



doi • 10.5578/tt.2025031113
Tuberk Toraks 2025;73(3):165-177
Received: 30.04.2025 • Accepted: 20.06.2025

RESEARCH ARTICLE

Enhancing tuberculosis diagnosis: A deep learning-based framework for accurate detection and quantification of TB *bacilli* in microscopic images

Dinesh Jackson Samuel
RAVINDRAN¹(ID)
Rajesh Kanna
BASKARAN²(ID)

¹ Department of Mathematics and Computer Science, Pittsburg State University, Kansas, United States of America

² Department of Computer Science, Rajiv Gandhi National Institute of Youth Development, Chennai, India

Cite this article as: Ravindran DJS, Baskaran RK. Enhancing tuberculosis diagnosis: A deep learning-based framework for accurate detection and quantification of TB bacilli in microscopic images. *Tuberk Toraks* 2025;73(3):165-177.

Address for Correspondence

Dr. Dinesh Jackson Samuel RAVINDRAN
Department of Mathematics and
Computer Science, Pittsburg State University
KANSAS-USA
e-mail: dravindran@pittstate.edu

This is an open-access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution, and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes (<http://creativecommons.org/licenses/by-nc/4.0/>).

Data Sharing Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

©Copyright 2025 by Tuberculosis and Thorax.
Available online at www.tuberktoraks.org

ABSTRACT

Enhancing tuberculosis diagnosis: A deep learning-based framework for accurate detection and quantification of TB *bacilli* in microscopic images

Introduction: Tuberculosis (TB), a highly contagious disease, remains one of the leading causes of death globally. The proposed computer-assisted TB detection system enhances diagnostic accuracy and efficiency by integrating deep learning and segmentation techniques.

Materials and Methods: It consists of two key subsystems: Automated field-of-view (FOV) recognition and TB bacilli segmentation. Using a motorized microscopic stage, the system systematically captures Ziehl-Neelsen-stained sputum smear images at 100x magnification. A customized Inception V3 model with transfer learning identifies FOVs containing TB bacilli, reducing variability and manual effort. Segmentation techniques, including coarse-level thresholding and shape descriptors like area, perimeter, and eccentricity, refine bacilli detection and eliminate artifacts.

Results: This study highlights the significant potential of deep learning and image processing techniques in advancing medical diagnostics, particularly TB detection. This framework has the potential to improve clinical outcomes and support global TB eradication efforts by providing a reliable tool for early TB diagnosis.

Conclusion: The system achieved a mean receiver operating characteristic score of 0.9505, a precision of 0.924, a recall of 0.882, and an F1 score of 0.902, demonstrating its potential to improve TB screening, particularly in resource-limited settings. By minimizing reliance on skilled technicians and enhancing diagnostic reliability, this approach offers a scalable solution for effective TB detection and severity assessment.

Key words: Tuberculosis detection; sputum smear analysis; field-of-view recognition; TB bacilli segmentation; automated microscopy

ÖZ

Tüberküloz tanısında doğruluğun artırılması: Mikroskopik görüntülerde TB bacilli tespiti ve sayımı için derin öğrenme tabanlı bir yöntem

Giriş: Son derece bulaşıcı bir hastalık olan tüberküloz (TB), küresel çapta önde gelen ölüm nedenlerinden biri olmaya devam ediyor. Önerilen bilgisayar destekli TB tespit sistemi, derin öğrenme ve segmentasyon tekniklerini entegre ederek tanılabilirliği artırıyor.

Materyal ve Metod: Sistem iki temel alt sistemden oluşur: Otomatik görüş alanı (FOV) tanıma ve TB bacilli segmentasyonu. Motorlu bir mikroskopik aşama kullanarak sistem, 100x büyütmede Ziehl-Neelsen boyalı balgam yayma görüntülerini sistematik olarak yakalar. Transfer öğrenimi olan özelleştirilmiş bir Inception V3 modeli, TB bacilli içeren FOV'ları belirleyerek değişkenliği ve manuel çabayı azaltır. Alan, çevre ve eksantriklik gibi kaba seviye eşikleme ve şekil tanımlayıcıları da dahil olmak üzere segmentasyon teknikleri, bacilli tespitini iyileştirir ve eserleri ortadan kaldırır.

Bulgular: Bu çalışma, derin öğrenme ve görüntü işleme tekniklerinin tıbbi teşhisleri, özellikle de TB tespitini ilerletmedeki önemli potansiyelini vurgulamaktadır. Bu çalışma, erken TB teşhisi için güvenilir bir araç sunarak, klinik sonuçları iyileştirme ve küresel TB eradikasyon çabalarını destekleme potansiyeline sahiptir.

Sonuç: Sistem, 0.9505'lik ortalama bir ROC puanı, 0.924'lük bir hassasiyet, 0.882'lik bir geri çağırma ve 0.902'lik bir F1 puanı elde ederek, özellikle kaynak kısıtlı ortamlarda TB taramasını iyileştirme potansiyelini gösterdi. Nitelikli teknisyenlere olan bağımlılığı en aza indirerek ve tanı güvenilirliğini artırarak, bu yaklaşım etkili TB tespiti ve ciddiyet değerlendirmesi için ölçeklenebilir bir çözüm sunar.

Anahtar kelimeler: Tüberküloz tespiti; balgam yayma analizi; görüş alanı tanıma; TB bacilli segmentasyonu; otomatize mikroskopi

INTRODUCTION

Tuberculosis (TB), caused by *Mycobacterium tuberculosis* (MTB), remains a major global health challenge despite medical advances. The disease, primarily transmitted via airborne droplets, predominantly affects the lungs, accounting for up to 85% of cases, but can also involve other organs such as the skin, bones, kidneys, and brain. According to the World Health Organization (WHO), approximately 10 million people develop active TB annually, resulting in nearly two million deaths. Accurate and timely diagnosis is vital to controlling the spread of TB. Conventional diagnostic methods like culture, Ziehl-Neelsen (ZN) staining, and tuberculin skin tests are reliable but often slow and labor-intensive. In resource-constrained settings, the demand for high-throughput screening further strains diagnostic workflows, increasing the likelihood of human error and delaying treatment.

Microscopy remains a cornerstone in TB detection. Sputum smears are stained and analysed manually by trained technicians, who scan randomly from 6.050 microscopic fields of view (FOVs) per specimen. As per WHO guidelines, two samples per patient must be screened to ensure diagnostic accuracy. This process is time-consuming, induces fatigue, and often leads to reduced sensitivity and specificity in high-prevalence regions. Acid-fast bacilli (AFB) count in FOVs helps classify TB severity: >10 AFBs/FOV indicates a 3+ level; 1-10 AFBs in 100 FOVs indicates 2+ or 1+ levels, while 1-9 AFBs across 100 FOVs denotes a scanty infection (1). Manual screening of

this scale is unsustainable under heavy workloads, demanding the need for automated, AI-driven diagnostic solutions.

Recent advances in image processing and machine learning offer promising alternatives (2). Object detection and classification techniques powered by neural networks have demonstrated high accuracy, especially when training and testing data share similar acquisition parameters (3). Protocol adherence and technician shortages, however, continue to limit traditional microscopy's diagnostic potential. Several machine learning models have been developed for TB detection. Kusworo et al. have applied Otsu thresholding to NTSC-transformed grayscale images for feature extraction (4). Osman et al. have used a pixel classifier and k-means clustering with a Hybrid Multi-Layered Perceptron for classification (5). Bayesian techniques, CY color models, and two-level segmentation strategies have also shown effectiveness in isolating and identifying TB bacilli (6-8). Convolutional Neural Networks (CNNs) have become a key technology in TB bacilli detection. Compared to shallow networks, CNNs require fewer parameters and provide higher classification performance (9). Transfer learning has further boosted their accuracy (10,11). Yang et al. combined CNN and logistic regression to classify AFB from stained tissues, while Panicker et al.'s DenseNet-based model achieved 99.7% accuracy on an open-source dataset using fivefold cross-validation (12,13). Additional approaches have incorporated fuzzy logic and hybrid models for enhanced segmentation and classification (14).

This study presents a CNN-based system designed to automate the detection and classification of TB *bacilli*, aiming to improve diagnostic efficiency and reliability in high-burden settings. The contributions of the work are presented below.

1. Automated microscopic imaging: A motorized microscopic stage with programmable scanning patterns (zigzag, spiral, etc.) automates FOV acquisition, increasing throughput and minimizing human intervention during specimen examination.
2. Deep learning-based recognition: A custom-tuned Inception V3 deep neural network combined with transfer learning and an Support Vector Machine (SVM) classifier effectively distinguishes TB-infected from non-infected FOVs, enhancing diagnostic accuracy.
3. Efficient feature extraction and classification: Transfer learning using pre-trained ImageNet weights allows for fast and accurate microbial classification. SVM parameters are optimized using grid search and heuristics for precise classification.
4. Robust *bacilli* detection and counting: The system integrates advanced preprocessing, two-level segmentation, morphological filtering, and shape descriptor analysis to accurately detect and count TB *bacilli*, aiding severity estimation.
5. Validated with diverse datasets: The system was trained and tested on over 1,200 annotated images sourced from real clinical specimens and public databases, achieving a high ROC value (0.9505) and demonstrating generalizability across different data sources.

MATERIALS and METHODS

Automated microscopy has significantly advanced digital image acquisition for medical diagnostics, particularly for diseases like TB. In microscopic sputum smear analysis, capturing multiple FOVs is essential to build a TB dataset. This is achieved using a digital camera mounted on the microscope eyepiece and a programmable motorized stage to automate FOV scanning. While commercial motorized stages are available, they are often costly and limited in flexibility. Earlier solutions, such as the 3-axis motorized stage introduced by Champbell et al., employed open-loop stepper motors for X, Y, and Z-axis movement but lacked the precision offered by

feedback mechanisms (15). Automated stages are also widely used in applications such as cell migration tracking and stereological measurements like cell counting and morphology analysis (16,17). For high-precision needs, systems with submicron resolution have been developed. Bhakti et al. enhanced microcontroller precision using the FT232RL chip and ATMEGA 8 software, achieving resolutions of approximately 0.198 $\mu\text{m}/\text{step}$ horizontally and 0.197 $\mu\text{m}/\text{step}$ vertically (18).

In addition to hardware advancements, deep neural networks have revolutionized image classification. AlexNet was a landmark CNN model that achieved top performance in the 2012 ImageNet challenge (19). It introduced deep convolutional layers and large filter banks to effectively learn hierarchical features. Subsequent developments, such as the Inception architecture, optimized local receptive fields and computation using parallel multi-scale filters (1×1 , 3×3 , and 5×5) and dimensionality reduction through 1×1 convolutions (20,21). Inception V3 further improved upon this with auxiliary classifiers to enhance convergence and accuracy (22,23). Recently, pulmonary TB diagnosis using an intelligent microscopy scanner and image recognition model for improved AFB detection in smears has been proposed by Chen et al., incorporating artificial intelligence and machine learning (24). Although computationally intensive, a study on detecting TB *bacilli* using the YOLOv8 model was also proposed. An additional detection head was added to the YOLO architecture specifically for identifying TB *bacilli*. The model demonstrated good performance, achieving a precision of 85.7% in *bacilli* detection (25). Fu et al. have proposed an AI-based TB detection system that also classifies the severity of the infection based on WHO grading. However, the automated scanning tool in their system had a relatively low sensitivity of 77.4% (26). An advanced fuzzy-based algorithm was proposed by Samuel et al. for categorizing the severity level of TB after detecting and counting TB *bacilli* (27).

The proposed system uses these advances to build a fully automated TB diagnosis pipeline. It integrates a programmable microscopic stage for autonomous specimen scanning with a deep learning-based recognition module (28). The scanning path is configurable, allowing the system to function without human intervention. For recognition, pre-trained models like VGG and Inception V3 are fine-tuned using domain adaptation and transfer learning. Following FOV

classification, a two-stage segmentation process is applied to detect AFB in ZN-stained slides. Initial RGB thresholding performs coarse segmentation, followed by Sauvola's local thresholding for fine segmentation (29). Morphological features, such as area, perimeter, convexity, and axis length are then extracted to accurately identify and count TB *bacilli*, providing clinically relevant diagnostic information.

Automated microscopic TB detection system

The proposed microscopic TB detection system is designed to assist in rapid disease diagnosis, particularly in regions where the high volume of sample examinations overwhelms technicians. To address this challenge, the system incorporates a motorized microscopic stage which reduces the dependency on humans and increases the throughput of sputum smear examinations. The system is intuitive enough for use by individuals with minimal knowledge of microscopy, as illustrated in Figure 1.

The system comprises three primary stages:

1. Data collection system
2. Recognition system
3. TB *bacilli* counting

Data collection system

The data collection system automates the capture of all FOVs during specimen scanning using a programmable microscopic stage. The stage is capable of

precise movements in horizontal and vertical directions, covering the specimen entirely according to the WHO standard for slide preparation (2 cm x 1 cm, equivalent to 20.000 square micrometers). A digital camera with CMOS sensor is attached to the top of microscope eyepiece captures high-resolution images or videos as shown in Figure 2. The motorized stage supports various scanning patterns, including inward square, circular, spiral, interlaced and zigzag, enhancing diagnostic sensitivity and specificity (28).

Components of the programmable microscopic stage:

1. Programmable user-driven interface: A 2D Computer-Aided Design software, DraftSight, defines the zig-zag scanning patterns, exporting movement instructions in .DFX format.
2. Microcontroller: Converts .DFX files into G-Codes using micro-computer and numeric control software, enabling machine-understandable movement commands.
3. Linear driving system: Utilizes an efficient servo motor-based closed-loop driving system for a precise X-Y directional movements.

This setup minimizes human intervention in acquiring all the field of views.

Recognition system

The recognition system classifies infected and non-infected FOVs using Deep Learning models, with transfer learning applied to Inception V3, a pre-trained

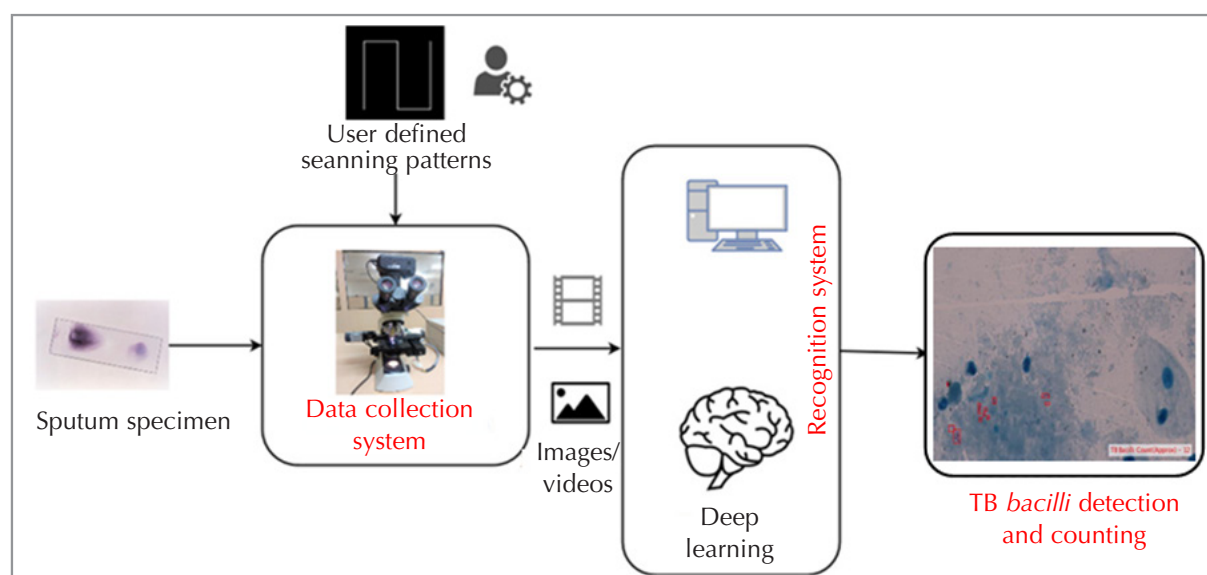


Figure 1. Workflow of the proposed TB *bacilli* detection system.



Figure 2. Automated digitalization of specimen using programmable microscopic stage.

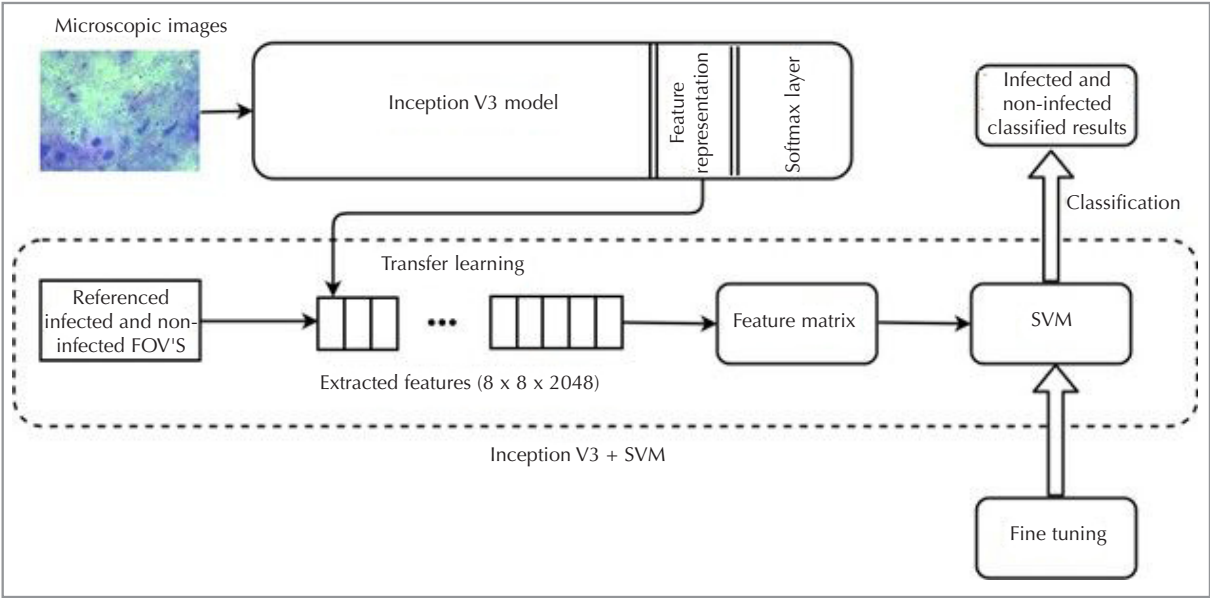


Figure 3. Recognition of infected field of view.

network as illustrated in Figure 3. The transfer learning and SVM-based classification streamline the microbial FOV detection. The Inception V3 architecture, with its efficient dimensionality reduction techniques, is optimized for computational speed and accuracy. By substituting the 5×5 convolution layers with 3×3 filters and employing 1×1 convolutions for dimensionality reduction, the model enhances feature extraction capabilities. Transfer learning, which applies knowledge from pre-trained models to new datasets, further improves classification efficiency while mitigating overfitting risks. The microbial dataset, comprising labelled images of infected and non-infected field views, undergoes cross-validation to determine optimal parameter settings for the SVM. The Support Vector Machine Optimization uses grid search and heuristics validate SVM parameters. And the Kernel functions optimize the decision boundary for better separation of microbial and non-microbial datasets.

Custom-tailored inception V3 model

Google's Inception V3 DeepNet architecture features 22 deep layers dedicated to training datasets, with the total number of layers, including pooling layers, reaching approximately 100. A linear layer is incorporated into this model for linear activation, while rectified linear activations are applied to all convolution layers within the Inception module. The model's receptive field is 229×229 in the RGB colour space, normalized to have a zero mean. In the Inception architecture, layers are stacked sequentially and concatenated to derive output correlations effectively. As spatial resolution decreases with each successive layer, higher layers capture more abstract features. To address computational inefficiencies, the Inception V3 model incorporates dimensionality reduction techniques. Specifically, 3×3 and 5×5 convolution filters are utilized in the higher layers; however, 5×5 convolutions are computationally expensive. To mitigate this, the architecture replaces 5×5 filters with multiple 3×3 filters and applies 1×1 convolution filters before complex convolutions to reduce dimensions further. This design aggregates visual information at each stage and passes it to subsequent layers, significantly improving computational efficiency. The Inception V3 model processes feature at a fraction of the computational cost, achieving speeds 3 to 10 times faster than non-inception models.

In our transfer learning approach, the early convolutional layers of the Inception V3 model is responsible for capturing low-level generic features, were frozen during training to retain their pre-trained weights. Fine-tuning was applied to the deeper layers, specifically the final inception block and the fully connected layers, to adapt the model to the domain-specific characteristics of ZN-stained sputum smear images. This selective training helped prevent overfitting and reduced computational requirements while allowing the network to specialize in recognizing TB *bacilli* features.

Transfer learning

Training a large dataset for recognizing TB-*bacilli* FOVs from the scratch is time-intensive and prone to overfitting. To address this, transfer learning is employed. Transfer learning methods are classified into supervised, semi-supervised, and unsupervised approaches, with knowledge transfer occurring through examples, representations, parameters, or relational structures. In this system, the pre-trained Google's Inception V3 network, trained on the ImageNet dataset weights is transferred for microbial classification tasks. These weights are fine-tuned to meet application-specific requirements, especially by importing pre-trained weights before the fully connected layers and passing them to a SVM for classification.

This process ensures accurate microbial dataset classification while significantly reducing computational complexity. Optimal SVM parameters are determined through empirical evaluation using cross-validation and heuristic approaches, enabling low-error classification. This transfer learning framework leverages existing knowledge effectively, ensuring fast and reliable performance.

Classification using support vector machine

Once knowledge is transferred, SVM parameters are optimized and validated for the target domain. Gradient-free numerical optimizers are employed to refine SVM parameters and reduce the search space. The SVM identifies the optimal separating hyperplane for linear and non-linear data by solving a quadratic programming problem. Parameters, such as σ and C , play a critical role in the placement and orientation of the hyperplane. Adjusting these parameters allows the SVM to balance margin maximization and error tolerance effectively.

To optimize the SVM classifier, a comprehensive grid search was conducted to identify the most effective combination of hyperparameters. The regularization parameter C was varied across the range (0.1, 0.5, 1.0, 10.0), and multiple kernel functions, including linear, polynomial, and Radial Basis Function (RBF) were evaluated. The RBF kernel consistently yielded superior performance in cross-validation, striking a balance between classification margin and misclassification tolerance. Parameter tuning was guided by 5-fold cross-validation on the training data, and the configuration yielding the highest mean accuracy was selected. This structured optimization process ensured a robust and generalizable decision boundary for TB classification. This hybrid deep learning-SVM approach was chosen to leverage the strength of deep feature extraction from Inception V3 while using SVM as an effective and light-weight classifier.

Higher values of parameter C result in narrower margins with fewer training errors, whereas smaller C values create wider margins with higher training errors. Using grid search algorithms, the system identifies the optimal parameters and kernel functions for SVM classification. Extracted pre-trained weights from the Inception V3 model, prior to the fully connected layers, are passed to the SVM for classification. By tuning and validating these parameters, the system achieves robust image classification in microbial datasets.

TB *bacilli* counting

Image preprocessing is a critical step in TB *bacilli* detection, enhancing image quality through contrast stretching and separation of RGB colour planes. Following enhancement, segmentation techniques distinguish the red-stained TB *bacilli* from the bluish background. A two-level segmentation approach—coarse and fine—separates *bacilli* from debris, using adaptive local thresholding methods like Sauvola's technique for precision. The segmented images undergo connected component analysis to eliminate non-*bacilli* objects, ensuring accurate feature extraction.

Morphological operations further refine the analysis by focusing on shape descriptors to identify TB *bacilli*. These descriptors include area, perimeter, circularity, major axis length, convex hull perimeter, and eccentricity, all of which characterize the rod-shaped morphology of TB *bacilli*. By integrating these geometrical and shape features, the system distinguishes TB *bacilli* from non-TB *bacilli* with high accuracy. The comprehensive automation and advanced recognition techniques employed in this system significantly enhance the diagnostic process, reducing human dependency and improving outcomes in TB detection. Thus, the microscopic examination of specimens is automated in the proposed system, significantly reduced the reliance on skilled technicians. By integrating programmable scanning stages and deep

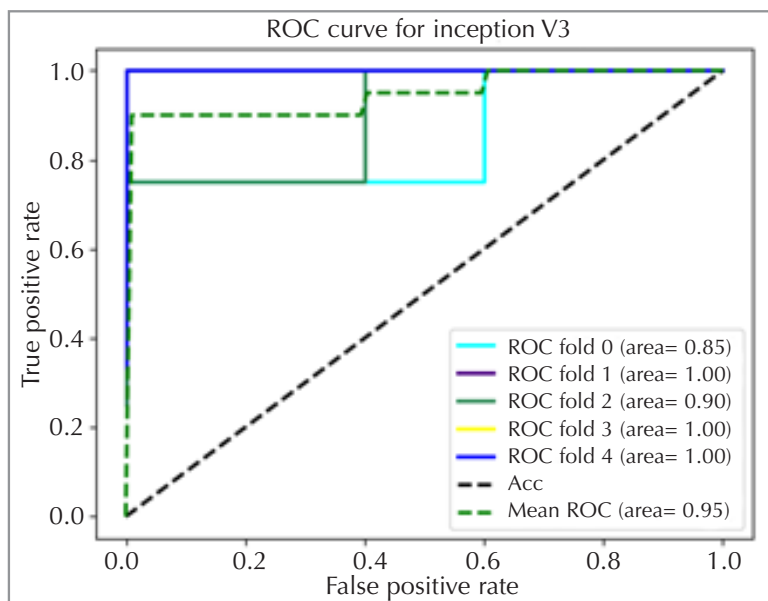


Figure 4. ROC curve for Inception V3 + SVM.

learning models, the system enhances sensitivity, specificity, and throughput. The combined segmentation, morphological operations, and feature extraction accurately detect and quantify TB *bacilli*, providing critical insights into disease severity and supporting clinical decision-making.

Experiments and discussion

This section details the collection of microscopic images of infected and non-infected *bacilli* FOVs and the training and testing process of the InceptionV3-SVM model for classifying AFB infection using the acquired image data.

TB digital corpus for learning

To develop and validate the InceptionV3-SVM deep learning model, a dataset comprising digital images and videos of TB *bacilli* (also known as AFB) in both infected and non-infected states was curated. ZN stained sputum smear specimens served as the source for these images, obtained from two main contributors. The Pondicherry Institute of Medical Sciences (PIMS), a multi-specialty hospital situated in southern India near Tamil Nadu and Pondicherry, provided anonymized ZN-stained samples from pulmonary TB patients. These samples were specially prepared to emphasize sputum regions for efficient pulmonary TB screening. As per the ethical norms, once the slides are digitalized, we returned the slides back to the hospital. Since 2016, PIMS has been sharing these specimens, facilitating the construction of a robust TB digital corpus consisting of images and videos. The specimens were analyzed using an Olympus C21i microscope integrated with a motorized stage. The zigzag scanning pattern ensured complete FOV coverage, with the camera mounted on the eyepiece capturing the views as images or videos. Videos were recorded at a resolution of 1920 x 1280 pixels, 25 fps, and 72 dpi. The video frames were later decomposed into distinct, non-overlapping images.

To enhance dataset diversity, additional ZN-stained sputum smear FOV images were obtained from the Ziehl-Neelsen Sputum Smear Microscopy Image Database (ZNSMiDB). Developed by Shah et al. in collaboration with Jaypee University of Information Technology and Indira Gandhi Medical College, India, the database contains over 1000 images (800x600 pixels, 72 dpi), categorized into TB FOV, non-TB FOV, auto-focused TB *bacilli*, and manually segmented TB *bacilli* images.

Fine-tuning the inception V3 model

A total of 1242 images were collected for training, comprising 620 FOV images with TB *bacilli* and 622 without TB *bacilli*. The Inception V3 model was fine-tuned using backpropagation to recognize TB-specific features, with input images resized to 229 x 229 pixels and converted to RGB. To improve computational efficiency, 5 x 5 convolution filters were replaced with 3 x 3 filters, and batch normalization auxiliary classifiers were removed. Training was conducted for 50 epochs using stochastic gradient descent with a batch size of 25, a learning rate of 0.045, and exponential decay every two epochs. Max pooling with a stride of 2 pixels was applied for dimensionality reduction, while dropout with a 0.5 rate in fully connected layers helped prevent overfitting. Pre-trained ImageNet weights were fine-tuned with TB images, and the final softmax layer was optimized through backpropagation to enhance classification accuracy.

Cross-validation and model evaluation

During evaluation, stratified K-fold cross-validation (K= 5) is employed, splitting the dataset into training, validation and testing subsets while ensuring balanced representation. The dataset of 1242 images is categorized into TB *bacilli* and non-TB *bacilli* FOVs, maintaining a 4:1 split ratio, with 40 samples used for training and 10 for testing per fold. The mean accuracy across all folds serves as the final evaluation metric. To further validate performance, the receiver operating characteristic (ROC) curve is analyzed, with a mean value of 0.9505, indicating high classification accuracy (Figure 4). The SVM classifier is used during cross-validation, where a grid search identifies the optimal C-value (0.5) and kernel function, with the RBF kernel yielding the best results.

The model's accuracy is defined as the percentage of correctly identified TB and non-TB FOV instances and is computed using the formula:

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})} \quad (1)$$

Here, TP, FN, FP, and TN represent the counts of true positives, false negatives, false positives, and true negatives, respectively. Table 1 summarizes the performance evaluation of both the fine-tuned Inception V3 model and the customized Inception V3+SVM model.

Table 1. Cross-validation results for inception v3 and SVM		
	Inception V3 + SVM (Accuracy ± STDV)	Inception V3 (Fine tuning)
Number of TB infected FOV samples	620	620
Number of non-infected FOV samples	622	622
One-fold validation	85.254 ± 0.295	--
Two-fold validation	99.567 ± 0.053	--
Three-fold validation	90.734 ± 0.054	--
Four-fold validation	100 ± 0	--
Five-fold validation	99.637 ± 0.012	--
Average accuracy	95.05	78.37
SVM: Support vector machine, STDV: Standart deviation, FOV: Field-of-view.		

To ensure robust evaluation and enhance the generalizability of the proposed model, the dataset was curated from two independent sources viz., the PIMS and the publicly available ZNSMiDB. This combination introduces variability in slide preparation, staining quality, imaging equipment, and clinical conditions, reflecting real-world heterogeneity commonly encountered in TB diagnostics. Such multi-source data composition helps mitigate biases associated with single-institution datasets and supports the development of a model resilient to diverse image characteristics.

Further model performance was assessed using stratified 5-fold cross-validation, a widely accepted internal validation method in medical imaging studies, which ensures balanced representation of infected and non-infected samples across folds. This approach provides a rigorous and stable estimate of the model’s classification accuracy while minimizing overfitting risks. The consistently high accuracy and low variance observed across folds indicate strong model stability. The inclusion of publicly sourced data within the validation process partially addresses generalizability concerns.

Segmentation

Once the infected field of view is identified, the image undergoes segmentation to separate the foreground from the background through a process known as image thresholding. In this scenario, the TB bacilli, which are characterized by their red coloration, are distinguished from the predominantly blue background. Both coarse and fine-level segmentation methods are utilized to isolate the *bacilli* from their surrounding background. During the coarse-level segmentation phase, pixels corresponding to the TB

bacilli are marked as the foreground, and a threshold parameter (T) is applied to differentiate them from the background, as shown in Figure 5. This procedure is iterated until the mean intensity values for the foreground and background pixels are calculated.

The fine-level segmentation phase employs adaptive local thresholding, which considers both the pixel’s intensity and its surrounding neighborhood. The segmented image from the coarse-level stage serves as the input for this phase, where a local threshold is determined for each pixel based on the intensity values of its neighboring pixels. The technique is implemented here from Sauvola’s method, which calculates a specific threshold for each pixel. Let $g(x, y)$ represent the intensity of a pixel at the coordinate (x, y) (30).

$$o(x,y)=m(x,y)\begin{cases} 0 & \text{if } g(x,y) \leq t(x,y) \\ 255 & \text{otherwise} \end{cases} \tag{2}$$

Following both the coarse-level and fine-level segmentation stages, the segmented outcomes are combined to detect a foreground containing TB *bacilli*. However, residual debris may still be present. To eliminate these extraneous elements, an 8-pixel connected component analysis is performed. The connected pixel groups are evaluated to classify them as either TB *bacilli* or non-TB *bacilli* objects.

Morphological operations

Once the image is segmented, morphological operations are applied to further refine the image and extract features. These operations help in identifying the shape descriptors of the TB *bacilli*, which is a rod-shaped object with defined parameters for size and structure.

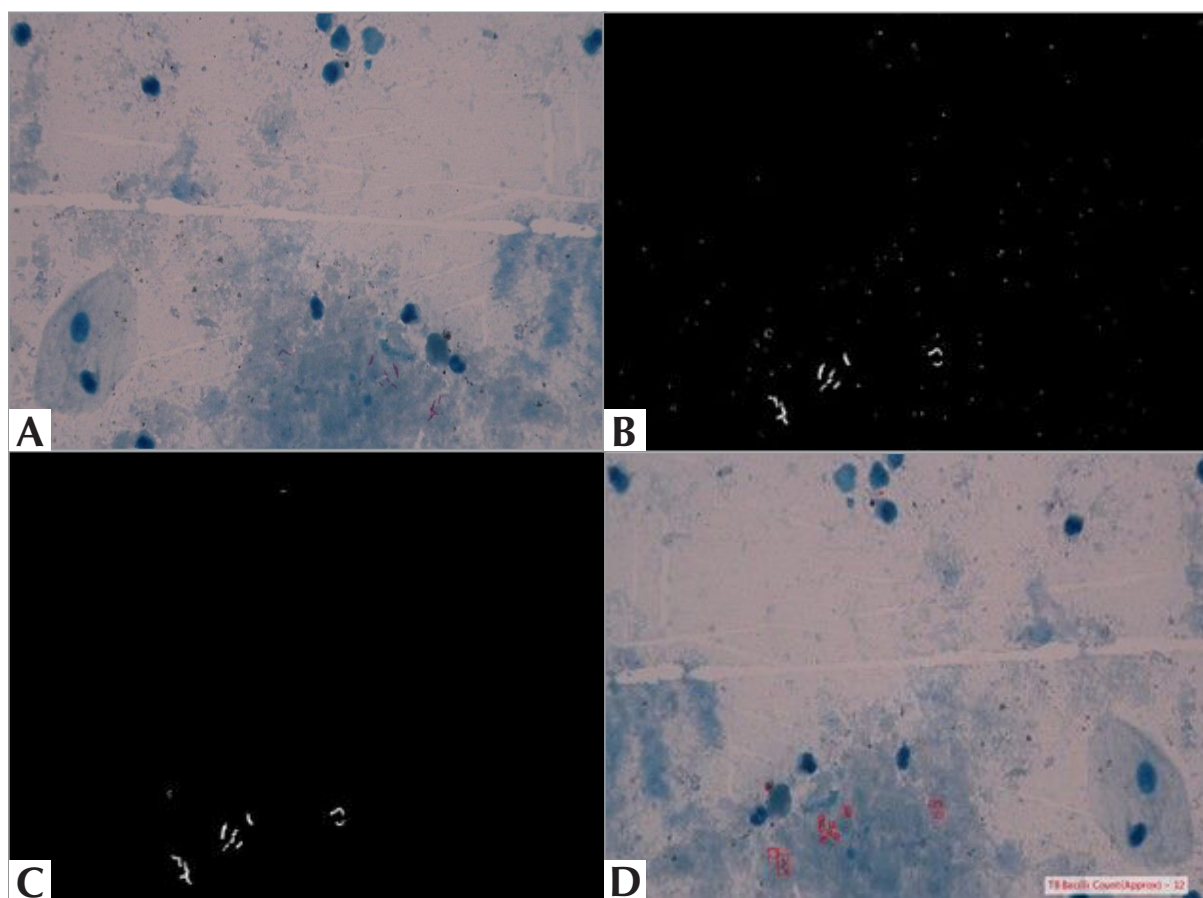


Figure 5. Shows the steps involved in the segmentation of TB *bacilli*. (A) Digitized FOV with TB *bacilli* in purple. (B) Coarse and fine-level segmentation using RGB thresholding and Sauvola's algorithm. (C) Debris removed after applying shape descriptors. (D) TB *bacilli* detected and counted using generated bounding boxes.

To classify the *bacilli* as TB or non-TB, various geometrical and shape features are extracted, which include:

- Area (A) and perimeter (P): The area quantifies the total number of pixels comprising an object, whereas the perimeter measures the length of its boundary or contour.
- Circularity or roughness: Circularity evaluates the object's roundness, ensuring it is neither overly elongated nor square-like. This metric is particularly helpful in identifying geometrical forms such as cylinders, spheres, and cones.
- Major axis length: This represents the longest diameter of the object, passing through its center and touching the farthest points along its boundary.
- Convex hull perimeter: The convex hull is the smallest boundary enclosing the object. The perimeter of this convex hull provides an estimate of the object's overall shape, with adjustments made to exclude unnecessary points and refine its contour.
- Eccentricity: This parameter measures the elongation of an object by comparing the longest axis to the perpendicular axis, aiding in distinguishing between elongated and circular shapes.

These geometrical features are crucial in identifying and classifying TB *bacilli* from other objects in the image.

Comparison with existing work

The proposed TB detection system demonstrates promising performance in identifying TB *bacilli* in digitized ZN-stained sputum smear images. By leveraging a fine-tuned Inception V3 architecture combined with an SVM classifier, the model achieved an average classification accuracy of 95.05% and an

F1-score of 0.902 for bacilli segmentation. These metrics indicate strong model generalization and detection capability. However, to fully appreciate the system's contribution, it is important to contextualize these results within the landscape of existing literature and to critically reflect on the challenges and future work needed for clinical adoption.

The studies in TB detection using machine learning and deep learning have reported performance metrics in the range of 85-93% accuracy and F1-scores ranging from 0.78 to 0.89. For instance, Lopes et al. used a CNN-based architecture for bacilli detection and achieved an F1-score of 0.88 on a similar ZN-stained dataset (31). Sharma et al. reported 91.4% accuracy using an ensemble CNN model, although it required a larger training dataset and longer inference times (32). Hwang et al. proposed a patch-based method using AlexNet, which showed high sensitivity (up to 94%) but suffered from a higher false-positive rate and slower scanning speed (33). Compared to these approaches, proposed method shows a balanced improvement in both computational efficiency and diagnostic performance. The integration of transfer learning (from ImageNet) and domain-specific adaptation through SVM optimization has allowed us to maintain high classification accuracy even with a relatively modest dataset size. The two-level segmentation approach combined with morphological operations achieves a precision of 0.924 and a recall of 0.882. These results demonstrate the effectiveness of the method in accurately detecting TB *bacilli*. Consequently, we obtained a high F1 score of 0.902, indicating a strong balance between precision and recall in the detection process. This performance highlights the robustness of the proposed technique for identifying TB *bacilli* in microscopic images.

The proposed TB detection system is promising but also has some limitations in clinical relevance. The dataset used though is limited in diversity, being sourced primarily from two sources. This lack of variation in staining techniques, imaging equipment, and patient demographics could hinder the model's generalizability to broader clinical scenarios. Another key limitation is the absence of independent external validation. While 5-fold cross-validation assesses internal consistency, it does not reflect true performance on unseen data. Independent testing on external datasets is critical in medical imaging to confirm a model's robustness and reliability across settings.

The implementation of the system in clinical practice also presents challenges. It requires specialized hardware, such as a programmable motorized microscope stage, which may not be accessible in all healthcare settings, particularly in low-resource areas where TB is most prevalent. Furthermore, some level of operator training and regulatory clearance would be essential before the system can be adopted in practice.

Future research prioritizes assembling a more diverse and representative dataset and testing the system in real clinical workflows. Optimizing the model for portable or embedded devices could also increase accessibility in resource-limited regions. In addition, integrating patient metadata and adding explainability features would enhance clinical trust and usability.

CONCLUSION

In this study, we presented a robust and efficient deep learning-based framework for the detection and classification of TB *bacilli* (AFB) from ZN stained sputum smear images. By leveraging the Inception V3 model in combination with a SVM classifier and utilizing transfer learning, we developed a highly accurate and computationally efficient system for distinguishing TB-infected and non-infected fields of view. The use of advanced segmentation techniques specifically, coarse and fine-level segmentation with adaptive local thresholding based method, enabled precise bacilli isolation from complex image backgrounds. Additionally, the extraction of key geometrical and shape descriptors further enhanced detection robustness.

The model demonstrated strong performance, achieving a mean ROC curve score of 0.9505, a precision of 0.924, a recall of 0.882, and an F1 score of 0.902. Stratified K-fold cross-validation confirmed the model's stability and generalizability across diverse imaging conditions. Beyond these technical results, the system offers practical advantages for TB screening, particularly in low-resource settings where access to trained personnel is limited. By automating bacilli identification and reducing human error, the system facilitates faster, more consistent diagnoses and can be integrated into existing diagnostic workflows. Its modular design also opens the possibility for remote diagnostics via telemedicine and supports deployment in field or mobile laboratory environments.

Looking ahead, future research will address several key areas to strengthen the clinical applicability of

this system. Expanding the dataset to include more diverse samples with variations in staining techniques, imaging equipment, and patient demographics will enhance generalizability and robustness. Refining the segmentation pipeline using more advanced, data-driven methods could further improve detection precision in challenging visual environments. In addition, efforts will be directed toward optimizing the system for real-time operation on portable or embedded devices to increase accessibility in remote regions. Integrating clinical metadata and explainable AI components will also be explored to support clinician interpretation, foster user trust, and comply with medical regulatory standards. Finally, these directions will help transform the proposed system from a research prototype into a practical diagnostic solution with real-world impact in global TB control efforts.

Acknowledgement

This research is supported by Vellore Institute of Technology. The authors also wish to show their gratitude to Dr. Anil Purty, for sharing the ZN-stained sputum smear specimen during the course of this research.

Declaration of generative AI in the writing process

During the preparation of this work, the authors used AI tools to improve the flow of the text, correct any potential grammatical errors, and improve the writing. After using the tool, authors reviewed and edited the content as needed and took full responsibility for the content of the publication.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflict of interest

The authors declare that they have no conflicts of interest related to this work.

Ethical Committee Approval: This study involves no direct patient contact and uses only anonymized archival data, posing minimal risk and requiring no procedures beyond standard clinical care; therefore, ethical approval was not required.

CONFLICT of INTEREST

The authors declare that they have no conflict of interest.

AUTHORSHIP CONTRIBUTIONS

Concept/Design: DJS

Analys/Interpretation: RKB

Data acquisition: DJS

Writing: DJS

Clinical Revision: RKB

Final Approval: RKB

REFERENCES

1. *Technical and Operational Guidelines for Tuberculosis Control*. Available from: <https://tbcindia-wp.azurewebsites.net/technical-and-operational-guidelines-for-tb-control-in-india-2016/> (Accessed date: 20.11.2016).
2. Forero M, Sroubek F, Cristóbal G. Identification of tuberculosis bacteria based on shape and color. *Real-Time Imaging* 2024; 10(4): 251-62. <https://doi.org/10.1016/j.rti.2004.05.007>
3. Osuna E, Freund R, Girosi F. Training support vector machines: an application to face detection, in *computer vision and pattern recognition* (pp. 130-136).
4. Kusworo A, Rahmad G, Aris S, Sofjan FK, Adi P, Ari B. Tuberculosis (TB) identification in the the Ziehl-Neelsen sputum sample in NTSC channel and support vector machine (SVM) classification, *International Journal of Innovative Research in Science, Engineering and Technology*, 2013; 9(2): 5030-5.
5. Osman MK, Mashor MY, Jaafar H. Detection of Tuberculosis Bacilli in Tissue Slide Images using HMLP Network Trained by Extreme Learning Machine. *Elektronika ir Elektrotechnika* 2012; 120(4): 69-74. <https://doi.org/10.5755/j01.eee.120.4.1456>
6. Sadaphal P, Rao J, Comstock GW, Beg MF. Image processing techniques for identifying Mycobacterium tuberculosis in ziehl-neelsen stains. *Int J Tuberc Lung Dis* 2008; 12(5): 579-82.
7. Osman MK, Mashor MY, Jaafar H. Tuberculosis bacilli detection in Ziehl-Neelsen- stained tissue using affine moment invariants and Extreme Learning Machine. In *Proc. of IEEE 7th International Colloquium on Signal Processing and its Applications* 2011 pp. 804-13. <https://doi.org/10.1109/CSPA.2011.5759878>
8. Samuel DJR, Rajesh Kanna B. A programmable microscopic stage: Design and development. *Microsc Res Tech* 2019; 82(4): 429-42. <https://doi.org/10.1002/jemt.23184>
9. Oquab M, Bottou L, Laptev I, Sivic J. Learning and transfer-

- ring mid-level image representations using convolutional neural networks. *IEEE Conf Comput Vis Pattern Recognit* 2014; 1717-24. <https://doi.org/10.1109/CVPR.2014.222>
10. Shao L, Zhu F, Li X. Transfer learning for visual categorization: A survey. *IEEE Trans Neural Netw Learn Syst* 2015; 26(5): 1019-34. <https://doi.org/10.1109/TNNLS.2014.2330900>
 11. Dinesh Jackson Samuel R, Rajesh Kanna B Cybernetic microbial detection system using transfer learning. *Multimed Tools Appl* 79, 5225-42 (2020). <https://doi.org/10.1007/s11042-018-6356-z>
 12. Yang M, Nurzynska K., Walts, AE, Gertych A. A CNN-based active learning framework to identify mycobacteria in digitized Ziehl-Neelsen stained human tissues. *Comput Med Imaging Graph* 2020; 84: 101752. <https://doi.org/10.1016/j.compmedimag.2020.101752>
 13. Panicker RO, Sabu MK. Automatic detection of tuberculosis bacilli from conventional sputum smear microscopic images using densely connected convolutional networks. *SN Computer Science* 2022; 3(4): 263. <https://doi.org/10.1007/s42979-022-01133-w>
 14. Mithra KS. Enhanced fuzzy Gaussian networks for sputum image based Mycobacterium detection. *The Imaging Science Journal*, 2023; 71(4): 313-22. <https://doi.org/10.1080/13682199.2023.2167553>
 15. Campbell RAA, Eifert RW, Turner GC. Openstage: A Low-Cost Motorized Microscope Stage with Sub-Micron Positioning Accuracy. *PLoS ONE* 2014; 9(2): e88977. <https://doi.org/10.1371/journal.pone.0088977>
 16. Meijering E, Dzyubachyk O, Smal I, van Cappellen WA. Tracking in cell and developmental biology. *Seminars in Cell and Developmental Biology* 2009; 20: 894-902. <https://doi.org/10.1016/j.semcdb.2009.07.004>
 17. Freere RH, Weibel ER. Stereologic techniques in microscopy. *J Roy Micr Soc* 1967; 87: 25-34. <https://doi.org/10.1111/j.1365-2818.1967.tb04489.x>
 18. Bhakti TL, Susanto A, Santosa PI, Widayati DT. Design of motorized moving stage with submicron precision. *IJERA* 2012; 2(6): 674-8.
 19. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *NIPS* 2012; 25: 1106-14.
 20. Arora S, Bhaskara A, Ge R, Ma T. Provable bounds for learning some deep representations. *CoRR* 2013; abs/1310.6343.
 21. Lin M, Chen Q, Yan S. Network in network. *CoRR* 2013; abs/1312.4400.
 22. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 2015; 1-9. <https://doi.org/10.1109/CVPR.2015.7298594>
 23. Lee CY, Xie S, Gallagher P, Zhang Z, Tu Z. Deeply supervised nets. preprint arXiv 2014; 1409.5185,.
 24. Chen WC, Chang CC, Lin YE. Pulmonary tuberculosis diagnosis using an intelligent microscopy scanner and image recognition model for improved acid-fast bacilli detection in smears. *Microorganisms* 2024a; 12(8): 1734. <https://doi.org/10.3390/microorganisms12081734>
 25. Chen H, Gu W, Zhang H, Yang Y, Qian L. Research on improved YOLOv8s model for detecting mycobacterium tuberculosis. *Heliyon* 2024b; 10(18). <https://doi.org/10.1016/j.heliyon.2024.e38088>
 26. Fu HT, Tu HZ, Lee HS, Lin YE, Lin CW. Evaluation of an AI-based TB AFB smear screening system for laboratory diagnosis on routine practice. *Sensors* 2022; 22(21); 8497. <https://doi.org/10.3390/s22218497>
 27. Samuel RDJ, Kanna BR. A fuzzy strategy to eliminate uncertainty in grading positive tuberculosis. *International Journal of Computational Intelligence and Applications* 2022; 21(01): 2250006. <https://doi.org/10.1142/S1469026822500067>
 28. Samuel DJ, Baskaran RK. Design to automate the detection and counting of Tuberculosis (TB) bacilli. arXiv preprint arXiv 2021: 2105.11432.
 29. Shafait F, Keyzers D, Breuel TM. Efficient implementation of local adaptive thresholding techniques using integral images. *Society of Photo-Optical Instrumentation Engineers. of Proceedings of SPIE*. 2008; 6815. <https://doi.org/10.1117/12.767755>
 30. Lazzara G, Géraud T. Efficient multiscale Sauvola's binarization. *Springer-Verlag Berlin Heidelberg*, 2013. <https://doi.org/10.1007/s10032-013-0209-0>
 31. Lopes FM, Catarino JS, Lima RR, Oliveira LS, Papa JP. Detection of tuberculosis bacilli in microscopy images using deep learning and region-based segmentation. *Computers in Biology and Medicine* 2020; 123: 103843.
 32. Sharma R, Verma OP, Srivastava A. Ensemble of deep convolutional neural networks for automatic detection of tuberculosis in chest X-rays. *Biomedical Signal Processing and Control*, 2021; 70: 102988.
 33. Hwang S, Kim HE, Jeong J, Kim HJ. A novel approach for tuberculosis screening based on deep convolutional neural networks. In *Medical Imaging 2016: Computer-Aided Diagnosis*. *International Society for Optics and Photonics* 2016; 9785: 97852W. <https://doi.org/10.1117/12.2216198>