2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA) | 978-1-7281-8470-8/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICMLA51294.2020.00065

Tuberculosis Bacilli Identification: A Novel Feature Extraction Approach via Statistical Shape and Color Models

Hamed Yousefi*
Dept. of Imaging Science
Washington Univ. in St. Louis
St. Louis, MO, USA
hamed.yousefi@wustl.edu

Farnaz Mohammadi*
Dept. of Bioengineering
UCLA
Los Angeles, CA, USA
farnazmdi@ucla.edu

Niloufar Mirian*
Dept. of Radiology
Yale University
New Haven, CT, USA
0000-0003-1623-578X

Navid Amini
Dept. of Computer Science
Cal. State Univ., Los Angeles
Los Angeles, USA
namini@calstatela.edu

** Hamed Yousefi, Farnaz Mohammadi, and Niloufar Mirian are co-first authors.*

***Abstract*— Tuberculosis (TB) is an infectious disease resulting in 1.5 M deaths worldwide every year. Early diagnosis can play an important role in reducing TB induced mortality rates. A common inexpensive approach to test for TB is a manual examination of sputum smear images under the light/fluorescent microscope. To enhance the process, computer-aided diagnostic devices and machine learning algorithms have been developed with the capability of identifying TB bacilli using image processing techniques. Here, we propose a novel statistical model of the shape and color of TB bacilli in Ziehl-Neelsen stained images under the light microscope to identify bacilli in these images. These simple statistical models are utilized as a general library for reconstructing any bacillus with various background colors and can overcome the difficulties associated with geometric feature extraction methods. We use various methods to classify the individual bacilli and overlapping bacilli from the rest of the image based on the eigenvalues of the shape and color models. Our results show that *the k*-NN classifier performs best among selected classifiers with an average accuracy of 82.7% for single bacilli and overlapping bacilli detection, and 99.1% accuracy of identifying individual bacilli alone from overlapping bacilli and other objects.**

Keywords*— *Ziehl-Nelsen stained imaging, classification, segmentation, feature selection, MSER, Gabor filter, YIQ color space, PCA, statistical shape model, statistical color model.

I. INTRODUCTION

Tuberculosis (TB) is one of the world's major health problems. Based on the World Health Organization (WHO) report, it is among the 10 deadliest diseases worldwide [1]. Almost 10 million people were diagnosed with TB and 1.5 million died only in 2018 [1]. It is infectious and could be spread by coughing and is caused by the bacillus *Mycobacterium tuberculosis* [2]. With an increasing rate of death and diagnosis of this disease, TB remains one of the fatal diseases and global concerns in the health realm [3] which can affect some organs of the body including kidneys, liver, bones, brain, nervous system, and most commonly the lung (pulmonary tuberculosis) [2].

Fortunately, TB is treatable if it is diagnosed promptly and monitored accordingly; thus, early detection is an important step to control the development of TB and reduce the mortality rate of it. There are a number of limitations associated with manual method including human error and fatigue, restriction of medical facilities, and lack of accuracy [4]. Computer-aided tools have been helping to minimize these limitations Computational approaches take in the images and employ supervised/unsupervised learning algorithms to identify the bacilli correctly.

Combining color features along with the shape characteristics have been an interesting area of focus for clustering the bacteria from images [5]. Forero et al. [6] employed invariant shape features with a thresholding operation on the chromatin channel to the sputum images and conducted k-means clustering for the purpose of classification. In another work [7], Forero and colleagues have proposed a technique based on the heuristics from the shape contour and color features for image segmentation and used classification trees to determine whether a sample is positive or negative. Makkapathi et al. [8] have exploited a hue color component-based approach to segment the bacilli in ZN-stained images. They have used length and width parameters to identify a valid single bacillus. Despite high accuracies provided in previous studies, such approaches are sensitive to characteristics of images, such as magnification, background color, type of the image, and the type of the microscope that the images are acquired with, some of which might result in failure of those models. Here, we propose a framework that is independent of the aforementioned characteristics, and yet, provides high accuracy in classifying the bacilli.

In this study, we develop a series of pre-processing techniques to enhance the images for segmentation by employing decorrelation stretch, transforming the color space of images, and exploiting the Gabor filter. The maximally stable extremal region (MSER) algorithm will then be employed for image segmentation. Our approach can be summarized into the following steps: (1) Pre-processing the images, (2) segmenting the pre-processed images, (3) labeling the objects as "single bacillus", "overlapping bacilli", and "other" by an expert, (4) extracting the statistical model of the shape and color of bacilli, (5) feature extraction using eigenvalues of the statistical models, and (6) classification and cross-validation of the images.

Statistical shape models are well-established to study geometrical features of a set of shapes, which all are assumed to be a deformed version of a reference shape [9] [10] [11].

978-1-7281-8470-8/20/$31.00 ©2020 IEEE
DOI 10.1109/ICMLA51294.2020.00065

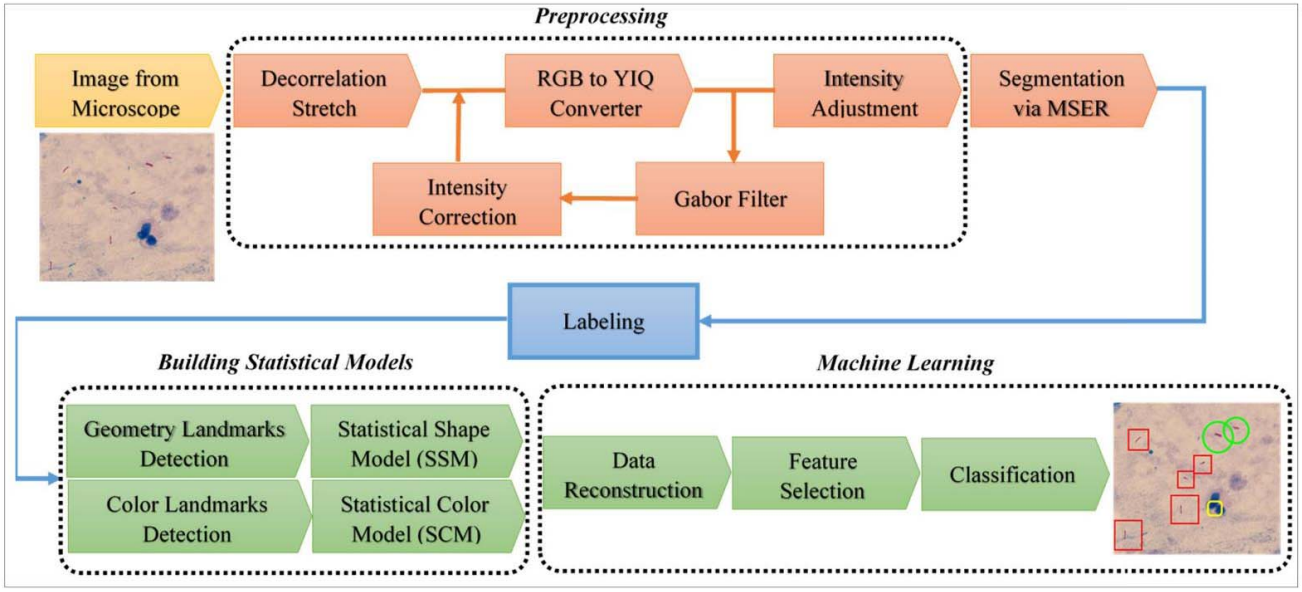

Fig. 1. Block diagram of the proposed technique.

They are built with statistical features such as mean and standard deviation; hence, the model is applicable to other spaces such as a color space. We utilize a number of eigenvalues of these models to select the features for classification. To the best of our knowledge, this study is the first instance of utilizing statistical shape and color model of the bacilli and their corresponding eigenvalues of the models as the feature selection approach for bacilli detection.

II. METHODS

The overall architecture of the proposed method is illustrated in Fig. 1. The top part shows the segmentation steps and the lower part indicates classification outputs using statistical models built. The following explains each step with detail.

A. Pre-processing

Decorrelation Stretch Transformation. Decorrelation Stretch is an enhancement technique for image processing, mostly used to enhance multi-spectral images [11][12]. In this technique, the covariance matrix of the colors is diagonalized by applying a Karhunen-Loeve transform [13]. To equalize the color variances, the contrast for each color is stretched, consequently, the colors become uncorrelated and cover the whole color space. In the end, the inverse transform is applied to map the colors to an approximation of the original image [11]. We apply decorrelation stretch to the images in RGB space to highlight the contrast of the color spectrum.

YIQ Transformation. YIQ is a color space that is obtained by transforming the RGB color matrix of the image using (1). The Y component represents brightness information, while I and Q represent the color information. The transformation matrix is as follows [14].

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

We are interested in the above color space since the bacilli in the microscopic images are in the color range between dark pink and purple, which is a part of the YIQ color space. By transforming an image from RGB into YIQ space, the result holds the enhanced contrast of the bacilli so that they are more significant and distinguishable. After applying decorrelation stretch the color transformation is applied to the images, and the third component (Q) is selected such that the bacilli are more recognizable.

Gabor Bank Filter. Gabor filter is used for segmenting objects with subtle boundaries such as fingerprint recognition [15] and is very sensitive to severe intensity changes, making it desirable as another enhancement step for these microscopic images [16]. Hence, in this step, the Gabor is applied to the Q component of transformed images. In this study, the 2D Gabor transform has been employed to distinguish the bacilli boundary from the background in microscopic images. In the image domain, this function sweeps the entire xy-plane and changes the parameters to find the maximum intensity difference in proximal pixels [17][18]. The Gabor 2-D filter over the (x, y) plane is defined as:

$$G(x,y) = e^{-\left(\frac{(x-x_0)^2}{2\sigma_x^2} + \frac{(y-y_0)^2}{2\sigma_y^2}\right)} \times e^{-2\pi i\left(u_0(x-x_0) + v_0(y-y_0)\right)} \quad (2)$$

in which (x_0,y_0) is the current location of the filter in the plane and (u_0,v_0) specifies the modulation with the spatial frequency, $\omega_0 = \sqrt{u_0^2 + v_0^2}$ and orientation $\theta_0 = tan^{-1}\frac{u_0}{v_0}$. Fig.2a

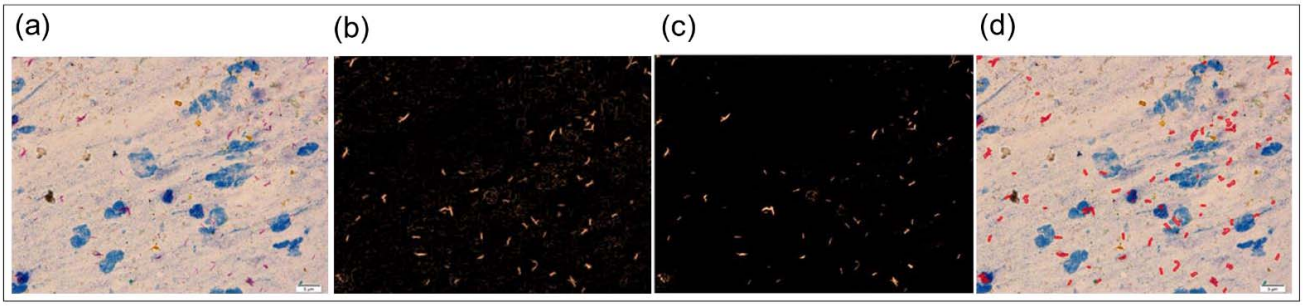

Fig. 2. Results of image enhancement. **(a)** one sample of raw images. **(b)** output of Gabor bank filter. **(c)** an enhanced image from intensity correction step **(d)** segmented objects via maximally stable extremal region (MSER) algorithm.

represents the raw microscopic images, and Fig. 2b shows three samples of Gabor filter output.

Intensity Adjustment. The three components of HSV color space are hue, saturation, and value (intensity). After transforming the image from RGB to HSV, intensity correction is achieved by replacing the intensity component of the image by the output of the Gabor filter from the previous step. The image is then transformed from HSV to RGB to retrieve the original images with enhanced intensity. Now that the intensity of the images is enhanced, we transform the images to YIQ color space since it includes a spectrum of pink to purple colors which makes the bacilli more distinguishable. As the last enhancement step before segmentation, we adjust the intensity of the image using Gamma correction [19] which is a nonlinear operator to adjust the luminescence of an image.

Since the number of bacilli in each image is different, it will influence the statistical moments of the image; making it impossible to define a global threshold for all the images. To avoid this problem, we find the mean and standard deviation of the third channel of the image color space (Q component), and based on the standard deviation, gamma is considered as the shape parameter of image adjustment. Based on the standard deviation, the gamma parameter is divided into four intervals (3), and larger values of gamma correspond to a higher standard deviation.

$$\gamma = \begin{cases} 1 & SD \leq 0.015 \\ 3 & 0.015 \leq SD \leq 0.03 \\ 5 & 0.03 \leq SD \leq 0.045 \\ 10 & otherwise \end{cases} \quad (3)$$

Fig. 2c depicts the output of intensity correction with different gamma parameters derived from standard deviation of the Q component in YIQ color space.

B. Image Segmentation and Labeling

Segmentation. We use maximally stable extremal region (MSER) segmentation, an unsupervised learning algorithm [20]. The MSER algorithm extracts regions of an image that remain stable for a number of intensity thresholds of a gray-scale image. This means if we a set of thresholds for a gray-level image, that span all the intensity range from white to black, there exist some regions that appear and persist for several thresholds while sweeping the intensity range. The algorithm picks these regions as objects. The only parameter of this method, Δ, controls the intensity intervals such that when the intensity is increased by Δ, the stability becomes the relative area variation of that region (see Fig. 2d).

Labeling. Our dataset includes 43 smear images. After the segmentation process and detecting the objects, we find 2 points with maximum Eucleadian distance on the perimeter of an object and draw a square with twice the size of that distance as the width and length of a bounding box, co-centered with the object. These squared boxes that include the objects, are cropped from the original image and are provided to an expert for labeling the objects as "bacillus" representing an individual bacillus, "overlap" representing multiple bacilli as one object, and "other" to show any detected object other than the first two classes. The labeling process by domain experts resulted in 738 "bacillus", 354 "overlap", and 335 "other" objects.

C. Statistical Models of the Shape and Color

This subsection describes the process of building the statistical shape and color model for the objects. In this study, we only build the model for the "bacillus" class.

Statistical Shape Model (SSM). To prepare the data for creating the model, the first step is to align the orientation of the objects. To do so, we find a bounding ellipse that best fits each rod-shaped object. Then the objects are rotated so that their major axis is aligned with the x-axis. The second step is finding a set of landmarks representative of each object. Since the images are 2D the boundary pixels, or the shape vector, of the object's contour are found in the 2D cartesian space as $\boldsymbol{x}=[x_1, x_2, \dots, x_{20}, y_1, y_2, \dots, y_{20}]^T$. Measurements in both dimensions are converted to a 1D signal. To find 20 points as the landmarks, spline interpolation is applied to each of the signals. To co-center the landmarks, we subtract all the x's and y's from their means. Building the SSM is based on Principal-

Component Analysis (PCA) which is a dimensionality reduction method. To train the model, 400 out of 738 samples are randomly chosen from the class of "bacillus", since we focus on distinguishing bacilli from non-bacilli objects as a post-processing step. Initially, the mean shape vector for the training dataset is calculated as $\bar{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x_i}$. The covariance matrix is calculated by $C = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{x_i} - \bar{\boldsymbol{x}})(\boldsymbol{x_i} - \bar{\boldsymbol{x}})^T$ which provides a basis for computing the principal components.

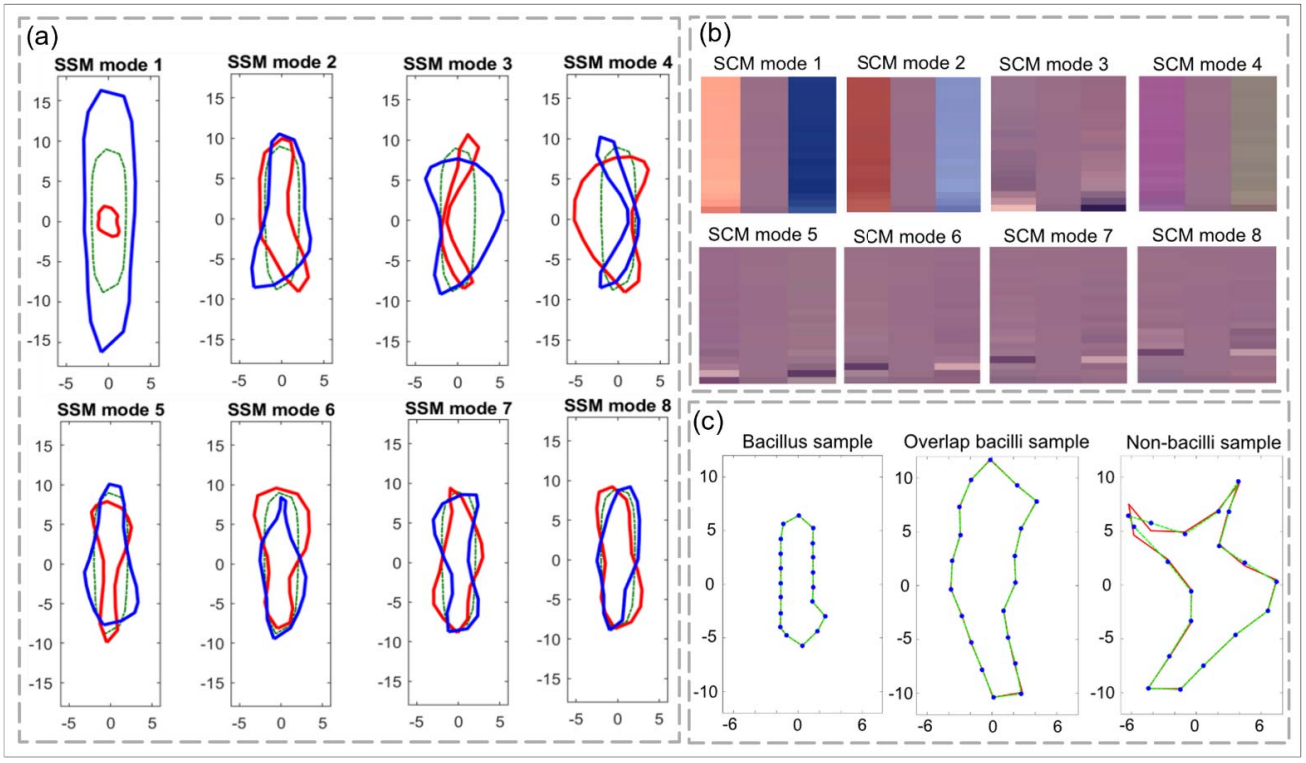

Fig. 3. (a) Statistical Shape Model (SSM) derived from 20 points around the boundary of each bacilli data. The dotted green object is related to mean model and variation across the mean model from -3σ (red objects) to +3σ (blue objects) visualized for the first 8 modes of variation **(b)** Statistical Color Model (SCM) derived uniformly from 20 color pixels of each bacilli data. Each image related to one of the first 8 modes variations from -3σ (left column) to +3σ (right column) and mean model (middle column) **(c)** Reconstructed shapes from SSM for bacilli (left), overlapping bacilli (middle) and non-bacilli object (right).

Afterward, the eigenvalues and the eigenvector are calculated and sorted in a descending order. With 40 points, we will have 39 eigenvectors $\boldsymbol{p_j}$ forming the matrix $\boldsymbol{P} = (\boldsymbol{p_1}|\boldsymbol{p_2}|\dots|\boldsymbol{p_{39}})$, as well as a 39-dimensional vector $\boldsymbol{b} = \boldsymbol{P}^T(\boldsymbol{x} - \bar{\boldsymbol{x}})$. Then, the landmarks of the aligned shapes are defined as:

$$\boldsymbol{x} = \bar{\boldsymbol{x}} + \boldsymbol{Pb} \tag{4}$$

As Fig. 3a depicts, eigenvectors with larger eigenvalues correspond to a more significant shape variation. The images show average of the model along with 3 standard deviations (±3σ). With the first 10 eigenvectors we were able to reconstruct 96% of the data correctly with the following approximation (see Fig. 3c):

$$\boldsymbol{x} \approx \bar{\boldsymbol{x}} + \boldsymbol{P'b'} \tag{5}$$

where, $\boldsymbol{P'} = (\boldsymbol{p_1}|\boldsymbol{p_2}|\dots|\boldsymbol{p_N})$ and $\boldsymbol{b'} = \boldsymbol{P'^T}(\boldsymbol{x} - \bar{\boldsymbol{x}})$.

Statistical Color Models (SCM). The color modeling procedure is similar to the shape modeling, except that we need to define the landmarks according to the red, green, and blue components for each pixel. The number of landmarks must be the same in all the objects that are being modeled since objects have a different number of pixels we select 20 representative pixels of each object based on the histogram of the pixels to be the landmarks for each of the RGB components. We start from the closest pixel to the mean to the farthest point in terms of Euclidean distance that forms the 20th point. After obtaining the landmarks, we find the color vector analogous to the shape vector in the SSM. The mean and covariance matrix for the training dataset is then calculated. Sorted eigenvectors form the and matrices to be used in the color reconstruction. The first eight variation modes built from SCMbacilli illustrated in Fig. 3b. Each image contains three columns which show data variation in its related eigenvalues from -3σ to +3σ across the mean model. Aside from the color model of the bacilli objects, we decided to create the color model of the background (SCM_{bkg}), so that we include the information about the neighborhood of the bacilli and thus, enhance the identification of the bacilli in any background. To do so, for each bacillus object, similar to the color model, we take 20 pixels as landmarks from the background and build the corresponding color model for the background. This will enable the color model of the bacilli to be independent of the background, which leads to a significant increase in the accuracy of classification.

D. Feature Selection and Classification

The features used for classification are the coefficients, b's in (5) across the eigenvalues extracted from the SSM, $SCM_{bacilli}$, and SCM_{bkg} which include the color model of the background. Since we have selected 39 eigenvalues from each model, we will have 117 features in total. To select the most contributing features for classification, we use the relief-based algorithm [21] to rank the features by assigning scores based on the importance of feature value among nearest neighbor instance pairs. The Relief algorithm takes in the data and the predictor matrix along with the number of nearest neighbors and computes ranks and

weights of the predictors with k nearest neighbors. We use the built-in function, Relief, in MATLAB for this purpose and use an optimum number ranked features for optimum performance of each classifier [22], [23].

As Fig. 4a and Table 1 depict, various methods are employed for the purpose of classification, such as, (1) k-nearest neighbors (k-NN) with k = 7 number of nearest neighbors; (2) fully connected neural network with backpropagation and regularization parameter (γ=0); (3) support vector machine (SVM) with radial basis function (RBF) kernel; (4) linear discriminant analysis (LDA) (5) naïve Bayes; and (6) decision tree classifier. For all cases, we set the number of classes equal to 3, for which we found out that the "bacilli" class is strongly distinguishable from "other" and "background" classes. In all classifications, the data matrix is standardized by centering each of their features and dividing by their standard deviations.

III. Results

From 43 smear images, we segment a total of 1427 objects into three sets of classes, namely, "bacillus", "overlap", and "other". As mentioned in the labeling section, the segmentation results in 738 objects belonging to the "bacillus" class, 354 "overlap", and 335 "other" objects. SSM and $SCM_{bacilli}$ of these objects are created, and coefficients for each eigenvalue of SSM, $SCM_{bacilli}$, and SCM_{bkg} are extracted as the feature space. Adding background information in the feature space improves the accuracy by 35.2±4.1% on average among all the classifiers. Using Relief feature selection for 10 nearest neighbors, we rank all the features. We employ various classifiers with 5-fold cross-validation and evaluate their performance by means of receiver operating characteristics (ROC) curves and confusion matrices, as summarized in Table 1 and Fig. 5. The three best performing classifiers are LDA, k-NN, and ANN, for which the optimal number of features are 73, 38, and 9, respectively.

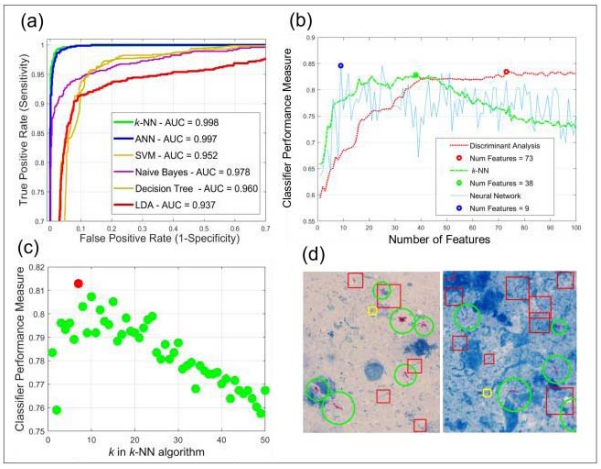

Fig. 4. (a) ROC for different classifiers. **(b)** Plot of performances versus the number of features for k-NN (green), ANN (blue), and LDA (red), **(c)** Best k-NN performance versus k number of neighbors. **(d)** Final output images, diagnosed individual bacillus (red rectangle), overlapping bacilli (green circle), and non-bacilli (yellow rectangle).

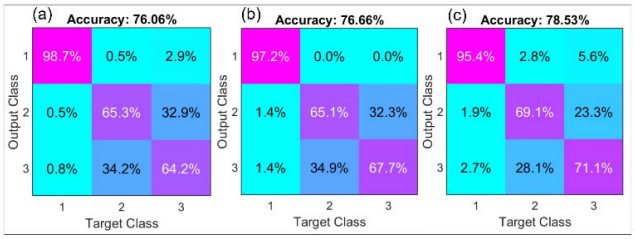

Fig. 5. Confusion matrix for **(a)** k-NN, **(b)** ANN, and **(c)** LDA classifiers. Classes 1, 2, and 3 are assigned to "single bacillus", "overlapping bacilli", and "non-bacilli", respectively.

Fig. 4b shows the classification performance with respect to the number of selected features for each classifier. Utilizing the Relief feature selection method increases the accuracy of k-NN and ANN by 17% and 60%, respectively; however, it does not substantially improve the other classification methods. The best performing classifier is the k-NN with k equal to 7, achieving 82.27% accuracy, 92.03% specificity, and 75.99% sensitivity (Fig. 4c). Based on the confusion matrices presented in Fig. 5, k-NN is also able to discriminate individual bacilli from the two other classes better than ANN and LDA. ANN delivers the performance of 81.12% (only 1.5% less than the accuracy of k-NN), utilizing only 9 features which is significantly lower than that of LDA and k-NN. Fig. 4d shows a sample final output of the proposed technique.

The proposed technique is novel in the sense that one does not need to extract color and shape features, such as eccentricity, compactness, area, perimeter, solidity, and so forth.

IV. Conclusion

Tuberculosis (TB) is among the most lethal infectious diseases in the world that kills one person every 21 seconds, according to the World Health Organization [1]. Early and accurate detection of TB is essential to prevent TB-associated deaths (currently ~1.5M a year globally). In this study, we propose a fully automatic framework to detect bacilli in smear images.

We design a segmentation module that identifies and separates objects in each smear microscopic image. The statistical shape and color models are developed for the bacillus class and the corresponding eigenvalues of the models are selected as features. Statistical models are used to reconstruct the shape and color of the bacilli objects with only 20 landmarks. Several classifiers are employed to distinguish bacillus from overlapping bacilli, and any other objects in the smear images. We conclude that k-NN classifier performs the best with an average accuracy of 82.27% for the 3 classes ("single bacillus", "overlapping bacilli", and "other") with the ability to identify a single bacillus with 98.6% accuracy. Incorporation of the color of bacilli objects' background pixels into the feature space improve the accuracy of the model by 38% in k-NN and 67% in ANN methods. In addition, exploiting the Relief algorithm for feature selection further enhances the accuracy of the model by 17% and 60% in k-NN and ANN, respectively.

TABLE 1: Results of classification using different classifier

	Accuracy (%)			Sensitivity (%)			Specificity (%)		
Features/ Classifier	**SSM +SCM**	**SSM +SCM +BKG**	**Relief (feature selection)**	**SSM +SCM**	**SSM +SCM +BKG**	**Relief (feature selection)**	**SSM +SCM**	**SSM +SCM +BKG**	**Relief (feature selection)**
***k*-NN (k=7)**	51.37	70.92	**82.27**	40.92	61.49	**75.99**	70.35	83.05	**92.03**
SVM (RBF)	55.36	76.45	76.73	42.20	70.16	70.30	71.04	88.93	88.42
Naive Bayes	46.95	74.91	74.70	38.30	69.80	68.04	69.18	88.49	87.92
Decision Tree	54.45	76.66	75.54	42.35	70.09	68.63	71.28	88.16	87.90
LDA	59.5	**83.25**	75.26	49.12	**78.12**	68.49	74.39	**92.03**	87.24
ANN	30.07	50.35	**81.12**	28.23	0.37	**76.18**	63.33	83.72	**91.57**

The MATLAB code for all the sections, including the segmentation, model implementation, and classification along with the ZN-stained microscopic images that have been used to train and test the models are publicly available on the GitHub page of the author [27].

REFERENCES

[1] W. H. Organization, "WHO | Global tuberculosis report 2019," *World Heal. Organ.*, 2020.

[2] H. B. Rachna and M. Mallikarjuna Swamy, "Detection of Tuberculosis bacilli using image processing techniques," *Int. J. Soft Comput. Eng.*, vol. 3, no. 4, 2013.

[3] K. Veropoulos, C. Campbell, G. Learmonth, B. Knight, and J. Simpson, "The Automated Identification of Tubercle Bacilli using Image Processing and Neural Computing Techniques," 1998.

[4] M. K. Osman, M. Y. Mashor, and H. Jaafar, "Tuberculosis bacilli detection in Ziehl-Neelsen-stained tissue using affine moment invariants and extreme learning machine," in *Proceedings - 2011 IEEE 7th International Colloquium on Signal Processing and Its Applications, CSPA 2011*, 2011.

[5] S. R. Reshma and T. Rehannara Beegum, "Microscope image processing for TB diagnosis using shape features and ellipse fitting," in *2017 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems, SPICES 2017*, 2017.

[6] M. G. Forero, F. Sroubek, and G. Cristóbal, "Identification of tuberculosis bacteria based on shape and color," *Real-Time Imaging*, 2004.

[7] M. Forero, G. Cristobal, and J. Alvarez-Borrego, "Automatic identification techniques of tuberculosis bacteria," in *Applications of Digital Image Processing XXVI*, 2003.

[8] V. Makkapati, R. Agrawal, and R. Acharya, "Segmentation and classification of tuberculosis bacilli from zn-stained sputum smear images," in *2009 IEEE International Conference on Automation Science and Engineering, CASE 2009*, 2009.

[9] C. Lindner, "Automated Image Interpretation Using Statistical Shape Models," in *Statistical Shape and Deformation Analysis: Methods, Implementation and Applications*, 2017.

[10] H. Yousefi, M. Fatehi, M. Bahrami, and R. A. Zoroofi, "3D statistical shape models of radius bone for segmentation in multi resolution MRI data sets," in *2014 21st Iranian Conference on Biomedical Engineering, ICBME 2014*, 2014.

[11] J. Harman, "Using Decorrelation Stretch to Enhance Rock Art Images," *Am. Rock Art Res. Assoc. Annu. Meet.*, 2006.

[12] H. Yousefi, M. Fatehi, and R. A. Zoroofi, "3d segmentation and volumetric of growth plate of radius bone for legal age determination in mri data set," in *2015 22nd Iranian Conference on Biomedical Engineering, ICBME 2015*, 2016.

[13] R. J. Clarke, "Relation between the Karhunen Loève and cosine transforms," *IEE Proc. F Commun. Radar Signal Process.*, 1981.

[14] B. Ahirwal, M. Khadtare, and R. Mehta, "FPGA based system for Color Space Transformation RGB to YIQ and YCbCr," in *2007 International Conference on Intelligent and Advanced Systems, ICIAS 2007*, 2007.

[15] S. Chavan, P. Mundada, and D. Pal, "Fingerprint authentication using Gabor filter based matching algorithm," in *Proceedings - International Conference on Technologies for Sustainable Development, ICTSD 2015*, 2015.

[16] F. Nabi, H. Yousefi, and H. Soltanian-Zadeh, "Segmentation of major temporal arcade in angiography images of retina using generalized hough transform and graph analysis," in *2015 22nd Iranian Conference on Biomedical Engineering, ICBME 2015*, 2016.

[17] F. Nabi, H. Yousefi, H. Soltanianzadeh, "Major temporal arcade separation in angiography images of retina using the Hough transform and connected components", 23rd Iranian Conference on Electrical and Engineering, IEEE, 2015.

[18] H. Yousefi, S. Askari, G. A. Dumont, and Z. Bastany, "Automated decomposition of needle EMG signal using STFT and wavelet transforms," in *2014 21st Iranian Conference on Biomedical Engineering, ICBME 2014*, 2011.

[19] A. Yanof, M. Skow, H. Tran, and A. Adby, "Digital image processing using white balance and Gamma correction," 7,236,190, 2007.

[20] H. Chen, S. S. Tsai, G. Schroth, D. M. Chen, R. Grzeszczuk, and B. Girod, "Robust text detection in natural images with edge-enhanced maximally stable extremal regions," in *Proceedings - International Conference on Image Processing, ICIP*, 2011.

[21] R. J. Urbanowicz, M. Meeker, W. La Cava, R. S. Olson, and J. H. Moore, "Relief-based feature selection: Introduction and review," *Journal of Biomedical Informatics*. 2018.

[22] H. Yousefi, F. Nabi, and S. Askari, "Fuzzy clustering and feature selection analysis toward improved identification of MUAP in needle EMG signal," in *2015 22nd Iranian Conference on Biomedical Engineering, ICBME 2015*, 2016.

[23] H. Ghasemzadeh, N. Amini, R. Saeedi, and M. Sarrafzadeh, "Power-aware computing in wearable sensor networks: An optimal feature selection," *IEEE Trans. Mob. Comput.*, 2015.

[24] M. El-Melegy, D. Mohamed, and T. ElMelegy, "Automatic Detection of Tuberculosis Bacilli from Microscopic Sputum Smear Images Using Faster R-CNN, Transfer Learning and Augmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019.

[25] M. El-Melegy, D. Mohamed, T. Elmelegy, and M. Abdelrahman, "Identification of tuberculosis bacilli in ZN-stained sputum smear images: A deep learning approach," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2019.

[26] R. O. Panicker, K. S. Kalmady, J. Rajan, and M. K. Sabu, "Automatic detection of tuberculosis bacilli from microscopic sputum smear images using deep learning methods," *Biocybern. Biomed. Eng.*, 2018.

[27] https://github.com/HamedYousefi-WashU/Tuberculosis-Identification