Proceedings of the International Conference on Biomedical Engineering and Systems
Prague, Czech Republic, August 14-15, 2014
Paper No. 140

Comparing Classification Techniques to Detect Breast Tumour

Passant Wahdan, Amani Saad
Computer Engineering Department
Arab Academy for Science and Technology
Alexandria, Egypt
passantwahdan89@gmail.com; amani.saad@aast.edu

Amin Shoukry
Computer and Systems Engineering Department
Alexandria University
Alexandria, Egypt
amin.shoukry@alexu.edu.eg

Abstract - Breast tumour detection in ultrasound images has been a challenge due to the presence of different kinds of noise caused by various factors. The focus of this research is the design, implementation andperformance evaluation ofseveral tumour detection systems based on different classifiers and using ultrasound breast images. First, Gaussian and anisotropic diffusion filters are applied to remove additive and speckle noise, respectively, and histogram equalization is used for image enhancement. Second, textural features are extracted from the input image followed by principal component analysis to reduce the dimensionality of the data set. Finally, the classification process is performed using two different classifiers includingsupport vector machine (SVM) andBootstrap aggregating (bagging) on REP tree. A comparison of the performance of these classifiers is presented.

Keywords: Breast tumor detection, Medical image processing, Ultrasound, Textural analysis, Classification.

1. Introduction

"Cancer is one of the leading causes of death worldwide. One in eight deaths worldwide is due to cancer as shown by Garcia et al. (2007).""More than 17 million people died in 2011 from different types of cancer. 1383500 women were diagnosed with breast cancer while 458400 died that year from the same cause. 89% of women diagnosed with breast cancer are still alive 5 years after their diagnosis. Dramatically, one-third of breast cancer death can be decreased if detected and treated early; this means that nearly 400, 000 lives could be saved every year as shown by Jemal et al. (2011)." Towards this goal, different techniques including thermography, mammography and diagnostic sonography (ultrasound imaging) have been developed. Ultrasound imaging is more reliable than mammography for women under 40. Using it, the exact location, shape and size of a tumour can be found, while in thermography only the presence of a tumour is indicated. Ultrasound medical imaging is an accurate and cost effective method of medical diagnosis. It uses low-power, high frequency sound waves to visualize the body's internal structures. "It is considered non-invasive, practical and harmless. Ultrasound imaging is the main focus of this research. The main problem with ultrasound imaging is the noise caused by imperfect instruments, the data acquisition process, and other natural interfering phenomena as shown by Guo. (2010)." In this paper we propose a model to detect tumors in ultrasound images. The rest of this paper is organized as follows: Section 2 presents the background. Section 3 presents the proposed model.Section 4 summarizes the experimental work. Finally, the conclusions and future work are presented in section 5.

2. Background

Several techniques related to breast tumors have been developed for detecting and differentiating between cancerous and benign tumors using image possessing techniques and ultrasound images. “A new CADsystemwas developed to classify breast tumors using support vector machines as shown by Huang et al. (2006).”“Another also aimed to classify breast tumors in ultrasound images using a hybrid classifier based on a multilayer perceptron network and genetic algorithms as shown by Alvarengal et al. (2007).” Later another computer-aided diagnosis system wasdeveloped to detect and segment the tumor regions.His detection algorithm works in two stages: tumor localization and tumor boundary delineation. In the first stage, the AdaBoost classifier using Haar-like features is employed. Next, an SVM classifier is applied as shown by Jiang et al. (2012).”

“The process is usually divided into three phases: the preprocessing phase, the detection phase and the classification phase. Different techniques areused in each phase. The preprocessing phase deals with different kinds of noise such as amplifier noise (Gaussian noise), Salt and pepper noise, Poisson noise and Speckle noise. Different kinds of filters can be applied to remove noise such as Mean filter, Median filter, Gaussian filter and anisotropic diffusion as shown by Michailovich et al. (2006).”“Ultrasound images contain mainly additive and speckle noise that is why Gaussian filter and anisotropic diffusion have been implemented as shown by Forst et al. (2009).”“Morphological operators and Histogram equalization are examples of enhancement techniquesas shown by Hummel (2012).”

3. Proposed Model

Comparing detection techniques of tumors in ultrasound images is the main focus of this research. The system is divided into three steps pre-processing, feature extraction and detection using a classification-based approach. A block diagram representing the proposed system is shown in figure 1. Ultrasound images contain much noise as shown in figure 2.

Fig. 1. Block diagram of the proposed system.

Fig. 2.Example of an original ultrasound image of a breast tumor.

3.1. Pre-processing Stage

The main concern in the pre-processing stage is applying the de-noising and enhancement techniques without destroying the useful information in an ultrasound image. The pre-processing stage is divided into two steps: filtering and enhancement. Gaussian filters are used to get rid of additive noise. “Also, anisotropic diffusion filter is used to overcome the major drawbacks of conventional spatial filters and improve the image quality by preserving important boundary information as shown by Minavathi(2012).” ”To further improve the image quality, histogram equalization technique is used for image enhancementas shown by Wang et al. (2008).”

3.1.1. Filtering

The main objective of the filtering step is de-noising the image as far as possible while preserving the important data in it. Gaussian smoothing is used to reduce image noise.The Gaussian filter is a non-uniform low pass filter. It removes high-frequency components from the image. It is used in order to enhance image structures at different scales as shown in figure 3. Mathematically, a Gaussian filter modifies the input signal by convolution with a Gaussian function, this transformation is also known as the Weierstrass transform. The Gaussian function is:

$$G(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2}{2\sigma^2}} \quad (1)$$

Where σis the standard deviation of the distribution and assuming that the distribution has a zero mean. Anisotropic diffusion reduces the speckle noise and also blurs the image without compromising its quality as shown in figure 4. "The main idea in anisotropic diffusion is to smooth the homogenous areas of the image while enhancing the edges. This creates a piecewise constant image from which the segmentation boundaries can be easily obtained as shown by Guo. (2010)." "The anisotropic diffusion is implemented using the derivation of speckle reducing anisotropic diffusion (SRAD) as shown by Yu et al. (2002)."

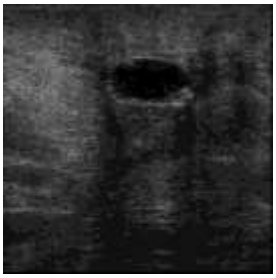

Fig. 3. Image after applying Gaussian filter.

Fig. 4. Image after applying Anisotropic diffusion.

Fig. 5. Image after applying Histogram equalization.

3.1.2. Enhancement

The aim of image enhancement is to improve the image quality, or to provide better input for other automated image processing techniques as shown in figure 5."Histogram equalization is known to adjust image intensities to enhance contrast without affecting the information content of the image that is why it is chosenas shown by Minavathi (2012)."Contrast enhancement is an important problem in both digital and analogue image processing.Histogram equalization helps in reducing differences among images from various ultrasonic systems. The equalized histogram for pixel S is defined as follows

$$S_K = \sum_{j=0}^{K} P_r(r_j) \quad (2)$$

Where r represent the gray level of the pixel to be enhanced, K= 0,1,2,3, L-1 and L is the total number of possible gray levels in the image.

3.2. Feature Extraction and Dimensionality Reduction

"Textural parameters calculated from the gray level co-occurrence matrix, of a preprocessed input image, helps in understanding the image content as shown by Gadkari (2014)."The Gray level Co-occurrence Matrix (GLCM) is a powerful measure used in texture classification. "Calculating textural parameters from the gray level co-occurrence matrix is an indicator of the overall image content and it helps understand the details of theimage as shown by Gadkari (2014)." In order to classify the image properly 16features are extracted from the GLCM. "These features are:

i. Autocorrelation which is the cross-correlation of a signal with itself.
ii. Contrast between a pixel and its neighbor correlation which is a measure of grey tone linear dependencies in the image.
iii. Cluster Prominence which represents the peakedness or flatness of the graph of the co-occurrence matrix with respect to values near the mean value.
iv. Cluster Shade which is a measure of skewness of the image.
v. Energy which is the sum of squared elements from the GLCM.
vi. Entropy which is a statistic measure of the disorder or complexity of an image.
vii. Homogeneity which measures image homogeneity as it assumes larger values for smaller gray tone differences in pair elements.
viii. Sum of squares variance is a measure of the total variability of a set of scores around a particular number, usually the mean of the set of scores.
ix. Sum variance this feature puts high weights on elements that differ from the average value of the image pixel.
x. Additional features including sum entropy, difference variance, difference entropy, information measure of correlation, inverse difference, inverse difference normalized and inverse difference moment normalized as shown by Haralick et al. (1973)."

"Principal component analysis (PCA) is used to reduce the dimensionality of the feature space of the data set, while retaining as much as possible of the variation present in it. "PCA is achieved by transforming the data set into a new set of uncorrelated variables which are ordered so that the first few retain most of the variation present in all of the original variables as shown by Abdi et al. (2010)." A total of 16 features have been extracted from the GLCM. Using PCA, 4 factors were found to contain most of the variation present in the original dataset.The use of a reduced set of uncorrelated features enhances the final classification stage which has been implemented using support vector machine and Bagging.

3.3. Classification

Two classification approaches have been used in the present work, support vector machine and bagging using REP trees.

3.3.1. Support Vector Machine

"SVM is a supervised learning technique that seeks an optimal hyper-plane to separate two classes of samples. Mapping the input data into a higher dimension space is done by using Kernel functionswith the aim of obtaining a better distribution of the data. Then, an optimal separating hyper-plane in the high-dimensional feature space can be easily chosen as shown by Chenga et al. (2009)." Cross-validation is applied to prevent data over fitting. The training set is divided into 10 folds of equal size. Sequentially, one fold is tested using the classifier trained on the remaining 9 folds. Thus, each instance of the whole training set is predicted once so the cross-validation accuracy is the percentage of data which are correctly classified.

3.3.2. Bagging

"Bagging is a technique used for improving the quality of estimators as shown by Breiman (1996)." It is a machine learning ensemble meta-algorithm used to improve the stability and accuracy of the REP tree. In other wordsit is usedtogrow an ensemble of trees and letting them vote for the most popular class. It also reduces the variance and helps avoiding over-fitting. "Bagging is a special case of the model averaging approach. It randomly distorts the data set by re-sampling it. Bagging seems to enhance accuracy when random features are used. REP tree is afast decision tree learner. It builds a decision/regression tree using information gain/variance and prunes it using reduced-error pruning (with backfitting) as shown by Hall et al. (2009)and Campo Avila et al. (2011)."

4. Experimental Work

The proposed method is applied on 107 ultrasound images obtained from different sources. Both the filtering and enhancement steps are implemented using matlabR2012a. Figure 6 shows the result of applying the pre-processing techniques. Principle Component Analysis and the three classification techniques are implemented using Weka 3.6.9. All classifiers have been trained and tested using the same training, validation and testing data, respectively.

Fig. 6.Breast ultrasound image after filtering and enhancement.

4.1 Support Vector Machine

The correctly classified instances were 75.5%. The performance measures of the SVM classifier are shown in Table.1. TP and FP rates correspond to true and false positive rates, respectively. Yes class represents the images defined by a physician as images with tumors and the No class were clean.

Table.1.The performance measures of the SVM classifier after applying image processing techniques.

	TP RATE	FP RATE	PRECISION	RECALL	F-MEASURE
YES CLASS	0.702	0.382	0.611	0.702	0.653
NO CLASS	0.618	0.298	0.708	0.618	0.66
WEIGHTED AVG.	0.657	0.337	0.664	0.657	0.657

4.2 Bagging

The correctly classified instances were 85.9% while the incorrectly classified instances were 14.1%. The performance measures of the Bagging classifier are shown in Table 2.

Table. 2. The performance measures of the Bagging classifier after applying image processing techniques.

	TP RATE	FP RATE	PRECISION	RECALL	F-MEASURE
YES CLASS	0.8	0.065	0.941	0.8	0.865
NO CLASS	0.935	0.2	0.782	0.935	0.851
WEIGHTED AVG.	0.858	0.124	0.872	0.858	0.859

5. Conclusions and Future Work

In this paper, comparisons between different systems for breast tumours detection using Ultrasound images are proposed. It can provide a second opinion for a physician to detect breast tumours. It consists of three main steps pre-processing, feature extraction and classification. Gaussian blurring, anisotropic diffusion and histogram equalization have been used to reduce additive noise, speckle noise and to enhance the image quality, respectively. The second step is feature extraction and dimensionality reduction. PCA has been used to reduce the dimension of the feature vector. The comparison was in the third step by using Support Vector Machines and Bagging ensemble classifier as two different classification techniques. It is applied to classify the images into image with/without tumour. In the future new cases will be added to the data set. More image processing and feature extraction techniques will be added to the system to improve its accuracy. We plan to train the designed classifiers using different features and use other classifiers as input to the ensemble classifier.

References

Abdi H., Williams L. J. (2010). Principal component analysis.

Alvarengal A., Pereira W., Infantosi A., Azevedo C. (2007). Classifying Breast Tumours on Ultrasound Images Using a Hybrid Classifier and Texture Features. pp. 1 – 6.

Breiman L. (1996). Bagging Predictors.

Campo-Avila J., Moreno-Vergara N., Trella-Lopez M. (2011).Analizying Factors to Increase the Influence of a Twitter User.

Chenga H.D., Shana J., Jua W., Guoa Y., Zhangb L. (2009). Automated Breast Cancer Detection and Classification Using Ultrasound Images Asurvey,

Frost V. S., Stiles, Abbott J., Shanmugan K.S., Holtzman J. (2009). A Model for Radar Images

Gadkari D. (2014).Image Quality Analysis Using GLCM.

Garcia M., Jemal A., Ward E., Center M., Hao Y., Siegel R., Thun M.(2007). Global Cancer Facts & Figureures.American Cancer Society.

Guo Y. (2010). Computer-Aided Detection of Breast Cancer Using Ultrasound Images.

Hall M., Frank E., Holmes G., Pfahringer B., Reutemann P., Witten I. (2009). The WEKA Data Mining Software: An Update.

Haralick R. M. ,Shanmugam K., Dinstein I. (1973). Textural Feartures for Image Classification.

Huang Y., Wang K., Chen D. (2006). Diagnosis of breast tumors with ultrasonic texture analysis using support vector machines." Neural Comput&Applic," April pp 164-169.

Hummel R. (1977). Image enhancement by histogram transformation.

Jemal A., Bray F., Center M. M., Ferlay J., Ward E., Forman D (2011).Global cancer statistics.

Jiang P., Peng J., Zhang G., Cheng E., Megalooikonomou V., Ling H. (2012). Learning Based Automatic Breast Tumor Detection and Segmentation in Ultrasoung Images.

Michailovich O.V., Tannenbaum A. (2006). Despeckling of medical ultrasound images.

Its Application to Adaptive Digital Filtering of Multiplicative Noise.

Minavathi, Murali S., Dinesh M. S. (2012). Classification of mass in breast ultrasound images using image processing techniques.

Wang, Chen Q., Shen D. (2008). Fast Histogram Equalization for medical Image Enhancement.

Yu Y., Acton S. T. (2002).Speckle Reducing Anisotropic Diffusion.