Journal of Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

Ziehl–Neelsen sputum smear microscopy image database: a resource to facilitate automated bacilli detection for tuberculosis diagnosis

Mohammad Imran Shah
Smriti Mishra
Vinod Kumar Yadav
Arun Chauhan
Malay Sarkar
Sudarshan K. Sharma
Chittaranjan Rout

SPIE.

Mohammad Imran Shah, Smriti Mishra, Vinod Kumar Yadav, Arun Chauhan, Malay Sarkar, Sudarshan K. Sharma, Chittaranjan Rout, "Ziehl–Neelsen sputum smear microscopy image database: a resource to facilitate automated bacilli detection for tuberculosis diagnosis," *J. Med. Imag.* **4**(2), 027503 (2017), doi: 10.1117/1.JMI.4.2.027503.

Ziehl–Neelsen sputum smear microscopy image database: a resource to facilitate automated bacilli detection for tuberculosis diagnosis

Mohammad Imran Shah,[a] Smriti Mishra,[a] Vinod Kumar Yadav,[a] Arun Chauhan,[a] Malay Sarkar,[b] Sudarshan K. Sharma,[c] and Chittaranjan Rout[a,*]
[a]Jaypee University of Information Technology, Department of Biotechnology and Bioinformatics, Waknaghat, Himachal Pradesh, India
[b]Indira Gandhi Medical College, Department of Pulmonary Medicine, Shimla, India
[c]Indira Gandhi Medical College, Department of Pathology, Shimla, India

Abstract. Ziehl–Neelsen stained microscopy is a crucial bacteriological test for tuberculosis detection, but its sensitivity is poor. According to the World Health Organization (WHO) recommendation, 300 viewfields should be analyzed to augment sensitivity, but only a few viewfields are examined due to patient load. Therefore, tuberculosis diagnosis through automated capture of the focused image (autofocusing), stitching of viewfields to form mosaics (autostitching), and automatic bacilli segmentation (grading) can significantly improve the sensitivity. However, the lack of unified datasets impedes the development of robust algorithms in these three domains. Therefore, the Ziehl–Neelsen sputum smear microscopy image database (ZNSM iDB) has been developed, and is freely available. This database contains seven categories of diverse datasets acquired from three different bright-field microscopes. Datasets related to autofocusing, autostitching, and manually segmenting bacilli can be used for developing algorithms, whereas the other four datasets are provided to streamline the sensitivity and specificity. All three categories of datasets were validated using different automated algorithms. As images available in this database have distinctive presentations with high noise and artifacts, this referral resource can also be used for the validation of robust detection algorithms. The ZNSM-iDB also assists for the development of methods in automated microscopy. [DOI: 10.1117/1.JMI.4.2.027503]

Keywords: tuberculosis; computer-aided diagnosis; autofocusing; autostitching; automated microscopy; conventional microscopy image database; bacilli segmentation.

Paper 17055RR received Mar. 2, 2017; accepted for publication Jun. 14, 2017; published online Jun. 30, 2017.

1 Introduction

Tuberculosis, alongside the human immunodeficiency virus (HIV) infection, is the leading cause of death from a single infectious agent.[1] Early diagnosis of tuberculosis is essential to attain better health outcomes.[2] A noninvasive and economical investigation such as a sputum smear microscopy test is considered as a crucial factor in misdiagnosis of this disease, especially in low- and middle-income countries.[3] This sputum smear test is generally performed using fluorescence microscopy (FM) or bright-field microscopy/conventional microscopy (CM). The latter is the most preferred test in low- and middle-income countries due to its accessibility, minimal bio-safety standard, and cost effectiveness.[4,5] This test is also used as the primary technique for detection of tuberculosis in remote areas.[4,5]

According to the World Health Organization (WHO) guidelines, 300 viewfields of a smear slide should be examined in CM within 24 h of collection of a specimen for accurate diagnosis.[6] As manual identification and counting of bacilli using CM is a very time consuming and labor intensive task, it takes 40 min to 3 h to analyze even 40 to 100 viewfield images from a single slide to diagnose a patient as tuberculosis positive or negative.[7] Therefore, the sensitivity of tuberculosis detection varies, and it relies on the experience of microbiologists.[8] The effectiveness of diagnosis is compromised to a significant extent for extrapulmonary, pediatric, or HIV-patients co-infected tuberculosis.[9] All of these shortcomings can be addressed through an automated microscope, which will not only increase the accuracy but also reduces the time of diagnosis.[10,11] Efforts were made to increase the sensitivity of this diagnostic test by incorporating automated methods.[12,13] However, the attainment of success is limited mainly because of the inadequacy of data and dependency of automated methods on image contents.[14]

The automated microscopy for bacilli detection requires efficient algorithms in the following three domains:

i. Autofocusing: In a stack of images captured from a single viewfield with different focuses, an image with the best average focus over the entire viewfield is defined as the focused one. The maximum value of the focus measure function (FMF) corresponds to the best-focused image.[11] This method would facilitate automated capturing of the best-focused image. In recent years, various autofocusing algorithms have been proposed and implemented in microscopy images for diverse biological applications.[11,15,16]
ii. Autostitching: This method stitches viewfields of a smear-slide to form a mosaic or slide map. The WHO recommended analysis of 300 viewfields can be achieved faster and efficiently by automated stitching of overlapping viewfields followed by detection of

*Address all correspondence to: Chittaranjan Rout, E-mail: chittaranjan.rout@juit.ac.in

2329-4302/2017/$25.00 © 2017 SPIE

bacilli in a mosaic (stitched-image) by segmentation methods.[17] Bacilli segmentation algorithms use bacilli shape and size as the potential features to segment the bacilli from other objects.[5,12] Autostitching also facilitates automatic detection of bacilli on the edge by joining half bacillus structures on the boundaries of two different viewfields. Though many autostitching methods were developed,[17,18–20] they were not validated on diverse datasets.

iii. Automatic bacilli segmentation and grading: It is a process of segmentation and counting of bacilli either from viewfield or mosaic. Pattern recognition and machine learning techniques have been used to detect the bacilli in images,[4,5,7,12,21,22] but their efficacy and scopes are limited due to their implementation on nonunified and limited datasets.

Databases and tools have already provided a better diagnosis of cancer and other diseases.[23,24] Databases were used to develop algorithms/methods, which were implemented in computer-aided diagnosis (CAD) systems to obtain a second opinion about a disease.[25–27] The databases and CADs are also effectively used for the early detection of diseases. Keeping in view these accomplishments of databases, the Ziehl–Neelsen sputum smears microscopy image database (ZNSM-iDB) has been developed.[28] This database contains diverse categories of image datasets with both medium and high noise backgrounds. It possesses datasets for all three processes required for automated microscope development. Standard protocols were used to acquire the images,[6] and the datasets were validated by various automated microscopy algorithms to establish their robustness. This database can be used to develop and evaluate efficient and robust algorithms related to automated microscopy.

2 Materials and Methods

2.1 *Data Collection*

Digital images of viewfields from 10 different ZN-stained sputum smear slides (belonging to tuberculosis positive patients) were acquired under the guidance of two expert microscopists. Triplicate data for each category were collected using three different microscopes at 100× magnification (Table 1 and Secs. 2.2.1–2.2.7). The objective lens with 100× magnification was used as the typical *Mycobacterium tuberculosis* bacilli width and length of about 0.5 and $2-4$ μm, respectively.[29] Images were acquired in RGB (red, green, and blue) color space with ".jpg" file-format. Image dimensions in pixels and DPI (dots per inch) are also provided. The dimensions of image represent the number of pixels in an image. Detailed configurations of all the microscopes and acquired image properties are mentioned below:

i. First datasets were acquired using a Labomed Digi 3 digital microscope (MS-1), which features an $L \times 400$ trinocular microscope and an iVu 5100 digital camera module 5.0 megapixel CMOS sensor. The acquired images were 800×600 pixels with bit depth and resolution of 24 (eight per channel) and 120 DPI, respectively, at 100× magnification. The physical size (pixel pitch) of each single pixel is 2.2 μm.

ii. Second datasets were acquired using a Motic BA210 digital microscope (MS-2), which features a Siedentopf type Binocular head and Moticam 2500 digital camera module 5.0 megapixel CMOS sensor. The acquired images were 1280×1024 and 2592×1944 pixels with bit depth and resolution of 24 and 96 DPI, respectively, at 100× magnification. The physical size of a pixel is 2.2 μm.

iii. Third datasets were acquired using an Olympus CH20i digital microscope (MS-3), which features a trinocular microscope and a smartphone digital camera with 16-megapixel BSI-CMOS sensor. The smartphone was attached to the microscope using HY0088 microscope mobile phone interface. The acquired images were 5312×2988 (16 MP), 3984×2988 (12 MP), and 2048×1152 (2.4 MP) pixels with bit depth and resolution of 24 and 72 DPI, respectively, at 100× magnification. The physical size of a pixel is 1.12 μm.

2.2 *Description of Datasets*

The architecture of ZNSM-iDB with its applications is presented in Fig. 1. This database contains seven different categories of

Table 1 Category-wise presentation of datasets available in ZNSM-iDB

Group	Category of data	No. of digital images from different microscope (MS)		
		MS-1	MS-2	MS-3
1.	Autofocusing dataset[a]	9 stacks	10 stacks	30 stacks
2.	Overlapping viewfields for autostitching	7 sets (50 to 90 images/set)	6 sets (50 images/set)	10 sets (50 images/set)
3.	Manually segmented bacilli in a viewfield	2 sets (50 images/set)	2 sets (50 images/set)	2 sets (50 images/set)
4.	Viewfields without bacilli	50	50	50
5.	Single or few bacilli	100	100	100
6.	Overlapping (occluded) bacilli	200	200	200
7.	Over-stained viewfields with bacilli and artifacts	250	250	250

[a]Each stack contains 20 images.

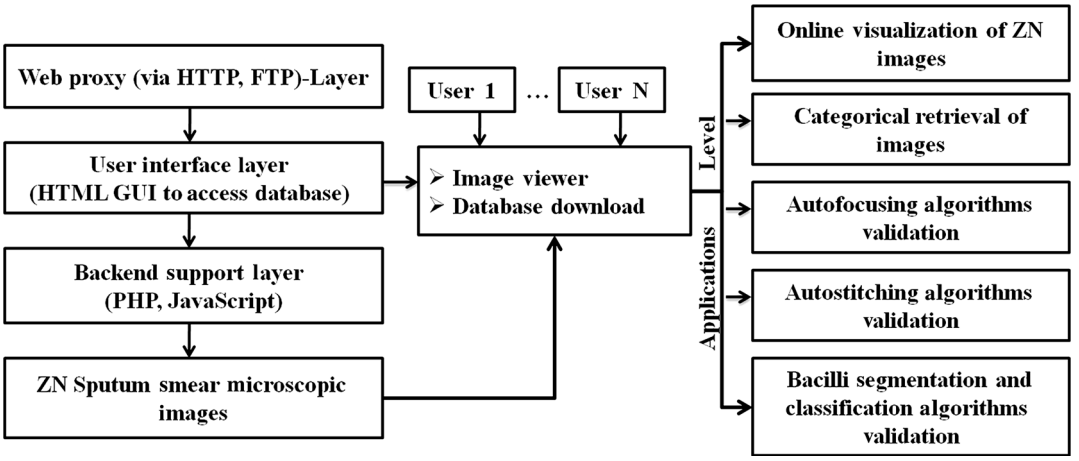

Fig. 1 Architecture and applications of ZNSM-iDB.

digital images in triplicate sets (one set from each microscope), which may be separately visualized or retrieved for further processing and applications (Table 1). Multiple and diverse datasets were provided so that the robustness of developed algorithms can be evaluated.

A detailed description for each category of data is given below.

2.2.1 *Autofocusing dataset*

In this category, every stack of images is restricted to a single viewfield. Each stack contains at least 20 images captured at different focus lengths, in which one is marked as the best focused, while the others are unfocused to different extents. The 10th image is the best-focused one in most of the stacks.

2.2.2 *Overlapping viewfields for auto-stitching*

Adjacent overlapping viewfields can be stitched to make the mosaic or slide map using image processing techniques. In this dataset, 10 overlapping viewfield images were acquired in a row and then the slide was moved left or right to take the images of the next row (Fig. 2).

2.2.3 *Manually segmented bacilli in a viewfield*

This dataset contains viewfields with manually segmented or marked bacilli. Different shapes were used for marking such as a circle or oval shape for single bacillus, square or rectangle for occluded bacilli, diamond for unclassified red structures, and hexagon for the artifacts [Fig. 3(a)].

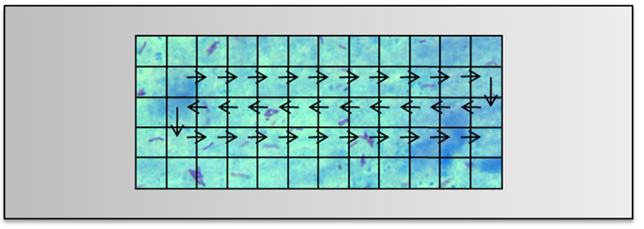

Fig. 2 A depiction of direction in which the images were acquired from a ZN-stained slide. Each square box corresponds to a viewfield.

2.2.4 *Viewfields without bacilli*

Images in this group range from medium to very high-density background depending upon the presence of over-staining

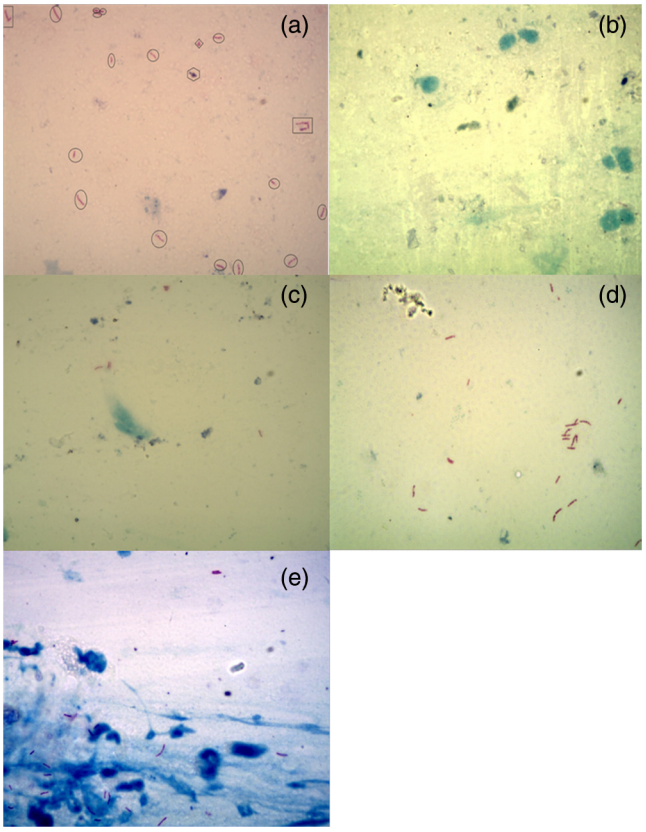

Fig. 3 Sample images of five different category datasets available in ZNSM-iDB. (a) Manually segmented viewfield, (b) viewfield without bacilli, (c) viewfield with single or few bacilli, (d) viewfield with occluded bacilli, and (e) over-stained viewfields with bacilli and artifacts.

and artifacts, but the viewfield does not contain any bacilli [Fig. 3(b)].

2.2.5 *Viewfields with a single or few bacilli*

In this group of images, the number of bacilli in each viewfield varies from 1 to 10. This group also has medium to high-density backgrounds [Fig. 3(c)].

2.2.6 *Viewfields with overlapping (occluded) bacilli*

In many instances, two or more bacilli are overlapped at the same position and form an occluded bacilli cluster. Images in this category are diverse in terms of medium to high-density backgrounds [Fig. 3(d)].

2.2.7 *Over-stained viewfields with bacilli and artifacts*

Sometimes, the quality of ZN-stained CM images is not good due to the presence of artifacts and over-staining. Therefore, bacilli detection is difficult using a segmentation method, and only robust methods can produce a better performance. In this category, more than 200 images from each microscope are provided that contain over-stained (blue) regions with artifacts and/or bacilli [Fig. 3(e)].

2.3 *Data Validation*

ZNSM-iDB resource contains autofocusing, autostitching, and bacilli segmentation and classification image datasets. Some algorithms/methods reported in these three domains were also implemented on the datasets.

Twenty-four FMFs were implemented on autofocusing datasets to determine the best-focused one in each stack of images.[30] The performance of FMFs in classifying a stack is determined using different criteria such as accuracy, focus error, false maximum, and full-width at half-maximum. The robustness of FMFs to the different imaging conditions (median filtering, noise addition, contrast reduction, saturation increment, and nonuniform illumination addition) was also determined.

Autostitching of overlapping viewfields datasets was performed and reported.[31] The overlapping subparts of the viewfields were stitched together using scale-invariant feature transform (SIFT) feature extraction and random-sample-consensus (RANSAC) selection method. The divide and conquer algorithm was implemented for faster stitching and mosaic formation. Comparisons of similarities between the original and stitched images were performed using correlation (COR), structural similarity (SSIM), and feature similarity (FSIM) measures.

Bacilli segmentation and classification using the watershed algorithm[32] was performed on ZNSM-iDB database. Forty images from three microscopes (MS-1, MS-2, and MS-3) were randomly extracted and grouped into medium- and high-density background datasets. The shape and size of the objects were determined to filter the true bacilli.[33] Similarly, the watershed algorithm was implemented on 30 randomly extracted images from the smartphone enabled microscope (MS-3) to segment bacilli.[34] In both studies, sensitivity and specificity of the watershed algorithm were calculated. In an another study, images were divided into four groups based on the infection level (Table 2),[35] and sensitivity and specificity of the watershed segmentation method for classifying an image as TB positive or negative were determined for each group.

Table 2 Grading of viewfields on the basis of infection level.

Number of bacilli	Number of viewfields to be examined	Grading
1 to 9 in 100 viewfields	100	Scanty[a]
10 to 99 in 100 viewfields	100	1+
1 to 10 in each viewfield	50	2+
>10 in each viewfield	20	3+

[a]Report exact number of bacilli present in the viewfields.

Sensitivity and the precision rate of this segmentation method for identifying true bacilli were determined for each group. Furthermore, the discordance rate was calculated for the watershed segmentation method to evaluate the percent of pairs where the observation with TB-positive has a lower predicted probability than TB-negative.[36,37] The predicted probability was calculated in a binary logistic regression model.[36]

3 Results and Discussion

3.1 *Applications of the Data Resource*

The ZN sputum smear microscopy image database (ZNSM-iDB) is a unified image resource that has the potential to assist in the development of algorithms and tools related to automated grading (computer aided detection of bacilli) of smear slides using image processing techniques (Fig. 4). The database is freely available from Ref. 28, and a tutorial is provided for effective exploration of diverse datasets (Table 1 and Fig. 5). As most of the tuberculosis diagnosis centers have a high patient load, false negative rates resulting in poor sensitivity is the stark reality.[7] The sensitivity of the CM varies from 0.32 to 0.94, whereas it ranges from 0.52 to 0.97 for FM.[38] The specificity is approximately similar in both microscopy tests ranging from 0.94 to 1.0. As examination of a few viewfields is considered as an important factor for the poor sensitivity of this bacteriological test, automation can improve the accuracy of a diagnosis.

In addition to automatic loading of the slide, an automated microscope for tuberculosis detection requires a combination of robust methods for automated capturing of the focused image (autofocusing), stitching of viewfields to form a mosaic (autostitching), and automatic bacilli segmentation and grading. Groups 1, 2, and 3 datasets can be used for the development and validation of bacilli grading algorithms (Table 1). The remaining four categories (groups 4, 5, 6, and 7) of datasets are provided to streamline the sensitivity and specificity of the bacilli detection method as images in these categories are diverse in terms of contents (medium to a high-density background, occluded bacilli, few bacilli, without bacilli, over-staining, etc.). As 300 viewfields should be analyzed for efficient diagnosis, automation can significantly improve the CM performance through analysis of a large number of viewfields.[12]

3.2 *Validation*

Autofocusing, autostitching, and bacilli segmentation and classification algorithms were implemented on ZNSM-iDB datasets.

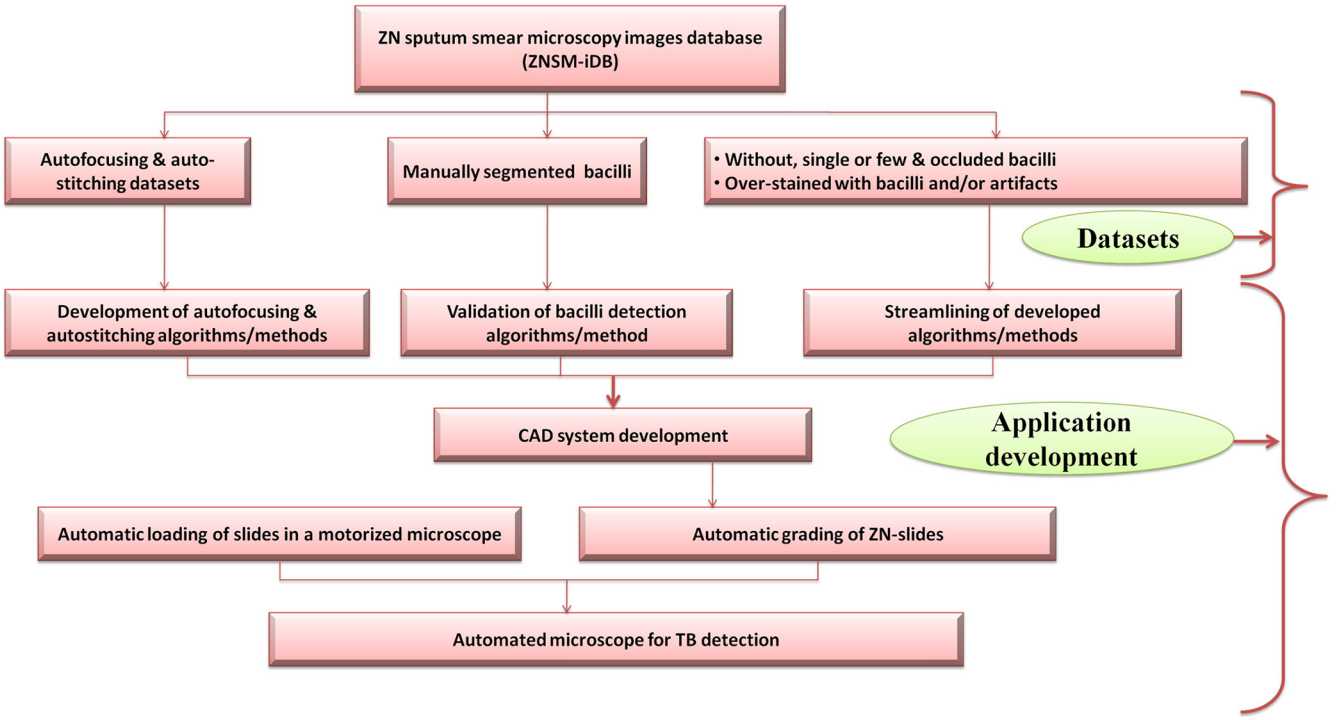

Fig. 4 Application of ZNSM-iDB database in automated microscopy.

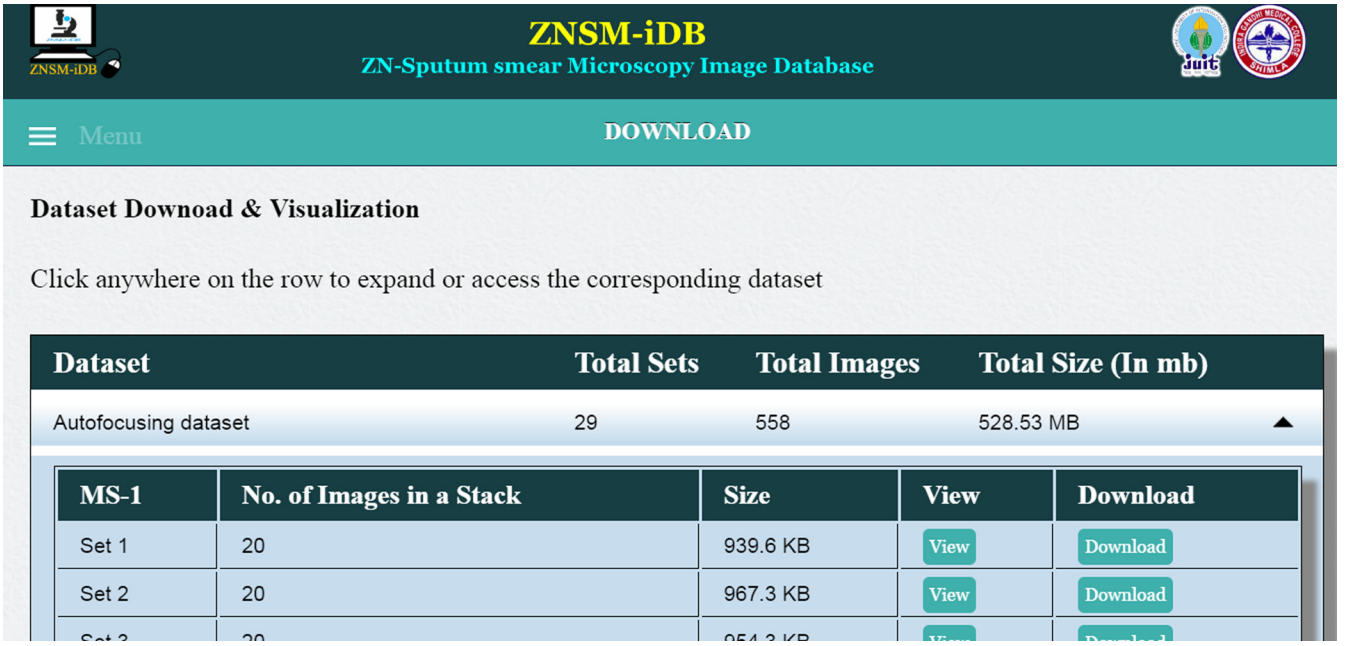

Fig. 5 Image visualization and data download page.

3.2.1 *Performance of focus measure functions*

Comprehensive and comparative analyses of 24 FMFs have shown that Gaussian derivative, Tenengrad gradient, steerable filters, and Hemli and Scherer's mean are the most robust and accurate FMFs in all three microscopes to determine the best focus in CM images. These four FMFs were also robust to image distortions (noise addition, contrast reduction, saturation increment, and nonuniform illumination addition).[30] In an earlier study, it was also reported that the Tenengrad method produced a better FMF for CM images.[11]

3.2.2 *Autostitching of viewfields*

Different autostitching methods were evaluated on ZNSM-iDB datasets acquired using MS-1. The SIFT feature extraction and RANSAC feature selection methods were used to generate a mosaic.[31] Similarity scores between the stitched and original

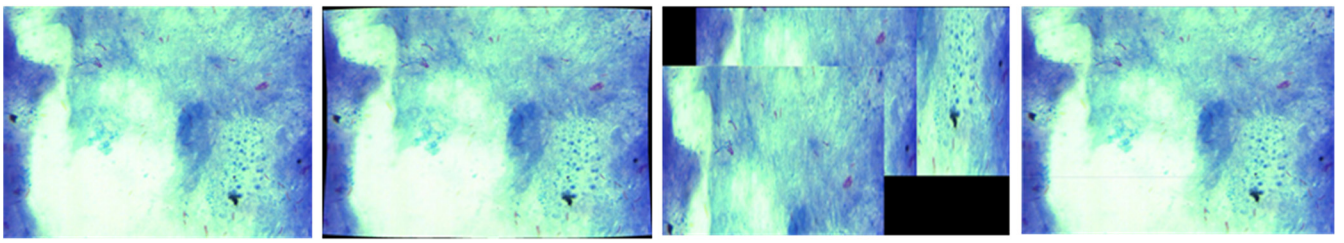

Fig. 6 Comparison between images of original viewfield with stitched mosaic. (a) Original image, (b) mosaic formed using Autostitch, (c) MicroMos, and (d) divide-and-conquer. Reproduced this figure from Ref. 31 with permission.

images were 0.997, 0.988, and 0.989 using COR, SSIM, and FSIM measures, respectively. These results were significantly better than the performances of Autostitch[20] and MicroMos[38] software (Fig. 6).

3.2.3 *Bacilli segmentation and classification*

Bacilli segmentation and classification was performed using a watershed segmentation method on ZNSM-iDB datasets.[33,34] The sensitivity and specificity of this method for classifying a medium density background image as tuberculosis positive or negative were 100% and 93%, respectively, while for a high-density background the sensitivity remained unchanged, but specificity was reduced to 72% due to over-staining and artifacts.[33] Similarly, the sensitivity and specificity of this segmentation method for classifying a smartphone enabled microscopic images (medium to high-density background) as tuberculosis bacilli positive or negative were 93.3% and 87%, respectively.[34]

The sensitivity, specificity, and precision rate of watershed segmentation on different infection levels (Scanty, 1+, 2+, and 3+) are depicted in Table 3. The performance of this segmentation method is related to the level of infection where a higher level of infection yields a better performance. Furthermore, this segmentation provided a discordant rate of 3.73% with respect to tuberculosis positive and negative cases.

Table 3 Performance of watershed segmentation method on different infection levels.

	TB positive or negative (%)		True bacilli (%)	
Grade	Sensitivity[a]	Specificity[a]	Sensitivity[b]	Precision[b]
Scanty	42.86	74.73	55.56	26.19
1+	81.25	75.29	55.56	52.4
2+	92.86	NA[c]	61.73	72
3+	100	NA	90.23	79.09

[a]Sensitivity and specificity of watershed segmentation method for classifying an image as TB positive or negative.
[b]Sensitivity and precision rate of watershed segmentation method for identifying true bacilli in the viewfield images.
[c]Datasets does not contain tuberculosis negative viewfield images.

4 Conclusion

The ZNSM-iDB is well diverse in terms of image acquisition technology and content. These images were acquired from three microscopes with different configurations and scopes. Since autofocusing, autostitching, and automated grading methods are necessary for automated microscope development, this resource may work as a standard platform to compare existing methods as well as to develop new algorithms (Fig. 4). CAD helps to achieve better sensitivity and specificity for complicated diseases.[39] An established CAD system, "TBDx," is available for tuberculosis screening in FM images,[40] but most of the tuberculosis-endemic countries are using CM due to its accessibility, minimal bio-safety standard, and cost-effectiveness. Kinyoun-stained sputum smear bright-field microscopic images are available for autofocusing and bacilli segmentation,[41] but bacilli in these images are less visible leading to poorer sensitivity than those in ZN-stained images.[42]

The ZBSM-iDB contains seven different categories of diverse viewfield images for the development of efficient and robust algorithms for automated capturing of focused image (autofocusing), stitching of viewfields to form mosaics (autostitching), and automatic bacilli segmentation and grading. The MS-3 image datasets can assist a smartphone-based inexpensive disease diagnosis system and have the potential to be used in the remote areas of TB-endemic countries where smartphones are widely available. Additionally, the advantage of using a smartphone camera is that it can be used simultaneously for automatic bacilli detection using image processing methods as well as maintaining an electronic health record.[43] Validation of ZNSM-iDB datasets was also performed using autofocusing, autostitching, and bacilli segmentation and classification methods. Performances of these algorithms suggested that the datasets are diverse and can be used for the development of automated CM. Diverse data available in this resource can facilitate the development of algorithms/model on one microscope, while the validation of the same method can be done on another microscope(s). The ZNSM-iDB is expected to serve as a referral resource for the development and validation of robust segmentation algorithms as the images possess high noise and artifacts. This resource also assists research communities in the development of methods in the domains of automated microscopy.

Disclosures

There is no funding involved in this work. The authors declare that they have no conflict of interest. Use of these patient samples was approved by the ethical committee of Indira Gandhi Medical College, Shimla, India.

Acknowledgments

We would like to thanks Jaypee University of Information Technology, Solan for providing a doctoral fellowship.

References

1. WHO, TB Global Report (2015).
2. C. T. Sreeramareddy et al., "Time delays in diagnosis of pulmonary tuberculosis: a systematic review of literature," *BMC Infect. Dis.* **9**(1), 91 (2009).
3. P. M. Small and M. Pai, "Tuberculosis diagnosis—time for a game change," *N. Engl. J. Med.* **363**(11), 1070–1071 (2010).
4. P. J. Tadrous, "Computer-assisted screening of Ziehl–Neelsen-stained tissue for mycobacteria algorithm design and preliminary studies on 2,000 images," *Am. J. Clin. Pathol.* **133**(6), 849–858 (2010).
5. R. Khutlang et al., "Automated detection of tuberculosis in Ziehl–Neelsen-stained sputum smears using two one-class classifiers," *J. Microsc.* **237**(1), 96–102 (2010).
6. F. Ba and H. L. Rieder, "A comparison of fluorescence microscopy with the Ziehl–Neelsen technique in the examination of sputum for acid-fast bacilli," *Int. J. Tuberc. Lung Dis.* **3**(12), 1101–1105 (1999).
7. M. Sotaquira, L. Rueda, and R. Narvaez, "Detection and quantification of bacilli and clusters present in sputum smear samples: a novel algorithm for pulmonary tuberculosis diagnosis," in *Int. Conf. on Digital Image Processing*, pp. 117–121, IEEE (2009).
8. A. Kimura et al., "Evaluation of autofocus functions of conventional sputum smear microscopy for tuberculosis," in *Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3041–3044 (2010).
9. D. Shingadia and V. Novelli, "Diagnosis and treatment of tuberculosis in children," *Lancet Infect. Dis.* **3**(10), 624–632 (2003).
10. W. E. Mesker et al., "Supervised automated microscopy increases sensitivity and efficiency of detection of sentinel node micrometastases in patients with breast cancer," *J. Clin. Pathol.* **57**(9), 960–964 (2004).
11. O. A. Osibote et al., "Automated focusing in bright-field microscopy for tuberculosis detection," *J. Microsc.* **240**(2), 155–163 (2010).
12. M. G. Forero, G. Cristóbal, and M. Desco, "Automatic identification of Mycobacterium tuberculosis by Gaussian mixture models," *J. Microsc.* **223**(2), 120–132 (2006).
13. M. G. Costa et al., "Automatic identification of mycobacterium tuberculosis with conventional light microscopy," in *30th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 382–385 (2008).
14. S. Pertuz, D. Puig, and M. A. Garcia, "Analysis of focus measure operators for shape-from-focus," *Pattern Recognit.* **46**(5), 1415–1432 (2013).
15. C. C. Gu et al., "Region sampling for robust and rapid autofocus in microscope," *Microsc. Res. Tech.* **78**(5), 382–390 (2015).
16. J. M. Mateos-Pérez et al., "Comparative evaluation of autofocus algorithms for a real-time system for automatic detection of Mycobacterium tuberculosis," *Cytometry Part A* **81**(3), 213–221 (2012).
17. F. Yang, Z. S. Deng, and Q. H. Fan, "A method for fast automated microscope image stitching," *Micron* **48**, 17–25 (2013).
18. X. Bai, X. Ning, and L. Wang, "Analysis and comparison of feature detection and matching algorithms for rovers vision navigation," in *8th IEEE Int. Symp. on Instrumentation and Control Technology (ISICT)*, pp. 66–71 (2012).
19. S. Ali and M. Hussain, "Panoramic image construction using feature based registration methods," in *15th Int. Multitopic Conf. (INMIC)*, pp. 209–214 (2012).
20. M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *Int. J. Comput. Vision* **74**(1), 59–73 (2007).
21. C. F. F. C. Filho et al., "Automatic identification of tuberculosis mycobacterium," *Res. Biomed. Eng.* **31**(1), 33–43 (2015).
22. E. Priya and S. Srinivasan, "Automated identification of tuberculosis objects in digital images using neural network and neuro fuzzy inference systems," *J. Med. Imaging Health Inf.* **5**(3), 506–512 (2015).
23. S. G. Armato, III et al., "The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans," *Med. Phys.* **38**(2), 915–931 (2011).
24. A. Depeursinge et al., "Building a reference multimedia database for interstitial lung diseases," *Comput. Med. Imaging Graphics* **36**(3), 227–238 (2012).
25. Y. Song et al., "Feature-based image patch approximation for lung tissue classification," *IEEE Trans. Med. Imaging* **32**(4), 797–808 (2013).
26. A. O. de Carvalho Filho et al., "Automatic detection of solitary lung nodules using quality threshold clustering, genetic algorithm and diversity index," *Artif. Intell. Med.* **60**(3), 165–177 (2014).
27. S. Diciotti et al., "Automated segmentation refinement of small lung nodules in CT scans by local shape analysis," *IEEE Trans. Biomed. Eng.* **58**(12), 3418–3428 (2011).
28. M. I. Shah et al., "Ziehl–Neelsen Sputum smear Microscopy image DataBase (ZNSM-iDB)," 2017, http://14.139.240.55/znsm (27 June 2017).
29. M. K. Osman et al., "Colour image segmentation of tuberculosis bacilli in Ziehl–Neelsen-stained tissue images using moving K-mean clustering procedure," in *Fourth Asia Int. Conf. on Mathematical/Analytical Modelling and Computer Simulation (AMS)*, pp. 215–220 (2010).
30. M. I. Shah et al., "Identification of robust focus measure functions for the automated capturing of focused images from Ziehl–Neelsen stained sputum smear microscopy slide (unpublished)," Jaypee University of Information Technology, India (2017).
31. A. Chauhan, "Development of efficient computer aided diagnosis methods for tuberculosis diagnosis (Unpublished doctoral thesis)," Jaypee University of Information Technology, India (2016).
32. S. Beucher and C. Lantuejoul, "Use of watersheds in contour detection," in *Int. Workshop on Image Processing, Real-Time Edge and Motion Detection/Estimation* (1979).
33. M. I. Shah et al., "Automatic detection and classification of tuberculosis bacilli from ZN-stained sputum smear images using watershed segmentation," in *Int. Conf. on Signal Processing (ICSP)* (2016).
34. M. I. Shah et al., "Automatic detection and classification of tuberculosis bacilli from camera-enabled Smartphone microscopic images," in *Fourth Int. Conf. on Parallel, Distributed and Grid Computing (PDGC)*, pp. 287–290, IEEE (2016).
35. Revised National Tuberculosis Control Programme, *Module for Laboratory Technicians*, p. 15, Directorate General of Health Services, Ministry of Health and Family Welfare, New Delhi (1999).
36. C. M. Dayton, "Logistic regression analysis," *Statistics* 474–574 (1992).
37. D. Bhalla, "Calculating concordant, discordant and tied pairs," 2014, http://www.listendata.com/2014/08/modeling-tips-calculating-concordant.html (11 April 2017).
38. F. Piccinini, A. Bevilacqua, and E. Lucarelli, "Automated image mosaics by non-automated light microscopes: the MicroMos software tool," *J. Microsc.* **252**(3), 226–250 (2013).
39. A. Depeursinge et al., "Near-affine-invariant texture learning for lung tissue analysis using isotropic wavelet frames," *IEEE Trans. Inf. Technol. Biomed.* **16**(4), 665–675 (2012b).
40. J. J. Lewis et al., "'Proof-of-concept' evaluation of an automated sputum smear microscopy system for tuberculosis diagnosis," *PLoS One* **7**(11), e50173 (2012).
41. M. G. F. Costa et al., "A sputum smear microscopy image database for automatic bacilli detection in conventional microscopy," in *36th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2841–2844 (2014).
42. T. L. Sawadogo et al., "Comparison of Kinyoun, auramine O, and Ziehl–Neelsen staining for diagnosing tuberculosis at the National Tuberculosis Center in Burkina Faso," *Med. Sante Trop.* **22**(3), 302–306 (2011).
43. D. N. Breslauer et al., "Mobile phone based clinical microscopy for global health applications," *PloS One* **4**(7), e6320 (2009).

Mohammad Imran Shah received his BSc degree in bioinformatics from DAVV University, Indore, and his MSc degree in bioinformatics from Manipal University, Karnataka. Currently, he is a PhD student in biotechnology and bioinformatics at Jaypee University of Information Technology, Solan, India. His present interests include image processing, medical image analysis, computer aided diagnosis, and bioinformatics databases and tools development.

Smriti Mishra received her BSc degree from DAVV University, Indore, India, and her MSc degree in bioinformatics from Manipal

University, Manipal, Karnatka, India. Currently, she is a PhD research scholar in the Department of Biotechnology and Bioinformatics at Jaypee University of Information Technology (JUIT). Her present research interests include algorithm development, artificial intelligence, machine learning, and bioinformatics databases and tools development.

Vinod Kumar Yadav received his BTech degree in bioinformatics from the Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Solan, India. His present interests include database development for biomedical data.

Arun Chauhan received his BSc degree in life-sciences from Lucknow University, Lucknow, his MSc degree in bioinformatics from Chhatrapati Shahu Ji Maharaj University, Kanpur, and his MTech degree in information technology from IIIT Allahabad. Currently, he is a PhD student in the Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Solan, India. His present interests include computer aided diagnosis, medical image analysis, and bioinformatics.

Malay Sarkar received his MBBS degree from RG Kar Medical College, Kolkata, India, and his MD degree in TB and chest diseases in 1998. Currently, he is working as a professor and head in the Department of Pulmonary Medicine, Indira Gandhi Medical College, Shimla, India. His research interests include computer aided diagnosis, chronic obstructive pulmonary disease, interstitial lung diseases, epidemiology, and development of clinical decision support systems.

Sudarshan K. Sharma received his MBBS and MD degrees from Indira Gandhi Medical College, Shimla, India. Currently, he is working as a professor in the Department of Pathology, Indira Gandhi Medical College, Shimla, India. His research interests include histopathology, cytology and computer aided diagnosis.

Chittaranjan Rout received his PhD in computational chemistry from the Department of Chemistry, University of Delhi, India. He worked as a research associate at the School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi. He is currently working as an associate professor at Jaypee University of Information Technology, Solan. His research interests include image processing, computer aided diagnosis, prediction of vaccine candidates and drug targets, computational drug development, and development of clinical decision support systems.