



HAL
open science

Segmentation and classification of benign and malignant breast tumors via texture characterization from ultrasound images

Mohamed Benaouali, Mohamed Bentoumi, Menad Touati, Abdelmalik Taleb-Ahmed, Malika Mimi

► To cite this version:

Mohamed Benaouali, Mohamed Bentoumi, Menad Touati, Abdelmalik Taleb-Ahmed, Malika Mimi. Segmentation and classification of benign and malignant breast tumors via texture characterization from ultrasound images. 2022 7th International Conference on Image and Signal Processing and their Applications (ISPA), May 2022, Mostaganem, Algeria. pp.1-4, 10.1109/ISPA54004.2022.9786350 . hal-03689545

HAL Id: hal-03689545

<https://hal.science/hal-03689545v1>

Submitted on 5 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Segmentation and classification of benign and malignant breast tumors via texture characterization from ultrasound images

Mohamed Benaouali
Signal and Systems Laboratory
University Abdelhamid Ibn Badis of
Mostaganem
Mostaganem, Algeria
mohamed.benaouali@univ-mosta.dz

Mohamed Bentoumi
Signal and Systems Laboratory
University Abdelhamid Ibn Badis of
Mostaganem
Mostaganem, Algeria
mohamed.bentoumi@univ-mosta.dz

Menad Touati
Signal and Systems Laboratory
University Abdelhamid Ibn Badis of
Mostaganem
Mostaganem, Algeria
menad_tt@gmail.com

Abdelmalik Taleb Ahmed
IEMN DOAE UMR CNRS 8520
Université Polytechnique des Hauts de
France
Valenciennes, France
abdelmalik.taleb-ahmed@uphf.fr

Malika Mimi
Signal and Systems Laboratory
University Abdelhamid Ibn Badis of
Mostaganem
Mostaganem, Algeria
malika.mimi@univ-mosta.dz

Abstract— The present paper deals with breast tumors classification from ultrasound images. The proposed procedure consists of four steps, namely preprocessing, segmentation, feature extraction and classification. To improve the quality of ultrasound images, the preprocessing step consists of anisotropic filtering and histogram equalization that are performed on the original images. The segmentation is performed on the preprocessed images using the Level Set method that allows to extract the region of interest (ROI) and to reduce its size at the same time. Two feature extraction methods are used in this work namely, the local binary pattern (LBP) method and the histogram of oriented gradients (HOG) method. The two methods (LBP and HOG) are techniques of textures analysis and allow to characterize the ROI. The extracted feature sets constitute the inputs for three classifiers namely, support vector machines (SVM), k-nearest neighbors (KNN) and decision trees (DT). In this work, the best results are obtained by the concatenation of the two feature vectors namely LBP and HOG associated to the SVM classifier. This allows to achieve an accuracy of 96%, a sensitivity of 97% and a specificity of 94%.

Keywords— *Ultrasound images, Breast tumor, Segmentation, HOG features, LBP features, Classification.*

I. INTRODUCTION

Medical imaging has become an indispensable tool for the diagnosis of several diseases including breast cancer in women. This type of cancer is one of the main causes of death in women worldwide. Its early detection is the most important factor in reducing mortality and the costs of its management. Ultrasound (US) imaging is the best alternative to mammographic imaging which is a painful examination for a woman. US imaging can also be considered as a complementary non-invasive examination for the breast cancer diagnosis [1]. However, the main weakness of ultrasound imaging is that the quality of the obtained image depends on the performance of the device employed by the radiologist. The US images interpretation requires specialized radiologists because of its difficulty and the appearance of a systemic noise named Speckle. Nevertheless, the implementation of a computer-aided diagnosis (CAD) system can help radiologists to overcome the US images drawbacks for the detection and classification of breast cancer purposes.

Although in the scientific literature there are many research studies addressing the problem of detection and classification of breast tumors using different imaging modalities such as mammography; ultrasound breast imaging has attracted less interest due to the US images quality and the lack of public expertise databases. In this work, we present a procedure to process ultrasound breast images for breast tumors classification purposes. This classification allows to distinguish between malignant and benign tumors.

The remainder of the paper is organized as follows. Section 2 presents a description of the different steps constituting the implemented procedure. The results obtained are presented and discussed in section 3. Finally, section 4 concludes the paper with some perspectives cited as well.

II. PROCEDURE PRESENTATION

Herein, we present our procedure for classification of benign and malignant breast tumors from ultrasound images. As reported in Fig. 1, the procedure can be decomposed into four steps represented by four blocks: A preprocessing block, a segmentation block, a feature extraction block, and a benign and malignant tumor classification block. The four steps are detailed below after a brief description of the used database.



Fig. 1. Illustrative block diagram of benign and malignant breast tumor classification procedure.

A. Data description

The considered database was collected from *Baheya hospital, Cairo, Egypt* in 2018 [2], for early detection and treatment of cancer in women. The image set is obtained from 600 patients. The database consists of 780 images with a standardized size of 500 x 500 pixels. The image set is divided into three classes, namely normal (133 images), benign (437 images) and malignant (210 images) (Fig. 2).

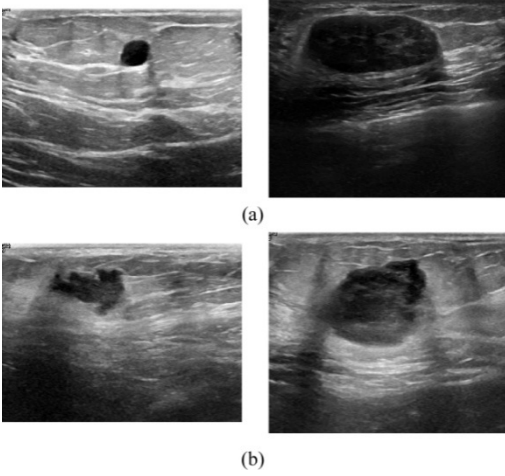


Fig. 2. Examples of US images of breast tumors obtained from Baheya hospital database [2]; (a): benign tumors, (b): malignant tumors.

B. Preprocessing step

Due to the presence of Speckle noise, the ability of an automatic system to fully extract texture will be certainly limited. For this purpose, all images are denoised by a Speckle reducing anisotropic diffusion filter (SRAD1) [3]. Then, the denoised images are enhanced so that improve their contrast. The result of this preprocessing is shown in Fig. 3. Compared to the original images, the enhanced and denoised images show a better contrast.

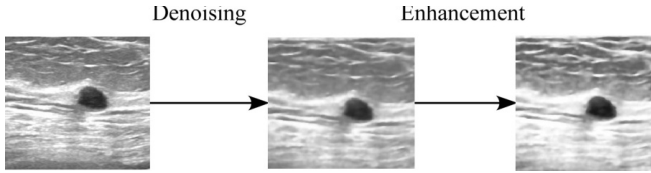


Fig. 3. An illustrative example of the performed image preprocessing: (a) original image, (b) denoised image and (c) denoised and enhanced image.

C. Segmentation step

The segmentation of breast ultrasound images is essential for the breast tumors diagnosis. The Level Set method used in this work is widely used for medical image segmentation. However, this remains a challenge for traditional segmentation methods as they cannot fully delineate tumor regions with complex texture. In this work, the Level Set method is used for segmentation. It is a contour tracking method that is widely used for medical image segmentation [11].

This step allows to extract the regions of interest (ROI) from the breast US images, which also allows to reduce the image size for the remaining processing.

D. Feature extraction

Feature extraction from the breast ultrasound image is a key step for classification of benign and malignant breast tumors. It is of some importance to use texture features for tumor image analysis because there are significant differences of texture information between benign and malignant tumors from ultrasound images. In this context, local binary patterns method (LBP) [4] and histogram of oriented gradients (HOG) method [5] are used to extract texture features from segmented ROI. Noting that, benign and malignant tumors often show differences in morphology. Generally, a benign tumor has a regular shape i.e., round or oval, and their contour is relatively

smooth. On the other hand, a malignant tumor has an ill-defined contour and its shape is irregular with lobules.

- **The LBP descriptor** was introduced in [4] for texture classification. It is computed on a 3×3 pixels region. An 8-bit ordered binary number is generated by comparing the central pixel value with the neighboring pixels values. The LBP operator is expressed as follows:

$$LBP(x, y) = \sum_{p=1}^8 2^{p-1} s(i_p - i_c) \quad (1)$$

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{else} \end{cases}$$

where i_c is the gray value of the central pixel (x, y) , p is the number of the adjacent pixel, i_p is the gray value of the adjacent pixel and $s(x)$ is the symbolic function.

- **The HOG descriptor** forms the feature by calculating the statistic histogram of the gradient direction in the image local area. In this work, we proceed as follows: First, the gradient of each pixel is calculated. Then, the image is divided into 32×32 pixels cells, and the gradient histogram of each cell is counted to form a descriptor. Finally, each 2×2 cell is concatenated to form a block, and then all blocks are concatenated to obtain the HOG feature vector [5].

The texture characteristics extracted by the two methods, namely, the LBP and HOG, constitute the feature vector for each ROI. Each obtained vector is used as input for classification.

E. Classification

Classification represents an important step in medical image diagnosis. A classifier is used to predict the class of the feature vector of an image [6]. In the following, only the basic concepts of classification based on SVM, KNN and decision tree (DT) are introduced.

- **The SVM classifier:** A support vector machines (SVM) are machine learning methods, introduced in the 1995s by V. Vapnik [7]. A linear SVM classifier attempts to find a separable linear hyperplane (an optimal hyperplane) by maximizing the margin of the classifier (the distance between two classes) while minimizing the sum of the classification errors.

For an SVM classifier, we consider a training set D of N examples (X_i, y_i) with $X_i \in \mathbb{R}^p$ belongs to a class labeled by $y_i \in \{+1, -1\}$. The separating hyperplane H can be defined by (2), where $(w \in \mathbb{R}^p)$ and $(b \in \mathbb{R})$ represent the parameters of the separating hyperplane.

$$H: \langle w, X_i \rangle + b \quad (2)$$

The w and b values are determined by learning by minimizing the criterion J under the following constraints:

$$\min_{w, b} J = \frac{1}{2} \|w\|^2 \quad (3)$$

under constraints:

$$y_i(\langle w, X_i \rangle + b) \geq 1 ; i = 1 \dots N \quad (4)$$

- **KNN classifier:** A k-nearest neighbor (KNN) classifier is a supervised learning algorithm that can be used to solve classification problems where the nature of the separating hyperplane is not taken into account. A KNN classifier is based on the majority vote of k nearest neighbors. The voting is based on the distance measure between the test

example X_i and the training examples. If the distance of the test example is closer to k other examples of the same class, then the test example is assigned the same class. The most commonly used distance measure is the Euclidean distance.

- **The decision tree (DT) [8]:** It is a supervised learning algorithm that can be used for image feature vectors classification. The decision tree builds classification or regression models in the form of a tree structure. In the learning phase, the decision tree is designed by decomposing the overall decision into a set of serially placed single decisions where each node (single decision) represents a "test" on a feature of the input vector. At the end of each decision branch, we find the predicted class of the example (the final decision).

III. EXPERIMENTATION AND RESULTS

A. Methodology:

The aim of this work was to study the procedure implementation feasibility for the segmentation and classification of breast tumors from the ultrasound images. Recall that is a four-steps based procedure (Fig. 1) organized as follows:

- **Preprocessing:**

This step is performed by two operations: an anisotropic filtering to eliminate the so-called Speckle noise followed by an enhancement by histogram equalization of the filtered image.

- **Segmentation:**

In this step, the Level Set method was used to segment the image. Then, the region of interest (ROI) is extracted according to the information from the expertise included in the considered image database.

- **Feature extraction:**

In this step, feature extraction consists of transforming the extracted ROI into a vector representing the segmented image. Two feature vectors are constructed, namely HOG feature vector and LBP feature vector. By concatenation of the two feature vectors obtained previously, we provide a third feature vector. Note that the segmented ROI have been resized to the same size in order to have feature vectors of the same dimension.

- **Classification**

The extracted features are used as inputs for the three classifiers, namely, SVM, KNN and decision tree (DT). In this study, to objectively highlight the performance of our procedure, we assess classification performance using three evaluation metrics, namely accuracy, sensitivity and specificity [9][10].

- **The accuracy** expresses the global rate of correct classification which is formally defined as follows:

$$accuracy = \frac{\text{number of images correctly classified}}{\text{number of all images}} \quad (5)$$

In this work, the accuracy is estimated by the k -fold cross-validation method. The whole image set is divided into k subsets. Then, we perform training on $k-1$ sets while the test (accuracy estimation) is performed on the remaining set. This operation is repeated k times and each time we change the test set. The global accuracy is the average of the k obtained rates.

We apply 10-fold cross validation to quantitatively evaluate our proposed procedure.

In the same way, we assess the classifiers with the two other metrics, namely sensitivity and specificity. These metrics are complementary and summarize the performance of a classifier by taking into account all the components of the confusion matrix. The latter is composed of 4 values (Table I) in the case of a two-class classification issue (positive class vs. negative class) [9][10].

TABLE I. CONFUSION MATRIX: THE OUTCOMES OF CLASSIFICATION INTO POSITIVE AND NEGATIVE CLASSES.

Actual class \ Predicted class	Positive	Negative
	Positive	TP: true positive
Negative	FP: false positive	TN: true negative

- **Sensitivity** is a metric that expresses the rate of positive examples correctly predicted by the classifier:

$$Sensitivity = \frac{TP}{TP+FN} \quad (6)$$

- **Specificity** is a metric that expresses the rate of negative examples correctly predicted by the classifier:

$$Specificity = \frac{TN}{TN+FP} \quad (7)$$

B. Results and Discussion

Accurate classification of benign and malignant tumors from breast ultrasound images is a difficult task in the presence of various artifacts such as variety of shapes, speckle noise, ill-defined contours and poor contrast. Table 2 shows the results obtained by considering accuracy, sensitivity and specificity metrics. The reported results are determined for different feature sets including the three investigated classifiers: SVM, KNN and DT.

From Table II, one can see that the best classification performances are obtained by using the concatenated texture feature vectors (LBP + HOG) in association with the SVM classifier. This is expressed by an accuracy reaching 96%, a sensitivity of 97% and a specificity of 94%. The second-best classification result was obtained from the HOG feature set in association with the SVM classifier achieving an accuracy of 94%, a sensitivity of 97% and a specificity of about 92%. For the LBP characterization associated with the SVM classifier, the accuracy reaches 93 %, the sensitivity is about 94% and the specificity is found to be equal to 93%. However, the use of the same feature set in association with the two other classifiers (KNN, DT) provides weak classification performances.

IV. CONCLUSION

In this work, we implemented a procedure for classification of benign and malignant tumors from ultrasound images. The texture feature concatenation is proposed to improve classification performance. First, the LBP and the HOG feature vectors are extracted from ultrasound images after proper preprocessing and segmentation so that to determine the region of interest from ultrasound images. Then, three different classifiers (SVM, KNN, DT) are used to predict the classes of the preprocessed images. The classification assessment allows us to conclude that the fusion of LBP and HOG features in association with the SVM classifier provides

the best performances in terms of discrimination between benign and malignant tumors with an accuracy reaching 96%, a sensitivity reaching 97% and a specificity of about 94%. This approves the feasibility and the efficiency of our proposed procedure for breast tumors classification based on ultrasound images.

In perspective, other feature extraction methods will be tested on a larger datasets of ultrasound images of breast tumors in order to improve the performance of the proposed procedure.

TABLE II. EXPERIMENT RESULTS

Feature set	Metric	Classifier		
		SVM	KNN	DT
HOG	Accuracy (%)	94	65	65
	Sensitivity (%)	97	100	100
	Specificity (%)	92	32	32
LBP	Accuracy (%)	93	79	79
	Sensitivity (%)	94	68	68
	Specificity (%)	93	90	90
Hog + LBP	Accuracy (%)	96	70	70
	Sensitivity (%)	97	100	100
	Specificity (%)	94	42	42

REFERENCES

- [1] A. T. Stavros, D. Thickman, C. L. Rapp, M. A. Dennis, S. H. Parker, and G. A. Sisney, Solid breast nodules: use of sonography to distinguish between benign and malignant lesions. *Radiology*, vol. 196, no. 1, pp.123–134, 1995.
- [2] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, Dataset of breastultrasound images,” Data Brief, vol. 28, Feb. 2020, Art. no. 104863. Available at: <http://www.sciencedirect.com/science/article/pii/S2352340919312181>
- [3] P. Perona and J. Malik. Scale space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Machine Intell.*, vol.12, pp. 629–639, 1990.
- [4] T. Ojala, M. Pietikäinen, and D. Harwood. A comparative study of texture measures with classification based on featured distribution. *Pattern Recognition*, vol. 29, no. 1, pp. 51–59,1996.
- [5] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol.1, pp. 886–893, San Diego, CA, USA, 2005.
- [6] P. Thamilselvan & Sathiaselvan, J.. A Comparative Study of Data Mining Algorithms for Image Classification. *International Journal of Education and Management Engineering*. 5. 1-9. 10.5815/ijeme. 2015.
- [7] V. Vapnik, The nature of statistical learning theory. *Springer science & business media*, 2013.
- [8] S. R. Safavian and D. Landgrebe. A survey of decision tree classifier methodology. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 21, no. 3, pp. 660–674, 1991.
- [9] M. Bentoumi, M. Daoud, M. Benaouali, A. Taleb Ahmed. Improvement of emotion recognition from facial images using deep learning and early stopping cross validation. *Multimedia Tools and Applications*, at: <https://dx.doi.org/10.1007/s11042-022-12058-0>, 2022.
- [10] Sokolova M., Japkowicz N., Szpakowicz S. Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation. In: *Sattar A., Kang B. (eds) AI: Advances in Artificial Intelligence. AI. Lecture Notes in Computer Science*, vol 4304. Springer, Berlin, Heidelberg., at: https://doi.org/10.1007/11941439_114, 2006.
- [11] Jiang, Xin & Zhang, Renjie & Nie, Shengdong. Image Segmentation Based on Level Set Method. *Physics Procedia*. 33. 840-845. 10.1016/j.phpro. 2012.