Ensemble Classifier for Benign-Malignant Mass Classification

Muhammad Hussain, Department of Software Engineering, King Saud University, Riyadh, Saudi Arabia

ABSTRACT
Mammography is currently the most effective imaging modality for early detection of breast cancer. In a CAD system for masses based on mammography, a mammogram is segmented to detect the masses. The segmentation gives rise to mass regions of interest (ROIs), which are either benign or malignant. There is a need to classify the extracted mass ROIs into benign and malignant masses; it is a hard problem because the texture micro-structures of benign and malignant masses have close resemblance. In this paper, a method for classifying mass ROIs into benign and malignant masses is presented. The key idea of the proposal is to build an ensemble classifier that employs Gabor features, consults different experts (classifiers) and takes the final decision based on majority vote. The system is evaluated on 512 (256 benign + 256 malignant) mass ROIs extracted from mammograms of DDSM database. The ensemble classifier improves the classification rate for the problem of the discrimination of benign and malignant masses to 90.64%. Comparison with state-of-the-art techniques suggests that the proposed system outperforms similar methods.

Keywords: Benign Masses, Breast Cancer Detection, CAD System, Ensemble Classifier, Malignant Masses

1. INTRODUCTION
Breast cancer is the second major form of cancer causing deaths among women all over the world. According to the statistics of National Cancer Institute, Surveillance, Epidemiology, and End Results (SEER) program, lifetime risk of developing breast cancer among American women is 12.2% (aka: one in eight), exceeded only by the lung cancer [1]; [2]. In the European Community, breast cancer represents 19% of cancer deaths and 24% of all cancer cases [3]; [4]. Women diagnosed between ages 40-49 years are the major victims having about 25% of all breast cancer deaths. The World Health Organization’s International Agency for Research on Cancer (IARC) has estimated more than one million cases of breast cancer to occur annually and reported that more than 400,000 women die each year from this disease [5].

Mammography is considered to be the most effective screening method for the early detection of breast cancer. A huge volume of mammograms to be screened in a routine checkup puts a heavy burden on the radiologists; due to human limitations some cases may be overlooked and a cancer case may be escaped to grow to an incurable state. The solution of this problem is to double check each case, but it needs more human resource which is financially

DOI: 10.4018/ijcvip.2013010106
unaffordable. The alternative is to use computer aided diagnosis (CAD) system, which can be used as a means for double check. By digitizing the mammograms and exploiting contemporary powerful image analysis techniques, CAD systems can be developed for effectively assisting the radiologists, but CAD systems could not get due attention because of high rate of false positives and false negatives, and further investigation is needed to come up with more robust and powerful image processing and pattern recognition tools. A CAD system for masses segments mammograms for extracting mass ROIs; these mass ROIs are either benign or malignant. The problem is to classify the extracted ROIs into benign and malignant masses; this is a hard problem and needs powerful representation for masses and a robust classification method. In this paper, we address this problem. The discrimination of mass ROIs as benign and malignant masses involves two main steps: feature extraction and classification procedure. We employ Gabor features for the description of ROIs, and introduce an ensemble classifier for classification.

Classification of mass ROIs into malignant and benign cases is a difficult problem, and a single classifier does not perform better. In our real world, matters that are complex to solve but have significant importance in our lives, we often seek the opinion from several experts before making a mindset to select an option from different available choices. These matters could be profoundly crucial where the decisions can have severe implications on the life of a patient, e.g. whether a patient is effected with breast cancer or not? We are going to utilize the same simple concept known with different names in the literature. Specifically, we are going to utilize a committee of classifiers for deciding whether or not a given mammogram image is effected with possible cancerous regions i.e. malignant masses. Considering a single classifier to be an expert, we have a team of experts for making the accurate decision. The experts are K Nearest Neighbors (KNN), Random Forest (RF) and Support Vector Machine with RBF kernel (SVM-RBF). The random forest consists of several decision trees and thus the number of actual experts is much larger than mentioned. The ensemble classifier improves the classification rate for the problem of the discrimination of benign and malignant masses, which is the hardest problem, to 90.64%.

The remainder of this paper is organized as follows. In Section 2, we present the proposed method. Section 3 gives the details of the database and the experimental setup used in our experiments. In Section 4, we present some experimental results to show the effectiveness of the proposed technique. Finally, Section 5 concludes this work.

2. RELATED WORK

Because of the impact of mass classification problem in detecting breast cancer, it has attracted the attention of many researchers, and many methods have been proposed. For a detailed review of these methods, an interested reader is referred to the review papers [17]; [18]; [19]; [20]. In the following paragraphs, we focus on the review the most related recent mass detection methods.

The main differentiating factors among different existing methods are the types of features that have been used for mass description, the way these features have been extracted, and the classification techniques. For classification different techniques have been used such as Multi-layer Perceptron Neural Network (MLP), Bayesian classifier, Support Vector Machine (SVM), Twin Support Vector Machine, Linear Discriminant Analysis (LDA).

Delogue et al. (2007) employed MLP for the characterization of mammographic masses. For validation, they used a data set consisting of 109 malignant and 117 benign masses, a total of 226 masses. Masses were represented using sixteen features extracted using gradient based segmentation algorithm. The architecture of the MLP used consisted of n input, three hidden and two output neurons, where n is the number of features; sigmoid activation function was used for both the hidden and output layers. For the
Related Content

**Stereo-Vision-Based Fire Detection and Suppression Robot for Buildings**
[www.igi-global.com/chapter/stereo-vision-based-fire-detection/60278?camid=4v1a](www.igi-global.com/chapter/stereo-vision-based-fire-detection/60278?camid=4v1a)

**Performance Analysis of Mobile Ad-Hoc Network Protocols Against Black Hole Attacks**
[www.igi-global.com/article/performance-analysis-of-mobile-ad-hoc-network-protocols-against-black-hole-attacks/87251?camid=4v1a](www.igi-global.com/article/performance-analysis-of-mobile-ad-hoc-network-protocols-against-black-hole-attacks/87251?camid=4v1a)

**Human Activity Recognition Using Gait Pattern**
[www.igi-global.com/article/human-activity-recognition-using-gait-pattern/95969?camid=4v1a](www.igi-global.com/article/human-activity-recognition-using-gait-pattern/95969?camid=4v1a)
Direct 3D Information Determination in an Uncalibrated Stereovision System by Using Evolutionary Algorithms
www.igi-global.com/article/direct-information-determination-uncalibrated-stereovision/55098?camid=4v1a