

# Chapter 20

## Feature Selection Using Random Forest Algorithm to Diagnose Tuberculosis From Lung CT Images

**Beulah Jeyavathana Rajendran**

*Saveetha School of Engineering, India & Saveetha Institute of Medical and Technical Sciences,  
Chennai, India*

**Kanimozhi K. V.**

*Saveetha School of Engineering, India & Saveetha Institute of Medical and Technical Sciences,  
Chennai, India*

### **ABSTRACT**

*Tuberculosis is one of the hazardous infectious diseases that can be categorized by the evolution of tubercles in the tissues. This disease mainly affects the lungs and also the other parts of the body. The disease can be easily diagnosed by the radiologists. The main objective of this chapter is to get best solution selected by means of modified particle swarm optimization is regarded as optimal feature descriptor. Five stages are being used to detect tuberculosis disease. They are pre-processing an image, segmenting the lungs and extracting the feature, feature selection and classification. These stages that are used in medical image processing to identify the tuberculosis. In the feature extraction, the GLCM approach is used to extract the features and from the extracted feature sets the optimal features are selected by random forest. Finally, support vector machine classifier method is used for image classification. The experimentation is done, and intermediate results are obtained. The proposed system accuracy results are better than the existing method in classification.*

DOI: 10.4018/978-1-6684-7544-7.ch020

## **INTRODUCTION**

Tuberculosis is one of the communicable bacterial diseases that is caused by the bacterium called *Mycobacterium tuberculosis* and that may affect any tissues of the body but it primarily disturbs the lungs. TB is one of the airborne pathogens that can be transmitted through air or by coughing or sneezing from one person to another. TB disturbs all age groups in all parts of the world. Tuberculosis bacteria are present in sputum samples and it is identified under a microscope.

In 2015, around 11 million people fell ill with TB and 2 million people were died from the disease. Over 95% of the deaths in TB occur in low and middle- revenue countries.

X-ray is not easily predicting the early stage of tuberculosis. Hence, because of this wrong prediction of tuberculosis, an automated detection of tuberculosis is used. To overcome the problems in existing methods, CT lung images are used for diagnosis of tuberculosis.

In image processing Feature extraction is an important step, which is a special form of dimensionality reduction. When the input data is too large to be processed and alleged to be redundant then the data is transformed into a reduced set of feature representations. Feature contains the information that is related to colour, shape, texture and context. Modified Random Forest Algorithm technique is based on optimization searching technique and it is used to find the optimal solutions. It is used for selecting the best features after the feature extraction process. This will continue until a needed solution is obtained. Classifying the images whether it is normal or abnormal by SVM classifier.

## **RELATED WORKS**

Les Folio (2014), presented the automated approach for detecting tuberculosis in conventional poster anterior chest radiographs. For the extracted region, set of texture features and shape features are computed, which enable the X-rays to be classified as normal or abnormal using a binary classifier. The pre- processing techniques is used to remove the noises and the feature extraction are done to extract the useful features in given image and the feature selection technique will optimize the top ranking features that are relevant for the image and the classifiers are employed to classify the images and the performance measures are found for the same (Sun et al., 2015). Laurens Hogeweg, Clara I. The performance is evaluated on a TB screening and a TB suspect database using both an external and a radiological reference standard. The systems to detect different types of TB related abnormalities and their combination is described. Yan Kang (2015), Wenbo Li, using a new adaptive VOI selection method. The improved GA algorithm to select the optimal feature combination from the feature pool to establish SVM classifier (Omisore, 2014). G. Vijaya, A. Suhasini identified the cancer tumor from lung CT images using edge detection and boundary tracing. To classify the lung cancer, by using the data mining, classification techniques like SMO (Sequential Minimal Optimization), J48 decision tree, Naive Bayes.

Once the classification is performed, we have to compare the experimental results of the above classification techniques, and determine which one gives accurate and correct answers (Girisha et al., 2013). Mumini Olatunji Omisore (2014) proposed the genetic neuro-fuzzy inferential model for the diagnosis of tuberculosis. Finally, SVM is used in the classification stage (Linguraru et al., n.d.). A. Zabidi, L.Y. Khuan IEEE International conference(2011) proposed the Binary Particle Swarm Optimization For Feature selection in Detection of Infants with Hypothyroidism. In this, he investigates the effect of feature selection with Binary PSO on performance of Multilayer perceptron classifier in discriminating

6 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/feature-selection-using-random-forest-algorithm-to-diagnose-tuberculosis-from-lung-ct-images/315055](http://www.igi-global.com/chapter/feature-selection-using-random-forest-algorithm-to-diagnose-tuberculosis-from-lung-ct-images/315055)

## Related Content

---

### The Use of AI and Structural Magnetic Resonance Imaging in the Diagnosis of Major Depressive Disorder

Deniz Tanirand Asli Beril Karakas (2023). *Machine Learning and AI Techniques in Interactive Medical Image Analysis* (pp. 167-187).

[www.irma-international.org/chapter/the-use-of-ai-and-structural-magnetic-resonance-imaging-in-the-diagnosis-of-major-depressive-disorder/313478](http://www.irma-international.org/chapter/the-use-of-ai-and-structural-magnetic-resonance-imaging-in-the-diagnosis-of-major-depressive-disorder/313478)

### P2P-COVID-GAN: Classification and Segmentation of COVID-19 Lung Infections From CT Images Using GAN

Nandhini Abirami, Durai Raj Vincentand Seifedine Kadry (2023). *Research Anthology on Improving Medical Imaging Techniques for Analysis and Intervention* (pp. 735-755).

[www.irma-international.org/chapter/p2p-covid-gan/315073](http://www.irma-international.org/chapter/p2p-covid-gan/315073)

### Automatic Screening of Diabetic Maculopathy Using Image Processing: A Survey

Shweta Reddy (2023). *Research Anthology on Improving Medical Imaging Techniques for Analysis and Intervention* (pp. 1623-1630).

[www.irma-international.org/chapter/automatic-screening-of-diabetic-maculopathy-using-image-processing/315122](http://www.irma-international.org/chapter/automatic-screening-of-diabetic-maculopathy-using-image-processing/315122)

### Engineering Applications of Artificial Intelligence

Ajanthaa Lakshmanan, R. Seranmadevi, P. Hema Sreeand Amit Kumar Tyagi (2024). *Enhancing Medical Imaging with Emerging Technologies* (pp. 166-179).

[www.irma-international.org/chapter/engineering-applications-of-artificial-intelligence/344668](http://www.irma-international.org/chapter/engineering-applications-of-artificial-intelligence/344668)

### Cardiac Image-Based Heart Disease Diagnosis Using Bio-Inspired Optimized Technique for Feature Selection to Enhance Classification Accuracy

Manaswini Pradhan (2023). *Machine Learning and AI Techniques in Interactive Medical Image Analysis* (pp. 151-166).

[www.irma-international.org/chapter/cardiac-image-based-heart-disease-diagnosis-using-bio-inspired-optimized-technique-for-feature-selection-to-enhance-classification-accuracy/313477](http://www.irma-international.org/chapter/cardiac-image-based-heart-disease-diagnosis-using-bio-inspired-optimized-technique-for-feature-selection-to-enhance-classification-accuracy/313477)