BIOCYBERNETICS AND BIOMEDICAL ENGINEERING XXX (2018) XXX–XXX

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/bbe

Original Research Article

Extraction of fuzzy rules at different concept levels related to image features of mammography for diagnosis of breast cancer

Mahsa Goudarzi, Keivan Maghooli *

Department of Biomedical Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran

ARTICLE INFO

Article history:
Received 20 March 2018
Received in revised form
30 July 2018
Accepted 16 September 2018
Available online xxx

Keywords:
Mammography
Breast cancer
Feature extraction
Fuzzy rule extraction
Look-up table method
Bagging algorithm

ABSTRACT

Mammography is an inexpensive and non-invasive method through which one can diagnose breast cancer in its early stages. As these images need interpretation by a radiologist, this may develop some problems due to fatigue, repetition, and need for a great deal of attention to details and other factors. Thus, a method capable of diagnosing breast cancer should be employed to help physicians in this regard.

In this paper, The mini Mammographic Image Analysis Society (mini-MIAS) database of mammograms is used. The aim is to distinguish between normal and abnormal classes. In the preprocessing stage, noise removal, removal of labels of images, heightening the contrast, and ROI segmentation are performed, and then compactness, entropy, mean, and smoothness are extracted from the images. In addition to classification, we have come to a new approach in order to create a complete knowledge base, which then we use this knowledge base for classification. We have a comprehensive knowledge base which covers all the conceptual levels.

The extracted features are referred to as fuzzy classifiers through the look-up table method. And, for evaluation of the results, the 10-fold method is used. Discretization operations are performed on training data across 2, 3, and 4 levels to develop concept hierarchy. Concept hierarchies reduce the data by replacing low-level concepts with higher-level concepts and the outcome is more meaningful and easier to interpret. Eventually, Bagging algorithm is used for finding out the majority vote and the final result of the discretization levels. The obtained accuracy is 89.37 ± 6.62.

1. Introduction

According to the World Health Organization (WHO), breast cancer has changed into the second common cause of death resulting from cancer in women [1]. Cancer treatment methods have a relatively high cost, but before any treatment, timely diagnosis is required. Although there is no effective way for preventing breast cancer, early diagnosis of this cancer is

* *Corresponding author at:* Department of Biomedical Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran.
E-mail addresses: Mahsa_Goudarzie@yahoo.com (M. Goudarzi), K_maghooli@srbiau.ac.ir (K. Maghooli).

https://doi.org/10.1016/j.bbe.2018.09.002
0208-5216/

essential for reducing mortality [2]. Thus, many attempts have been made to diagnose cancer in its primary stages. Imaging technologies play an important role in early diagnosis, and mammography has been widely accepted as a diagnostic and screening tool for diagnosing breast cancer in its early stages [3].

There are two types of tumors. neoplasms are in two types: malignant and benign. A tumor is benign when it is microscopic and apparel characteristics are almost harmless. In other words, it remains limited and topical and cannot metastasize to other parts of the body. Therefore, a surgical excisional removal can usually save a patient's life. Malignant tumors are called the cancer. This disease can invade and destroy adjacent structures. In addition, it can metastasize to further body parts and cause death.

Nowadays there are variety type of imaging producers for breast cancer such as magnetic resonance imaging (MRI) with different type and resolution, ultrasound or sonography, digital mammogram (DM), microscopic (histological) images, and infrared thermography (IRT) [4].

Mammographic screening allows for early diagnosis of a tumor with an inexpensive and noninvasive approach. However, there remains the problem of analysis of mammographic images by a radiologist. This analysis depends on the experience of the radiologist, and is also affected by fatigue and other factors. In addition, this is a repetitive task, which needs a great deal of attention to image details. Accordingly, the need for automatic systems for assisting the physician for early diagnosis has been strongly felt. These software systems are not a substitute for physician. Rather, they are used for providing a second view. Among the advantages of these methods are faster speed and lower costs [5]. The main objective of this project is to develop a novel method for diagnosing breast cancer. Various researches have been conducted for diagnosis and classification of breast cancer. Some of them are described below:

A study by Dora et al. [6] presents a Gauss–Newton representation based algorithm (GNRBA). This study has been conducted on the Wisconsin Diagnosis Breast Cancer (WBCD) database adapted from university of California (UCI) and its objective is to help in classification through finding optimal weights for the training samples. Bhardwaj et al. [7] presented a genetically optimized neural network algorithm (GONN) for classification of breast cancer, in which GA has been used for development of the optimized neural network architecture. Hassan et al. [8] performed the classification through the Hybrid Hidden Markov (HMM)-Fuzzy method. First, the features are selected based on the area under curve (AUC) of the region of interest (ROC) curve using the training data and then HMM generates the Log-Likelihood values for development of minimized fuzzy rules, which have been optimized by the Gradient Descent algorithm for better classification. Using the hybrid machine learning method, Şahan et al. [9] helped in breast cancer diagnosis, where the diagnosis has been made by hybridizing a fuzzy-artificial immune system and k-nearest neighbor (K-NN) algorithm on WBCD data. Keleş et al. [10] presented a method called Expert System for diagnosis of breast cancer (EXDBC). The fuzzy rules used in the inference engine of this system have been obtained using the neuro-fuzzy method. Karabatak et al. [11] presented a method for diagnosis of breast cancer, which included association rules (AR) system and neural network. AR has been used for reducing the dimensions of the breast cancer database, while neural network (NN) has been employed for its intelligent classification. Miranda et al. [12] used aspects of fuzzy logic for improving the presentation of associated features or image in order to make them a concept. The user feeds some information about the geometrical shape, contour, and density, and then the system offers a suggestion for BI-RAD classification. Nilashi et al. [13] used a knowledge-based system for classification of cancer data, which includes clustering, noise removal, and classification techniques. Expectation Maximization (EM) has been used for clustering data in similar groups. Classification and Regression Trees (CART) has been used for generation of fuzzy rules for classification in the fuzzy knowledge-based system. Ghosh et al. [14] presented a neuro-fuzzy classification method for data mining, whereby the inputs in this classification have been fuzzified by applying a bell-shape membership function. They also presented a fuzzification matrix in which the input patterns had a membership degree across different classes and based on the value of that membership degree, the pattern is allocated to a certain group or class. This has been done on ten data of the UCI Machine Learning database for classification and the aim is to analyze this method and compare it with evaluation of the performance of the two classification methods with Radial Basis Function Neural Network (RBFNN) and Adaptive Neuro-Fuzzy Inference System (ANFIS) supervision. This is a better method in comparison to these two methods. Töpfer et al. [15] performed a comparison between Hierarchical look-up table and local linear model tree (LOLIMOT) methods, both of which were identified as suitable for detection of complex nonlinear systems. LOLIMOT is a local linear neuro-fuzzy model, which is preferentially used for offline simulations. The look-up table performs online simulations of specified nonlinear systems under certain conditions.

The aim of this research is to create a complete knowledge base for classification. The more complete the knowledge base becomes, the less uncertainty there will be. We have a comprehensive knowledge base which covers all the conceptual levels.

This paper is organized as follows: Section 2 presents the methods. Section 3 reports the results which are discussed in Section 4 and conclusions are reported in Section 5.

2. Methods

Considering the importance of cancer diagnosis, the main aim of this research is to develop a conceptual hierarchy in a feature space from images and expression of fuzzy rules, given the uncertainty. Data discretization can reduce the number of values by dividing the range of the attribute into intervals and afterward can be performed recursively on the attribute to provide a hierarchical partition of the attribute values, known as concept hierarchy. A concept hierarchy for a numerical attribute defines a discretization of the attribute. Concept hierarchies reduce the data by replacing low-level concepts with higher-level concepts and the outcome is more meaningful and easier to interpret. The issues of pre-processing and

Please cite this article in press as: Goudarzi M, Maghooli K. Extraction of fuzzy rules at different concept levels related to image features of mammography for diagnosis of breast cancer. Biocybern Biomed Eng (2018), https://doi.org/10.1016/j.bbe.2018.09.002