Egyptian Informatics Journal (2014) **15**, 69–77

Cairo University

Egyptian Informatics Journal

www.elsevier.com/locate/eij
www.sciencedirect.com

ORIGINAL ARTICLE

A new fusion model for classification of the lung diseases using genetic algorithm

C. Bhuvaneswari [a,*], **P. Aruna** [a], **D. Loganathan** [b]

[a] *Department of Computer Science and Engineering, Annamalai University, Chidambaram, Tamil Nadu, India*
[b] *Department of Computer Science and Engineering, Pondicherry Engineering College, Puducherry, India*

Received 8 January 2014; revised 11 April 2014; accepted 5 May 2014
Available online 6 June 2014

KEYWORDS

Computed tomography;
Classification;
Fusion;
Genetic algorithm;
Training

Abstract Automatic classification of lung diseases in computed tomography (CT) images is an important diagnostic tool for computer-aided diagnosis system. In this study, we propose a new image based feature extraction technique for classification of lung CT images. A novel fusion based method was developed by combining the Gabor filter and Walsh Hadamard transform features using median absolute deviation (MAD) technique and hence, it possesses the advantages of both models. The proposed system comprises of three stages. In the first stage, the images are preprocessed and features are extracted by novel fusion based feature extraction technique, followed by second stage, in which extracted features are selected by applying genetic algorithm which selects the top ranked features. In the final stage, classifiers namely decision tree, K nearest neighbor (KNN), Multi layer perceptron Neural Networks (MLP-NN) are employed to perform classification of the lung diseases. A total of 400 datasets for the diseases bronchitis, emphysema, pleural effusion and normal lung were used for training and testing. The classification accuracy of above 90% is accomplished by multilayer perceptron neural network classifier. The system has been tested with a number of real Computed Tomography lung images and has achieved satisfactory results in classifying the lung diseases.

1. Introduction

In the recent years, medical CT Images have been applied in clinical diagnosis widely. It assists physicians to detect and locate pathological changes with more accuracy. Computed tomography images can be distinguished for different tissues according to their different gray levels. Lung diseases can be caused by infection, an exposure at the workplace, medications and various disorders. X-ray chest radiography and computer tomography (CT) are two common anatomic imaging modalities that are routinely used in the detection and diagnosis of a variety of lung diseases.

* Corresponding author. Mobile: +91 9994249954.
E-mail address: bhuvana.csdept@gmail.com (C. Bhuvaneswari).

Peer review under responsibility of Faculty of Computers and Information, Cairo University

Production and hosting by Elsevier

1110-8665 © 2014 Production and hosting by Elsevier B.V. on behalf of Faculty of Computers and Information, Cairo University.
http://dx.doi.org/10.1016/j.eij.2014.05.001

Medical images play a vital role in patient diagnosis, therapy, surgical, medical reference, and training. The Digital Imaging and Communications in Medicine (DICOM) standard [1,17] allows storing textual descriptions, known as metadata, along with the images. It was the most important breakthrough since the discovery of the X-rays, and CT has remained a cornerstone of diagnostic radiology throughout the years. An excellent overview of its technology and applications is given by Kalender [2]. CT imaging had evolved far enough in terms of speed and resolution to make it a valuable tool in the imaging of the lungs [3]. CT lung image classification with correction for perfusion gradient was suggested by Chabat et al., [4].

However, the large number of features actually represents a problem. It leads to the ''dimensionality curse'' problem [5], where the indexing structures degrade and the significance of each feature decreases, making the process of storing, indexing and retrieving extremely time consuming. Pathologies are clearly identified using automated CAD system [6]. It helps the radiologist in analyzing the digital images to bring out the possible outcomes of the diseases. The medical images are obtained from different imaging systems such as MRI scan, CT scan, Ultrasound scans.

A new process was proposed by automated method to quantification of affected lung parenchyma intensity distortion caused by gravity dependent perfusion gradient. Associated with the characteristics of exploitation and exploration, GAs can efficiently deal with large search spaces, and hence are less prone to get stuck into a local optimum solution when compared to other algorithms. This derives from the GAs ability to handle multiple concurrent solutions (individuals) in the search space and apply probabilistic genetic operators [7–9]. From the image processing point of view, it is important to gather as much features as possible to represent the images, yielding vectors with hundreds or even thousands of features to represent the images. Ginneken [10] has classified the lung regions extraction approaches into two different categories: either rule-based or pixel classification based category. Most of the proposed approaches belong to rule-based category [11,12], where a sequence of steps, tests and rules are used in the extraction process.

The computerized tomography has been found to be the most reliable method for early detection of tumors. Chiou et al. [12] proposed application of neural network based hybrid system for lung nodule detection, which based on artificial neural network architectures were developed for improving diagnostic accuracy and speed for lung cancerous pulmonary radiology. The configuration of the HLND system included the following processing phases; data acquisition and pre-processing, in order to reduce and to enhance the figure-background contrast, quick selection of nodule suspects based upon the most prominent feature of nodules, the disk shape and completed features pace determination and neural classification of nodules [13–15].

Genetic algorithms (GAs) are among the most used techniques to perform feature selection due to GAs ability to obtain either exact or approximate solutions in very large search spaces within tractable time. GA performs adaptive searching, following the standard concepts from natural genetics and evolution based on natural selection [16]. Owing to their potential, GA is employed in this work to perform feature selection.

The proposed work is organized as follows, the input of the system is CT lung images for which preprocessing is done to enhance the image, feature extraction is carried out by the proposed fusion median absolute deviation techniques which fuse the features of the gabor filter and Walsh Hadamard transform, feature selection is through genetic algorithm and the classification of the images is done through the classifiers and the performance measures, classification accuracy of each methods is discussed in this work.

2. Materials and methods

CT images of Lung diseases such as Emphysema, Bronchitis, Pleural effusion and normal lung are considered for classification in this work. The patient's age ranging from 15 to 50 comprising of both male and female are taken in this work. The images are obtained from the 16 slice Philips MX 16 evo CT scanner from Sri Manakula vinayagar medical college and hospital, Madagadipet, Puducherry and are anamolized by the radiologist. The original image is split into non-overlapping blocks of 16×16 windows and gray scale image is extracted.

3. Proposed work

The flow diagram of the proposed work is given in Fig. 1. The figure explains the block diagram for the proposed work that is represented in various steps as follows.

3.1. Preprocessing

Median filter and morphological smoothening filter techniques are applied to remove the noise from the images and enhance the image. The median filter will remove the salt and the pepper noises and produces the enhanced image. The dilation and erosion process is done for morphologically smoothening of the images. Erosion involves the alteration (removal) of pixels at the edges of regions, i.e., exchanging binary 1 value to 0, while dilation is the reverse process with regions growing out from their boundaries. The Erosion followed by dilation is known as Opening operation which suppresses the bright details smaller where dilation followed by erosion is closing operation which suppresses the dark details are computed. Herein this work the darker details are suppressed where the dilation is performed followed by erosion. The preprocessed image is obtained by applying the median filter and morphological smoothening.

3.2. Feature extraction

The feature extraction refers to creating a subset of new features by combination of the existing features. The purpose of feature extraction is to reduce original dataset by measuring certain features that distinguish one region of interest from another. Each subimage is taken from top left corner of the original image and the texture features and pixel coefficient values are extracted. The analysis and characterization of textures present in the medical images can be done by using statistical feature extraction method. Texture analysis is a quantitative method that can be used to quantify and detect structural abnormalities.