

# Diagnosis of Smear-Negative Pulmonary Tuberculosis using Ensemble Method: A Preliminary Research

Rusdah  
Faculty of Information Technology  
Universitas Budi Luhur  
Jakarta, Indonesia  
rusdah@budiluhur.ac.id

Mohammad Syafrullah  
Faculty of Information Technology  
Universitas Budi Luhur  
Jakarta, Indonesia  
mohammad.syafrullah@budiluhur.ac.id

**Abstract**— Indonesia is one of 22 countries with the highest burden of Tuberculosis in the world. According to WHO's 2015 report, Indonesia was estimated to have one million new tuberculosis (TB) cases per year. Unfortunately, only one-third of new TB cases are detected. Diagnosis of TB is difficult, especially in the case of smear-negative pulmonary tuberculosis (SNPT). The SNPT is diagnosed by TB trained doctors based on physical and laboratory examinations. This study is preliminary research that aims to determine the ensemble method with the highest level of accuracy in the diagnosis model of SNPT. This model is expected to be a reference in the development of the diagnosis of new pulmonary tuberculosis cases using input in the form of symptoms and physical examination in accordance with the guidelines for tuberculosis management in Indonesia. The proposed SNPT diagnosis model can be used as a cost-effective tool in conditions of limited resources. Data were obtained from medical records of tuberculosis patients from the Jakarta Respiratory Center. The results show that the Random Forest has the best accuracy, which is 90.59%, then Adaboost of 90.54% and Bagging of 86.91%.

**Keywords**— classification, high-tuberculosis burden country, medical data mining, tuberculosis data, Tuberculosis in Indonesia

## I. INTRODUCTION

Tuberculosis (TB) is one of the world's biggest threats. In the WHO report [1], in 2014, there were 9.6 million new TB cases, and 10% of them occurred in Indonesia. Indonesia is one of 22 countries with the highest burden of TB in the world [1]. In the National Guidelines for the Control of Indonesian Tuberculosis, it is reported that one of the causes of the increased burden of TB is the failure of the TB program. One of the causes of this failure was due to case finding or nonstandard diagnosis of TB and inadequate management of TB diagnosis [2].

TB diagnosis is difficult for several reasons [3], [4]. First, because TB symptoms, such as fever, cough, phlegm cough, bloody cough, and weight loss, are similar to the symptoms of lung cancer [5], pneumonia [6], acute respiratory infections, asthma and Chronic Obstruction Lung Disease [7]. Second, in pediatric patients because of the non-specific clinical symptoms of TB and the chest radiograph (CXR) is difficult to interpret [8] and sputum is difficult to obtain [1]. Third, in the case of a small number of germs (paucibacillary TB) [8], [9]. Fourth, in the case of extrapulmonary TB because patients often do not show symptoms of TB [10]–[12]. Fifth, in HIV-positive patients, the results of chest radiography can be atypical due to other infections [13]. Sixth, in the case of Smear-Negative Pulmonary Tuberculosis [14]–[17].

Pulmonary TB with negative smear and positive culture can be diagnosed by sputum culture, but it takes 6 to 8 weeks. Besides, sputum culture equipment is rarely found in developing countries [1], [14], [17], so the use of culture is limited and rarely recommended. Methods such as nucleic acid amplification tests, can provide faster results, but require high costs and the equipment is rarely found in many developing countries. Hence, in conditions of a limited resource, the diagnosis of SNPT is confirmed by symptoms, results of physical examination, and results of the investigation (at least with CXR). Unfortunately, the results of CXR examination between active and inactive tuberculosis are difficult to distinguish. For this reason, a final diagnosis is needed to reduce costs, and the possibility of underdiagnose and over treat cases [5], [17].

In previous studies, the diagnosis of tuberculosis was conducted using several types of input. These inputs include digital images of tissue samples [18], [19], images of TB bacteria [16], [20], the aroma of sputum with electronic nose [21], coughing sounds [22], pulmonary sound waves [23], [24], medical record data [9], [25], [26] and blood [27].

Unfortunately, digital images of network samples are hard to obtain because they are expensive and require equipment that rarely owned by health service in developing countries [28]. Similarly, data in the form of images of TB bacteria were obtained using a digital fluorescence microscope [29]. The smell of sputum, the sound of coughing, pulmonary, and blood sound waves is not a standard of TB case finding and diagnosis [2]. The focus of this study is to use medical record data as an input for TB diagnosis. In addition to the fact that medical record data is easily available and cheap, medical record data is also in accordance with standard case finding and TB diagnosis.

According to the Guidelines and Management of Indonesian Tuberculosis [2], the TB diagnosis process begins with the Suspect TB Discovery (TB screening) using symptom data and physical examination. Then a sputum examination (direct microscopic) is performed to determine a final diagnosis. If the result of a microscopic examination is positive, the patient is diagnosed with positive AFB pulmonary TB. If it is negative, the additional investigation must be conducted, at least by chest X-ray examination to determine if the patient is diagnosed with negative smear pulmonary TB [2].

Mello [14] showed that the dataset resulted from feature selection process (typical x-ray, sputum, weight loss, and age) can be used to predict SNPT. Unfortunately, sensitivity was

between 64-71% and specificity between 58-76%. The importance of using inputs that are in accordance with diagnostic standards is also shown in the research of Santos et al. [30]. The results of his study suggested that it is necessary to use investigations such as CXR and new methods to diagnose SNPT. The non-optimal sensitivity showed that the existing diagnosis model of SNPT could increase the number of under-diagnosis cases and the non-optimal specificity can increase the number of over-diagnosis cases.

This study is preliminary research that aims to determine the ensemble method with the highest accuracy in the diagnosis model of SNPT. This model uses input in the form of medical record data collected from the Jakarta Respiratory Center in *Perkumpulan Pemberantasan Tuberculosis Indonesia* (JRC-PPTI). The model determines whether a patient is diagnosed with pulmonary TB, SNPT, or not TB. The data used are symptoms, physical examination, and investigation in the form of phlegm examination and CXR. The output of this study can be an input for further research in the diagnosis of pulmonary tuberculosis using data mining techniques. This study is organized into four sections. The first section briefly describes tuberculosis, the background to the problem of tuberculosis diagnosis, the objective of the study, and the significance of the output of the research. The second section deals with previous studies related to tuberculosis diagnosis using classification techniques. The methodologies used to conduct this study are discussed in the third section. Finally, the fourth section is the experimental results and recommendations for future work.

## II. RELATED WORK

Previous studies using medical record data as an input for diagnosing SNPT used ANN [15], [30], [31], BNN [30] and Classification and Regression Tree (CART) [14]. Further examinations in addition to physical examination were used, such as HIV testing [14], [30], [31], sputum examinations [32] and other laboratory examinations [15], [31]. Mello [14] showed that the dataset of feature selection (typical x-ray, sputum, weight loss, and age) could be used to predict SNPT. Unfortunately, sensitivity is still low, which is between 64-71% and specificity between 58-76%. The importance of using inputs that are in accordance with diagnostic standards is also shown in the research of Santos et al. [30]. The results of his research suggest that it is necessary to use investigations such as CXR and new methods to diagnose SNPT.

Santos et al. [32] combined the Bayesian Network method and Artificial Neural Network (ANN) into the Bayesian Neural Network (BNN). The research showed that BNN accuracy increased by 6% compared to ANN. Asha et al. [33] compared the ensemble method and concluded that Bagging (97%) had the best performance, followed by Adaboost (96%) and Random Forest (93%). Asha et al. [34] compared the performance Support Vector Machine (SVM) and the ensemble method. They concluded that SVM had comparable performance with Random Forest and was better than Bagging and Random Forest. Random Forest is balanced with Adaboost in terms of accuracy but is more resistant to handling errors and outliers [35]. The Random Forest is not sensitive to the number of selected attributes in each split, so it is effective for extensive databases. It can be faster than bagging or boosting. In 2012, Asha et al. used the associative

classification method and produced accuracy that was comparable to their previous research [36].

## III. RESEARCH METHOD

### A. Data Collection

Data used in the research were patients' real data taken from JRC-PPTI from 2010 to 2014. The data were taken from patients aged over 15 years and had been declared cured (for the case of positive Acid-fast bacilli) or completed treatment (for the case of negative Acid-fast bacilli). The data were in hard copy format and manually recorded in an Excel format. Each record represents a patient's medical record. The attributes of patient data were selected based on directions from the doctors of JRC-PPTI. The data collected consists of 1423 records with 17 attributes. Table I shows all attributes along with their data types. Type N indicates numerical, and type C indicates a categorical attribute.

TABLE I. LIST OF ATTRIBUTES AND THEIR DATATYPES

Name of Attribute	Data Types
Sex	C
Age	N
Duration of a cough	C
Hemoptysis	C
Sputum	C
Fever	C
Weight loss	C
Loss of Appetite	C
Sweating at night	C
Breathless	C
Chest pain	C
Additional complaints	C
Weight	N
Preliminary diagnosis	C
Initial CXR	C
Initial Sputum Test	C
Final Diagnosis (Class)	C

### B. Data Exploration

Based on the type of illness, there are 49 general patients and 1,374 TB patients. The TB patient consists of 1,008 patients with negative smear and 366 patients with positive smear. Among patients with negative smear, there are 458 patients (45.4%) with a final diagnosis of SNPT.

Based on gender, there are 616 female patients (43%) and 807 male patients (57%). The age range of patients is between 15 and 89 years. Cases of SNPT occur more frequently in productive age (15-44 years), which is 374 cases (81.7%). The number of male patients (59.2%) in cases of SNPT in productive age is higher compared to female patients (40.8%).

### C. Data Preparation

Pre-processing activities conducted on dataset were data reduction and discretization of categorical attributes. Data reduction was conducted by removing 49 records of general patient data with a final diagnosis of Non-TB because they do

not have data on sputum examination results and CXR. The data used are 1374 records, consisting of 366 (26.64%) patients with positive smear and 1,008 (73.36%) patients with a negative smear (SNPT). Among 1,008 negative smear patients, there were 458 patients (45.44%) with a final diagnosis of SNPT and 550 patients (54.56%) with a final diagnosis of non-TB.

Discretization is carried out on Initial Sputum Test attribute to five categories, namely Negative, 1-9BTG, +1, +2, and +3. The discretization is based on the International Union Against Tuberculosis and Lung Disease (IUATLD) scale [37]. There are 28 records which are categorized into 1-9BTG. Besides, in the Additional Complaints attribute, there are seven other disease records generalized to None. Those are gastroesophageal reflux disease (GERD), CA, Chronic obstructive pulmonary disease (COPD), decoin, and three other records are pneumonia.

Attribute selection is conducted using the wrapper function at WEKA. The datasets consist of 1374 records and 17 attributes. Several attempts to attribute selection performed on Dataset2 using the wrapper function are presented in Table II.

TABLE II. COMPARISON OF DIFFERENT CLASSIFIERS' ACCURACY ON DATASETS WITH OR WITHOUT FEATURE SELECTION

Classifier	Attributes	Accuracy			
		without FS		with FS	
		ND	WD	ND	WD
C4.5	Preliminary diagnosis, Initial CXR, Initial Sputum Test	96,69%	96,69%	<b>97,38%</b>	<b>97,38%</b>
Naive Bayes	Preliminary diagnosis, Initial CXR, Initial Sputum Test, Age, Fever, Breathless	96,87%	96,94%	97,20%	97,42%
SVM	Preliminary diagnosis, Initial CXR, Initial Sputum Test, Sex, Age	<b>97,38%</b>	<b>97,38%</b>	<b>97,38%</b>	<b>97,38%</b>
KNN	Preliminary diagnosis, Initial CXR, Initial Sputum Test	90,39%	90,39%	96,94%	<b>97,38%</b>

Determination of training data and test data was carried out by stratified random sampling. First, the data is sorted by Final Diagnosis class, namely Non-TB, Pulmonary TB, and SNPT. Then the percentage of each class is calculated. Training data are taken as much as 70% of the total data, and 30% for testing data. Data were selected randomly but still took into account the percentage of each class (Fig 1).

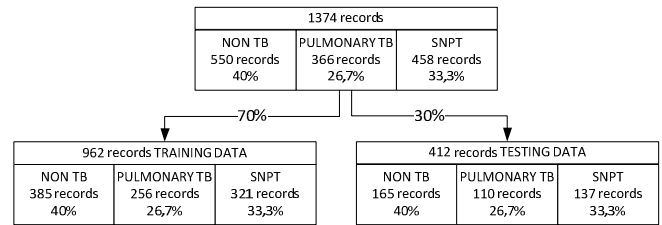


Fig. 1. The composition of training data and testing data.

The evaluation of the diagnosis model of SNPT was done using the confusion matrix. The purpose was to find out the average accuracy, micro averaging precision ( $precision_{\mu}$ ), and micro averaging sensitivity ( $sensitivity_{\mu}$ ) of the model. The testing data used amounts to 412 records. Determination of test data using stratified random sampling. Calculation of average accuracy using equation (1), micro averaging precision using equation (2), and micro averaging sensitivity using equations (3) [38].

$$Average Accuracy = \frac{\sum_{i=1}^l \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}}{l} \times 100\% \quad (1)$$

$$Precision_{\mu} = \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l (FP_i + TP_i)} \times 100\% \quad (2)$$

$$Sensitivity_{\mu} = \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l (TP_i + FN_i)} \times 100\% \quad (3)$$

where,

$TP_i$  the *True Positive* of  $i^{th}$  class, which is the amount of positive data of  $i^{th}$  class that classified correctly by the classifier;

$TN_i$  the *True Negative* of  $i^{th}$  class, which is the amount of negative data of  $i^{th}$  class that classified correctly by the classifier;

$FN_i$  the *False Negative* of  $i^{th}$  class, is the amount of negative data but is incorrectly classified by the classifier;

$FP_i$  : *False Positive* of  $i^{th}$  class, is the amount of positive data but is incorrectly classified by the classifier;

$l$  is a number of class;  $\mu$  is micro averaging.

#### D. Modeling

The study aims to model the diagnosis of SNPT by exploring three ensemble methods, namely random forest, AdaBoost, and Bagging. Weka 3.7 is used to build the model. The development of the SNPT diagnosis model used 962 records of training data.

Data sets resulted through attribute selection were trained with several ensemble learning methods using 10-fold Cross Validation [33], [34], [39]. Exploration of several combinations of the single classifier, AdaBoost and bagging was done to determine the best performing classification method. The classification results in the form of 3 classes, namely pulmonary TB, SNPT Pulmonary TB, SNPT, and Non-TB.

#### E. Evaluation

At this stage, evaluation is carried out on the classification model of SNPT diagnosis using 412 records of testing data. The purpose is to measure the multi-class classification model

using the value of average accuracy, micro averaging sensitivity, and micro averaging precision of the model [38].

#### IV. EXPERIMENTAL RESULTS

In this section, experimental results are presented, which includes the data set used in our experiments and the reported accuracy. The experimental results are presented in Table III.

TABLE III. COMPARISON OF DIFFERENT ENSEMBLE METHODS' ACCURACY ON TRAINING AND TESTING DATASETS

Dataset	Classifier		
	Random Forest	Bagging with C45	Adaboost with C4.5
Training Data	91,79	91,79	91,79
Testing Data	90,59	86,91	90,54

Table III shows that by using stratified random sampling in determining training data and test data, the accuracy of the model during training and testing is comparable. Random Forest has the best accuracy, which is 90.59%, followed by Adaboost (90.54%) and Bagging (86.91%). The confusion matrix is presented in Table IV.

TABLE IV. CONFUSION MATRIX

		Predicted as		
		Non TB	PTB	SNPT
Actual	Non TB	134	0	34
	PTB	0	112	0
	SNPT	5	0	127

Table IV shows that out of 168 non-TB patients, 34 of them were predicted as SNPT. Whereas from 132 SNPT patients, five patients that incorrectly predicted as Non-TB. All classes of pulmonary TB are correctly classified as pulmonary TB, which is in accordance with the rules for TB diagnoses established in the National TB Control Guide in Indonesia (P2PL, 2014).

Analysis of incorrectly classified data is done by comparing TB diagnoses in medical. The fact is if the results of negative sputum examination and chest X-ray give an infiltrate picture, and the results of the initial diagnosis are Non-Suspected TB, then the final diagnosis is Non-TB. If the results of negative sputum examination and chest X-ray give an infiltrate picture, and the results of the initial diagnosis are suspected TB, then the final diagnosis is SNPT.

Whereas, in data testing, out of five SNPT patients who were incorrectly classified, had negative sputum examination results and chest X-ray gave an infiltrate picture, and the results of the initial diagnosis were Non-Suspected TB. They were diagnosed by doctors as SNPT patients, whereas according to the medical rules, they should be diagnosed as Non-TB. Besides, 34 non-TB patients who were incorrectly classified had negative sputum examination results, and a chest X-ray gave an infiltrate picture, and the results of the initial diagnosis were suspected of TB. Those data, according to the rules, are diagnosed as non-TB, but the doctor determined them as SNPT patients. In this case, the doctor's consideration when making a diagnosis of the patient becomes something that cannot be explained in this study.

Evaluation of the model used three parameters, these are average accuracy using equation (1), micro averaging precision using equation (2), and micro average sensitivity using equation (3). The results of the evaluation model are presented in Table V.

TABLE V. EVALUATION RESULT

Model	Average Accuracy	Precision <sub>μ</sub>	Sensitivity <sub>μ</sub>
Random Forest	90,59%	90,53%	90,53%

The evaluation results in Table V show that the model has high accuracy in classifying classes of pulmonary TB, SNPT, and non-TB. The sensitivity of 90.53% means that the proposed model has a high true positive level so that it can classify pulmonary TB and SNPT well. High sensitivity indicates that the model can reduce the occurrence of cases underdiagnosed. Precision 90.53% states that the proposed model has a right level of accuracy in classifying a class.

The diagnosis of SNPT is established by TB trained doctors based on the results of the initial sputum examination and at least the chest X-ray results as a follow-up examination. However, in conditions of limited resources, this model can be used as a diagnostic tool.

#### V. CONCLUSION

The attributes used for diagnosing SNPT are obtained using the wrapper method, namely Initial Sputum Test, Initial CXR, Preliminary Diagnosis, and Final Diagnosis. These are in accordance with the National Guidelines for TB Control in Indonesia.

Determination of training and testing data using stratified random sampling makes the performance of the model during the training and testing phase have comparable performance.

Random Forest has a high sensitivity, which is 90.53%. The sensitivity shows the ability of the Random Forest to classify Pulmonary TB and SNPT well, thereby reducing the occurrence of underdiagnosed cases.

This is preliminary research. The results of this study can be used as the basis for the development of the final diagnosis model of SNPT. As a result of exploring feature selection methods, C4.5 has a comparable performance with SVM and KNN. So, it is necessary to develop a diagnosis model as future research using the stacking and voting method.

#### REFERENCES

- [1] WHO, WHO Global Tuberculosis Report 2015, 20th ed. WHO Library Cataloguing-in-Publication Data, 2015.
- [2] K. R. D. P2PL, Pedoman Nasional Pengendalian Tuberkulosis. Kemenkes RI, 2014.
- [3] A. A. Imianvan and J. C. Obi, "Fuzzy Cluster Means Expert System for the Diagnosis of Tuberculosis," Glob. J. Comput. Sci. Technol., vol. 11, no. 6, 2011.
- [4] G. Alvarez-Uria, J. M. Azcona, M. Midde, P. K. Naik, S. Reddy, and R. Reddy, "Rapid Diagnosis of Pulmonary and Extrapulmonary Tuberculosis in HIV-Infected Patients. Comparison of LED Fluorescent Microscopy and the GeneXpert MTB/RIF Assay in a District Hospital in India," Tuberc. Res. Treat., pp. 1-4, 2012.

- [5] M. Bhatt, R. Bhaskar, and S. Kant, "Pulmonary Tuberculosis as Differential Diagnosis of Lung Cancer," *South Asian J. Cancer*, vol. 1, no. 1, pp. 36, 2012.
- [6] F. M. E. Uzoka, J. Osuji, and F. O. Aladi, "A Framework for Cell Phone Based Diagnosis and Management of Priority Tropical Diseases," in *IST-Africa 2011 Conference*, 2011, pp. 1–13.
- [7] Tuberculosis Coalition for Technical Assistance, *International Standards for Tuberculosis Care (ISTC)*, Second. The Hague, 2009.
- [8] H. M. S. C. Kusuma, "Diagnostik Tuberkulosis Baru," *Sari Pediatr.*, vol. 8, no. 4, pp. 143–151, 2007.
- [9] A. Nesredin, "Mining Patients' Data for Effective Tuberculosis Diagnosis: The Case of Menelik II Hospital," *Addis Ababa University*, 2012.
- [10] R. U. K. R. M. Radzi, W. Mansor, and J. Johari, "Review of Mycobacterium Tuberculosis Detection," in *2011 IEEE Control and System Graduate Research Colloquium*, 2011, pp. 189–192.
- [11] A. Jain, "Extra Pulmonary Tuberculosis: A Diagnostic Dilemma," *Indian J. Clin. Biochem.*, vol. 26, no. 3, pp. 269–73, Jul 2011.
- [12] M. Bahadori and M. H. Azizi, "Common Challenges in Laboratory Diagnosis and Management of Tuberculosis," *Iran. Red Crescent Med. J.*, vol. 14, no. 1, pp. 3–9, 2012.
- [13] H. F. Swai, F. M. Mugusi, and J. K. Mbwambo, "Sputum Smear Negative Pulmonary Tuberculosis: Sensitivity and Specificity of Diagnostic Algorithm," *BMC Res. Notes*, vol. 4, pp. 475, Jan 2011.
- [14] F. C. D. Q. Mello, L. G. do V. Bastos, S. L. M. Soares, V. M. Rezende, M. B. Conde, R. E. Chaisson, A. L. Kritski, A. Ruffino-netto, and G. L. Werneck, "Predicting Smear Negative Pulmonary Tuberculosis with Classification Trees and Logistic Regression: A Cross-sectional Study," *BMC Public Health*, vol. 6, no. 43, pp. 1–8, 2006.
- [15] Y. Benfu, S. Hongmei, S. Ye, L. Xiuhui, and Z. Bin, "Study on the Artificial Neural Network in the Diagnosis of Smear Negative Pulmonary Tuberculosis," in *2009 World Congress on Computer Science and Information Engineering*, 2009, pp. 584–588.
- [16] R. Santiago-Mozos, F. Perez-Cruz, M. Madden, and A. Artes-Rodriguez, "An Automated Screening System for Tuberculosis," *IEEE J. Biomed. Heal. Informatics*, no. 99, pp. 1–8, Okt 2013.
- [17] C. M. Muvunyi and F. Masaisa, "Diagnosis of Smear-Negative Pulmonary Tuberculosis in Low-Income Countries: Current Evidence in Sub-Saharan Africa with Special Focus on HIV Infection or AIDS," in *Understanding Tuberculosis - Global Experiences and Innovative Approaches to the Diagnosis*, P.-J. Cardona, Ed. InTech, 2006, pp. 127–146.
- [18] M. K. Osman, F. Ahmad, Z. Saad, M. Y. Mashor, and H. Jaafar, "A Genetic Algorithm-Neural Network Approach for Mycobacterium Tuberculosis Detection in Ziehl-Neelsen Stained Tissue Slide Images," in *2010 10th International Conference on Intelligent Systems Design and Applications*, 2010, pp. 1229–1234.
- [19] M. K. Osman, M. H. M. Noor, M. Y. Mashor, and H. Jaafar, "Compact Single Hidden Layer Feedforward Network for Mycobacterium Tuberculosis Detection," in *2011 IEEE International Conference on Control System, Computing and Engineering*, 2011, pp. 432–436.
- [20] I. Siena, K. Adi, R. Gernowo, and N. Mirnasari, "Development of Algorithm Tuberculosis Bacteria Identification Using Color Segmentation and Neural Networks," *Int. J. Video Image Process. Netw. Secur. IJVIPNS-IJENS*, vol. 12, no. 4, pp. 9–13, 2012.
- [21] V. S. Kodogiannis, "Point-of-Care Diagnosis of Bacterial Pathogens in Vitro, Utilising an Electronic Nose and Wavelet Neural Networks," *Neural Comput. Appl.*, 2013.
- [22] B. H. Tracey, G. Comina, S. Larson, M. Bravard, J. W. López, and R. H. Gilman, "Cough Detection Algorithm for Monitoring Patient Recovery from Pulmonary Tuberculosis," in *33rd Annual International Conference of the IEEE EMBS*, 2011, pp. 6017–6020.
- [23] R. Lestari, M. Ahmad, B. Alisjahbana, and T. Djatmiko, "The Lung Diseases Diagnosis Software: Influenza and Tuberculosis Case Studies in The Cloud Computing Environment," in *International Conference on Cloud Computing and Social Networking*, 2012, pp. 1–7.
- [24] K. W. Becker, C. Scheffer, M. M. Blanckenberg, and A. H. Diacon, "Analysis of Adventitious Lung Sounds Originating from Pulmonary Tuberculosis," *35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pp. 4334–4337, 2013.
- [25] T. Uçar and A. Karahoca, "Predicting Existence of Mycobacterium Tuberculosis on Patients using Data Mining Approaches," *Procedia Comput. Sci.*, vol. 3, pp. 1404–1411, Jan 2011.
- [26] T. Uçar, A. Karahoca, and D. Karahoca, "Tuberculosis Disease Diagnosis by using Adaptive Neuro Fuzzy Inference System and Rough Sets," *Neural Comput. Appl.*, vol. 23, no. 2, pp. 471–483, Apr 2013.
- [27] M. Lauria, "Rank-based miRNA Signatures for Blood-based Diagnosis of Tuberculosis," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, 2015, pp. 4462–4465.
- [28] R. Khan, S. Abid, W. Jaffri, Z. Abbas, K. Hameed, and Z. Ahmad, "Diagnostic Dilemma of Abdominal Tuberculosis in Non-HIV Patients: An Ongoing Challenge for Physicians," *World J. Gastroenterol.*, vol. 12, no. 39, pp. 6371–6375, 2006.
- [29] A. Goyal, M. Roy, P. Gupta, M. K. Dutta, S. Singh, and V. Garg, "iMedPub Journals Automatic Detection of Mycobacterium tuberculosis in Stained Sputum and Urine Smear Images Abstract," *iMedPub Journals Arch. Clin. Microbiol.*, vol. 6, no. 3:5, pp. 1–4, 2015.
- [30] A. M. Dos Santos, B. B. Pereira, J. M. Seixas, F. C. D. Q. Mello, and A. L. Kritski, "Neural Networks: An Application for Predicting Smear Negative Pulmonary Tuberculosis," in *Advances in Statistics in the Health Sciences*, Birkhäuser Boston, 2007, pp. 275–287.
- [31] A. A. El-solh, C. Hsiao, S. Goodnough, J. Serghani, and B. J. B. Grant, "Predicting Active Pulmonary Tuberculosis Using an Artificial Neural Network," *Chest J.*, vol. 116, no. 4, pp. 968–973, 1999.
- [32] A. M. Dos Santos, A. C. Silva, and L. D. O. Martins, "A Neural-Bayesian Approach for Predicting Smear Negative Pulmonary Tuberculosis," in *Congresso Brasileiro de Redes Neurais*, 2007, no. 1, pp. 1–5.
- [33] T. Asha, S. Natarajan, and K. N. B. Murthy, "Diagnosis of Tuberculosis using Ensemble methods," in *3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT)*, 2010, pp. 409–412.
- [34] T. Asha, S. Natarajan, and K. N. B. Murthy, "Effective Classification Algorithms to Predict the Accuracy of Tuberculosis-A Machine Learning Approach," *Int. J. Comput. Sci. Inf. Secur.*, vol. 9, no. 7, pp. 89–94, 2011.
- [35] J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques*, 3rd ed., vol. 1. Morgan Kaufmann Publishers, 2012.
- [36] T. Asha, S. Natarajan, and K. N. B. Murthy, "Data Mining Techniques in the Diagnosis of Tuberculosis," in *Understanding Tuberculosis - Global Experiences and Innovative Approaches to the Diagnosis*, P.-J. Cardona, Ed. InTech, 2012, pp. 333–352.
- [37] R. Lumb, A. Van Deun, I. Bastian, and M. . Fitz-Gerald, *Laboratory Diagnosis of Tuberculosis by Sputum Microscopy: The handbook*, Global. Adelaide: SA Pathology, 2013.
- [38] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, 2009.
- [39] T. Asha, S. Natarajan, and K. N. B. Murthy, "A Data Mining Approach to the Diagnosis of Tuberculosis by Cascading Clustering and Classification," *J. Comput.*, vol. 3, no. 4, 2011.