Mammogram Image Segmentation Using Bioinspired Novel Bat Swarm Clustering

David González-Patiño[1], Yenny Villuendas-Rey[2], Amadeo J. Argüelles-Cruz[1]

[1] Instituto Politécnico Nacional, Centro de Investigación en Computacion,
Mexico City, Mexico

[2] Instituto Politécnico Nacional, Centro de Innovación y Desarrollo Tecnológico en Cómputo,
Mexico City, Mexico

yenny.villuendas@gmail.com, davidglezp-92@hotmail.com,
aarguelles@ipn.mx

Abstract. Segmentation is one of the main tasks related to breast cancer classification. Automatic and semiautomatic algorithms have been proposed lately, and in this paper, a new method to segment mammography images is proposed using novel bat algorithm (NBA) and unsupervised metric measures as objective function. Results showed a useful method to segment mammographies using bioinspired algorithms based on bat optimization.

Keywords: Mammogram, bioinspired, novel bat algorithm, swarm, breast cancer.

1   Introduction

Recent studies have showed that breast cancer is one of the main causes of death in women around the world [1]. In Mexico, breast cancer is the main reason of deaths in adult women in ages among 30 and 54 years [2][3]. It is also one of the currently most common cancers in Mexico, with high risk. Risks are estimated based on new cancer cases between people free of this disease [4].

Detecting breast cancer using x-ray mammograms in early phases is the best way to prevent breast cancer deaths. This is the reason why mammographies are recommended in order to obtain an early detection of breast cancer among women older than 26 years and it is highly recommended to obtain a mammography study for women from 40 to 49 years old [1].

In Mexico it is very important to improve screening analysis to promote early detection for breast cancer because only 5-10% cases of breast cancer in Mexico are detected in early phases [1].

Mammograms are images obtained by mammography. Typically, these images are gray scaled, and the cancer lesion are differentiated due to their texture and color [5]. Automatic detection of lesions in mammogram includes the segmentation of the region of interest, in order to separate the lesion from the rest of the mammary tissue. To

accomplish this task, several segmentation techniques have been used [6–9]. However, development of simple and accurate segmentation algorithms is still a challenge in digital image processing [7].

Segmentation algorithms can be divided in two main categories: region and edge based [7]. The first category assigns each pixel to a region while the other category constructs an image based on the boundaries of a specific object.

Region based segmentation can be modeled as an optimization problem, considering the search space as all possible categories for each pixel, and the objective function as a quality measure for the segmentation.

Metaheuristics algorithms are inspired in nature processes and are used in a lot of optimization problems due to their power and easy implementations. They also try to find the optimal and practical solution to nonlinear or very complex problems [10].

In this paper we focus on this topic, implementing a new method to segment mammograms to later assist diagnosis using segmentation based on regions. Our method is based on the recently proposed Novel Bat Algorithm (NBA) [11]. Novel Bat Algorithm is a swarm intelligence algorithm based on the echolocation behavior of bats and this characteristic can be associated with an objective function.

The paper is organized as follows. Section 2 addresses some previous works in the field of automatic or semi-automatic breast cancer segmentation for mammograms. Section 3 explains the algorithm used for mammogram segmentation, while section 4 presents the experimental results. The paper finalizes with conclusions and future work.

2   Previous Works

Around the world, mammograms has been used to detect breast cancer, however there are another techniques used for diagnosis of breast cancer, being the most commonly used: ultrasounds [12], Magnetic Resonance Imagery (MRI) [13] and other more invasive methods like biopsies [14].

Semi-automatic and automatic segmentation is useful for the specialist because it provides a simpler image and the lesion can be detected easily. The algorithms proposed for mammogram segmentation are varied and some of them have been proposed since 1998 [15] using a clustering algorithm extracting texture features and matrices based on lengths.

In 2012 Rahmati et al. proposed a method to segment lesions in mammographies, based on a maximum likelihood active contour [9]. This algorithm separates the image in 2 parts: lesion and background.

Recently was proposed a method for mammographies segmentation using Deep Neural Networks, this method was proposed in 2016 by Dubrovina et al. [16] and uses a new model for classification in mammograms.

Similarly Sargent and Park [8] designed in the same year a method for segmentation of medical images such as mammograms, tomograms and scanned images using side-by-side images comparisons.

Until now, there is no method proposed based on Novel Bat Algorithm for breast cancer mammograms segmentation. In this paper, we proposed a method to use this algorithm using metrics based on lengths as optimization function.

3   Proposed Segmentation Algorithm

Segmentation algorithms are used to split the background and the region of interest in images, in the following, we will explain the proposed method based on Novel Bat Algorithm proposed by Xian-Bing Meng et al. in 2015 [11] based on the Bat Algorithm proposed by Xin-She Yang in 2012 [17].

Using the variables called loudness and pulse rate, which represent the proximity to other bats; the Novel Bat Algorithm generates random solutions using the currently optimal solutions. This algorithm was used in this paper because it explores a lot of solutions because of the randomly generated solutions in order to find the best solution.

This algorithm is inspired in the echolocation of bats, a characteristic that helps bats to sense distances and provides them knowledge about the objects of interest and backgrounds.

The pseudoalgorithm of Novel bat algorithm uses an objective function for the optimization process. The pseudoalgorithm adapted to the segmentation problem is showed below.

The frequency of each bat is defined in equation 1, and is calculated using minimum and maximum frequencies and a random value between 0 and 1 given by a uniform distribution:

$$f_i = f_{min} + (f_{max} - f_{min})\beta. \tag{1}$$

Velocity and position for each bat is defined in equation 2 and 3 respectively, position is given by previous position and velocity, and on the other hand velocity is given by the previous velocity and the difference between actual position and currently best position defined as x_*:

$$v_i^t = v_i^{t-1} + (v_i^t - x_*)f_i , \tag{2}$$

$$x_i^t = x_i^{t-1} + v_i^t . \tag{3}$$

The flow chart in Figure 1 was done according to the pseudoalgorithm proposed by Xian-Bing Meng [17].

We need to initialize the bat population with random values in the range from 0 to 255 because these are the values of possible gray levels.

The algorithm iterates a defined number of times and for each iteration the algorithm will generate solutions adjusting frequencies and velocities for each bat.

New solutions are generated randomly and they are accepted if they are better than previous solutions evaluating them in the objective function.

David González-Patiño, Yenny Villuendas-Rey, Amadeo J. Argüelles-Cruz
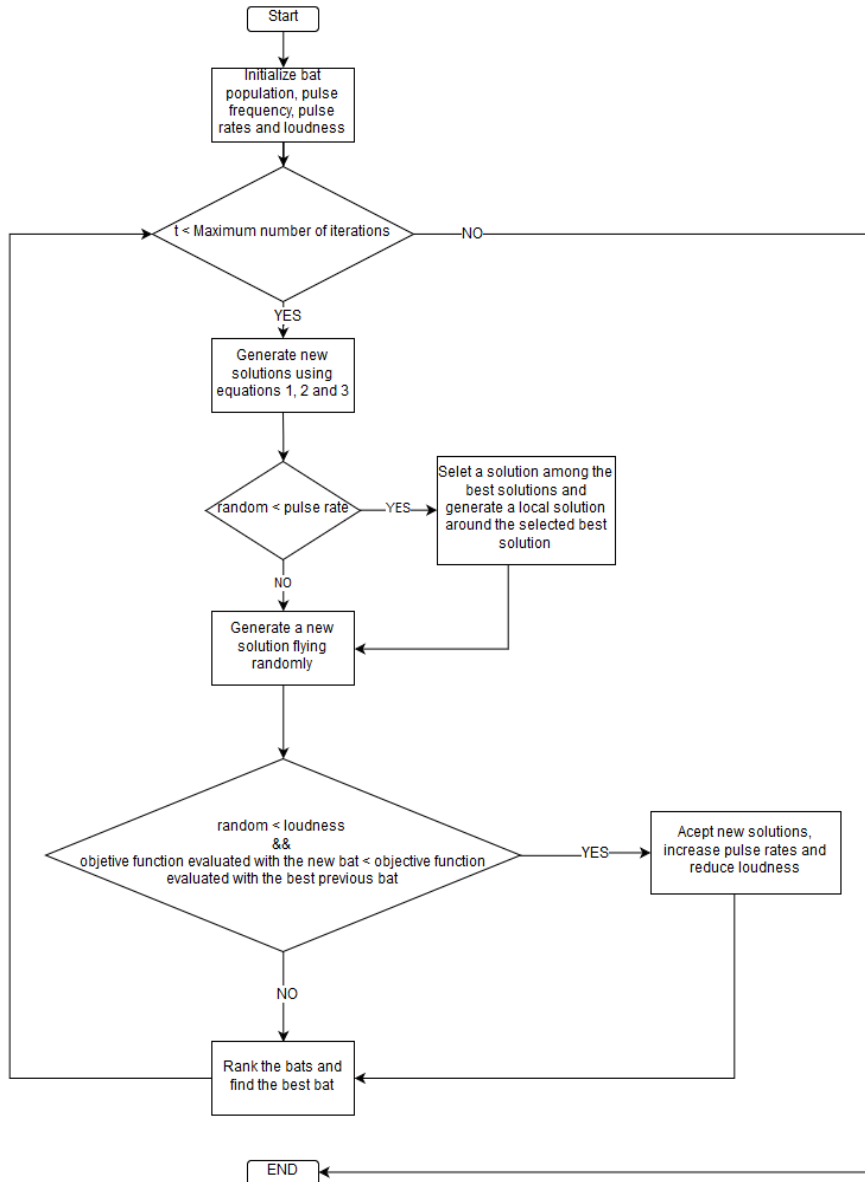
Fig. 1. Flow chart for Novel Bat Algorithm.

The Novel bat algorithm was originally programmed in Matlab by Xian-Bing Meng in 2015 [11], this algorithm was tested in Matlab and for this paper the modified algorithm was programmed in C# using Microsoft Visual C# 2010 based on the same algorithm.

For the implementation presented in this paper each bat represents a gray level and a partition of the image as represented in figure 2.
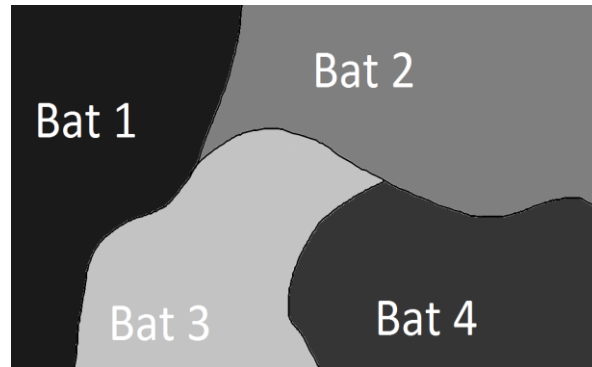
Fig. 2. Representation of gray levels for segmentation, each bat represents a region of the image.

The algorithm was modified for processing mammographies images using formula 4 as objective function:

$$\frac{\text{Minimum distance between pixels in different group}}{\text{Maximum distance between pixels in same group}}. \tag{4}$$

The metric used to calculate the distance between each bat was the Minkowski distance for order 1. The bounds used for this implementation were based on gray levels; this means only numbers from 0 to 254 were used.

The segmentations were evaluated according to the work presented by Hui Zhang et al. in 2007 [18]. That work is a survey of evaluations in image segmentations using unsupervised methods. To evaluate the segmentations we used average squared color error (F, equation 5) which penalizes over-segmentations; the survey proposed an improved measure to evaluate average squared color error (F', equation 6).

F' is an extension of F and it penalizes segmentations that have a lot of small regions of the same size:

$$\text{F} = \sqrt{N} \sum_{j=1}^{N} \frac{ej^2}{\sqrt{Sj}}, \tag{5}$$

$$\text{F}' = \frac{1}{1000 * SI} \sqrt{\sum_{b=1}^{MaxArea} N(b)^{1+\frac{1}{b}} \sum_{j=1}^{N} \frac{ej^2}{\sqrt{Sj}}}, \tag{6}$$

where N represents number of regions in the image and Sj are the quantity of pixels in the region j.

SI represents the area of the image and N(b) is number of regions of the segmented image that have exactly b units of area.

In equation 2 and 3 the squared color error of region j ej^2 is defined in equation 7:

$$ej^2(Rj) = \sum_{p \in Rj} (Cx(p) - Ĉx(Rj))^2.$$

(7)

Cx(p) represents the value of component x for pixel p. And finally the average value of component x in the region j is defined in equation 8:

$$Ĉx(Rj) = \frac{\sum_{p \in Rj} Cx(p)}{Sj}.$$

(8)

These errors measures were used in the performance tests.

4   Experiments

The algorithm proposed in this paper was tested using the dataset Breast Cancer Digital Repository [19] provided by the Faculty of Medicine in the University of Porto, in Portugal.

This dataset contains 200 lesions proven by biopsies of 190 women, also including 362 segmentations in the mammograms. The images used were gray scaled. The boundaries for each lesion were identified by expert radiologists.

We used the NBA based segmentation algorithm proposed in this paper (SNBA) for segmentations in two and three regions. Each region was assigned with a color and colors are not representative to their original pixel color.

Several experiments were tested for this paper and all images were processed using two and three bats. The measures were calculated for all the segmented images and compared to the same measures calculated using the segmentations made by the radiologist.
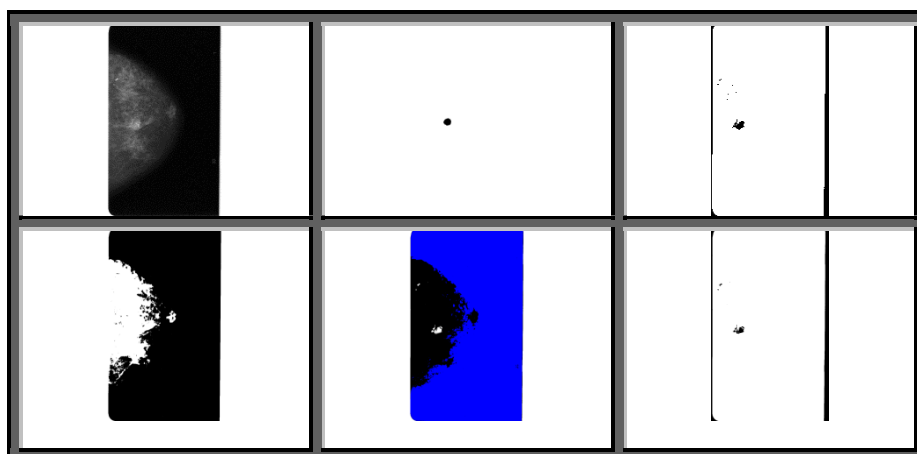
Fig. 3. (Top left) Original mammogram, (Top middle) Segmented image by expert radiologist, (Top right) Segmented image using OTSU method, (Bottom left) Image after processing using the algorithm proposed in this paper using two bats for the segmentation, (Bottom middle) Image after processing using the algorithm proposed in this paper using three bats for the segmentation, (Bottom right) Image after obtaining the segmentation cluster of interest in the segmentation for three bats algorithm.
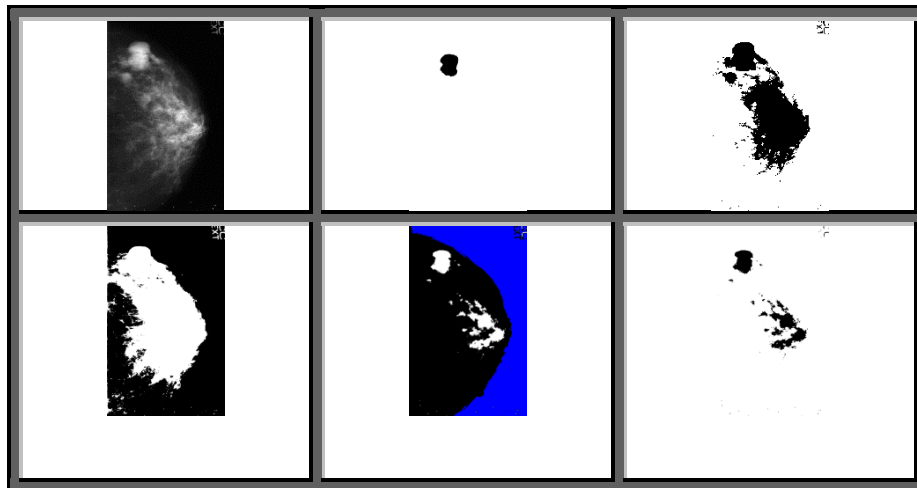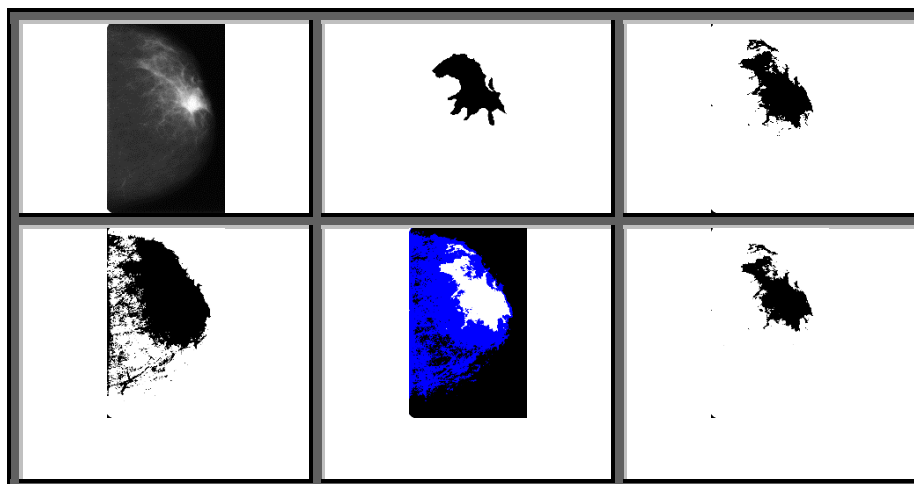
Fig. 4. (Top left) Original mammogram, (Top middle) Segmented image by expert radiologist, (Top right) Segmented image using OTSU method, (Bottom left) Image after processing using the algorithm proposed in this paper using two bats for the segmentation, (Bottom middle) Image after processing using the algorithm proposed in this paper using three bats for the segmentation, (Bottom right) Image after obtaining the segmentation cluster of interest in the segmentation for three bats algorithm.
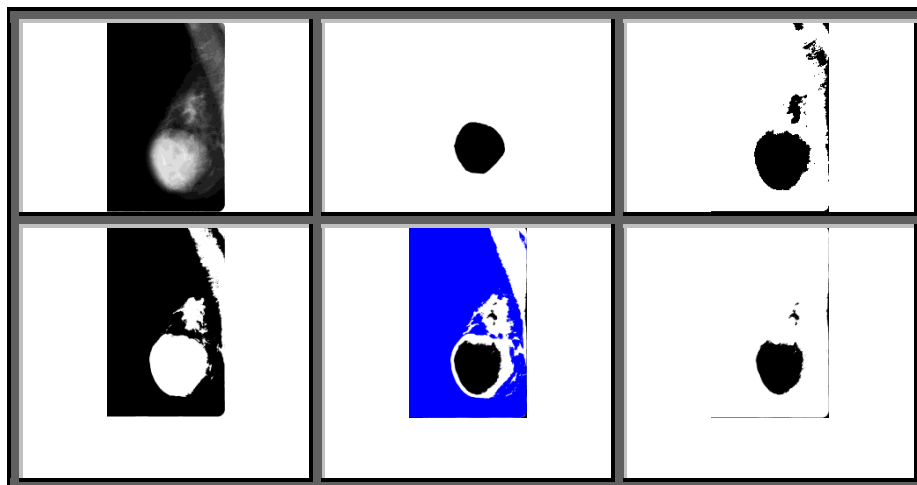
Fig. 5. (Top left) Original mammogram, (Top middle) Segmented image by expert radiologist, (Top right) Segmented image using OTSU method, (Bottom left) Image after processing using the algorithm proposed in this paper using two bats for the segmentation, (Bottom middle) Image after processing using the algorithm proposed in this paper using three bats for the segmentation, (Bottom right) Image after obtaining the segmentation cluster of interest in the segmentation for three bats algorithm.

Table 1. Measures of error for segmented images.

Figure	Measure of error	Segmented image using two bats (SNBA)	Segmented image using three bats (SNBA)	Cluster of interest (SNBA)	Segmented image using OTSU method
3	F	3.20039×10^5	1.58467×10^5	1.58467×10^5	1.58611×10^5
	F'	2.693×10^{-4}	1.333×10^{-4}	1.333×10^{-4}	1.334×10^{-4}
4	F	1.940457×10^6	2.414146×10^6	2.414146×10^6	2.155042×10^6
	F'	16.29×10^{-4}	20.27×10^{-4}	20.27×10^{-4}	18.09×10^{-4}
5	F	9.985477×10^6	9.296394×10^6	8.606007×10^6	9.751196×10^6
	F'	84.03×10^{-4}	78.23×10^{-4}	72.42×10^{-4}	82.06×10^{-4}
6	F	9.985477×10^6	9.296394×10^6	8.606007×10^6	8.268497×10^6
	F'	84.03×10^{-4}	78.23×10^{-4}	72.42×10^{-4}	69.58×10^{-4}

Fig. 6. (Top left) Original mammogram, (Top middle) Segmented image by expert radiologist, (Top right) Segmented image using OTSU method, (Bottom left) Image after processing using the algorithm proposed in this paper using two bats for the segmentation, (Bottom middle) Image after processing using the algorithm proposed in this paper using three bats for the segmentation, (Bottom right) Image after obtaining the segmentation cluster of interest in the segmentation for three bats algorithm.

Measures are presented in Table 1. As we can observe segmented images using the algorithm proposed in this paper (SNBA) showed a smaller error in comparison to other segmentations used.

As shown in Table 1, the proposed SNBA algorithm is able to obtain a good segmentation by using only three bats (representing the regions corresponding to cancer lesion, mammary tissue and background, respectively). However, some noisy pixels are marked as cancer lesion, and further processing will be needed to remove them. Some columns showed very similar errors representing no significant changes in the segmented images.

5   Conclusions

This paper presents a new method to segment images using bioinspired Novel Bat Algorithm (NBA). As many bioinspired algorithms, the NBA used here has some parameters that can be adjusted to improve the performance of the segmentation process. In a similar way, the quantity of bats can be adjusted to obtain better segmentations. The adjustment of parameters can improve the convergence rate of this algorithm and it is a good line of investigation. In addition, the investigation detects some noisy pixels marked as cancer lesion, and further processing will be needed to address this issue.

References

1. Knaul, F.M., Nigenda, G., Lozano, R., Arreola-Ornelas, H., Langer, A., Frenk, J.: Breast cancer in Mexico: a pressing priority. Reprod. Health Matters, 16, pp. 113−123 (2008)
2. Borja-Aburto, V.H., Dávila-Torres, J., Rascón-Pacheco, R.A., González-León, M., Fernández-Gárate, J.E., Mejía-Rodríguez, I., de Jesús González-Izquierdo, J., los Ríos, P.M.: Cancer mortality in the Mexican Social Security Institute, 1989-2013. Salud Publica Mex. 58, pp. 153−161 (2016)
3. Lozano, R., Knaul, F.M., Gómez-Dantés, H., Arreola-Ornelas, H., Méndez, O.: Tendencias en la mortalidad por cáncer de mama en México, 1979-2006. Observatorio de la Salud. Doc. Trab. Compet. y Salud, Fund. Mex. para la Salud (2008)
4. Wun, L.-M., Merrill, R.M., Feuer, E.J.: Estimating lifetime and age-conditional probabilities of developing cancer. Lifetime Data Anal. 4, pp. 169−186 (1998)
5. Yaffe, M.J.: Digital mammography. In: PACS, pp. 363−371 (2006)
6. Byrd, K., Zeng, J., Chouikha, M.: Performance assessment of mammography image segmentation algorithms. In: 34th Applied Imagery and Pattern Recognition Workshop

(AIPR'05), p. 6 (2005)

7. Timp, S., Karssemeijer, N.: A new 2D segmentation method based on dynamic programming applied to computer aided detection in mammography. Med. Phys. 31, pp. 958–971 (2004)

8. Sargent, D., Park, S.Y.: Automatic segmentation of mammogram and tomosynthesis images. In: SPIE Medical Imaging (2016)

9. Rahmati, P., Adler, A., Hamarneh, G.: Mammography segmentation with maximum likelihood active contours. Med. Image Anal. 16, pp. 1167–1186 (2012).

10. Doerner, K.F., Gendreau, M., Greistorfer, P., Gutjahr, W., Hartl, R.F., Reimann, M.: Metaheuristics: progress in complex systems optimization. Springer Science & Business Media (2007)

11. Meng, X.-B., Gao, X.Z., Liu, Y., Zhang, H.: A novel bat algorithm with habitat selection and Doppler effect in echoes for optimization. Expert Syst. Appl. 42, pp. 6350–6364 (2015)

12. Kelly, K.M., Dean, J., Comulada, W.S., Lee, S.-J.: Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts. Eur. Radiol. 20, pp. 734–742 (2010)

13. Warner, E., Plewes, D.B., Hill, K.A., Causer, P.A., Zubovits, J.T., Jong, R.A., Cutrara, M.R., DeBoer, G., Yaffe, M.J., Messner, S.J., others: Surveillance of BRCA1 and BRCA2 mutation carriers with magnetic resonance imaging, ultrasound, mammography, and clinical breast examination. Jama, 292, pp. 1317–1325 (2004)

14. Albertini, J.J., Lyman, G.H., Cox, C., Yeatman, T., Balducci, L., Ku, N., Shivers, S., Berman, C., Wells, K., Rapaport, D., others: Lymphatic mapping and sentinel node biopsy in the patient with breast cancer. Jama, 276, pp. 1818–1822 (1996)

15. Sahiner, B., Chan, H.-P., Petrick, N., Helvie, M.A., Goodsitt, M.M.: Computerized characterization of masses on mammograms: the rubber band straightening transform and texture analysis. Med. Phys. 25, pp. 516–526 (1998)

16. Dubrovina, A., Kisilev, P., Ginsburg, B., Hashoul, S., Kimmel, R.: Computational mammography using deep neural networks. Comput. Methods Biomech. Biomed. Eng. Imaging Vis, pp. 1–5 (2016)

17. Yang, X.-S., Hossein Gandomi, A.: Bat algorithm: a novel approach for global engineering optimization. Eng. Comput. 29, pp. 464–483 (2012)

18. Zhang, H., Fritts, J.E., Goldman, S.A.: Image segmentation evaluation: A survey of unsupervised methods. Comput. Vis. image Underst, 110, pp. 260–280 (2008)

19. Moura, D.C., López, M.A.G.: An evaluation of image descriptors combined with clinical data for breast cancer diagnosis. Int. J. Comput. Assist. Radiol. Surg. 8, pp. 561–574 (2013)