Expert Systems with Applications 28 (2005) 713–723

Expert Systems with Applications

www.elsevier.com/locate/eswa

Breast cancer diagnosis system based on wavelet analysis and fuzzy-neural

Rafayah Mousa, Qutaishat Munib*, Abdallah Moussa

Department of Computer Information Systems, University of Jordan, Amman 11942, Jordan

Abstract

The high incidence of breast cancer in women has increased significantly in the recent years. The most familiar breast tumors types are mass and microcalcification. Mammograms—breast X-ray—are considered the most reliable method in early detection of breast cancer. Computer-aided diagnosis system can be very helpful for radiologist in detection and diagnosing abnormalities earlier and faster than traditional screening programs. Several techniques can be used to accomplish this task. In this paper, two techniques are proposed based on wavelet analysis and fuzzy-neural approaches. These techniques are mammography classifier based on globally processed image and mammography classifier based on locally processed image (region of interest). The system is classified normal from abnormal, mass for microcalcification and abnormal severity (benign or malignant). The evaluation of the system is carried out on Mammography Image Analysis Society (MIAS) dataset. The accuracy achieved is satisfied.

Keywords: Digital mammogram classifier; Breast cancer; Mass tumor; Microcalcification; Wavelet analysis; ANFIS

1. Introduction

The interpretation and analysis of medical images represent an important and exciting part of computer vision and pattern recognition. Developing a computer-aided diagnosis system for cancer diseases, such as breast cancer, to assist physicians in hospitals is becoming of high importance and priority for many researchers and clinical centers. It is a complex process to develop a computer vision system to perform such tasks.

The high incidence of breast cancer in women has increased significantly in the recent years. It is the cause of the most common cancer death in women. It is a leading cause of fatality in women, with approximately 1 in 12 women affected by the disease during their lifetime (Spence, Parra, & Sajda, 2001). In Australia, approximately 1 of 13 women develops the disease (Verma & Zakos, 2000). A report from the National Cancer Institute (NCI) estimates that about one in eight women in the United States (approximately 12.5%) will develop breast cancer during their lifetime (Arun, 2001). Early detection plays a very important factor in cancer treatment and allows better recovery for most patients. The required medical image for the diagnosing process of breast cancer, mammogram (breast X-ray), is considered the most reliable method in early detection (Arun, 2001; Verma & Zakos, 2000).

Due to the high volume of images to be analyzed by radiologists, and since senior radiologists are rare, reliable radiological diagnosis is not always available and the accuracy rate tends to decrease. A statistics shows that only 20–30% of breast biopsies are proved cancerous (Zaiane, Maria-Luiza, & Alexandru, 2002), and 10% of all cases of breast cancer go undetected by mammography (Bird, Wallace, & Yankaskas, 1992). Moreover, digital mammograms are among the most difficult medical images to be read according to the differences in the types of tissues and their low contrasts. Important visual clues of breast cancer include preliminary signs of masses and microcalcification clusters (Roberts, Kahn, & Haddawy, 1995). Unfortunately, at the early stages of breast cancer, these signs are very subtle and varied in appearance, making diagnosis even difficult to specialists. Therefore, automatic reading of digital medical images becomes highly desirable. It has proven that double reading of the mammogram, by two radiologists, increases the accuracy, but at high costs

* Corresponding author. Tel.: +962 7749 8959; fax: +962 6535 5511.
E-mail address: maq@ju.edu.jo (Q. Munib).

0957-4174/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2004.12.028

(Zaiane, Maria-Luiza, & Alexandru, 2001). Therefore, the motivation of the computer-aided diagnosis systems (Mendez, Tahoces, Lado, & Souto, 1998; Taylor, 1995; Woods, 1994; Yin, Giger, Vyborny, Doi, & Schmidt, 1993) is to assist medical staffs to achieve high efficiency and accuracy.

Radiologists basically look for two types of patterns in mammography: micromicrocalcifications and masses (Arun, 2001). The diagnosis result of tissue is classified into three categories: normal which represents mammogram without any cancerous cell, benign which represents mammogram showing a tumor, but not formed by cancerous cells and malign which represents mammogram showing a tumor with cancerous cells (Verma & Zakos, 2000). It is difficult to distinguish a benign from one that is malignant. Consequently, many unnecessary biopsies are often undertaken due to the high positive false rate (Arun, 2001; Wang & Karayiannis, 1998).

Interpreting medical images that used for diagnosing process involves preprocessing and detection of regions of interest (Arun, 2001). Preprocessing stage deals with image enhancement and noise removal. The enhanced image is then scanned for selected region of interest. Histogram equalization is one of many techniques that used to enhance mammograms. The next stage is to extract features from the region of interest. These features then passed to classifier to decide whether this mammogram normal or abnormal. Many approaches are used to build such classifiers for digital mammograms such as neural network and data mining. The neuro-fuzzy approach is a typical approach for the developing of such types of systems (Mitra & Hayashi, 2000; Nieto & Torres, 2003; Russo & Jain, 2001; Verma & Zakos, 2000). Neural network provides algorithms for learning and classification, whereas fuzzy logic deals with issues reasoning on a higher semantic level.

In this paper, two techniques for building a computer-aided diagnosis system for classification of abnormality in digital mammograms are designed and evaluated. The first one is a neuro-fuzzy classifier based on features extracted from the wavelet analysis of the image. It consists of preprocessing, features extraction and classification stages. Histogram equalization and gray level thresholding techniques are applied for enhancing the images. Features are extracted from the whole image, which represents the unit of classification. We called this technique mammography classifier based on globally processed image. In the classification stage, we apply the Adaptive Neuro-Fuzzy Inference System (ANFIS) (Jang, 1993). The purpose of the system is to classify normal mammogram from abnormal one and to determine abnormal severity in the abnormal one. It could be mass (benign or malign) or microcalcification (benign or malign). In this paper the abnormal cases: mass (circumscribed and speculated and microcalcification) and micocalcification are considered.

Many studies have been made on the problem of breast cancer diagnosing based on digital mammograms (Qian, Sunden, Sjostrom, Fenger-Krog, & Brodin, 2002; Sameti & Ward, 1996; Verma & Zakos, 2000; Woods, 1994). Verma and Zakos (2000) developed a system based on fuzzy-neural and feature extraction techniques for detecting and diagnosing microcalcifications' patterns in digital mammograms. They investigated and analyzed a number of feature extraction techniques. The following 14 features were used for the proposed method: average histogram, average gray level, energy, modified energy, entropy, modified entropy, number of pixels, standard deviation, modified standard deviation, skew, modified skew, average boundary gray level, difference and contrast. The formula for entropy, energy, skew, and standard deviation were modified so that the iterations started with the first pixel of the pattern and ended at the final pixel and found that a combination of three features, entropy, standard deviation and number of pixels, is the best combination to distinguish a benign microcalcification pattern from a malignant one. The fuzzy technique only detects the center pixel of a microcalcification area. Therefore, it detects other areas that look like a microcalcification. It is up to the user to decide whether the resulting detection is a microcalcififcation or some other area. The back-propagation technique was used for classification of features into benign or malignant.

Zaiane et al. (2001) used neural network and data mining techniques for detection and classification of digital mammograms. Histogram equalization are used to enhance the images. The proposed methods classified the digital mammograms in two categories: normal and abnormal. The data collection they used in their experiments was taken from MIAS (Suckling et al., 1994). The extracted features that used are two existing features (type of the tissue and position of the breast), and four statistical parameters. In their experiments they used 90% of the dataset—322 images—for training the systems and 10% for testing them. The success rate obtained using the neural network (back-propagation algorithm) is 81% on average. On the other hand, it is 69% on average for association rule classifier. In the following research for Zaiane et al. (2002) the data mining classifier is enhanced by applying two pruning methods of rules. They are eliminating the specific rules and keep only those that are general and with high confidence, and prune some rules that could introduce errors at the classification stage. All the extracted features presented in (Zaiane et al., 2002) have been computed over smaller windows of the original image. The classification rate increased to 80%.

Research into the detection of microcalcifications using the wavelet transform has been carried out by McLeod and Parkin (1996). Extraction of possible microcalcifications is firstly achieved by wavelet decomposition of the mammogram using Deubechies wavelets to three levels. This research showed that microcalcifications are mostly prominent in the high-pass subbands of levels 2 and 3, with level 1 containing mostly noise and fine structural detail.

Wang and Karayiannis (1998) proposed an approach for detecting microcalcifications in digital mammograms