Lesion Detection in Breast Ultrasound Images Using a Machine Learning Approach and Genetic Optimization

Fabian Torres[1,2(✉)], Boris Escalante-Ramirez[1,2], Jimena Olveres[2], and Ping-Lang Yen[3]

[1] Centro Virtual de Computación, Universidad Nacional Autónoma de México, 04510 México CDMX, Mexico
fabian_torres@cvicom.unam.mx.com
[2] Departamento de Procesamiento de Señales, Facultad de Ingeniería, Universidad Nacional Autónoma de México, 04510 México CDMX, Mexico
[3] Department of Bio-Industrial Mechatronics Engineering, National Taiwan University, Taipei, Taiwan

Abstract. Breast ultrasound has become one of the most important and effective modalities for early detection of breast cancer and it is most suitable for large scale breast cancer screening and diagnosis in low-resource countries. Breast lesion detection is a crucial step in the development of Computer Aided Diagnosis and Surgery systems based on ultrasound images, since it can be used as a seed point to subsequently initialize segmentation methods such as region growing, snakes or level-sets. Because of inherent artifacts of the ultrasound images, such as speckle, acoustic shadows and blurry edges, the detection of lesions is not an easy task. In this work we propose a machine learning based approach to locate lesions in breast ultrasound images. This approach consists on the classification of image pixels as lesion or background with a Random Forest optimized with genetic algorithms to generate candidate regions. After pixel classification the method chooses the correct lesion region by discriminating false positives using a new proposed probability approach. The pixel classification and region discrimination steps are compared with other methods, showing better results in the detection of lesions. The lesion detection was evaluated using the True Positive Fraction and the False Positives per image, having results of 84.4% and 15.6% respectively.

Keywords: Breast lesion · Ultrasound · Random Forest · Genetic algorithms

1 Introduction

Although mammography is the most used imaging method for breast tumor analysis, ultrasound has been used as one of the gold standard techniques for breast cancer imaging, since mammography may miss over 1/3 lesions in dense breast. Currently ultrasound is responsible for about 1/5 of all diagnostic images and has become an important and effective modality for early detection of breast cancer and it is most suitable for large scale breast cancer screening and diagnosis in low-income countries,

© Springer Nature Switzerland AG 2019
A. Morales et al. (Eds.): IbPRIA 2019, LNCS 11867, pp. 289–301, 2019.
https://doi.org/10.1007/978-3-030-31332-6_26

but the visualization of lesions in breast ultrasound (BUS) images is a difficult task due to some intrinsic characteristics of the images, like speckle, acoustic shadows and blurry edges [1].

Computer Aided Diagnosis and Surgery (CAD/CAS) systems based on BUS images have been developed to help the physician to have better visualization of the lesion by overcoming the considerable inter and intra-variability, since it directly affects the performance of the quantitative analysis and diagnosis of the lesions. Lesion detection is an initial state of CAD/CAS systems as a seed point to subsequently initialize segmentation methods such as region growing, snakes or level-sets. Because of the mentioned inherent artifacts in BUS images, the detection of lesions is not an easy task. Several semi-automatic and automatic methods have been proposed. Three main categories of methods for breast lesion detection can be identified: (1) Classical approaches; (2) Graph based approaches; and (3) Machine-Learning approaches [2]. Due to the challenging nature of the task, just using a single image processing technique cannot achieve desirable results. Most successful approaches employ hybrid techniques and model biological priors using pixel intensity, texture, and spatial information [3].

Machine learning methods thrive in BUS lesion detection in the last decade because they provide a good approach to integrate different levels of features. In this work we use a machine learning based approach to locate a lesion in BUS images. This approach consists on the classification of image pixels as tumor or background to generate candidate regions, and then choose the correct region where the tumor is located. To achieve this, we compare several machine learning methods, concluding that the Random Forest algorithm outperforms other machine learning algorithms in the pixel classification task. After optimization using genetic algorithms, an accuracy of 82.92% was achieved using a set of 19 texture descriptors (2 histogram, 3 co-occurrence, 4 run-length and 10 Hermite coefficients), 850 trees with $m = 10$ and a maximum of samples per leaf of 1 pixel. After pixel classification, false positive regions must be removed. Several methods have been proposed to find the correct region where the lesion is located, here we propose a new discrimination method based on a probability image build with the computed probabilities of each pixels by the RF and compare it with different approaches, by evaluating the True Positive Fraction and the False Positives per image, having results of 84.4% and 15.6% respectively.

1.1 Modeling Lesions in Breast Ultrasound Images

Successful approaches for lesion detection in BUS should model domain related priors appropriately. The main features used for modeling breast tumors in ultrasound images are: intensity; internal echo pattern (Texture); and spatial distribution.

Ultrasound gray-level intensity provide helpful information about the density of the different tissues found in the image and helps to differentiate and identify different structures. The main disadvantage of medical ultrasound images is the poor quality due to speckle noise. Speckle reduction in ultrasound images is usually done by techniques that are applied directly to the original image domain. Several methods have been proposed to address the problem of speckle noise. The speckle reduction anisotropic (SRAD) filter is an edge-sensitive diffusion for speckled images. This filter has a large potential in assisting segmentation techniques and has been used in BUS images to obtain more homogeneous regions while preserving edges [4]. On the other hand, contrast enhancement in ultrasound images has the purpose to adjust the display contrast and increase the edge sharpness of objects. Histogram equalization and stick filtering have been widely used in ultrasound breast images to improve the contrast between the lesion and the background [5].

Internal echo pattern can be described using texture. Texture information provides a way to differentiate the lesion from other objects that have similar gray intensities, like acoustic shadows. The main texture descriptors used for lesion detection in BUS images are extracted from histogram and co-occurrence matrices. First-order texture descriptors are extracted from the pixel image values. These descriptors do not consider the spatial relationship with neighborhood pixels. The most frequently used first-order texture descriptors in BUS images are central moments of the histogram. The gray-level co-occurrence matrix describes how frequently two gray-levels appear in a window separated by a given distance and a given angle. Second-order texture descriptors computed from the analysis of the co-occurrence matrices have been proposed by Haralick [6]. Some of these texture descriptors have been used for the segmentation and classification of breast tumors [7]. Although these descriptors consider the spatial relationship between pixels, the computational cost of computing the co-occurrence matrix is very high compared to first order descriptors. Another method to characterize texture that also takes into account the spatial relationship between pixels is based on run-lengths of image gray-levels, where the run-length matrix of an image is defined as the number of runs with pixels of equal gray level and a given run-length [8]. Although these descriptors have not been widely used as an effective texture classification and analysis method, it has been demonstrated by Tang et al. that there is rich texture information contained in this matrices [9]. On the other hand, methods that resemble the human visual system have increased in popularity because they allow images to expand into a local decomposition that describes intrinsic attributes and highlights structures that are useful for segmentation. The main advantage of the Hermite transform is the easy extraction of important details as lines, edges and texture information by applying a decomposition scheme. Hermite-based texture descriptors have been used in the segmentation of ultrasound images successfully [10].

The spatial distribution of BUS is widely used in lesion detection approaches to discriminate lesions from other tissue such as fat and glands. Breast lesions are usually located in the center of the images, while subcutaneous fat, glandular tissue and skin typically appear in the upper portion of the image. Modern ultrasound systems can acquire high-resolution images which may include other structures such as ribs, pectoral muscle or the air in the lungs, making the lesion detection more difficult. Nowadays it is no longer necessary to place the suspected lesion at the center of the image for better visualization, hence, methodologies that assume that the lesion is centered in the image fail in more cases when using modern ultrasound systems [2].

1.2 Lesion Detection in Breast Ultrasound Images

The improvement of the performance of lesion detection is an increasingly challenge that has reach a bottleneck, and only a few new approaches were published in the last several years. BUS segmentation is a crucial step in CAD/CAS systems, since it directly affects the performance of the quantitative analysis and diagnosis of tumors. Different kinds of automatic and semiautomatic methods have been developed. Three main categories of breast tumor segmentation can be identified [11]:

Classical Approaches. Most of the classical approaches are quite simple, fast and efficient to conduct initial segmentation of BUS images using simple low-level features, but these methods are vulnerable to low image quality due to noise, inhomogeneity and low contrast. Thresholding is the most intuitive, simple and fast of these methods and it has been successfully used for BUS lesion detection [12]. Region growing methods grow regions defined by a set of pixels (seed) to bigger regions using a growth criterion. The seed can be chosen manually or automatically, and the main challenge of this techniques is to find a growth criterion that adjust correctly to noisy images. Madabhushi et al. proposed a region growing based method for breast tumor segmentation with automatic seed selection [5]. Watershed is a powerful segmentation method with better results than thresholding and region growing that could integrate domain-related priors. The main problem of watershed is finding the markers, because using the local minimum gradient as a marker usually results in over segmentation and region merging should be involved [2].

Graph Based Methods. These methods are among the earliest techniques for breast lesion detection and provide a simple way to organize task-related priors and image information in a unified framework; they are flexible, computationally efficient and suitable for expressing soft constrains between random variables, like pixels. Markov Random Fields with maximum a posteriori optimized with Iterated Conditional Model is a flexible framework for image partition. The performance of methods based on this framework usually is good but has a shortcoming because they only obtain local optimum solutions. The approaches based on graph cuts focus on designing well-defined boundary problems by using more comprehensive data and smoothness terms to deal with contrast and inhomogeneity. The main disadvantage of these approach is that they tend to generate a much shorter boundary than the real one (shrink problem). Normalized cut methods avoid the shrink problem, but it cannot integrate semantic information and user interaction is needed to achieve good performance. Although graph-based methods account for the second largest portion of BUS segmentation, these techniques fade away due to the successful application of other powerful approaches [2].

Machine Learning-Based Approaches. Image segmentation can also be viewed as a two-class classification problem (classifying pixels into lesion and background). Supervised and unsupervised learning methods have been employed in lesion detection in BUS images. The unsupervised methods aim to partition the image in disjoint regions as a preprocessing step [2]. Supervised learning methods integrate different levels of features and can learn the relation between the inputs and the target outputs. The most common supervised learning approaches used in BUS are: Naive Bayes Classifier; Support Vector Machines; and Artificial Neural Networks [3, 13–16]. These methods are used for pixel classification using a set of texture features, where the main difference between them is the set of texture descriptors used. Other machine learning methods that may be suitable for BUS images, like Random Forest, have been used for the segmentation and classification of breast tumors in mammograms. Classification methods usually cannot produce accurate tumor boundary and refinement is usually necessary [2]. After pixel classification a binary image containing the lesion region and false positive regions is obtained, several methods have been used for the discrimination of false positive regions and the detection of the lesion region. Several methods have been proposed to select the correct region including spatial information, considering that the lesion usually is found near the center of the image in the parenchyma of the breast [5, 17]; however, this assumption does not apply for images acquired with modern ultrasound systems since they may include pectoral muscles and ribs information [2]. Because of this, machine learning methods such as SVM have been used to discriminate false positive and find the correct lesion region, but extracting new features for region classification is not an easy task [18].

1.3 Optimization of Machine Learning Methods with Genetic Algorithms

Dimensionality reduction is a step commonly used in machine learning, especially when dealing with a high dimensionality space of features. A high number of features may slow down the methods while giving similar results as obtained with a smaller subset; also, not all the features used to describe the problem are necessarily relevant and beneficial for the learning task. Dimensionality reduction is usually performed by constructing a new dimension space through feature transformation or by selecting a subset of the original dimensions, like principal component analysis and feature selection respectively. Different feature selection methods have been developed and applied in machine learning following different search strategies like: Forward selection (start with an empty set and greedily add features one at a time); Backward elimination (start with a feature set containing all features and greedily add or remove features); Random mutation (start with a feature set containing randomly selected features and add, or remove randomly selected features). Also, the feature selection methods can be divided into: Filters that use an evaluation function independent of the learning algorithm; Wrappers that use the same machine learning algorithm that will be used for modeling; and Embedded approaches that perform feature selection during the model generation [19]. Genetic algorithms (GA) have been used as a Wrap random mutation approach for feature selection. GAs make it possible to explore greater range of possible solutions to a problem under controlled and well understood conditions. It has been proved theoretically and empirically that these algorithms provide a global near-optimal solution for various complex optimization problems [20].

Besides feature selection, setting the parameters of a classifier has an important influence on its classification accuracy. A common used parameter search approach is the grid search, but its search ability is low. The optimal classification accuracy of a classifier can be obtained by feature selection and optimal parameters setting. The trend in recent years is to simultaneously optimize feature selection and parameter optimization. GAs have the potential to generate both feature selection and parameter optimization at the same time [21]. Although GAs have been previously used to find the optimal features for classifying and segmenting breast tumors in ultrasound images they have not been used for parameter optimization.

2 Materials and Methods

In this work we propose an automatic lesion detection method in BUS images. The method consists on a three-step approach: (1) Preprocessing; (2) Pixel classification; and (3) Identification of the lesion region. Feature selection and parameter optimization of the machine learning method for pixel classification was performed using a simple GA to improve the accuracy of the classification.

2.1 Dataset

A public data base of 58 BUS images with a lesion, provided by the Department of Radiology of Thammasat University and Queen Sirikit center of Breast Cancer of Thailand, was used for the evaluation of the proposed method; the database includes the ground truth hand-drawn by leading radiologists of these centers [22].

2.2 Random Forest

Random Forests are a combination of tree predictors, usually CARTs, such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest [23]. RFs trains an ensemble of individual decision trees based on samples, their class designation and variables; every tree is built using a random subset of samples and variables [24].

Suppose a forest of decision trees $(RF = \{T_1(X_1), T_2(X_2), \ldots, T_B(X_B)\})$ is constructed with a training data set X with N instances, where $X = \{x_1, x_2, \ldots x_M,\}$ is a M-dimensional vector of features associated with an instance in the data set and B is the number of trees in the forest. From the training data, a randomly sampled set, X_i, with replacement (bootstrap), is extracted to grow each tree in the forest. This bootstrap set of size n, where $n < N$, usually contains about two-third of the samples in the original training set; also, a random set of features of size m, where $m < M$, extracted from X is used for each bootstrap; where the size m of the random set is usually $\sqrt{M}$ or $\sqrt{M}/2$ [25]. To classify an instance every tree in the forest records a vote for the class to which the instance belong and it is labeled as a member of the class with the most of votes. One characteristic of the random forest classifier is that not only the class of the object can be computed but also the probability of the object belonging to that class could be obtained. The standard approach to probability estimation in many areas of science relies on logistic regression. However, it is almost impossible to guarantee that a logistic model is well-specified when modern data sets with nonlinear or high dimensional structure are used. Random forest has become a widely used tool for probability estimation. After fitting a RF to training data, it is a common practice to infer conditional class probabilities for test point by simply counting the fraction of trees in the forest that vote for a certain class [26].

2.3 Genetic Algorithms

GAs are based on natural selection and sexual reproduction. Natural selection determines which members of a population survive to reproduce, and sexual reproduction ensures the mixing and recombination of the genes of their offspring. A string of bits corresponding to the presence or absence of specific features and parameter values (in binary representation) are used in GAs to describe different members of the population (individuals) [27]. Each individual in a generation is tested, looking for the optimization of an objective function as a measure of fitness. The individual fitness $F(x_i)$ is computed as the individual performance $f(x_i)$ relative to that of the whole population.

The reproduction operator used in a simple GA is a single point crossover; where individuals with a high fitness value are paired at random, exchanging a random segment between individuals to create two offspring. The mutation operation consists of flipping each bit of the individuals with lowest fitness value [28]. This is repeated through several generations until a predefined condition is satisfied.

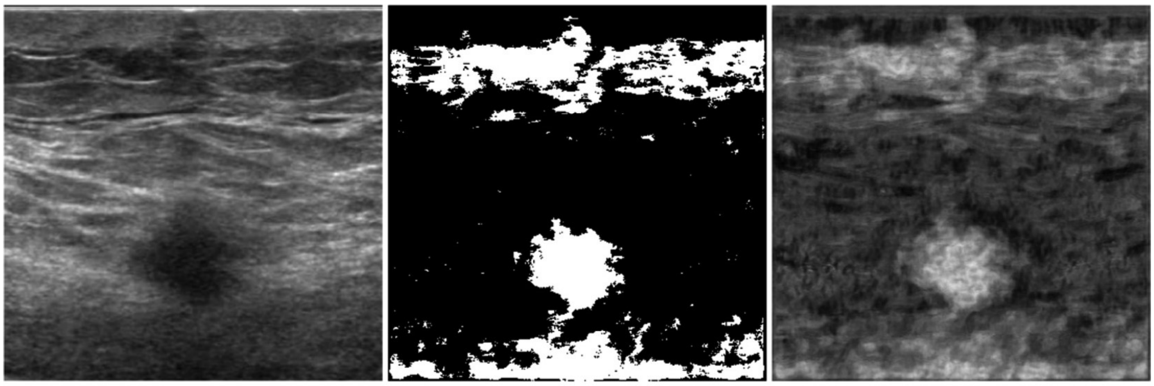
2.4 Proposed Method

The proposed method consists in three steps:

Preprocessing. The preprocessing step consists on extracting descriptive features from BUS images that could help in the classification of pixels into lesion and background classes. An enhanced intensity image is obtained using the SRAD filter and histogram equalization. To obtain texture images that could describe the internal echo pattern of the lesion and the background a total of 29 texture descriptors extracted from histogram, co-occurrence matrices, run-length matrices and Hermite transformation were computed; using the original intensity image without any pre-processing to avoid elimination of any texture related information.

Pixel Classification. A supervised machine learning method is used to classify pixels into lesion or background classes. Gray-level intensity values, extracted from the original and preprocessed images, are used as features for the classification. During the training of the classifier a GA was used to find the optimal subset of features and parameter optimization. After pixel classification a binary image is obtained, but this binary image usually contains false positive regions and a further discrimination must be made. Along with the binary classification image a lesion probability image is generated with the probability estimated with the RF classifier. The classification and probability images are shown in Fig. 1.

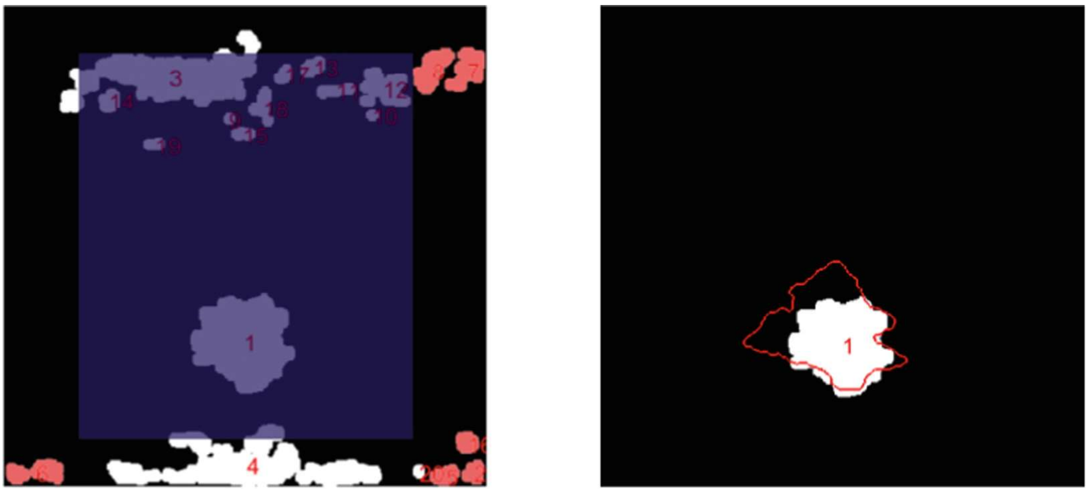
Lesion Region Detection. After pixel classification the discrimination of false positive regions must be made to find the localization of the lesion region. In this work we propose a new discrimination method based on a probability image build with the computed probabilities of each pixels by the RF. First basic mathematical morphology (dilation and erosion) is applied to the classification image in order to eliminate small regions and disconnect weak connected regions. After applying mathematical



a) b) c)

Fig. 1. Results of random forest classifier: (a) Original; (b) Classification and (c) Probability images

morphology a deletion of the candidate regions connected with the boundary of the image is done as in [17], excluding the regions that are connected with a window about half the size of the whole image and centered at the image center. After the connected-boundary regions are deleted the probability of each region is computed as the mean of all the pixels inside the region using the gray-level values of the probability image obtained with the RF classifier. The region with the highest probability is choose as the detected lesion region. This step is illustrated in Fig. 2.



a) b)

Fig. 2. Proposed method for lesion detection after classification: (a) Mathematical morphology of image 1a, deleted boundary-connected regions to be deleted are marked in red; and (b) maximum probability region chosen as lesion region. (Color figure online)

3 Experiments and Results

The proposed method was evaluated and compared against other state of the art methods for pixel classification and lesion detection. The results of these steps are explained in this section.

3.1 Pixel Classification Using Random Forest

Several machine learning methods were tested to find the classifier that has better results in the classification of pixels into lesion and background classes. A set of 31 features (original, enhanced intensity and 29 texture descriptors) were used for pixel classification. A set of pixels were extracted from the original and preprocessed images and labeled as lesion or background. A k-fold cross-validation (with k = 4) was used to find the accuracy (Eq. 1), sensitivity (Eq. 2) and specificity (Eq. 3) of the classification.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{2}$$

$$Specificity = \frac{TN}{TN + FP} \tag{3}$$

where TP, TN, FP and FN are the true positives, true negatives, false positives and false negatives pixels found in the classification process. The results in terms of accuracy of the classification with different classifiers is shown in Table 1.

The RF with default parameters (500 trees and $m = \sqrt{M}/2$) has a better accuracy than the other tested classifiers. After finding the best classifier a GA was used to find the optimal set of features and parameters that improves the outcome of the classification. A run of a simple GA was computed during 400 generation with a population of 15 individuals and a generational gap (number of individuals that survive to the next generation) of 5%. The accuracy error ($1 - Accuracy$) of the RF method was used as the individual performance and it was computed using the same cross-validation method as mentioned before but using different characteristics and parameter values for the classification according to the GA individuals. After the GA run, an accuracy of 83.92% was achieved with a set of 19 features (2 histogram, 3 co-occurrence, 4 run-length and 10 Hermite coefficients), 850 trees with $m = 10$ and a maximum of samples per leaf of 1 pixel. The sensitivity and specificity are often used to complement the evaluation of segmentation algorithms; sensitivity is used to measure how many pixels in the region of interest are correctly segmented, it does not tell anything about how many pixels in the background are going to be segmented as tumor (FP); the specificity measures how many pixels in the background are correctly excluded and does not tell if a tumor pixel is going to be correctly segmented as tumor (FN). The sensitivity and specificity of the optimized RF method were 82.23% and 82.61% respectively. It can be seen in Table 1 that the tree-based methods (CART, ABoost, LBoost and RF) have better balance in terms of sensitivity and specificity.

Table 1. Pixel classification accuracy.

Classifier	Accuracy	Sensitivity	Specificity
Logistic regression	73.30 ± 0.28%	69.61%	76.97%
SVM (Gauss Kernel)	55.28 ± 0.60%	99.88%	10.86%
Naïve Bayes	68.08 ± 0.32%	62.20%	73.94%
KNN	77.22 ± 0.34%	79.26%	75.18%
CART	73.54 ± 0.40%	73.84%	73.24%
Aboost	74.86 ± 0.52%	74.81%	74.90%
LBoost	74.59 ± 0.68%	74.41%	74.59%
RF	81.14 ± 0.43%	81.00%	81.28%
RF+GA	82.92 ± 0.52%	82.23%	82.61%

3.2 Evaluation of Lesion Detection

In lesion detection current practice, a radiologist annotates a rectangular region of interest (ROI) where the lesion is located. Most of the BUS lesion detection methodologies in the literature evaluate their algorithms using the seed point as detection criterion [2]. After lesion detection with the proposed algorithm a bounding box that comprises the detected lesion region is generated. The lesion detection is considered a true positive if the center of the bounding box is placed within the bounding box of an expert radiologist and considered a false positive when the center is outside the bounding box. The True Positive Fraction (TPF, Eq. 4), and the False Positives per image are used as quantitative measurements of the sensitivity of the lesion detection technique (FPs, Eq. 5).

$$TPF = \frac{TP}{Total\,number\,of\,images}, \tag{4}$$

$$FPs = \frac{FP}{Total\,number\,of\,images}, \tag{5}$$

The proposed method results are shown in Table 2. Different methods to discriminate false positive regions were tested for comparison of the method, the results are also shown in Table 2. The result of the lesion detection in three ultrasound images using the proposed approach are shown in Fig. 3.

It can be seen in Table 2 that the proposed lesion detection method outperforms the methods used for comparation. It is important to notice that the Madabhushi [5] and Shan [17] methods relies in the assumption that the lesion is located near the center of the image and this assumption is not always true, especially when using modern ultrasound systems for acquisition. On the other hand, Yang [16] and Jiang [18] methods use machine learning to classify the regions, CART and SVM respectively. Extracting new characteristics from BUS images for region classification is not an easy task, Jiang use the results of a k-means pixel clustering algorithm as features for the classification, but this method shows poor results in the lesion detection compared with

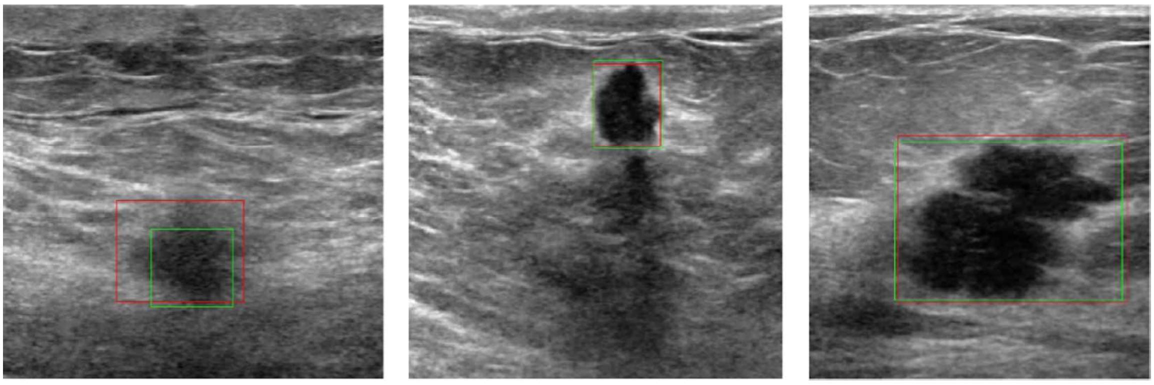


Fig. 3. Results of lesion detection in BUS images. The bounding box annotated by the experts is marked in red and the chosen by the proposed algorithm is marked in green. (Color figure online)

Table 2. Lesion detection evaluation

Classifier	TPF	FPs
Proposed method	84.48%	15.52%
Madhabushi	74.14%	25.86%
Shan	65.52%	34.48%
Jiang	68.97%	31.03%
Yang	75.86%	24.14

the proposed method. Yang proposed morphology characteristics such as size, compactness, region ratio and width height ratio as characteristics for classification; this method is not a good approach, as seen in Table 2, since breast lesions does not have a defined morphology.

4 Conclusion

In this work we present a new method for lesion detection in BUS images. The proposed method consists of three steps. In the first step preprocessing is used to extract an enhanced intensity image and texture images to be used as features for pixel classification. The second step consists of pixel classification using a random forest classifier and the extracted features from the preprocessing step. The random forest classifier was compared to other machine learning classification methods, showing better results in the classification of pixels into tumor or background classes. Also, the pixel classification method is improved using a simple GA to find an optimal subset of features and parameters. After pixel classification a false positive region discrimination must be done. In this work we proposed a new method based on a probability image generated using the probability of each pixel to belong to a lesion using the RF classifier in the second step. The proposed method was compared with four methods found in the literature, showing better results in finding the lesion region location. While lesion detection is an important step in the development of CAD/CAS systems, the segmentation of tumor boundaries could be more helpful to assist physicians in the diagnosis and treatment of breast cancer. Emerging methods such as Deep Learning could be used for feature extraction (preprocessing step), pixel classification, lesion detection and segmentation with high accuracy. Although the increasing computational power of hardware and parallel computing techniques, the development of BUS lesion detection methods using modern deep neural networks represent a challenge in terms of computational time and size of the training data sets, since modern neural networks need thousand of images for training and it is a difficult task to collect this amount of data, especially in low-income countries, and no public databases with the required amount of data are available [12, 29, 30].

Acknowledge. This work has been sponsored by UNAM grants PAPIIT IA103119 and UNAM PAPIIT IN116917. The DGAPA financial support of the postdoctoral fellowship program in the Facultad de Ingeniería is gratefully acknowledge.

References

1. Stöblen, F., Landt, S., Stelkens-Gebhardt, R., Sehouli, J., Rezai, M., Kümmel, S.: First evaluation of the diagnostic accuracy of an automated 3d ultrasound system in a breast screening setting. Int. J. Cancer Res. Treat. (2011)
2. Xian, M., Zhang, Y., Cheng, H.D., Xu, F., Zhang, B., Ding, J.: Automatic breast ultrasound image segmentation: a survey. Pattern Recogn. **79**, 340–355 (2018)
3. Liu, B., Cheng, H.D., Huang, J., Tian, J., Tang, X., Liu, J.: Fully automatic and segmentation-robust classification of breast tumors based on local texture analysis of ultrasound images. Pattern Recogn. **43**(1), 280–298 (2010)
4. Sivakumar, R., Gayathri, M.K., Nedumaran, D.: Speckle filtering of ultrasound b-scan images - a comparative study of single scale spatial adaptive filters, multiscale filter and diffusion filters. Int. J. Eng. Technol. **2**(6), 514 (2010)
5. Madabhushi, A., Metaxas, D.N.: Combining low-, high-level and empirical domain knowledge for automated segmentation of ultrasonic breast lesions. IEEE Trans. Med. Imaging **22**(2), 155–169 (2003)
6. Haralick, R.M.: Statistical and structural approaches to texture. Proc. IEEE **67**(5), 786–804 (1979)
7. Liao, Y.Y., Wu, J.C., Li, C.H., Yeh, C.K.: Texture feature analysis for breast ultrasound image enhancement. Ultrason. Imaging **33**, 264–278 (2011)
8. Selvarajah, S., Kodituwakku, S.R.: Analysis and comparison of texture features for content based image retrieval. Int. J. Latest Trends Comput. **2**(1), 108–113 (2011)
9. Tang, X.: Texture information in run-length matrices. IEEE Trans. Image Process. **7**(11), 1602–1609 (1998)
10. Estudillo-Romero, A., Escalante-Ramirez, B., Savage-Carmona, J.: Texture analysis based on the Hermite transform for image classification and segmentation, vol. 8436, p. 843619 (2012)
11. Huang, Q., Luo, Y., Zhang, Q.: Breast ultrasound image segmentation: a survey. Int. J. Comput. Assist. Radiol. Surg. **12**(3), 493–507 (2017)
12. Yap, M.H., Edirisinghe, E.A., Bez, H.E.: A novel algorithm for initial lesion detection in ultrasound breast images. J. Appl. Clin. Med. Phys. **9**(4), 2741 (2008)

13. Huang, S.-F., Chen, Y.-C., Woo, K.M.: Neural network analysis applied to tumor segmentation on 3D breast ultrasound images. In: 2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Proceedings, ISBI, pp. 1303–1306 (2008)
14. Chen, D.-R., Chang, R.-F., Kuo, W.-J., Chen, M.-C., Huang, Y.-L.: Diagnosis of breast tumors with sonographic texture analysis using wavelet transform and neural networks. Ultrasound Med. Biol. **28**(10), 1301–1310 (2002)
15. Yankaskas, B.C.: Epidemiology of breast cancer in young women. Breast Dis. **23**, 3–8 (2006)
16. Yang, M.-C., Huang, C.-S., Chen, J.-H., Chang, R.-F.: Whole breast lesion detection using naive bayes classifier for portable ultrasound. Ultrasound Med. Biol. **38**(11), 1870–1880 (2012)
17. Shan, J., Cheng, H.D., Wang, Y.: A novel automatic seed point selection algorithm for breast ultrasound images. In: 2008 19th International Conference on Pattern Recognition, pp. 1–4 (2008)
18. Jiang, P., Peng, J., Zhang, G., Cheng, E., Megalooikonomou, V., Ling, H.: Learning-based automatic breast tumor detection and segmentation in ultrasound images. In: 2012 9th IEEE International Symposium on Biomedical Imaging (ISBI), pp. 1587–1590 (2012)
19. Mladenić, D.: Feature Selection for Dimensionality Reduction, pp. 84–102. Springer, Heidelberg (2006)
20. Tsai, C.-F., Eberle, W., Chu, C.-Y.: Genetic algorithms in feature and instance selection. Knowl.-Based Syst. **39**, 240–247 (2013)
21. Zhao, M., Fu, C., Ji, L., Tang, K., Zhou, M.: Feature selection and parameter optimization for support vector machines: a new approach based on genetic algorithm with feature chromosomes. Expert Syst. Appl. **38**(5), 5197–5204 (2011)
22. Rodtook, A., Makhanov, S.S.: Multi-feature gradient vector flow snakes for adaptive segmentation of the ultrasound images of breast cancer. J. Vis. Commun. Image Represent. **24**(8), 1414–1430 (2013)
23. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
24. Touw, W.G., et al.: Data mining in the Life Sciences with Random Forest: a walk in the park or lost in the jungle? Brief. Bioinform. **14**(3), 315–326 (2013)
25. Azar, A.T., Elshazly, H.I., Hassanien, A.E., Elkorany, A.M.: A random forest classifier for lymph diseases. Comput. Methods Programs Biomed. **113**(2), 465–473 (2014)
26. Olson, M.A., Wyner, A.J.: Making sense of random forest probabilities: a kernel perspective, December 2018
27. Holland, J.H.: Genetic algorithms. Sci. Am. **267**(1), 66–72 (1992)
28. Fleming, P.J., Purshouse, R.C.: Genetic algorithms in control systems engineering. Control Syst. Robot. Autom. **XVII** (1993)
29. Han, S., Kang, H.-K., Jeong, J.-Y., Park, M.-H., Kim, W., Bang, W.-C., Seong, Y.-K.: A deep learning framework for supporting the classification of breast lesions in ultrasound images. Phys. Med. Biol. **62**(19), 7714–7728 (2017)
30. Antropova, N., Huynh, B.Q., Giger, M.L.: A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets. Med. Phys. **44**(10), 5162–5171 (2017)