The diagnostic performance of leak-plugging automated segmentation versus manual tracing of breast lesions on ultrasound images

Ultrasound
2017, Vol. 25(2) 98–106
© The Author(s) 2017
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1742271X17690425
journals.sagepub.com/home/ult

Hui Xiong, Laith R Sultan, Theodore W Cary, Susan M Schultz, Ghizlane Bouzghar and Chandra M Sehgal

Abstract

Purpose: To assess the diagnostic performance of a leak-plugging segmentation method that we have developed for delineating breast masses on ultrasound images.

Materials and methods: Fifty-two biopsy-proven breast lesion images were analyzed by three observers using the leak-plugging and manual segmentation methods. From each segmentation method, grayscale and morphological features were extracted and classified as malignant or benign by logistic regression analysis. The performance of leak-plugging and manual segmentations was compared by: size of the lesion, overlap area (O_a) between the margins, and area under the ROC curves (A_z).

Results: The lesion size from leak-plugging segmentation correlated closely with that from manual tracing (R^2 of 0.91). O_a was higher for leak plugging, 0.92 ± 0.01 and 0.86 ± 0.06 for benign and malignant masses, respectively, compared to 0.80 ± 0.04 and 0.73 ± 0.02 for manual tracings. Overall O_a between leak-plugging and manual segmentations was 0.79 ± 0.14 for benign and 0.73 ± 0.14 for malignant lesions. A_z for leak plugging was consistently higher (0.910 ± 0.003) compared to 0.888 ± 0.012 for manual tracings. The coefficient of variation of A_z between three observers was 0.29% for leak plugging compared to 1.3% for manual tracings.

Conclusion: The diagnostic performance, size measurements, and observer variability for automated leak-plugging segmentations were either comparable to or better than those of manual tracings.

Keywords

Breast cancer, ultrasonography, lesion segmentation, computer-aided analysis

Date received: 14 July 2016; accepted: 8 December 2016

Introduction

Despite advances in detection and treatment, breast cancer is still one of the most common and lethal cancers worldwide, with more than 1.7 million new cases among women and over 0.5 million deaths from breast cancer a year.^{1,2} Incidence rates vary greatly worldwide from 19.3 per 100,000 women in Eastern Africa to 89.7 per 100,000 women in Western Europe.² The American Cancer Society estimates that about one in eight women in the United States will develop invasive

breast cancer during their lifetimes.³ With advances in technology, ultrasound (US) is being used increasingly to differentiate malignant from benign solid breast masses with high sensitivity.^{4–6} Several studies have

Department of Radiology, University of Pennsylvania, Philadelphia, PA, USA

Corresponding author:

Laith R Sultan, Ultrasound Research Lab, Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104, USA.
Email: lsultan@mail.med.upenn.edu

investigated automatic computer algorithms for characterizing breast masses on US images to improve the diagnostic value.^{7–10} The computer-based analysis involves lesion segmentation, feature extraction, and predictive modeling by learning and testing. While feature extraction and pattern recognition are easily automated, segmentation is often performed manually by users determining margins of the lesion. To automate the analysis, various investigators proposed computerized methods to define lesion margins.^{11–20}

Breast lesions on US images are often characterized by discontinuous margins. Through these discontinuities the isoechoic regions within the lesion leak into the surrounding regions during automated segmentation. To overcome this limitation, we have developed a multiscale semiautomated leak-plugging (LP) algorithm to segment solid breast masses on US images.¹¹ Starting from a seeded region, the algorithm grows candidate regions by decreasing the disk radius iteratively. At each iteration, leakage in the lesion margin is identified and plugged by masking out pixels at the leak site to ultimately yield a high-resolution margin. The metrics of evaluation of LP segmentation were based on assessing reproducibility and overlap between segmented and manual tracings. These metrics, while important, do not provide direct insights into how well the segmented margins perform on the task of performing diagnosis. For an automated method to be useful, the detected margins should not only exhibit high reproducibility and correlation with manual tracings, but must also demonstrate that the delineated margins have sufficient details to yield diagnostic performance comparable to or better than that of the manual tracing. In this study, we compare the diagnostic performance of LP segmentations with manual tracings from three observers. Diagnosis from images is a multistep process involving segmentation, feature extraction, and classification. Each comparison between LP segmentation and manual tracing was made using the same features and classification scheme. Variability in margins from repeated experiments of LP segmentation was compared with that of manual tracings. The goal was to demonstrate that LP segmentation could achieve a diagnostic performance that is equivalent to manual tracings.

Materials and methods

Image acquisition

The study was approved by the institutional review board and informed consent was not required for this retrospective study. Ultrasound images of 52 breast lesions (26 benign and 26 malignant) were selected from the institutional database of images. These

images were different from the images used in developing the algorithm.¹¹ No specific inclusion or exclusion criteria were used to select the images other than that masses should be biopsy proven with known mammographic BI-RADS.

The images were acquired using a broadband 12–5 MHz transducer and a Philips ATL 5000 scanner. Four to five images for each mass were selected for lesion segmentation by LP or manual tracing and the same cases were used for both segmentation methods.

Image segmentation

Three observers (O1, O2, and O3) consisting of two radiologists with 3 years' experience and one senior sonographer with more than 15 years' experience participated in the study by drawing the boundaries for each breast lesion. The participants were blind to the histological classifications of the lesions and to the margins drawn by other observers. The lesions were also segmented by the LP algorithm,¹¹ which is described and illustrated in Figure 1. LP is a multiscale region growing algorithm. Starting from a seed drawn by a non-expert approximately in the center of the lesion to sample the interior grayscale distribution, the algorithm grows candidate regions using disk structuring elements at multiple scales. Region growing is started with a large structuring element. With a larger structuring element, the segmentation does not leak, but the detected margin is coarse and lacks the detail necessary for differentiating lesions. To detect a more detailed boundary, the disk radius is decreased in steps of one pixel. With each decrement, leakage in the margin is detected by comparing the size of the region with that detected with the previous structuring element. If a leak is identified, it is plugged by masking out pixels at the leak site. Thus, by iteratively decreasing the size of the structuring element and plugging the leaks, margins with fine details are obtained. For this study, the starting structuring element was 20 pixels; the process of decrement was repeated until the radius of the structuring element reached a lower bound of four pixels (1.06 mm). To evaluate the observer variability in LP segmentation, the seed selection and segmentation were repeated on the same set of images by three observers.

Feature extraction and classification

From each LP and manual segmentation, margin grayscale and morphological features were extracted. The grayscale and morphological features were: brightness difference at the margin, margin sharpness, angular variation in brightness at the margin, depth-to-width ratio, axis ratio, tortuosity, radius variation, and elliptically normalized skeleton.^{7,21} The features, F , were

Figure 1. Block diagram describing the leak plugging algorithm. Step (a): Seeding: user defines a seed region in the center of the lesion. Step (b): Region growing: the seed region is grown using a disk element of radius $r(\max)$ where $r(\max)$ represents the largest possible leak in the margin. In this study, leaks in the margin were assumed to be smaller than 20 pixels. The disk radius is reduced by one pixel and the region growing is repeated. Step (c): Leak through the margin is determined by measuring the difference in the size of two consecutive regions determined by disks of decreasing size. Leak is defined as a pool that fits a structuring disk one pixel larger in radius than the current radius. Step (d): If no leak is detected, region growing is repeated with a smaller disk element. Steps (e, f): If a leak is detected it is plugged before growing the region with a smaller disk element of decreasing radius. Step (g): The process of region growing is repeated iteratively by successively decreasing the size of the disk element until a minimum radius of the disk element $r(\min)$ is achieved; $r(\min)$ defines the resolution of the margin detected.

used with logistic regression to determine the probability of malignancy ($P(M|F)$). A round-robin (leave-one-out) approach was used to assess the discriminating capability of the extracted lesion features: $N-1$ samples of the N samples in the data were trained to predict the behavior of the remaining sample, and the process was repeated until each sample had been the test case. The predicted probabilities, $P(M|F)$, were evaluated against the biopsy results to generate the ROC curve. The area under the ROC curve (A_z) for each LP and manual tracing trial was determined as a measure of its diagnostic performance.

Metrics of assessment

Evaluation of the segmented margin is a two-part problem. The first part deals with assessing similarity and differences between segmentations and the second part with the diagnostic efficacy of the segmented margin. Previous studies have evaluated the segmentation performance by assessing the overlap between the segmented margins with the manual tracing. While these metrics are important, they do not provide assessment of how well the segmented margin performed in differentiating malignant and benign masses. In this study, we measured both aspects of segmentations. Three metrics were used to compare the performance of LP segmentations and manual tracings. They included: (1) size (area) of the lesion L_a , (2) overlap between the areas enclosed within the margins, O_a , also known as the Jaccard coefficient or index, and (3) area A_z under the

ROC curve for lesion differentiation of each segmentation method. The purpose for L_a was to assess observer variation in the estimated size of the lesion, a measurement often used to evaluate breast lesions. Since two equal areas may not cover the same region, L_a does not assess the segmentation performance. The differences in segmentation performance were assessed by measuring O_a . The purpose for measuring A_z was to evaluate variation in the diagnostic performance, i.e. ability to differentiate malignant and benign masses.

Linear regression was used to compare areas of the lesions for both manual and LP boundary definition. Margin overlap for the manual tracing and LP segmentation was calculated both for malignant and benign lesions to assess the difference between the two methods. A two-tailed Student's t -test was used to determine the significance of the results. Finally, the area under the ROC curve A_z for each LP trial was compared to the area under the curve for the three manually drawn lesion margins. The difference was considered significant at $p \leq 0.05$.

Results

General results

Among the 26 benign lesions, 12 cases were found to be fibroadenomas, 12 cases identified as miscellaneous fibrocystic changes, and the remaining two cases were diagnosed histopathologically as intraductal papilloma and ductal hyperplasia. Of the malignant masses, 22

cases were found to be invasive ductal carcinoma, which was the most common diagnosis, one case was diagnosed as invasive mucinous carcinoma, one case as adenocarcinoma, and the remaining two cases were determined as poorly differentiated carcinoma. The sizes of the lesions did not show a difference between malignant and benign tumors ($p=0.125$); however, ages of patients were significantly higher ($p=0.009$) in malignant (59.5 ± 12.7 years) compared to benign (46.5 ± 12.9 years).

Evaluation of segmentation

Lesion size. The areas of lesions determined from the LP segmentation correlated closely with the areas obtained from the manual tracing (Figures 2 and 3). A least-squares linear regression fit of the automated segmentation and the manual tracing areas for the three observers yielded a correlation coefficient R^2 of 0.91 with a slope 1.068. The coefficient of variation (CV) for area assessment for individual cases was larger for the manual tracings compared to the LP segmentations (Figure 4). The average CV for the manual tracings was $19 \pm 19\%$ compared to $11.4 \pm 19\%$ for LP for three observers. The difference was significant ($p=0.04$).

Overlap area. Table 1 summarizes the overlap (O_a) between the areas determined by different observers using LP segmentations and manual tracings. The overlap was higher for LP than for the manual tracings: 0.92 ± 0.01 and 0.86 ± 0.02 for benign and malignant masses, versus 0.80 ± 0.01 and 0.73 ± 0.01 for the manual tracings. The difference between the two

segmentation approaches was significant for both benign and malignant cases ($p < 10^{-7}$).

Evaluation of diagnostic performance. Figure 5 shows examples of LP segmentation alongside manual tracing. Although there is a close match between the lesion shapes found by the two methods (Figure 5(b) and (c)), LP detected margins with finer resolution. The ROC curves for the LP segmentation experiments by three different observers were very similar and were closely superimposed (Figure 6). Similarly, the ROC curves for the manual tracings also superimposed but showed notable differences in parts of the curve (Figure 7).

Figure 3. Bland-Altman plot showing the differences between leak plugging segmentation and manual tracing area measurements (mm^2).

Figure 2. The correlation between leak plugging (LP) and manual segmentation methods of lesion area measurement. The solid line is the least squares fit of the data to the model $y=mx$. The dotted line with the lower slope represents perfect agreement between LP and manual tracing. ROI: Region of Interest.

Figure 4. Coefficient of variation (CV) of the segmented area for leak plugging (LP) and manual (Man) segmentations.

Table 1. Inter- and intraobserver assessments of margin overlap for three observers (O1, O2, O3) of leak plugging segmentation compared to that of manual tracings

Interobserver					Intraobserver		
	Leak plugging		Manual tracing				
Observer	Benign	Malignant	Benign	Malignant	Observer	Benign	Malignant
O1 ↔ O2	0.91 ± 0.01	0.82 ± 0.05	0.81 ± 0.01	0.77 ± 0.03	O1 ↔ O1	0.78 ± 0.18	0.76 ± 0.13
O1 ↔ O3	0.96 ± 0.01	0.96 ± 0.07	0.81 ± 0.01	0.70 ± 0.02	O2 ↔ O2	0.81 ± 0.12	0.70 ± 0.17
O2 ↔ O3	0.91 ± 0.01	0.81 ± 0.05	0.80 ± 0.09	0.71 ± 0.02	O3 ↔ O3	0.78 ± 0.13	0.72 ± 0.13
Overall	0.92 ± 0.01	0.86 ± 0.06	0.80 ± 0.04	0.73 ± 0.02	Overall	0.79 ± 0.14	0.73 ± 0.14

Figure 5. (a) Sonogram of a malignant breast tumor. (b) Manual segmentation by user-defined boundary. (c) Computer-automated segmentation of the breast tumor by the leak plugging algorithm.

The area under the ROC curve (A_z) for LP was consistently higher than that for manual tracings (Table 2): 0.910 ± 0.003 compared to 0.888 ± 0.012 . When paired comparison was made between LP segmentations and manual tracings (Table 2), the difference between the two groups was not significant ($p = 0.95$). The coefficient of variation of A_z between observers was 0.29% for automated segmentation

compared to 1.3% for manual tracings thereby indicating less variation in the diagnostic performance for the LP method between observers. The sensitivity for LP segmentation for the three observers was 92.3% and specificity ranged from 84.6% to 88.5%, while for manual segmentation sensitivity ranged from 76.9% to 88.5% and specificity from 84.6% to 96.2% for the three observers (Table 3). The number of cases

Figure 6. ROC curves showing the diagnostic performance of computer-automated leak plugging segmentation of breast lesions on ultrasound by three observers (O1, O2, O3).

Figure 7. ROC curves showing the diagnostic performance of manual tracing segmentation of breast lesions on ultrasound by three observers (O1, O2, O3).

misidentified as malignant and benign by the three observers using both manual and LP segmentation methods are shown in Table 4. The Youden index was used as an operational point to compare the error rate. The results showed that fewer malignant cases were misidentified with LP segmentation compared to the manual tracing.

Table 5 summarizes the relative magnitude of logistic regression weights for each feature used in manual and LP segmentations. Since these weights are

Table 2. Comparison of the diagnostic performance of leak plugging and manual tracing segmentation. CV stands for coefficient of variation. O1, O2, O3 refer to observers 1, 2, and 3, respectively

Observer	ROC area (A_z) – leak plugging	ROC area (A_z) – manual tracing	p -Value
O1	0.911 ± 0.047	0.884 ± 0.049	0.98
O2	0.907 ± 0.047	0.879 ± 0.049	0.98
O3	0.912 ± 0.047	0.901 ± 0.045	0.91
CV	0.29%	1.3%	

Table 3. Shows the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) for leak plugging and manual segmentations. O1, O2, O3 refer to observers 1, 2, and 3, respectively

	Leak plugging			Manual tracing		
	O1	O2	O3	O1	O2	O3
Sensitivity (%)	92.3	92.3	92.3	88.5	84.6	76.9
Specificity (%)	88.5	88.5	84.6	96.2	84.6	92.3
PPV (%)	88.9	88.9	85.7	95.8	84.6	90.9
NPV (%)	92.0	92.0	91.7	89.3	84.6	79.9

Table 4. The number of cases misidentified as malignant (M) and benign (B) by the three observers with both manual and LP segmentation methods. O1, O2, O3 refer to observers 1, 2, and 3, respectively

Observer	Manual tracing			Leak plugging		
	M	B	Total	M	B	Total
O1	3	1	4	2	3	5
O2	4	4	8	2	3	5
O3	6	2	8	2	4	6

determined by fitting the data to a logistic regression equation, they are specific to the dataset studied. For the data studied, circularity, radius variation, and depth-to-width ratio were found to be the most important features of segmentation methods.

Table 5. The relative magnitude of logistic regression weights for each feature used in both segmentation methods. Each value in the table represents average of the coefficients for the three observers (O1, O2, and, O3)

	Circularity	Radius Variation	Depth to width	Bright difference	Margin sharpness	Angular variation margin	Angular variation in	Intercept
Man	30.99 ± 18.60	28.46 ± 28.98	-4.13 ± 3.84	0.37 ± 0.29	0.02 ± 0.13	-0.23 ± 0.91	0.09 ± 0.50	-31.82 ± 23.90
LP	10.49 ± 1.50	6.51 ± 2.81	-1.86 ± 0.73	0.40 ± 0.02	0.01 ± 0.04	0.16 ± 0.15	-0.79 ± 0.07	-11.54 ± 3.17
P-value	0.20	0.32	0.41	0.87	0.87	0.54	0.09	0.28

Discussion

Delineation of lesion margin plays an important role in the diagnosis of breast masses. The shape of the lesion provides clues as to whether it is malignant or benign, and margin segmentation is needed to determine the size of the lesion. In general, malignant lesions have irregular margins compared to well circumscribed boundaries for benign cases.²² Lesion margins are usually defined manually by a user with medical expertise who draws the tissue-lesion border. The process, however, is labor intensive and prone to significant variation due to differences in selection criteria. Automated segmentation of the lesions offers an alternative to manual tracing in which users provide minimal input that requires little medical knowledge and software training. Automated methods rely on quantitative assessment of image grayscale distribution, so they are by design more objective and less user dependent. As they also require less user knowledge and time, automated methods are a cost-effective option for lesion analysis. Such approaches can be especially useful in low-resource environments with limited access to medical experts for analyzing US images.

Ultrasound images offer some unique challenges to automated segmentation of lesions. Shadowing, non-orthogonal incidence of the ultrasound on the lesion surface, and speckle interference are some of the many factors that cause lesion margins to be discontinuous. Partially missing lesion-tissue boundaries often cause segmentation methods that work on other modes of imaging to fail on US images at leaks in the margin. In addition, because the 3D volume of a tumor is essentially constructed by a sequence of 2D images, LP of segmentation can be extended to analyzing 2D images of the tumor volume. One can also envision extending the current method of segmentation by using spheres in place of disks to plug leaks. The LP algorithm addresses these limitations and the method only requires a manually identified seed at the approximate center of the lesion as an input.

Although the LP algorithm has shown high correlation with manual tracing, its utility for differentiating malignant from benign breast masses has not been

previously demonstrated. In this study, we assessed the algorithm not just by its ability to produce reliable breast mass segmentations, as determined by the overlap area, but also by the size measurements (as determined by the lesion area measurements) and the performance of these segmentations for differentiating malignant from benign breast masses (as determined by the area under the ROC curve). The high correlation of the LP segmentation area with the manual tracing area suggests that the two methods are similar in comparing sizes of the lesions. The slope of the linear regression graph (Figure 2) indicates that LP estimated the lesion area to be higher by 7% compared to manual tracing. Although the reason for this systematic bias is not known, the higher values could be due to the added area from finer resolution of the LP margins (Figure 5(b) and (c)). The results also show that the measurements made on individual cases with LP had much lower (three times) variation than observed for manual tracings (Figure 4). The lower variation in measurements made using LP also indicate that the method is relatively independent of seed placement.

The results also show that LP experiments by different investigators lead to consistent diagnostic performance: the ROC curves for the experiments were nearly identical (Figure 6) with an average area under the curve A_z of 0.91 ± 0.003 with only 0.29% coefficient of variation. The sensitivity and specificity measured for both segmentation methods were also comparable, with higher sensitivity values obtained for LP. The high sensitivity and specificity for automated segmentation could demonstrate the clinical validity of the classification taking into account the incidence of malignant and benign lesions in the study. The area under the ROC curve for the manual tracings was consistently lower for each observer compared to LP segmentation (Table 2). Although paired comparison of diagnostic performance between LP and manual tracing did not show a statistical difference, a closer inspection (Figure 7) reveals that the ROC curves of the three observers did not superimpose as well as the LP curves in Figure 6. A marked difference in the diagnostic performance is observed in different parts of the

manual ROC curves. For instance, Observer O3 had the largest A_z of the three observers but had substantially weaker performance in the early part of the curve; this observer was better at differentiating malignant masses that look benign. The weak performance of O1 in this region was more than compensated for by better performance in the later part of the curve. O2 meanwhile was strongest in the middle of the curve, so was best at differentiating the difficult, indeterminate cases, but performed worse with easier cases. While such differences were observed for manual tracings, LP provided the same consistent results for all three experiments. In addition, the number of cases misidentified as malignant or benign using LP segmentation was less compared to manual tracings (Table 4). In addition, LP was able to identify malignant cases better than manual tracings. This means that differentiating lesions using LP has superiority in identifying the difficult malignant cases with irregular margins over the manual tracing.

In logistic regression methods, the goal is to make decisions by optimization of various coefficients using the maximum likelihood method. The optimized coefficients are likely to be data dependent. The purpose of machine learning is to determine the relative weights on a training set and apply them to the test sets. Since these weights are determined by fitting the data to a logistic regression equation, they are specific to the dataset studied. Although all the features contributed towards diagnosis, circularity, radius variation, and depth to width ratio were found to be most important contributors for the segmentation methods (Table 5).

This study also assessed the overlap between LP segmentation experiments and observer tracings (Table 1). Overlap between regions is the ratio of their intersection area to their union area, and is a measure of congruence, or true positive fraction. In comparison to our previous studies,¹¹ inter- and intraobserver overlap assessment was higher in this study possibly due to the difference in observer experiences and the study sample size. Overlap was greater between LP experiments than between manual tracings for benign and malignant masses showing a statistically significant difference. The overlap for the malignant cases was lower in both segmentation approaches implying that it was more difficult to delineate malignant masses compared to the benign masses. The better agreement (overlap) with LP segmentations indicates that automated segmentation is more precise in identifying irregular difficult margins.

The study also has some limitations. There are currently several approaches for segmentation,^{12–20} each with its strengths and weaknesses. However, this study addresses the issue for one approach, i.e. leak plugging. Due to differences in the methods, the results of the present study cannot be generalized to be

applicable to other approaches discussed in the literature. Another limitation is that the image data were acquired from a single scanner. Although LP does not use any scanner related properties other than the gray level values, there is reason to believe that the method should be suitable across different ultrasound platforms. A larger study with different scanners should provide more insight into the issue. In addition, future studies with more data will allow the comparison of the diagnostic performance of the two segmentation methods according to various histopathological subtypes of breast lesions, which is currently limited by the study of small sample size. Another limitation in our study is that a single-seeded region was used for segmentation. Increasing the number of seeded regions could further improve the accuracy and robustness of the LP segmentation algorithm.

Conclusion

In conclusion, LP compared favorably with manual tracing in diagnostic performance. Predictive models learned from LP segmentations were either just as good as or slightly better than models learned from manual tracings at differentiating malignant from benign masses. By several measures, LP segmentation showed less variability than manual tracing. The results on diagnostic performance further suggest that LP is a viable alternative to user-defined expert segmentation. Multiscale region growing techniques like LP, tailored to the properties of US images, could prove valuable for computer-aided diagnosis of breast cancer by ultrasound, potentially reducing, time, costs, and patient trauma while improving diagnostic power. Future studies with larger datasets are indicated to validate the technique.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported in part by NIH 598 grant CA130946.

Ethical approval

The study was approved by the institutional review board and informed consent was not required for this retrospective study.

Guarantor

CMS

Contributors

HX, LRS, and CMS researched literature and conceived the study. TWC designed the software. HX, LRS, CMS, GB, and SMS did the data analysis. HX, LRS, TWC, and CMS wrote the first draft of the manuscript. LRS, HX, and CMS wrote the final version of the manuscript. HX and LRS contributed equally as first authors to the manuscript. All authors reviewed and approved the final version of the manuscript.

Acknowledgement

We thank Karen Apadula for help with image acquisition and data analysis.

References

1. International Agency for Research on Cancer (IARC) and World Health Organization (WHO). GLOBOCAN 2012: Estimated cancer incidence, mortality and prevalence worldwide in 2012. <http://globocan.iarc.fr/Default.aspx> (accessed 31 December 2016)
2. World Health Organization (WHO). Global Health Estimates, 2013. http://www.who.int/healthinfo/global_burden_disease/en/ (accessed 31 December 2016)
3. The American cancer society. Key statistics about breast cancer, 2013. <http://www.cancer.org/cancer/breastcancer/detailedguide/breast-cancer-key-statistics>. (accessed 31 December 2016)
4. Mendelson EB, Berg WA and Merritt CR. Toward a standardized breast ultrasound lexicon, BI-RADS: ultrasound. *Semin Roentgenol* 2001; 36: 217–225.
5. Raza S, Goldkamp AL, Chikarmane SA, et al. Ultrasound of breast masses categorized as BI-RADS 3, 4, and 5: Pictorial review of factors influencing clinical management. *Radiographics* 2010; 30: 1199–1213.
6. Hong AS, Rosen EL, Soo MS, et al. BI-RADS for sonography: positive and negative predictive values of sonographic features. *AJR Am J Roentgenol* 2005; 184: 1260–1265.
7. Sehgal CM, Cary TW, Kangas SA, et al. Computer-based margin analysis of breast sonography for differentiating malignant and benign masses. *J Ultrasound Med* 2004; 23: 1201–1209.
8. Kashikura Y, Nakayama R, Hizukuri A, et al. Improved differential diagnosis of breast masses on ultrasonographic images with a computer-aided diagnosis scheme for determining histological classifications. *Acad Radiol* 2013; 20: 471–477.
9. Song JH, Venkatesh SS, Conant EA, et al. Comparative analysis of logistic regression and artificial neural network for computer-aided diagnosis of breast masses. *Acad Radiol* 2005; 12: 487–495.
10. Sahiner B, Chan HP, Roubidoux MA, et al. Computer-aided diagnosis of malignant and benign breast masses in 3D ultrasound volumes: Effect on radiologists' accuracy. *Radiology* 2007; 242: 716–724.
11. Cary TW, Conant EF, Arger PH, et al. Diffuse boundary extraction of breast masses on ultrasound by leak plugging. *Med Phys* 2005; 32: 3318–3328.
12. Chang RF, Wu WJ, Moon WK, et al. Automatic ultrasound segmentation and morphology based diagnosis of solid breast tumors. *Breast Cancer Res Treat* 2005; 89: 179–185.
13. Cui J, Sahiner B, Chan HP, et al. A new automated method for the segmentation and characterization of breast masses on ultrasound images. *Med Phys* 2009; 36: 1553–1565.
14. Tian JW, Ning CP, Guo YH, et al. Effect of a novel segmentation algorithm on radiologists' diagnosis of breast masses using ultrasound imaging. *Ultrasound Med Biol* 2012; 38: 119–127.
15. Liu Y, Cheng HD, Huang J, et al. An effective approach of lesion segmentation within the breast ultrasound image based on the cellular automata principle. *J Digital Imaging* 2012; 25: 580–590.
16. Huang QH, Lee SY, Liu LZ, et al. A robust graph-based segmentation method for breast tumors in ultrasound images. *Ultrasonics* 2012; 52: 266–275.
17. Gao L, Yang W, Liao Z, Liu X, et al. Segmentation of ultrasonic breast tumors based on homogeneous patch. *Med Phys* 2012; 39: 3299–3318.
18. Gomez W, Leija L, Alvarenga AV, et al. Computerized lesion segmentation of breast ultrasound based on marker-controlled watershed transformation. *Med Phys* 2010; 37: 82–95.
19. Cheng JZ, Chou YH, Huang CS, et al. ACCOMP: Augmented cell competition algorithm for breast lesion demarcation in sonography. *Med Phys* 2010; 37: 6240–6252.
20. Pons G, Martí J, Martí R, et al. evaluating lesion segmentation in breast ultrasound images related to lesion typology. *J Ultrasound Med* 2013; 32: 1659–1670.
21. Sehgal CM, Cary TW, Cwanger A, et al. Combined Naïve Bayes and logistic regression for quantitative breast sonography. In: *2012 IEEE international ultrasonics symposium (IUS)*, Dresden, Germany, 2012, pp. 1686–1689.
22. American College of Radiology (2013). *Breast imaging reporting and data system: BI-RADS Atlas*. 5th ed. Reston, VA: ACR, 2013.