

Chapter 5

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: 10.4018/978-1-7998-3092-4.ch005

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

7 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/271748

Related Content

Prototyping Smart Assistance with Bayesian Autonomous Driver Models

Claus Moebus and Mark Eilers (2011). *Handbook of Research on Ambient Intelligence and Smart Environments: Trends and Perspectives* (pp. 460-512).

www.irma-international.org/chapter/prototyping-smart-assistance-bayesian-autonomous/54671

Self-Organization and Peirce's Notion of Communication and Semiosis

João Queiroz and Angelo Loula (2011). *International Journal of Signs and Semiotic Systems* (pp. 53-61).

www.irma-international.org/article/self-organization-peirce-notion-communication/56447

Supporting Structured Group Decision Making through System-Directed User Guidance: An Experimental Study

Harold J. Lagroue III (2010). *Methodological Advancements in Intelligent Information Technologies: Evolutionary Trends* (pp. 261-278).

www.irma-international.org/chapter/supporting-structured-group-decision-making/38530

Ontology Alignment Overview

José Manuel Vázquez Naya, Marcos Martínez Romero, Javier Pereira Loureiro and Alejandro Pazos Sierra (2009). *Encyclopedia of Artificial Intelligence* (pp. 1283-1289).

www.irma-international.org/chapter/ontology-alignment-overview/10405

Modeling Fault Tolerant and Secure Mobile Agent Execution in Distributed Systems

H. Hamidi and K. Mohammadi (2006). *International Journal of Intelligent Information Technologies* (pp. 21-36).

www.irma-international.org/article/modeling-fault-tolerant-secure-mobile/2395