If we reach the threshold before analyzing 300 images, we stop early.

With this strategy (tables not shown) we get the same decisions than with the Bayesian approach (see below). For the TB-positive test subjects we get 7 out of 13 correct decisions. The test finishes early in the 7 subjects with 6 or more bacilli detections (subject 6 in Table IV has a total of 123 bacilli declarations in 151360 patches

TABLE II
TB-NEGATIVE TRAINING SUBJECTS. NUMBER OF PATCHES
DECLARED AS BACILLI, AND OBSERVED PROBABILITY OF
DECLARING A BACILLUS.

Subject ID	#patches	#bacilli	$p_{\text{decl. bac.}}$
30207	425184	0	0
30304	426560	0	0
30881	467840	0	0
30986	443072	0	0
31060	415552	0	0
31547	410048	0	0
32642	443072	0	0
32748	448576	0	0
51523	470592	1	2.125e-6
30261	456832	1	2.189e-6
30619	447200	1	2.236e-6
32550	436192	1	2.293e-6
30825	434816	1	2.300e-6
30911	421056	1	2.375e-6
31230	408672	1	2.447e-6
32955	401792	1	2.489e-6
30880	454080	2	4.405e-6
32688	445824	2	4.486e-6
30855	418304	2	4.781e-6
32750	418304	2	4.781e-6
31549	412800	2	4.845e-6
32111	410048	2	4.877e-6
31245	436192	3	6.878e-6
30262	432064	3	6.943e-6
29994	425184	3	7.056e-6
32497	410048	3	7.316e-6
31522	401792	3	7.467e-6
30783	392160	3	7.650e-6
30877	433440	4	9.228e-6
31934	429312	5	1.165e-5
30998	412800	5	1.211e-5
38841	225664	3	1.329e-5
30214	410048	10	2.439e-5
30663	392160	16	4.080e-5

TABLE III
TB-POSITIVE TRAINING SUBJECTS.

Subject ID	#patches	#bacilli	$p_{\text{decl. bac.}}$
7	100448	0	0
157512	194016	1	5.154e-6
31087	432064	3	6.943e-6
52707	400416	3	7.492e-6
48940	418304	6	1.434e-5
97649	90816	4	4.405e-5
8	111456	6	5.383e-5
32741	441696	57	1.290e-4
46989	390784	573	1.466e-3
10	155488	327	2.103e-3
9	121088	433	3.576e-3

but the Bayesian test finishes after considering 2752 patches); on the rest, the test runs until 300 images are examined (the number of bacilli detections for these later subjects are shown in Table IV [column #bacilli] but for subject 56036, where the number of bacilli detected in 300 images increases to 1). For the TB-negative test subjects we get 15 out of 15 correct decisions. The test runs until 300 images are analyzed (the bacilli detections for these subjects are the same than in Table V but for subject 36856 where the number of bacilli detected in 300 images increases to 3).

2) *Bayesian Approach*: In the Bayesian approach we can easily incorporate the information of all our training subjects. We propose to use a beta distribution as posterior for each class (TB-positive, TB-negative) and compute the maximum likelihood estimates of its parameters from the maximum likelihood estimates of the *declaring-a-bacillus* probabilities of the subjects that belong to that class. Therefore, from the numbers of Table II we obtain $p(\theta_0|\text{all TB-negative subjects})$ as a $\text{beta}(\theta_0|0.20753, 35033)$ (having replaced zero values by 1×10^{-6} to avoid numerical problems) and plug it in (3). We calculate

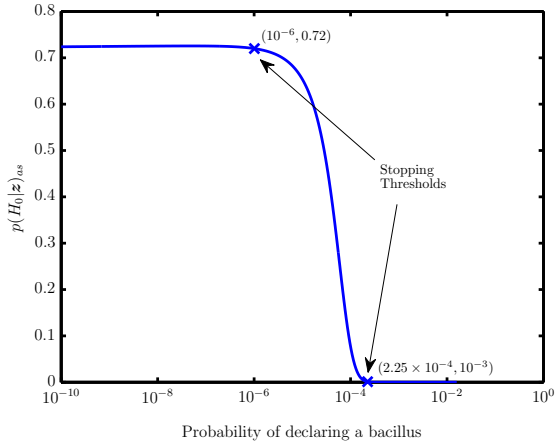


Fig. 3. Probability of being a TB-negative subject as function of the asymptotic probability of declaring a bacillus.

from Table III $p(\theta_1|\text{all TB-positive subjects})$ as a $\text{beta}(\theta_1|0.20583, 305.71)$. We fix the a priori probability of subject being TB-negative, $p(H_0) = 0.5$, in order to not favor any class.

We first explore our training data by obtaining the asymptotic posterior of being a TB-negative subject $p(H_0|\mathbf{z})_{as}$, shown in Figure 3, where we can observe that the maximum of $p(H_0|\mathbf{z})_{as}$ is bounded for the TB-negative subjects (low probability of declaring a bacillus): it does not rise much beyond 0.72. On the other hand, for TB-positive subjects (high probability of declaring a bacillus), $p(H_0|\mathbf{z})_{as}$ drops exponentially from probabilities of declaring a bacillus greater than 10^{-5} . Therefore, with the available training data and detector capability, we can attain high confidence in declaring a subject TB-positive, but not so much in declaring one TB-negative. This is because the posterior of θ_1 for TB-positive subjects captures the fact that some of them have no or very few bacilli declarations, as is the case with TB-negative subjects. On the other hand, when a sufficient number of bacilli are declared, the test is very confident in deciding that a subject is TB-positive.

Figure 3 helps to select the probability thresholds for stopping the Bayesian test. But we have to be careful as the asymptotic procedure assumes that an infinite number of samples have been analyzed. Therefore, we test a sequence with no bacilli declarations \mathbf{z}_{zero} , to see which probability threshold is adequate to stop analyzing a TB-negative subject. Figure 4 shows that after around 2×10^5 samples, the probability $p(H_0|\mathbf{z})$ reaches the value 0.72. This corresponds to 180 images without a bacilli declaration and seems an appropriate moment to stop. Therefore, we set $T_H = 0.72$ as the first stopping threshold, which corresponds to an asymptotic probability of declaring bacilli of 10^{-6} . We set the stopping threshold for a TB-positive subject to $T_L = 10^{-3}$, which seems a good compromise between early stopping and decision quality. Figure 3 shows that this threshold is equivalent to stop analyzing any subject with asymptotic probability of declaring a bacillus higher than 2.25×10^{-4} . The minimum number of patches analyzed is 1376, which corresponds to one image. Otherwise, the test stops when the maximum number of images is reached. See Figure 5 for a summary of the patient classifier procedure.

The results of this procedure for the test subjects are presented in Tables IV and V, where “stop” column informs whether the test has stopped before analyzing all the available images; “#patches” are the number of patches analyzed by the test (which is different to the number of patches available); and “#bacilli” the number of detected bacilli in the analyzed patches (again, this can be lower than the number of bacilli in the available images). On the one hand, Subjects 50304 and 55798 in Table IV have a similar number of detections, but the first has a higher rate of bacilli and its probability touches the TB-positive threshold whereas the second has its detections in the last images, which makes the

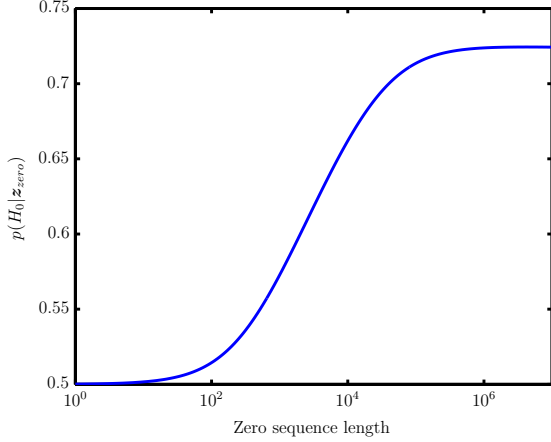


Fig. 4. Probability of TB-negative subject as function of the number of samples using an all zeros sequence.

procedure PATIENT CLASSIFIER SETUP

Define a_0, b_0 for the posterior of θ_0 . ▷

II-C2, III-B2..

Define a_1, b_1 for the posterior of θ_1 .

Define T_L, T_H with $p(H_0|z)_{as}$ (4) and $p(H_0|z_{zero})$.

return T_L, T_H ▷ Stopping Thresholds.

end procedure

procedure PATIENT CLASSIFIER($z, p(H_0), T_L, T_H$)

while $T_L < p(H_0|z) < T_H$ **do** ▷ See (1)

Analyze another image.

end while

return $p(H_0|z)$ ▷ Decision quality.

end procedure

Fig. 5. Patient Classifier Algorithm. If $p(H_0|z) < 0.5$ we decide subject is TB-positive, otherwise we decide subject is TB-negative.

test to be more cautious. Similarly, subjects 6 is decided a TB-positive patient after only two images because its 5 detections, but the test needs more images (not available) for subject 2 to meet the requirement of $p(H_0|z) \leq T_L$

TABLE IV

TB-POSITIVE TEST SUBJECTS. WE SHOW THE PROBABILITY OF BEING TB-NEGATIVE FOR $p(H_0) = 0.5$, THE NUMBER OF PATCHES USED BY THE TEST AND THE NUMBER OF BACILLI DECLARATIONS IN THESE PATCHES. WE DECIDE A SUBJECT IS TB-NEGATIVE IF $p(H_0|z) \geq 0.5$.

Subject ID	stop	$p(H_0 z)$	#patches	#bacilli
32743	1	0.720	238048	0
40279	1	8.95e-4	265568	65
50304	1	8.68e-4	92192	25
51862	0	0.708	436192	1
55165	1	0.720	238048	0
55798	0	0.190	399040	29
56036	1	0.720	238048	0
1	1	0	1376	319
2	0	0.208	105952	8
3	0	0.714	101824	0
4	1	3.71e-50	1376	37
5	0	0.713	99072	0
6	1	5.53e-6	2752	5

and it is decided to be TB-positive but with less certainty. Table V shows the effect of detections on early stopping: with the available data and threshold choices, the test only stopped for subjects with no detections, but those with one detection were about to stop. We also see how detections lower the confidence on the TB-negative decision. Globally, the hard decisions for the subjects, i.e., the result of thresholding $p(H_0|z)$ at 0.5, are the same than in the count-and-threshold case. We observe that TB-positive subjects can attain a high confidence value. However, without a high a priori probability for H_0 increasing the probability for the TB-negative subjects, our confidence for one of these will never be high, as we have already observed in the asymptotic analysis.

TABLE V
TB-NEGATIVE TEST SUBJECTS.

Subject ID	stop	$p(H_0 z)$	#patches	#bacilli
30257	0	0.693	443072	2
30266	1	0.720	238048	0
30295	0	0.675	423808	3
30665	0	0.706	377024	1
30725	0	0.675	422432	3
31228	0	0.656	416928	4
31523	0	0.706	388032	1
31534	1	0.720	238048	0
31684	0	0.659	430688	4
31819	1	0.720	238048	0
32240	1	0.720	238048	0
32633	0	0.707	408672	1
32781	0	0.649	470592	5
32956	0	0.706	393536	1
36856	1	0.720	238048	0

IV. DISCUSSION

A. Image Acquisition & Database

In this application, we face a challenging problem because the bacilli rate can be very low in some TB-positive subjects (e.g. training subject with ID 7 has been declared TB-positive by culture method but no bacilli were detected by our bacillus classifier). There are several reasons for this: bacilli live in a 3D volume and they can be orthogonal to the slide; the image acquisition process does not capture bacilli, or only very few of them; lack of focus in both manual (because of lack of training or excess of work) and automatic image acquisition; bacilli weakly tinted due to non-uniform tincture application or degradation of it, or due to antibiotics that degrade bacilli walls and less tincture get attached to them. These factors and the bacilli classifier capabilities condition the maximum performance of the system, but the proposed subjects classifier accounts for these sources of uncertainty in a principled way. It does

not matter whether the detector fails to detect bacilli or they have not been captured, the proposed subject classifier deals with all previous system errors.

The magnification used in this database is $20\times$; this is the magnification used in the standard practice among microbiologists in Madrid as a good trade-off between the time required to analyze a subject and the discrimination capability ($40\times$ is the recommended setting by WHO [8]). Other authors have used larger magnifications like $630\times$ [13], $250\times$ [14], or $100\times$ [16]. It is possible that the magnification of $20\times$ is lower than optimum for automatic recognition, and better results could be achieved with larger magnifications.

B. Subjects Classifier

The methodology presented here is general and can be applied to problems where the classes overlap and are defined by several collections of samples. The Bayesian approach demonstrates how to incorporate the uncertainty of the hypotheses in a closed-form expression that can be updated sequentially using sums for each new test sample. We also show that the maximum performance of the Bayesian test is limited by these uncertainties.

Common practice in TB screening systems is to declare a positive smear if the number of TB bacilli detections exceeds a threshold [9], [19]. In the case of humans, this number depends on the experience of the microbiologist and the type of tincture/microscope. On automated systems, this number is fixed to classify the subjects in the database balancing sensitivity and specificity of the decision about the smear. In real databases of sputum smears some TB-positive subjects might have less bacilli detections than some TB-negative subjects, therefore, some training subjects are going to be misclassified, which must be noted in the specificity and sensitivity of the global decision. Real detectors also

have a false alarm probability, which must be considered because it might render the system decisions useless. Our methodology takes into account all these issues and bounds the maximum quality of the system decisions from the available training data and, at the same time, avoids the need of fixing a threshold, by incorporating all the available information of the training patients in the decision process. In addition, the sequential nature of the decision mitigates to some extent the deficiencies of the detector and the image acquisition process. We have also developed a frequentist approach for this problem [33].

C. Experiments

The count and threshold approach, obtains the same decisions than the Bayesian approach in our test patients but gives no quality measures with the decisions. In our experiments, by using the former approach with a threshold of 6 detections we ignore the quality (i.e., if we can trust) of a TB-negative decision with 5 detections or a TB-positive decision with 6 detections. In addition, with this approach, the system is not capable of early stopping in a negative test or examining a different number of images for cases with different number of detections, which does not happen with sequential analysis. The Bayesian procedure is straightforward to adjust, and only the probability thresholds for the sequential test and the a priori probability of being TB-negative has to be specified (for example, $p(H_0) = 0.92$ is the a priori probability of a TB-negative subject in Madrid hospitals). The experiments show that the global performance can be estimated from the training data. They stress the importance of having a good bacilli classifier and show how to penalize our confidence in the decisions, taking into account all the steps of the decision process.

V. CONCLUSION

Screening systems need to provide not only a decision, but also an estimate of its quality. We have shown how to provide this quality and how to bound its maximum from the training data and bacilli classifier performance. This allows knowing what the strengths and weakness of the decision system are, and points toward further improvement in its parts. The described decision methodology is general and can be applied to any screening system. In addition, our subject classifier is straightforward to use and its computing requirements are negligible as it has a closed-form expression, which can be sequentially updated.

We have illustrated this framework with a screening system for tuberculosis, which can be trained for any type of microscope/tincture combination. The proposed system combines a bacilli classifier amenable for real-time operation (processing images at the rate they are captured) with sequential analysis, which stops the screening when the quality of the decision has reached the requirements. The evaluation of the quality of the decisions of this system on real data shows that the system is only confident in declaring TB-positive subjects. This opens up the opportunity of further research for the bacilli classifier.

ACKNOWLEDGEMENT

This work was supported in part by the Ministerio de Economía of Spain (under Projects ‘DEIPRO’, id. TEC2009-14504-C02-01, and ‘COMONSENS’, id. CSD2008-00010, ‘ALCIT’, id. TEC2012-38800-C03-01, ‘COMPREHENSION’, id. TEC2012-38883-C02-01), the Ministerio de Industria of Spain (project ‘CDTEAM’ of CENIT-Ingenio 2010), the EU Marie Curie Transfer of Knowledge Fellowship (CT-2005-029611), the Marie Curie ITN FP7-PEOPLE-2012-ITN-

316861 (Machine Learning for Personalized Medicine 2012), and Universidad Carlos III (CCG10-UC3M/TIC-5304). The work of R. Santiago-Mozos was supported in part by Ministerio de Educación of Spain (EX2009-0681) and Juan de la Cierva Program of the Ministerio de Ciencia e Innovación of Spain (JCI-2011-11150)

We are grateful to Dr. Manuel Desco for his considerable support in all this project and to Dr. María Jesús Ruiz Serrano for her supervision in the gathering of the database. We would also like to thank the Irish Centre for High-End Computing (ICHEC) for allowing us to use their computing facilities.

REFERENCES

- [1] World Health Organization, “Global tuberculosis report,” Tech. Rep. WHO/HTM/TB/2012.6, 2012.
- [2] V. Kumar, A. K. Abbas, and J. C. Aster, *Robbins Basic Pathology*. Saunders Philadelphia, 2012.
- [3] R. S. Wallis, M. Pai, D. Menzies, T. M. Doherty, G. Walzl, M. D. Perkins, and A. Zumla, “Biomarkers and diagnostics for tuberculosis: progress, needs, and translation into practice,” *The Lancet*, vol. 375, no. 9729, pp. 1920–1937, Jun. 2010.
- [4] R. Diel, R. Loddenkemper, K. Meywald-Walter, R. Gottschalk, and A. Nienhaus, “Comparative performance of tuberculin skin test, QuantiFERON-TB-Gold In Tube assay, and T-Spot.TB test in contact investigations for tuberculosis.” *Chest*, vol. 135, no. 4, pp. 1010–8, Apr. 2009.
- [5] A. Zwerling, S. van den Hof, J. Scholten, F. Cobelens, D. Menzies, and M. Pai, “Interferon-gamma release assays for tuberculosis screening of healthcare workers: a systematic review.” *Thorax*, vol. 67, no. 1, pp. 62–70, Jan. 2012.
- [6] M. Pai, J. Minion, H. Sohn, A. Zwerling, and M. D. Perkins, “Novel and improved technologies for tuberculosis diagnosis: progress and challenges.” *Clinics in chest medicine*, vol. 30, no. 4, pp. 701–16, viii, Dec. 2009.
- [7] C. C. Boehme, P. Nabeta, D. Hillemann, M. P. Nicol, S. Shenai, F. Krapp, J. Allen, R. Tahirli, R. Blakemore, R. Rustomjee, A. Milovic, M. Jones, S. M. O’Brien, D. H. Persing, S. Ruesch-Gerdes, E. Gotuzzo, C. Rodrigues, D. Alland, and M. D. Perkins, “Rapid molecular detection of tuberculosis and rifampin resistance.” *The New England journal of medicine*, vol. 363, no. 11, pp. 1005–15, Sep. 2010.
- [8] D. Boyle and M. Pai, “Tuberculosis: Diagnostic Technology Landscape,” UNITAID Secretariat, World Health Organization, Tech. Rep., Jul. 2012.
- [9] K. Steingart, M. Henry, V. Ng, P. Hopewell, A. Ramsay, J. Cunningham, R. Urbanczik, M. Perkins, M. Aziz, and M. Pai, “Fluorescence versus conventional sputum smear microscopy for tuberculosis: a systematic review.” *Lancet Infect Dis*, vol. 6, no. 9, pp. 570–81, 2006.
- [10] M. Pai, J. Minion, K. Steingart, and A. Ramsay, “New and improved tuberculosis diagnostics: evidence, policy, practice, and impact.” *Current opinion in pulmonary medicine*, vol. 16, no. 3, pp. 271–84, May 2010.
- [11] J. A. Robledo, G. I. Mejia, N. Morcillo, L. Chacon, M. Camacho, J. Luna, J. Zurita, A. Bodon, M. Velasco, J. C. Palomino, A. Martin, and F. Portaels, “Evaluation of a rapid culture method for tuberculosis diagnosis: a latin american multi-center study,” *International Journal of Tuberculosis and Lung Disease*, vol. 10, no. 6, p. 613, 2006.
- [12] B. Yu, F. B. Dazzo, R. Srivatsa, T. Zhang, and A. K. Jain, “A computer-aided system for image analysis of morphological diversity, abundance and spatial distribution,” Department of Computer Science, Michigan State University, East Lansing, Michigan, Tech. Rep. MSU-CPS-97-24, Jul. 1997.
- [13] K. Veropoulos, G. Learmonth, C. Campbell, B. Knight, and J. Simpson, “The automated identification of tubercle bacilli in sputum: A preliminary investigation.” *Analytical and Quantitative Cytology and Histology*, vol. 21, no. 4, pp. 277–281, Aug. 1999.
- [14] M. G. Forero, G. Cristobal, and M. Desco, “Automatic identification of mycobacterium tuberculosis by gaussian mixture models,” *Journal of Microscopy*, vol. 223, no. 2, pp. 120–132, 2006.
- [15] R. Khutlang, S. Krishnan, A. Whitelaw, and T. Douglas, “Automated detection of tuberculosis in Ziehl-Neelsen-stained sputum smears using two one-class classifiers,” *Journal of Microscopy*, vol. 237, no. 1, pp. 96–102, 2010.
- [16] R. Khutlang, S. Krishnan, R. Dendere, A. Whitelaw, K. Veropoulos, G. Learmonth, and T. S. Douglas, “Classification of Mycobacterium tuberculosis in images of ZN-stained sputum smears,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 4, pp. 949–957, 2010.
- [17] J. Chang, P. Arbeláez, N. Switz, C. Reber, A. Tapley, J. L. Davis, A. Cattamanchi, D. Fletcher, and J. Malik, “Automated tuberculosis diagnosis using fluorescence images from a mobile microscope.” in *MICCAI*, vol. 15, Jan. 2012, pp. 345–52.
- [18] World Health Organization, “Treatment of tuberculosis: guidelines – 4th ed.” 2009.
- [19] H. L. Rieder and A. V. Deun, “Proposal for a

- revision of the case definition of “sputum smear-positive tuberculosis”,” World Health Organization, WHO Policy background document, 2007. [Online]. Available: http://www.who.int/entity/tb/laboratory/proposed_reduction_hravd.pdf
- [20] R. Santiago-Mozos, R. Fernández-Lorenzana, F. Pérez-Cruz, and A. Artés-Rodríguez, “On the uncertainty in sequential hypothesis testing,” in *Fifth IEEE International Symposium on Biomedical Imaging (ISBI '08)*, Paris, France, 14–17 May 2008, pp. 1123–1126.
- [21] R. Santiago-Mozos and A. Artés-Rodríguez, “Uncertainty-based censoring scheme in distributed detection using learning techniques,” in *Proceedings of the Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU)*, Jul. 2006, pp. 2027–2034.
- [22] J. F. Canny, “A computational approach to edge detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.
- [23] F. Mindru, T. Tuytelaars, L. V. Gool, and T. Moons, “Moment invariants for recognition under changing viewpoint and illumination,” *Computer Vision and Image Understanding*, vol. 94, no. 1, pp. 3–27, 2004.
- [24] R. C. González and R. E. Woods, *Digital image processing*. Pearson Prentice Hall, 2008.
- [25] S. C. Pei and C. N. Lin, “Image normalization for pattern recognition,” *Image and Vision computing*, vol. 13, no. 10, pp. 711–723, 1995.
- [26] B. E. Boser, I. M. Guyon, and V. N. Vapnik, “A training algorithm for optimal margin classifiers,” in *COLT '92: Proceedings of the fifth annual workshop on Computational learning theory*. New York, NY, USA: ACM, 1992, pp. 144–152.
- [27] B. Schölkopf and A. Smola, *Learning with Kernels*. Cambridge, MA, USA: MIT Press, 2001.
- [28] E. L. Lehmann, *Testing statistical hypotheses*, 2nd ed. Springer, 1997.
- [29] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. Springer-Verlag New York, Inc., 1985.
- [30] C. C. Chang and C. J. Lin, *LIBSVM: a library for support vector machines*, 2001, software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [31] P. Viola and M. Jones, “Robust real-time object detection,” in *Second International Workshop on Statistical and Computational Theories of Vision - Modeling, Learning, Computing, and Sampling*, Jul. 2001. [Online]. Available: http://www.stat.ucla.edu/~sczhu/Workshops/sctv01/Viola_Jones.ps.gz
- [32] S. S. Keerthi, O. Chapelle, and D. DeCoste, “Building support vector machines with reduced classifier complexity,” *The Journal of Machine Learning Research*, vol. 7, pp. 1493–1515, 2006.
- [33] R. Santiago-Mozos, M. G. Madden, F. Pérez-Cruz, and A. Artés-Rodríguez, “On the screening of tuberculosis,” Universidad Carlos III de Madrid, Tech. Rep., 2012. [Online]. Available: <http://www.tsc.uc3m.es/~rsmozos/screening.pdf>