



BREAST ULTRASOUND IMAGE CLASSIFICATION BASED ON MULTIPLE INSTANCE LEARNING

Dr.E.Geetha,^[1] S Irshath,^[2]

Assistant Professor^[1], PG Scholar^[2]

Department of Medical Electronics Engineering,

Sengunthar College of Engineering, Tiruchengode, Tamil Nadu, INDIA

^[1]geethumeest@gmail.com, ^[2]irshathsd@gmail.com

Manuscript History

Number: IJIRAE/RS/Vol.07/Issue03/Special Issue/30.MRAESCE10109

Received: 15, February 2020

Final Correction: 27, February 2020

Final Accepted: 10, March 2020

Published: 14, March 2020

Editor: Dr.A.Arul Lawrence selvakumar, Chief Editor, IJIRAE, AM Publications, India

Copyright: ©2020 This is an open access article distributed under the terms of the Creative Commons Attribution License, Which Permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Abstract: Breast ultrasound (BUS) image segmentation is a very difficult task due to poor image quality and speckle noise. In this paper, local features extracted from roughly segmented regions of interest (ROIs) are used to describe breast tumors. The roughly segmented ROI is viewed as a bag. And sub regions of the ROI are considered as the instances of the bag. Multiple-instance learning (MIL) method is more suitable for classifying breast tumors using BUS images. However, due to the complexity of BUS images, traditional MIL method is not applicable. In this paper, a novel MIL method is proposed for solving such task. First, a self-organizing map is used to map the instance space to the concept space. Then, we use the distribution of the instances of each bag in the concept space to construct the bag feature vector. Finally, a support vector machine is employed for classifying the tumors. The experimental results show that the proposed method can achieve better performance.

INTRODUCTION

Breast cancer is the most common cancer found in women both in the developed and developing countries. One of ten new cancers diagnosed worldwide each year is breast cancer [1]. Breast cancer is also the principal cause of death from cancer for the female population globally [1, 2]. Ultrasound examination, which is noninvasive and non-radioactive, is more convenient and suitable for palpable tumors in daily clinical practice [3]. Differential diagnosis of breast lesions can be acquired from ultrasound images. However, ultrasonography is operator dependent, and reading breast ultrasound (BUS) images requires well-trained. Even experts may have inter-observer variation; therefore, computer-aided diagnosis (CAD) system is needed to help radiologists in breast cancer detection and classification [4]. CAD system acts as a second reader that assists radiologists in medical decision-making process to reduce diagnosis error. For breast ultrasound CAD systems, the tumor region is located as a region of interest (ROI), and the features are extracted from the ROI; finally, the tumor is classified as benign or malignant. The most predictive features of a benign mass image are oval or round shape, circumscribed margins, and homogeneous internal echoes. The most predictive features of a malignant mass image are speculated or microlobulated margins, irregular shape, ill-defined margins, and heterogeneous internal echoes [5, 6].

Segmentation is a key step for most of CAD systems. There are some problems in ultrasound (US) images such as attenuation, speckle, shadows, and signal dropout. In addition, the contrast of BUS images is very low. These characteristics make auto-segmentation of BUS images very difficult and cause large difference between the auto-segmented result and real ROI [7]. Such difference will directly affect the final classification accuracy because the features (shape, margin, etc.) are dependent on correctly located ROIs.

IJIRAE: Impact Factor Value – Mendeley (Elsevier Indexed); Citefactor 3.8 (2019); SJIF: Innospace, Morocco (2019): 5.276 | PIF: 5.469 | Jour Info: 6.085 | ISRAJIF (2019): 6.456 | Indexcopernicus: (ICV 2019): 198.35

Diseases tend to change tissue scatter properties which reflected by texture variations in BUS images. Different tissues have different textures; therefore, the texture of BUS image is an effective feature for differentiating benign and malignant breast tumors [3, 8]. Auto-covariance [9], fractal dimension [10], - co-occurrence matrix [11], run-length matrix [12], and wavelet coefficients [13] have been widely utilized to derive discriminant features. In method of [14], the local texture features of the sub- regions in roughly segmented ROIs were used to characterize lesions. But when modeling ROIs with local features, traditional supervised learning methods are not suitable, since in learning phase, we only know the label of ROI, while the labels of the subregions in ROI are unknown. It is not appropriate to simply assign the label of the ROI to its subregions. For solving such problem, ROI can be considered as a bag and its subregions can be viewed as the instances of the bag. Then, the problem is transformed to a multiple-instance learning problem [15]. Multiple-instance learning (MIL) was proposed to solve learning problems with incomplete information about the labels of the data. For traditional supervised learning, each training example is represented by a fixed-length vector of the features with known label. However, in MIL, each example is called a bag and represented by multiple instances. The number of instances in each bag can be different. In other words, bags are represented by variable- length vectors. Labels are only provided for the training bags, and the labels of instances are unknown. The MIL task is to learn a model to classify new bags [16, 17]. drug activity prediction [15], diverse density (DD) for stock market pre- diction [18], natural scene classification [19], content-based image retrieval [20], MIL support vector machine for image classification [21], Citation-*k*-nearest neighbor (kNN) for web mining [22], etc. The paper presents a novel method for automatic breast tumor detection and classification of BUS images based on local texture features and multiple instance learning method.

PREVIOUS WORK REVIEW

In our previous work [14], the image is enhanced in fuzzy domain [23]. The S-function is used as the fuzzy membership function [24]. Then, the enhanced image is divided into non- overlapping subregions. The co-occurrence matrix [25] was utilized to describe the texture information of subregions. A support vector machine (SVM) was applied to classify subregions into tumor or normal tissues. And the prior knowledge about BUS images was used to obtain a rough ROI. An example was shown in Fig. 1. In Fig. 1, the original BUS image (Fig. 1a) was enhanced and divided (Fig. 1b). When the classifier was applied to subregions, the rough location of the tumor was segmented (Fig. 1c). For extracting the local features of the ROI, the points evenly distributed in the ROI were selected as the checkpoints, and their neighborhood windows were constructed as shown in Fig. 2. For each checkpoint, five distances ($d_0, 1, 2, 3, 4$), four directions ($000^\circ, 45^\circ, 90^\circ, \text{ and } 135^\circ$), and five windows ($W_0, 1, 2, 3, \text{ and } 4$) are used to establish the co- occurrence matrices. From a co-occurrence matrix, four descriptors are calculated: entropy, contrast, sum average, and sum entropy [25]. These descriptors are employed to represent the local texture features of the checkpoints. Finally, an SVM classifier is trained to classify the check- points. A threshold is selected according to the experiments. When the ratio of the checkpoints in a ROI classified as malignant is higher than the threshold (50 %), the tumor is classified as malignant. The problem is more suitable for MIL which views the ROI as bag and its subregions as instances.

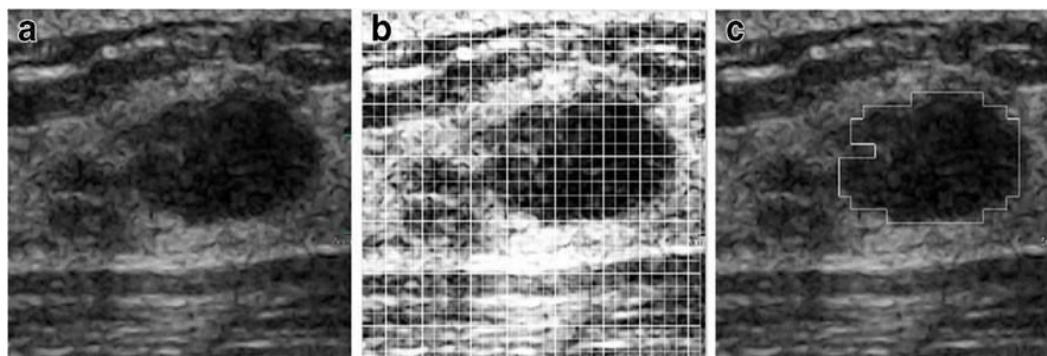


Fig. 1 a Original image. b Enhanced and divided image. c Rough ROI

The traditional MIL assumes that the positive bag has at least one positive instance and the negative bag has no positive were 83 mammary fibroadenomas, 8 intraductal papillomas, 3 mammary lipomas, and 2 benign phyllodes tumors. The privacy of the patients has been well protected. The original BI-RADS ratings of these lesions are assessed by experienced radiologists according to [27] and are listed in Table 1. for benign lesions. Informed consents to the Among malignant nodules, there were 61 invasive duct carcinomas, 7 invasive lobular carcinomas, 2 ductal carcinomas in situ, 1 mucinous carcinoma, and 1 medullar carcinoma. Among benign nodules, there instance.

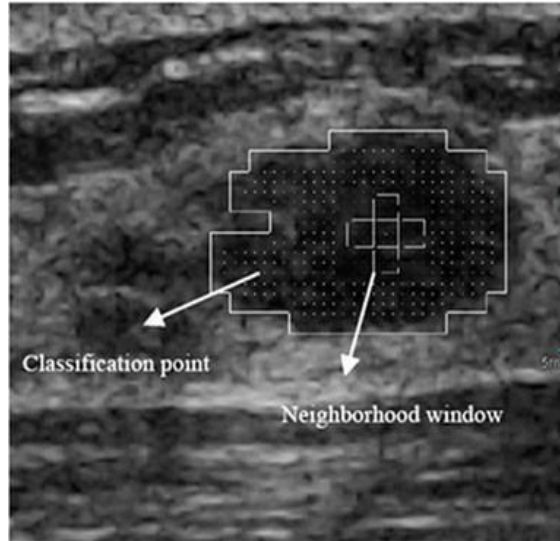


Fig. 2 Classification point and neighborhood window

However, such assumption is not suitable for classifying breast cancers. A malignant tumor is a group of cancer cells that may grow into surrounding tissues or spread to distant areas of the body. Invasive ductal carcinoma is the most common type of breast cancer. About 80 % of all breast cancers are invasive ductal carcinomas. It starts from a milk duct of the breast, breaks through the wall of the duct, and grows into the fatty tissue instance and the negative bag has no positive instance. However, such assumption is not suitable for classifying breast cancers. A malignant tumor is a group of cancer cells that may grow into surrounding tissues or spread to distant areas of the body. Invasive ductal carcinoma is the most common type of breast cancer. About 80 % of all breast cancers are invasive ductal carcinomas. It starts from a milk duct of the breast, breaks through the wall of the duct, and grows into the fatty tissue of the breast. It also may spread to the breast. It also may spread to Second Affiliated Hospital of Harbin Medical University. All images were collected by using a VIVID 7 with a 5–14- MHz linear probe and captured directly from the video signals. Pathologic validation was used for cancers and biopsy was used protocol were obtained from all patients in this study for benign lesions.

PROPOSED METHOD

The image database used in this research contains 168 cases (72 malignant and 96 benign). The average size of the tumors was 2.74 ± 1.3 cm (range, 0.77–5.64 cm). The ultrasound images were acquired by the Department of Ultrasound, $j = 1, \dots, n_i$, n_i is the number of instances of the bag X_i and $y_i \in \{-1, +1\}$ is the known label of X_i [28]. The key challenge of MIL is the ambiguity in the labels of the instances. The standard assumption of MIL is that the positive bag has at least one positive instance and the negative bag has no positive instance [17].

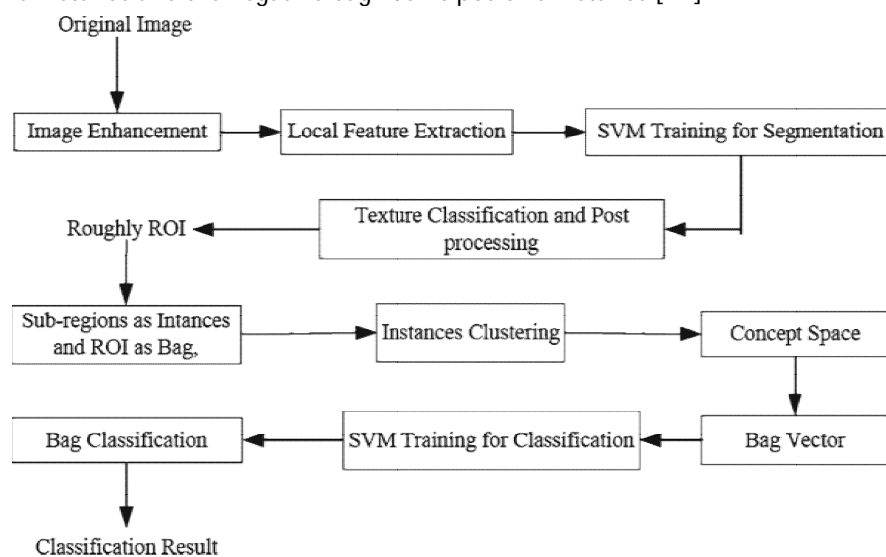


Fig.3 Flowchart of the proposed system

The generalized multiple-instance learning (GMIL) was proposed. It extends the assumption of how a bag's label is determined by its instances. In GMIL, a set of underlying concepts are considered; each concept may contribute to the classification. It assumes that a number of instances in each concept could determine the bag's label [29]. For our proposed approach, the GMIL idea was adopted. First, the concepts were learned from the instance space. Then, the bag was projected to the concept space and the projection could be used to construct the bag features. Finally, a trained SVM is employed to classify bags. The flowchart of the proposed approach is presented in Fig. 3. When a rough ROI is obtained and the local features of subregions are extracted, the classification task can be converted into an MIL task. The subregions of ROI can be viewed as the instances and the ROI can be considered as a bag. The key difficulty of MIL is the ambiguity of the instance labels. One method is to find which instance is like positive most by learning from instances statistically, such as DD algorithm [19]. The other method is to map the instance space to other spaces in which traditional learning algorithms can be employed, such as Citation- kNN algorithm [22]. In our proposed algorithm, to learn concepts from instances, the instances of all the bags are put together and clustered. Each cluster can be viewed as a concept in the concept space. The clusters can also reflect the distribution pattern of the instances. All the clusters can be considered as the concept space constructed from the instance space, i.e., by clustering, the instance space can be transformed to the concept space. Each instance of an ROI can be mapped to a concept in the concept space. Then, a bag can be represented by a fixed-length vector in the concept space. The element of the vector can be represented by the number of instances in a concept normalized by the number of instances of the bag. It reflects the distribution characteristics of a bag in the concept space. After transformation, traditional supervised learning algorithm can be employed in concept space. And f is a traditional supervised classifier learned from R^m , which is used to classify the bags. When a new bag is input, each instance of the bag is assigned to a concept (cluster) in the concept space learned from the training bags. Then, the bag can be represented by a vector whose length is equal to the number of concepts (clusters). And the vector is input to the classifier learned using training bags to classify new bags. A cluster method can be used to construct the concept space. In this paper, self-organizing map (SOM), or K -means, or fuzzy C -means is employed to transfer the instance space to the concept space, and their corresponding results will be compared. We will briefly discuss them below. & Self-organizing map SOM [30] is an unsupervised learning method. SOM represents input data by a small number of neurons and still preserves the topology of input data. The goal of SOM is to map arbitrary dimension input data to a one- or two- dimensional discrete map [31].

K-means

K -means clustering algorithm is one of the simplest unsupervised learning algorithms to partition a data set into k groups. Given a set of instances (x_1, x_2, \dots, x_n) , where each instance is a d -dimensional vector, k -means clustering aims to partition n instances into k sets ($k < n$) and $SO\{S_1, S_2, \dots, S_k\}$ to minimize the within-cluster sum of squares [32].

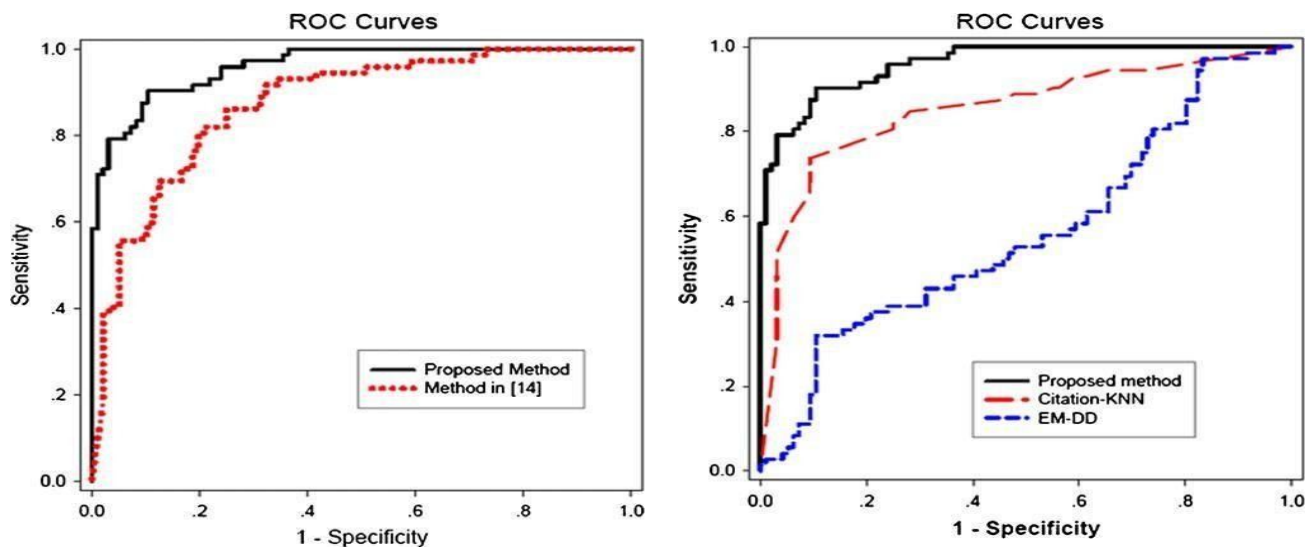
Fuzzy C-means

In fuzzy C -means (FCM) algorithm [33], each instance has a degree of belonging to clusters, rather than completely belonging to just one cluster. It iteratively classifies the data into optimal c partitions. All the instances of training bags.

EXPERIMENTAL RESULTS

In experiments, k -fold cross-validation approach is used. All the images are randomly divided into ten groups. Each time, one group is chosen for testing and the others are used for training. The performance of the proposed feature extraction and classification strategy is evaluated by the classification accuracy. Define the number of correctly and incorrectly classified malignant tumors as true positive (TP) and false negative (FN), and the number of correctly and incorrectly classified benign tumors as true negative (TN) and false positive (FP), respectively; the classification accuracy (ACC) is defined as: $(TP + TN) / (TP + TN + FP + FN)$. For comparison, the number of clusters in SOM, or K -means, or fuzzy C -means is chosen from 4 to 100. For determining the parameters C and λ , the grid search method is used [36]. $C=0.8$ and $\lambda=0.4$ are determined by experiments. The results are shown in Table 2. From Table 2, we can see that, in most cases, the method proposed by this paper has better performance than that of [14]. And the best performance is reached when 49 neurons were employed by SOM. The difference of ACCs between SOM and fuzzy C -means is significant ($p < 0.001$). But the difference of ACCs between SOM and k -means is not significant ($p > 0.05$). This confirms that the local features can be used to classify the tumors into benign and malignant well. When local features are modeled and extracted, the problem is more suitable for utilizing MIL than utilizing traditional supervised learning as described in [14]. The receiver operator characteristic (ROC) curves are also utilized to evaluate the performance of the proposed method as shown in Fig. 4. The area under curve (AUC) of the proposed method is higher than that of the method in [14] (0.96 vs. 0.87, $p < 0.005$). The ROC curves compared with that of Citation- KNN and EM- DD are shown in Fig. 5.

The AUC of the proposed method is higher than that of Citation- KNN (0.96 vs. 0.85, $p < 0.0001$) and EM-DD (0.96 vs. 0.56, $p < 0.0001$). The DD algorithm [19] tries to learn “true” concept from feature space. EM-DD is an improved DD algorithm. It considers the label of the instance as a missing attribute and estimates the label by using EM approach [34]. EM- DD algorithm is still under the assumption of traditional MIL, i.e., there is at least one positive instance in a positive bag and there are all negative instances in a negative bag. But this assumption is not suitable for BUS image classification.



The classification problem can be transformed from traditional supervised learning to MIL. The proposed novel MIL method is utilized to classify tumors into benign and malignant. The tumor region is located by rough segmentation and the local features are used to describe tumors. The instance space is mapped to the concept space and a classifier is trained to differentiate the tumors. The experimental results show that the proposed method has much better performance, and it will be useful for CAD systems of BUS images. The AUC of the proposed method is higher than that of the method in [14] (0.96 vs. 0.87, $p < 0.005$). The AUC of the method is also higher than that of Citation-KNN (0.96 vs. 0.85, $p < 0.0001$) and EM- DD (0.96 vs. 0.56, $p < 0.0001$). There are two limitations in this study. First, we just concentrated on local texture features. The global features (such as shape, margin, etc.) are also important. The combination of global features and local features will be studied in the future. Second, most cases in our study are invasive duct carcinomas mammary fibroadenomas which are more common. Other types should be studied in future research.

REFERENCES

1. Jemal A, Siegel R, Xu J, Ward E: Cancer statistics, 2010. *CA Cancer J Clin* 60(5):277– 300, 2017
2. Ferlay J, Héry C, et al: Global burden of breast cancer. *Breast Cancer Epidemiology*. C. Li, Springer, New York, 2010, 1–19
3. Huang Y, Wang K, Chen D: Diagnosis of breast tumors with ultrasonic texture analysis using support vector machines. *Neural Comput Appl* 15(2):164–169, 2017
4. Cheng HD, Shan J, Ju W, Guo Y, Zhang L: Automated breast cancer detection and classification using ultrasound images: a survey. *Pattern Recognit* 43(1):299–317, 2016
5. Stavros AT, Thickman D, Rapp CL, Dennis MA, Parker SH, Sisney GA: Solid breast nodules: use of sonography to distinguish between benign and malignant lesions. *Radiology* 196:123–134, 2015
6. Rahbar G, Sie AC, Hansen GC, et al: Benign versus malignant solid breast masses: US differentiation. *Radiology* 213:889–894, 2014
7. Noble JA, Boukerroui D: Ultrasound image segmentation: a survey. *IEEE Trans Med Imaging* 25(8):987–1010, 2016.
8. Jemal A, Siegel R, Xu J, Ward E: Cancer statistics, 2010. *CA Cancer J Clin* 60(5):277– 300, 2017
9. Ferlay J, Héry C, et al: Global burden of breast cancer. *Breast Cancer Epidemiology*. C. Li, Springer, New York, 2010, 1–19
10. Huang Y, Wang K, Chen D: Diagnosis of breast tumors with ultrasonic texture analysis using support vector machines. *Neural Comput Appl* 15(2):164–169, 2017



- 11.**Dr.R.Satish Kumar and Dr.K.Umadevi " A Novel peak torque Excitation Technique for Torque Improvement in Exterior Rotor Permanent Magnet Brushless DC Motor", International journal of Innovative research in Advanced Engineering, Vol. 1, 2014, pp. 227-236 (Impact factor 1.311).
- 12.**Dr.R.Satish Kumar and Dr.K.Umadevi " Torque Improvement for an Exterior Rotor PermanentMagnet Brushless DC Motor", International journal of Innovative research in Advanced Engineering, Vol. 1, 2014, pp. 1-5 (Impact factor 1.311).
- 13.**Dr.R.Satish Kumar and Dr.K.Umadevi " Novel Technique for Measurements of Dielectric Properties and Microwave Heating of In-Shell Eggs without Explosions in Microwave Oven for Pasteurization", International journal of Innovative research in Advanced Engineering, Vol. 2, , 2015, pp. 69-77 (Impact factor 1.311).
- 14.**Dr.R.Satish Kumar and Dr.M. Y. Sanavullah" Theoretical and experimental study of cooking regions for shell eggs in a domestic Microwave oven", International Conference on Electronics Computer Technology, 2011, <https://doi:10.1109/ICECTECH.2011.5941909>
- 15.**Raja, G P& Mangai, S 2018, 'Investigation On Optimization, Prioritizing and Weight Allocation Techniques for Load Balancing and Controlling Multimedia Traffic in Wireless Mesh Network', International Journal of Business Information Systems, SCOPUS Indexed Journal (Inderscience) - (P ISSN No: 1746-0972). Published Online: 10th Feb 2020, DOI: 10.1504/IJBIS.2020.105161.IF: 0.72.
- 16.**Raja, G P& Mangai, S 2017, 'Firefly Load Balancing Based Energy Optimized Routing for Multimedia Data Delivery in Wireless Mesh Network', Cluster Computing-The Journal of Networks Software Tools and Applications, SCOPUS Indexed Journal (Springer) - (E ISSN No: 1573-7543).Published Online: 27th Dec 2017, <https://doi.org/10.1007/s10586-017-1557-1>. IF: 2.040.