

Role of Artificial Intelligence in Diagnosis of Tuberculosis: An Investigation

Avree Ito-Fujita¹ and Shayna Katz²

^{1,2}*Nancy Atmospera-Walch School of Nursing, University of Hawaii at Mānoa,
Honolulu, USA
skatz@hawaii.edu²*

Abstract

*Tuberculosis (TB) is a worldwide problem that poses a severe threat to public health. Pathology is among the most essential methods for diagnosing tuberculosis in clinical practice. Identifying specifically colored tuberculosis bacilli under a microscope is essential for confirming tuberculosis as the diagnosis. Even for expert pathologists, it is time-consuming and laborious due to the tiny size and quantity of bacilli. This strenuousness often results in poor detection rates and incorrect diagnoses. TB is the most contagious human disease, driven by the bacteria *Mycobacterium Tuberculosis (MTB)*. It is an infectious illness that spreads from one person to another via the air. Pulmonary TB (PTB) is a tuberculosis that primarily affects the lungs. It may, however, affect other body regions, including the brain, bones, and lymph nodes. It is also known as Extrapulmonary Tuberculosis (EPTB). Because tuberculosis has common symptoms, it may be difficult to tell whether a patient has it without adequate testing. Employing image analysis and artificial intelligence-based classification techniques, an accurate and unique system for detecting tuberculosis has been developed in this work. There are two stages to the intended system. Three different Artificial Intelligence-based classification algorithms are applied once the X-ray image is initially processed utilizing pretreatment, classification, and feature extraction.*

Keywords: *Tuberculosis, Artificial intelligence, Mycobacterium tuberculosis, Machine learning, Deep learning*

1. Introduction

Tuberculosis (TB), a chronic contagious illness driven by the bacteria *Mycobacterium tuberculosis*, is a worldwide problem that poses a severe threat to global welfare. In recent days, TB has resurfaced in many parts of the globe after being effectively suppressed for several years. The occurrence of tuberculosis in China is said to be the 2nd highest in the world. During 2010, the incidence of active tuberculosis in persons aged fifteen and above was 460 per 100000, as per the 5th domestic epidemiological sampling study of tuberculosis. Medical experts have always had to deal with treating and preventing tuberculosis.

Pathology is among the most essential methods for diagnosing tuberculosis in clinical practice. Finding specifically dyed TB bacilli under a microscope and recognizing a set of morphologic alterations in the surrounding tissue is necessary for confirming TB as the diagnosis. Since the bacilli are so minute, looking for and identifying them under the microscope requires high-power fields, which only give a narrow viewing region across an

Article history:

Received (September 2, 2023), Review Result (October 14, 2023), Accepted (November 16, 2023)

entire tissue segment [1]. Furthermore, the amount of bacilli is generally insignificant. As a result, even for experienced pathologists, it is a time-consuming and challenging job, and this strenuousness often leads to poor identification rates and incorrect diagnoses. Numerous novel methods, such as PCR and RNA scopes, have been developed to increase the accuracy and sensitivity of the identification of tuberculosis bacilli. Still, they have yet to prove to be trustworthy and generally adopted.

The morphology of tuberculosis bacillus is comparatively basic compared to the complicated and variable form of tissues and cells, like a thin rod with a height of around 4 μm and a diameter of 0.5–1.0 μm . After acid-fast staining, the waxy lipid in the bacilli's cell wall appears purple-red, contrasting sharply with the blue background [2]. The detection of bacilli with such form and color after staining is unique to the diagnosis of tuberculosis. A Convolutional Neural Network (CNN) model describes a unique approach for automatically detecting stained TB bacilli in this research. The research strategy outperforms the competition in specificity and sensitivity while remaining constant and economical.

Since the 1970s, scientists have been developing Computer-Aided Diagnostic (CAD) tools to enhance Tuberculosis diagnosis throughout the screening procedure. CAD devices can support radiologists remotely and have an automated diagnostic procedure. Recent advances in Artificial Intelligence (AI) technology and methodologies have resulted in significant advancements in automatic computer picture identification. CAD systems employ artificial intelligence to analyze radiological pictures to diagnose anomalies and alleviate the scarcity of radiologists, particularly acute in underdeveloped countries. CAD systems are frequently used in the identification of breast cancer, bone suppression, neurological problems, lung cancer, and skin lesions by analyzing medical photographs. The two most extensively utilized AI approaches for constructing CAD systems suitable for evaluating Chest X-ray CXRs are ML and DL. With the progression of science and know-how and the advancement of medical imaging modalities, the interpretation and analysis of clinical visual information have become more complex for radiologists. When bacteriological tests fail to offer a suitable response, radiography becomes crucial.

The bacteria *Mycobacterium tuberculosis* causes tuberculosis. Fever, cough, night sweats, chest discomfort, weight loss, exhaustion, and anorexia are all common signs of active TB. Quality The World Health Organization (WHO) recommends chest X-ray pictures and laboratory-based diagnostic assays for early tuberculosis analysis and to close the tuberculosis finding breach [3]. Most people who exhibit symptoms of tuberculosis are examined using a computer-aided diagnostic system to evaluate their chest X-ray. According to the WHO, CXR is a suitable method for orderly showing patients with TB signs and identifying those requiring microscopic sputum examination. The primary assessment for an unexplained cough is a chest X-ray posteroanterior picture of the lung. Although TB illness is most often associated with the lungs, it may affect every organ system in the body [4]. In chest radiography, CAD systems may identify pneumothorax, pulmonary consolidation, cardiac hypertrophy, pleural effusion, infiltration, nodules, atelectasis, and emphysema, among other disorders. The anatomical anatomy of the chest is seen in [Figure 1].



Figure 1. Chest X-ray image

a) Chest anatomy b) A healthy chest X-ray image c) Tuberculosis with multiple cavitation X-ray images

2. Proposed technique

The KERAS approach was applied to extract the image attributes employed as categorization factors. KERAS is a Python-based open-source neural network framework that includes the ResNet50 design, which aids in removing picture features using arrays.

In this investigation, three categorization approaches were used.

1. A supervised adaptive knowledge with related procedures that examine the evidence and detect forms is the first approach used, which is centered on SVMs. Support vectors are the reference facts nearest to the hyperplane or the components in a data set that, if removed, would shift the position of the separating hyperplane. They can, therefore, be quantified as severe data-collection components. SVM, for short, is a supervised ML algorithm that can be used for regression and classification. Since classification issues are where SVMs are most frequently employed, this study concentrates on its efforts [5]. The SVM conducts linear and nonlinear categorization when a dataset is separated into two categories, as well as the Kernel function is utilized for nonlinear classifications; the kernels in nonlinear categorization are uniform polynomial and complicated regression analysis.
2. The 2nd technique is centered on Logistic Regression (LR), a categorization ML algorithm for predicting the likelihood of a dichotomous categorical-dependent variable. The logistic model is one of the most essential predictive methods for calculating the probability of a specific classification or occurrence, like success or failure. In contrast, logistic regression uses a set of forecast variables that might be either quantitative or unconditional. This can be employed in several scenarios, such as evaluating whether a picture comprises a tiger, cat, fish, or other animal. A chance between 0 and 1 will be given to each identified item in the picture, making a total of one. The Logit model, often known as the broad sense entropy classifier, is another term for "logistic regression ."Logistic regression is one of the supervised ML procedures for "categorization" problems. Its extraordinary capacity to discover embezzlers has earned it an especially favorable reputation in the banking industry during the previous two decades [6]. The network diagram depicts the usual usage of logistic regression and other standard linear classifiers to extract the properties of the pictures that will be used as categorization attributes

for creating the maps. As a result, a logistic regression needs a binary dependent variable. In addition, the desired value should be represented by the level 1 factor. Only essential variables should be used as autonomous variables and must be independent.

The third technique is based on closest neighbors, an ML algorithm based on supervised-type examples. The reason nearest-neighbor techniques have remained popular in practice is their long-term empirical success. Nevertheless, this argument may need to be more complex. Researchers concentrate on four aspects of nearest-neighbor methods critical to their continued success. Firstly, the freedom to select what "near" implies in closest neighbor prediction enables us to manage ad hoc distances or leverage current representations and range learning machinery to handle ad hoc distances, like "deep networks or decision tree-based ensemble learning approaches ."Secondly, due to the computational effectiveness of various approximation closest neighbor search techniques, nearest neighbor projection can scale to big HD datasets that are common in today's applications [7]. Finally, nonparametric techniques depend on data to make minimum modeling assertions instead of directly enabling the facts to drive projections. Nearest neighbor techniques may eventually be understood: they display the identified closest neighbors as evidence for their predictions. This approach is very beneficial for categorizing fresh samples and forecasting or projecting future values. It simply looks for the most comparable data points learned during the training phase and produces new point recommendations based on that categorization.

This study used the Montgomery database; the X-ray images were from the UNDP's tuberculosis control program in Iraq. The collection comprises 138 radiographs, 80 of which relate to healthy individuals and 58 of which display tuberculosis signs. All images are in DICOM format and have been de-identified [8]. A broad range of irregularities, such as spilling patterns, are included in the collection. In the form of a text document, the dataset comprises radiological data. Every image contains a label that can be utilized to recognize it. The designations might be TB-positive, TB-negative, or TB-free.

The photographs were pre-processed before being utilized. Pre-processing is divided into two parts: padding and scaling. They join the ResNet50 system after completing such processes. The channel's procedure is shown in [Figure 2].

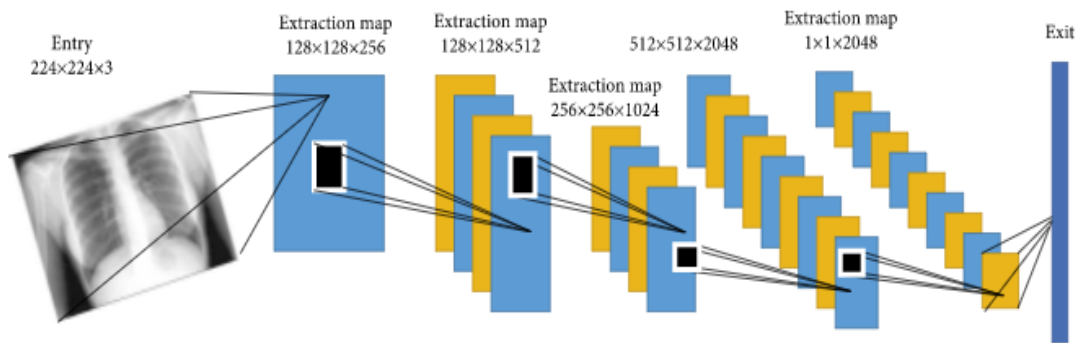


Figure 2. ResNet50 system procedure

They enter the system for the last pretreatment round to extract the features of the pictures that will be utilized as categorization attributes. The system accepts the matrix as an input and operates convolutions on it in every layer, generating removal maps. A vector of dimension

2048 is created in the channel's penultimate layer, which comprises the image's overall attributes like saturation, brightness, intensity, and others.

2.1. Procedure for cross-validation

The labels are made in a text file to name every one of the photos that will be utilized for the program's training after the setups with the features of the images for tuberculosis and Normal are received. Labels and characteristics are called inside the processing software, and links between names and features are generated to be translated into arrangements later.

The labels are made in a text file to name every one of the photos that will be utilized for the program's training after the setups with the features of the images for tuberculosis and regular are received [9]. Within the processing program, labels and attributes are called, and connections between names and features are created to be converted into arrangements later.

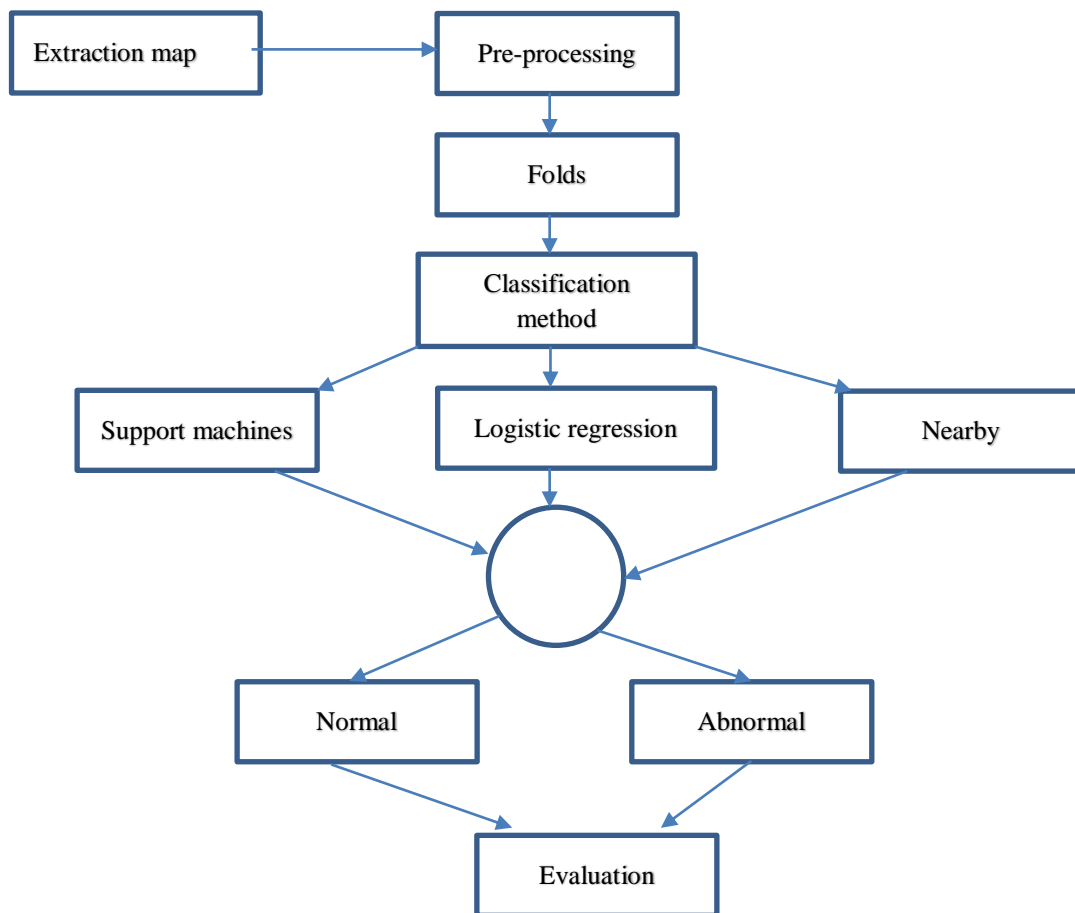


Figure 3. Cross-validation diagram

2.2. Procedure for exercise as well as test set

Training and test sets were generated for the 2nd state. The test pictures were never viewed by the training set since 80 percent of the photos were also utilized for training, and the remaining 20 percent were used for testing. We kept track of the following assessment

metrics for each categorization: precision, accuracy, recall, and F-measure. In the latest advancements, these metrics are often referred to by their English names.

3. Models and architecture

3.1. Convolutional neural network-centered cataloging model

A discriminative classifier created from a multilayer perceptron, the Convolutional Neural Network, is meant to detect specific patterns directly from picture pixels with little preparation. Convolutional Neural Network is also called shift-invariant ANNs because of its hierarchical system for spreading shift-invariant categorization. It is a top option for tackling the most difficult circumstances because of its uncanny purpose of extracting global features and contextualizing knowledge from data's intrinsic qualities.

Traditional CAD systems that use ML approaches like SVMs and other techniques have helped radiologists increase their diagnosis accuracy; nevertheless, many of these previous technologies required radiologists to extract disease signs manually [10]. Furthermore, lesions' dynamic nature and multiple properties make reapplying features composed in earlier studies to fresh patient data complex. As a result, typical ML procedures could be more appropriate for long-term success. With new improvements in radiology medical equipment, innovation, extensive data diagnostics, and many medical pictures being created daily, radiologists' workload has expanded dramatically [11]. Regarding this, using CNN in different detection techniques, rather than conventional methods, is valuable and practical for its capacity to mechanically extract complicated pathological aspects from visual data while meeting the inherent necessity of vast information.

The Google Inception Model, VGGNet, and ResNet, which differ in their segmental construction and the number of layers evaluated for picture categorization, vying to attain superior performance in identifying everyday items, were investigated. Researchers improved their presentation by adding a refinement phase to the drilling phase by incorporating a uniform alteration to the architecture of the final few layers of such systems before the output [12]. Furthermore, combining an ensemble system based on upgraded CNN models may improve the CAD system's diagnostic performance and general reliability. All CNN models were built in PyTorch and trained using the Adam optimizer.

3.2. Transfer learning

Transfer learning uses prior knowledge to rapidly develop appropriate predictions for novel specialized tasks. It is a method that retains already acquired information to be deployed to a linked activity [13]. Deep and intricate model architectures in computer vision are costly to train due to the large dataset size requirements and expensive hardware like graphics processing units.

As a result, all CNN models utilized in our studies were pre-trained on the Imagenet database to categorize everyday items into 1,000 different categories. The starting base for tuberculosis identification was derived using features learned from each layer. A considerable amount of time will be conserved by not having to start from scratch [14]. Before studying the chest radiographs in this work, ImageNet, a state-of-the-art framework employed by numerous researchers for deep CNNs, was previously trained on 1.2 million ordinary color photos with 1,000 categories. Although Anthimopoulos' research is well-documented for illness prediction using a CNN model, it needs more generality and transferability. Correspondingly, Rajpurkar attempted to improve precision by expanding the number of

layers in their employment, which is not a foolproof method [15]. This highlights the significance and the requirement of establishing an efficient hybridized ensemble and an optimized model for the highest accuracy effectiveness through convergence rate, which has yet to be discussed.

3.3 The ensemble model

An ensemble model is developed using the principle of ensemble learning, which combines numerous classifiers and provides the concluding productivity grounded on the findings supplied by the combined classifiers to improve output [16]. Classifiers employed for ensemble purposes must retain adequate variety to capture diverse characteristics from the same target. Building an ensemble model entails two stages: (a) generating a range of simple models from the original data, and (b) aggregating the population into a single system.

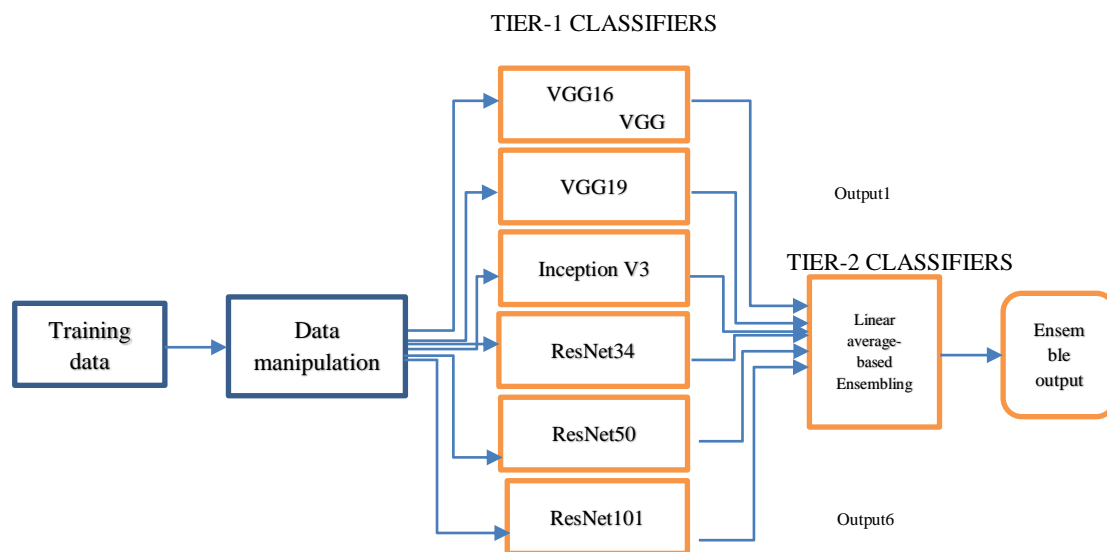


Figure 4. For tuberculosis diagnosis and localization, a proposed feature selection model structure was applied

3.4. Localization scheme

The inner link between the likelihood of categorization outcome and the weights of the final few layers in a CNN model will scramble to find areas in the picture that are associated with the prediction outcomes. According to Bossuyt et al., illness localization is determined by class activation mapping [17].

Localization in a convolutional neural network model with a global average pooling layer is shown in [Figure 5]. To construct a vector, extracted features obtained from the final convolutional layers are treated using global average pooling. The acquired vector will be utilized to calculate a weighted summary using the characteristics of the fully-linked layer to give a categorization result [18][19]. As a result, by linking the values from the last layer before the output to the pooling layer, the masses from the previous coating before the outcome may be predictable back to the characteristic charts to highlight locations where the model is estimated as showing significant information.

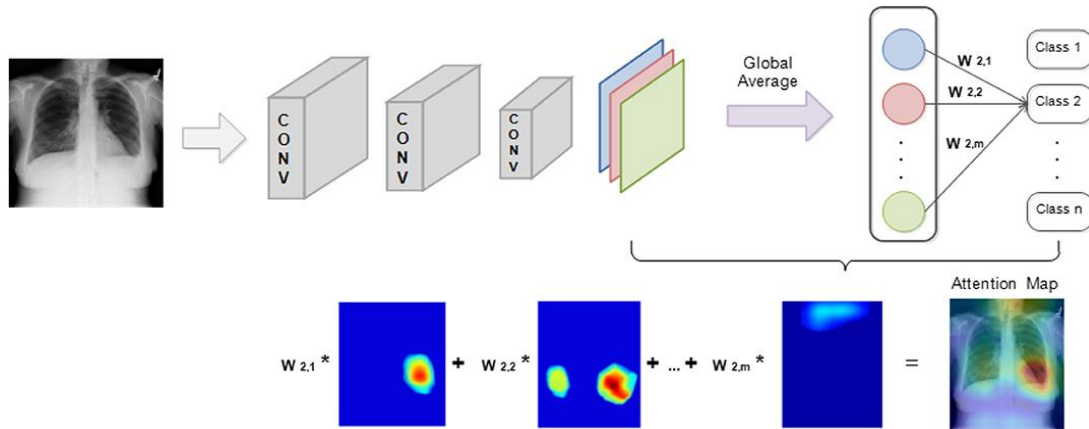


Figure 5. Tuberculosis localization employing class initiation charting

The preceding strategy fully leveraged Convolutional Neural Network representations' pattern detection and localization capabilities to attention map the target picture. Two separate purposes, image classification and object localization, may be effectively combined on the very same model by just manipulating the internal parameters inside a Convolutional Neural Network [20]. Depending on the input picture, the created interest map highlights the areas constituting the model's primary predictive criterion throughout the classification phase.

5. Conclusion

DL and the RESNET50 neural network are used to retrieve characteristics for categorization. The two categorization scenarios employed were "cross-validation" and training and examination data generation. The circumstance that produced the most outstanding results was one where the training and test sets were constructed with a reliability of more than 85 percent. SVM is the classification approach that performs best in the two situations used in this study. This can be observed in the findings acquired in the current study. They significantly outperform probability and enable effective picture categorization.

This study included a CT scan of the belly, a CT scan of the head, an MRI of the brain, and an MRI of the spine. After transforming such four sets of medical photographs to JPEG format, our proposed CNN architecture can effectively identify them by visual modality and anatomic location. The overall categorization performance was excellent in both the verification and test sets.

The information gathered allows us to analyze the viability of the approaches used. It also enables us to determine the optimal categorization scenario and ML approach for radiographs with and without TB categorization. In conclusion, a unified adjustment to the deep CNN conceptual framework and fine-tuning of the network through the ABC algorithm during the model training process has been presented to effectively predict the identification of TB-related disorders and the localization of particular disease manifestations. This study used multiple deep CNN models with diverse module topologies and various layers to assess the proposed method. A linear average-based ensemble structure built of those upgraded CNN models is constructed and used to enhance the overall efficiency.

This study's findings suggest that the quality of deep CNN models improves when the optimization stages are superimposed. Structure adjustments provide the most significant

increase in predictive performance for single CNN systems among the three phases. The categorization performance of CXRs is increased further by combining the various CNN algorithms. Furthermore, each model performs volatily on multiple datasets and for different classification objectives; however, classification accuracy and resilience are considerably enhanced when ensemble models are used. Even for the illness localization challenge, our suggested ensemble model may provide a satisfactory outcome by correctly presenting an awareness map to suspicious sick lung area highlight areas. Quantitative and qualitative data show that our technique performs very well compared to other state-of-the-art algorithms.

Reference

- [1] G. Chartrand, P. M. Cheng, and E. Vorontsov, “Deep learning: a primer for radiologists,” *Radiographics*, vol.37, pp. 2113-2131, (2017)
- [2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Adv Neural Inf Process Syst*, pp.1097-1105, (2012)
- [3] WHO, *Global Tuberculosis Report 2020*. World Health Organization, Geneva, (2020)
- [4] R. R. Nathavitharana, C. Yoon, and P. Macpherson, “Guidance for studies evaluating the accuracy of tuberculosis triage tests,” *J Infect Dis.*; vol.220, pp.116-125, (2019)
- [5] M. Rahman, M. S. Flora, and M. M. Husain, “National tuberculosis prevalence survey 2015–2016,” Ministry of Health and Family Welfare, Dhaka, (2016)
- [6] Goodfellow I, Bengio Y, Courville A, *Deep learning*. MIT Press, Cambridge, MA, (2016)
- [7] Z. Z. Qin, T. Naheyan, M. Ruhwald, “A new resource on artificial intelligence-powered computer-automated detection software products for tuberculosis programmers and implementers,” *Tuberculosis (Edinb)*, 127102049, (2021)
- [8] The Stop TB Partnership, FIND. Resource center on computer-aided detection products for the diagnosis of tuberculosis. <https://www.ai4hlth.org/> Date: 2020 Date accessed: September 10, (2020)
- [9] WHO consolidated guidelines on tuberculosis: module 2: Screening: systematic screening for tuberculosis disease. World Health Organization, Geneva, (2021)
- [10] K. Murphy, S. S. Habib, and S. M. A. Zaidi, “Computer-aided detection of tuberculosis on chest radiographs: an evaluation of the CAD4TB v6 system,” *Sci Rep.*, 105492, (2020)
- [11] S. M. A. Zaidi, S. S. Habib, and B. Van Ginneken, “Evaluation of the diagnostic accuracy of computer-aided detection of tuberculosis on chest radiography among private sector patients in Pakistan,” *Sci Rep.*; 812339, (2018)
- [12] T. Pande, C. Cohen, M. Pai, and F. Ahmad Khan, “Computer-aided detection of pulmonary tuberculosis on digital chest radiographs: A systematic review,” *Int J Tuberc Lung Dis.*; vol.20, pp.1226-1230, (2016)
- [13] D. B. Larson, H. Harvey, D. L. Rubin, N. Irani, R. T. Justin, and C. P. Langlotz, “Regulatory frameworks for development and evaluation of artificial intelligence-based diagnostic imaging algorithms: Summary and recommendations,” *J Am Coll Radiol.*, vol.18, pp.413-424, (2021)
- [14] M. Nash, R. Kadavigere, and J. Andrade, “Deep learning, computer-aided radiography reading for tuberculosis: A diagnostic accuracy study from a tertiary hospital in India. *Sci Rep.*; vol.10, pp.210, (2020)
- [15] Z. Z. Qin, M. S. Sander, and B. Rai, “Using artificial intelligence to read chest radiographs for tuberculosis detection: A multi-site evaluation of the diagnostic accuracy of three deep learning systems,” *Sci Rep.*, 915000, (2019)
- [16] A. P. Bradley, “The use of the area under the ROC curve in the evaluation of machine learning algorithms,” *Pattern Recognit.*, vol.30, pp.1145-1159, (1997)
- [17] P. M. Bossuyt, J. B. Reitsma, and D. E. Bruns, “Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative,” *Croat Med J.*, vol.44, pp.635-638, (2003)

- [18] S. Banu, F. Haque, and S. Ahmed, “Social Enterprise Model (SEM) for private sector tuberculosis screening and care in Bangladesh,” *PLoS One*, 15e0241437, (2020)
- [19] S. M. McKinney, M. Sieniek, and V. Godbole, “International evaluation of an AI system for breast cancer screening. *Nature*, vol.577, pp.89-94. (2020)
- [20] T. Saito and M. Rehmsmeier, “The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets,” *PLoS One.*, 10e0118432, (2015)