International Journal of Natural and Engineering Sciences
E-ISSN: 2146-0086 15(1): 1-11, 2021
Research Article

APPLICATION OF FIREFLY ALGORITHM IN BREAST CANCER DETECTION IN MAMMOGRAPHIC IMAGES

Gazal Mardanian[1,a], Neda Behzadfar[1,2,b,*]

[1]*Department of Electrical Engineering, Najafabad Branch, Islamic Azad University, Najafabad, Iran*
[2]*Digital Processing and Machine Vision Research Center, Najafabad Branch, Islamic Azad University, Najafabad, Iran*

**Corresponding Author:*
E-mail: n.behzadfar@pel.iaun.ac.ir

(Received 11th January 2021; accepted 02th April 2021)

a: ORCID 0000-0000-0000-0000, b: ORCID 0000-0002-3679-5664

ABSTRACT. Breast cancer is one of the most common cancers in women. In many cases, there are no obvious symptoms in patients with it. Accurate diagnosis of this type of cancer in the early stages is essential to reduce mortality. Mammography is a standard method used to diagnose breast diseases. To prevent mental analysis of mammographic images by radiologists and increase the accuracy of breast cancer detection, artificial intelligence systems have received much attention in recent years. In this paper, a method based on combining firefly algorithm and applying appropriate pre-processing on the image to detect breast cancer in mammographic images is presented. To do this, mammographic images in the mammography screening (DDSM) digital data set are used. Three performance criteria of accuracy, sensitivity and precision (93.4%, 91%, 95%) indicate the proper performance of the proposed method compared to existing studies. This work shows better performance when compared to existing work in literature.

Keywords: *breast cancer, digital mammography images, dilation, firefly algorithm, morphology*

INTRODUCTION

Breast cancer is one of the most common cancers [1]. Typically, one in eight women has breast cancer. If this disease is diagnosed in time, it can be easily treated [2, 3].

Cancer is the third leading cause of death in Iran and between 101 and 105 people die of cancer every day. Abnormalities such as the presence of glands in the breast, changes in the size and dimensions of the breast, differences in the color of the breast skin, chest pain, etc. are among the physical symptoms of breast cancer [4, 5]. Currently, about 75% of people with breast cancer in the second stage and only 18% of people in the early stages of the disease seek treatment [6, 7]. The methods used to diagnose breast cancer are usually based on image processing. In this method, cancer is diagnosed using mammography or thermography images [8]. Compared to other diagnostic methods such as ultrasound or magnetic resonance imaging (MRI), mammography is the most accurate method for early detection of breast cancer [9]. During mammography, some X-rays are absorbed in proportion to the tissue and other rays pass through it, which is proportional to the amount of energy absorbed in proportion to the nature of the tissue. The amount of signal emitted from the tumor tissue is different from normal tissue, and based on the amount of changes, it can be determined whether the tissue has a tumor or not.

Mammographic imaging processing helps the physician and radiologist in the easy diagnosis of the disease and thus protects the patient from irreversible risks [10].

In mammographic images, changes in gray areas in different parts of the image are minimal. This makes it difficult to segment areas containing the tumor through light levels alone. Mammographic images also include high-frequency components and different levels of noise that make it difficult to detect the tumor [11, 12].

In several studies, many methods have referred to the characteristics of optimization based on collective intelligence [13, 14]. Among the methods based on collective intelligence, we can mention artificial bee colony [15], ant colony optimization [16] and bee particle optimization method [17, 18]. The best methods based on collective intelligence currently available are the firefly algorithm, cuckoo search and bat algorithm [19, 20].

The ability of electrical impedance spectroscopy to roughly localize and clearly distinguish cancers from normal tissues and benign lesions is presented in [21], which localization of these lesions is confirmed by simultaneous, in register digital breast tomosynthesis mammography or 3-D mammograms. A computer aided detection system to help reduce reading time and prevent errors is developed in [22], which in the multi-stage system segmentations of the breast, the nipple and the chestwall are performed, providing landmarks for the detection algorithm.

In this paper, a combination of incandescent algorithm and morphology is used to find the tumor area. The tumor has a high light intensity in mammographic images, so using the light worm algorithm, one can find the area where the tumor is most likely to be present. The firewall algorithm at its lowest level focuses on generating solutions within a search space and selects the best solution for survival. Random search avoids getting caught in the local optimal trap. In general, this algorithm has the following three main advantages, and according to the problem of this research, which is of multi-quality type, this algorithm will be used:

- Automatic segmentation of the entire population into subgroups
- Ability to deal with multi-quality optimization
- Variety in solutions

For this method to be successful, proper pre-processing must be performed on the images. Then, using the morphology algorithm, the desired area can be expanded and the entire tumor area can be extracted. Using the morphology algorithm, it is possible to compare the uniformity of brightness and structure of each point with respect to the extracted tumor region, and if it is close, that point is added to the tumor region. In the second part of the article, the study of firefly algorithm and morphology is presented. Details of the proposed method are given in the third section. In the fourth section, the results are presented and finally the conclusion and discussion are expressed in the fifth section of the paper.

ALGORITHMS USED IN RESEARCH

Firefly algorithm

Fireflies live in nature in groups and their main characteristic is their blinking light [23, 24]. These lights have two basic functions: attracting the opposite sex and alerting the enemy. The dimmer cream always moves towards the brighter cream [25, 26]. The blinking behavior of fireflies is based on the criteria of attractiveness, brightness and distance [27, 28].

Morphology

Morphology is a wide range of nonlinear image processing operations that process images based on shape. In a morphology operation, the value of each pixel in the output image is set based on the value of other pixels in its neighborhood. In fact, morphological operations use neighborhood pixel values. Morphological operators include expansion, erosion, opening and closing [29].

SUGGESTED METHOD

This section describes the steps of the proposed method. First, the images are collected. An appropriate pre-processing is then applied to the collected images, after which the main processing will be performed. Fig. 1 shows the algorithm of the proposed method.

Data collection	From the DDSM database
Preprocessing	1. Reduce the size of images and remove empty edges from the movie 2. Remove background and noise from images (a) Thresholding algorithm (b) Area growth algorithm
Processing	1. Find the probable point of the tumor site using the firefly algorithm 2. Expand the starting point found using morphology operators 3. Determining the boundaries of the tumor area in mammographic images

Fig. 1. *Proposed method algorithm*

Collect pictures

This study uses mammographic images from a digital database for mammography screening (DDSM) prepared at the University of South Florida. The DDSM database consists of 2620 items, each of which is a collection of mammographic images and information corresponding to a patient's chest examination. In this database, the range of each suspicious area is specified. There is also background information on the location of the tumor determined by at least two experienced radiologists [30, 31]. Each item contains two images of each breast with tail (CC) and oblique (MLO) (two angles). In addition, more information such as examination time, breast tissue density, patient age, type of pathology (pathology), number of abnormal cases, file name, file type, scan date, etc. are also available [32].

Pre-processing

Images in the DDSM database contain noise and common elements in mammograms that may interfere with the results of the proposed method used. These adverse elements include patient profile, type of test performed, pre-image pixels, and noise generated due to defects in image production or scanning processes. The purpose of the pre-processing step is to remove unwanted elements. Fig. 2 shows an example of images in the DDSM database that contain various information [33].

Resize images and remove blank edges from the movie

The first step in the pre-processing process is to reduce the size of the original images and create images with a height of 1024 pixels. A similar reduction operation is also applied to the width of the images. The purpose of this step is to reduce the processing time, because if you work with the original size of the images (average height of 6000 pixels) the processing time will be very long. It should be noted that this reduction does not lead to data loss [34, 35]. Then all points are removed at a distance of 30 pixels from the vertical border and 60 pixels from the horizontal border. These omissions are intended to remove the area between the edges of the film and the empty spaces without the film.

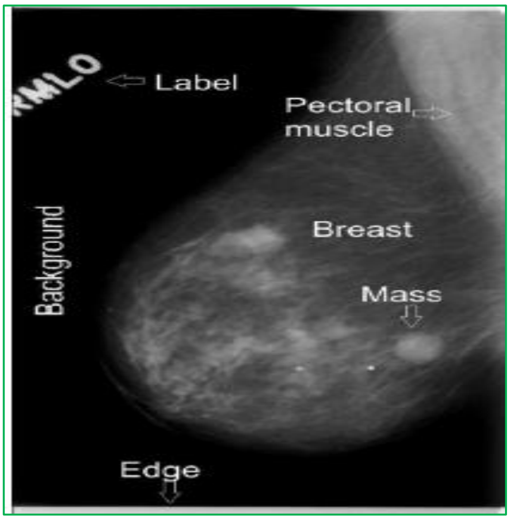


Fig. 2. *An example of the information contained in DDSM database images*

REMOVE BACKGROUND PIXELS

To remove background pixels, the threshold and area growth algorithm is used.

Resize images and remove blank edges from the movie

The background has near-black gray surfaces that are removed using the threshold algorithm. First, the image is divided into two halves and the sum of the gray surface values of each half of the image is calculated [36, 37].

The gray surfaces in the background are close to the black surfaces and have values close to zero, so the part with the lowest total contains the background part.

A medium filter is used to remove noise. This filter uses the mean (μ) and standard deviation (σ) of gray surfaces and calculates the L threshold based on it. The image is considered as a n×n matrix in the following form:

$$\begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nn} \end{bmatrix} \tag{1}$$

The mean (μ) and standard deviation (σ) are calculated from Equations (2) and (3), where m is the number of pixels in the image:

$$\mu = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} x_{ij}}{m} \tag{2}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} (x_{ij} - \mu)^2}{m}} \tag{3}$$

The value of L is calculated as follows:

$$L = \mu + \sigma \tag{4}$$

If the value of L is less than 50, it is used as a threshold. Otherwise μ (calculated mean value) is used. This threshold is to separate the background area, thus preserving the contour of the chest.

By thresholding, higher-intensity pixels, such as chest pixels and its detection marks, are placed in one cluster (white) and lower-intensity pixels, such as background and noise pixels, are placed in another (black) cluster. Therefore, background pixels are removed from the image.

Area growth algorithm

The area growth algorithm uses uniformity in the area. In this algorithm, each pixel is compared to its adjacent pixels and the degree of similarity between the pixels is evaluated. If the absolute value of the difference between their gray values is less than the threshold value, they are called two similar pixels.

By selecting a pixel as the starting point, the regional growth algorithm examines and controls the pixels adjacent to the starting pixel. If a pixel is large enough to resemble the pixel in the target area, the pixels in that area are added. This operation continues until it covers all the pixels in the image, or in other words, another pixel does not have the conditions to be added to the area.

In the image from the threshold operation, there are still components outside the chest, such as test identification labels and some noise. These components are removed from the image by the area growth algorithm, which separates areas not related to the image. Only the area with the largest area, which is the breast tissue, is selected.

PROCESSING

The main processing is performed on the final image obtained from the pre-processing stage. In this step, first the firefly algorithm is applied on the image and then using morphological operations, the location of the tumor in the images is detected.

Firefly algorithm

As mentioned, the tumor has a high brightness in mammography images, so using the firefly algorithm, we can find the point where the tumor is more likely to be present. Details of the firefly algorithm used are shown in Fig. 4.

In this process, 25 fireflies have been used. 25 fireflies are randomly placed on the 2D view of the image (Fig. 5). Then the firefly algorithm is applied to the image 100 times. Finally, by applying the firefly algorithm, the fireflies are located at the site of the tumor and are absorbed (Fig. 6). Fireflies are attracted to each other based on the intensity of light and attractiveness. (Creams that are less bright and attractive are attracted to creams that are more attractive and bright.)

Morphological operations

In morphological operations, the desired point obtained from the implementation of the firefly algorithm is expanded and the entire tumor area is extracted. In morphological operations, the degree of luminosity and structure uniformity of each point relative to the extracted tumor area is compared.

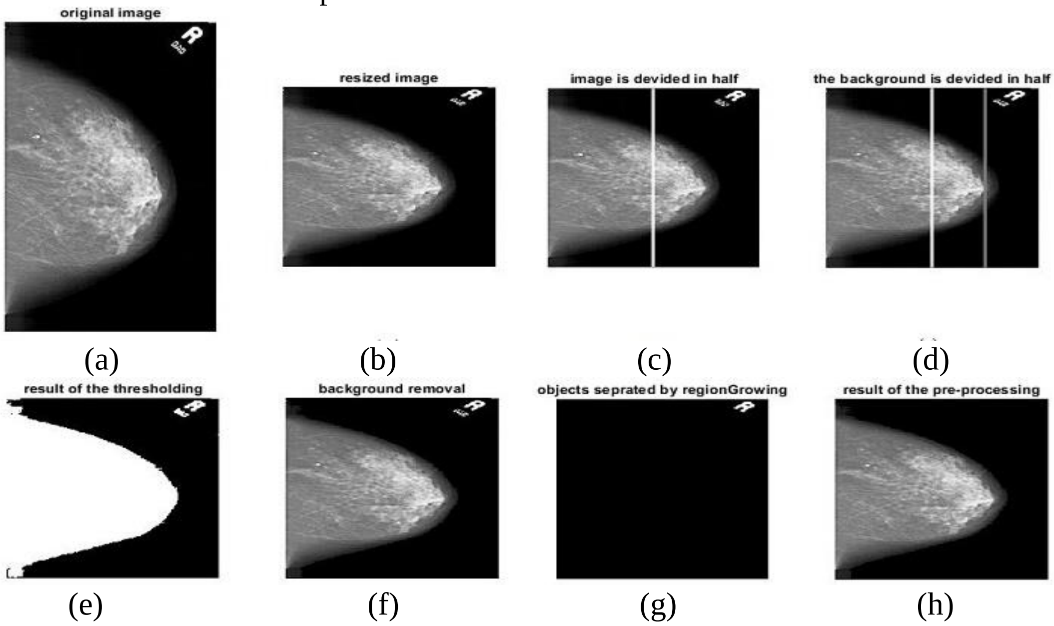


Fig. 3. *Eliminate background and noise (a) Original images, (b) Diminished image, (c) White line halving image, (d) Gray line dividing the background content of the image, (e) Thresholding results, (f) Delete background, (g) delete objects unrelated to area growth algorithm and (h) Pre-processing results*

Objective function f(x), $X=(x_1,\ldots,x_d)^T$
Initialize population of fireflies x_i (i=1,2,…,n)
Define light absorption coefficient γ
While (t<MaxGeneration)
For i=1:n all n fireflies
For j=1:I all n fireflies
Light intensity I_i at x_i is determined by $f(x_i)$
If ($I_j > I_i$)
Move firefly I towards j in all d dimensions
End If
Attractiveness varies with distance r via exp[$-\gamma r$]
Evaluate new solutions and update light intensity
End for j
End for i
Rank the fireflies and find the current best
End while
Postprocess results and visualization

Fig. 4. *Firefly Algorithm*

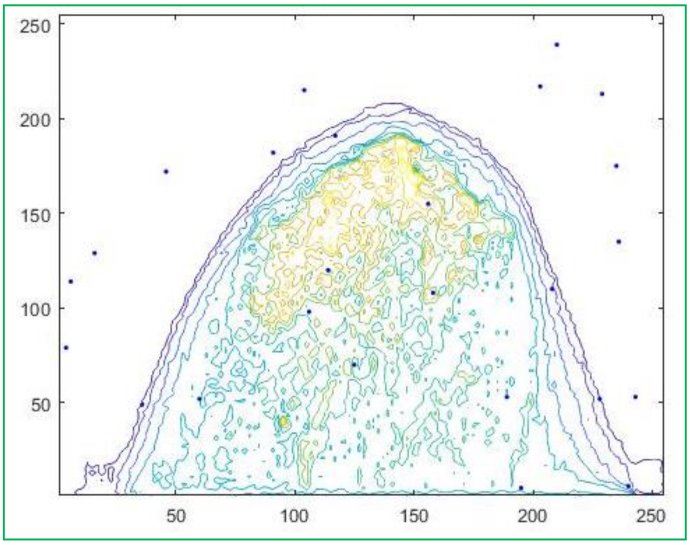


Fig. 5. *Initial locations of 25 fireflies*

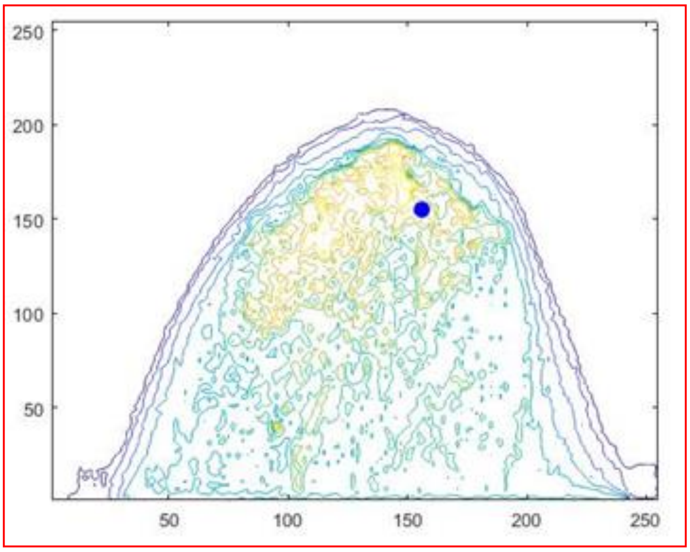
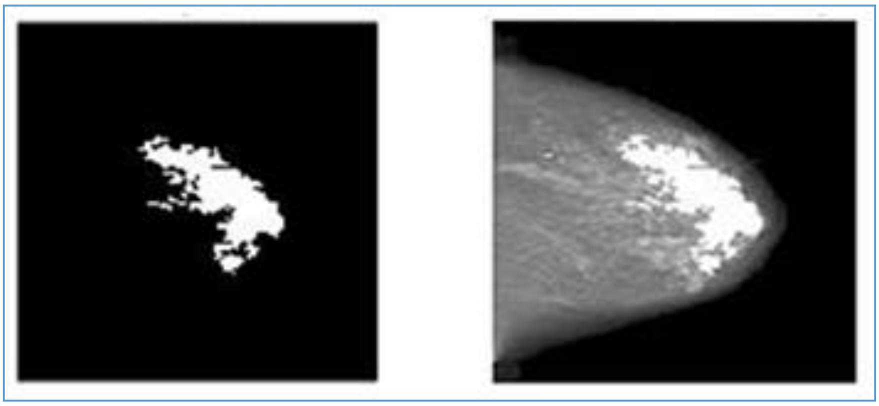


Fig. 6. *Final locations after 100 iterations*



(a) Gained *(b) Extracted*

Fig. 7. *Tumor area obtained and extracted from breast mammography image*

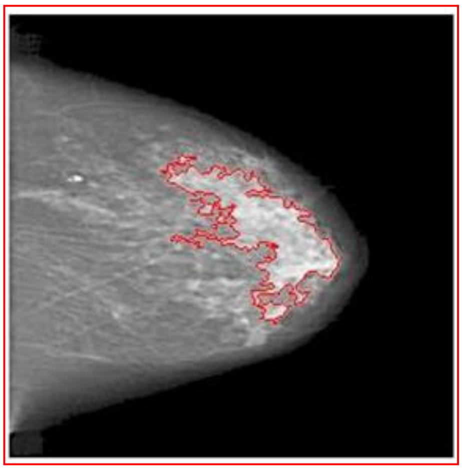


Fig. 8. *Determining the boundaries of the tumor area in mammography*

If similar, that point is added to the tumor area. As mentioned, morphological operators such as opening and closing, erosion and expansion are used to extract the entire tumor area, and the entire tumor area is obtained according to Fig. 7. In Fig. 8, the borders of the extracted tumor area are shown in red.

RESULTS AND DISCUSSION

Mammographic images are considered in two groups of healthy and tumor. The applied algorithm is evaluated based on whether it can correctly detect the tumor in the images or not. Four parameters to evaluate the performance of the algorithm are defined as follows:

If the applied algorithm can correctly identify the images that contain the tumor, they will be considered as real positives. Images that do not have a tumor and are not correctly detected by the system are also considered negative. The system can also be classified and evaluated according to the following classification:

True Positive (TP): Equals the number of tumor images that the applied algorithm has correctly identified as tumor.

True Negative (TN): Equals the number of healthy images that the applied algorithm has correctly identified as tumorous.

False Positive (PF): Equivalent to the number of tumor images that the applied algorithm has erroneously identified as healthy.

False Negative (FN): Equals the number of healthy images that the applied algorithm has incorrectly identified as tumorous.

With the help of defined parameters (TP, TN, FP, FN), the performance of the proposed algorithm can be evaluated using sensitivity, specificity and accuracy. The results of the algorithm evaluation are presented in Table (1).

$$\text{sensitivity} = \frac{TP}{TP + FN} \times 100 \tag{5}$$

$$\text{specificity} = \frac{TN}{TN + FP} \times 100 \tag{6}$$

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \tag{7}$$

Table 1. *The function of the proposed algorithm.*

Algorithm performance	Evaluation percentage
allergy	93.4%
being special	91%
Accuracy	95%

CONCLUSION

In this paper, using the firefly algorithm, the texture of the image was extracted based on the uniformity of light intensity and structure of each pixel compared to neighbouring pixels. The tumor has a higher brightness in mammographic images. Due to the nature of the firefly algorithm, fireflies are attracted to the area with greater light intensity. Of course, light intensity alone is not the only consideration, and the tumor area should have uniform characteristics in the whole area and be significantly different from the adjacent tissue, which is examined by the morphology algorithm in this article. In studies, the sensitivity and accuracy of tumor diagnosis in breast tissue is about 80 to 90%. The sensitivity and accuracy of the proposed method in tumor diagnosis were evaluated as 93.4% and 95%, respectively, which indicates the high efficiency of the proposed algorithm. According to the evaluation criteria obtained from the proposed algorithm, it can be said that the use of luminosity cream algorithm and morphology can increase the accuracy of breast cancer mass detection.

REFERENCES

[1] Tang, J., Rangayyan, R. M., Xu, J., El Naqa, I., & Yang, Y. (2009). Computer-aided detection and diagnosis of breast cancer with mammography: recent advances. *IEEE transactions on information technology in biomedicine*, *13*(2), 236-251.

[2] Kiarashi, N., Lo, J. Y., Lin, Y., Ikejimba, L. C., Ghate, S. V., Nolte, L. W., ... & Samei, E. (2014). Development and application of a suite of 4-D virtual breast phantoms for optimization and evaluation of breast imaging systems. *IEEE transactions on medical imaging*, *33*(7), 1401-1409.

[3] Zain, N. M., Seriramulu, V. P., Chelliah, K. K., Soin, N., & Arasaratnam, S. (2016, December). Association of bone mineral density and mammographic breast density in premenopausal and postmenopausal Malaysian Women: study on women with and without breast cancer. In *2016 International Conference on Bio-engineering for Smart Technologies (BioSMART)* (pp. 1-4). IEEE.

[4] Mardanian, G., & Behzadfar, N. (2020). A new method for detection of breast cancer in mammography images using a firefly algorithm. *Journal of Intelligent Procedures in Electrical Technology*, *10*(40), 23-32.

[5] Sannasi Chakravarthy, S. R., & Rajaguru, H. (2020). Detection and classification of microcalcification from digital mammograms with firefly algorithm, extreme learning machine and non-linear regression models: A comparison. *International Journal of Imaging Systems and Technology, 30*(1), 126-146.

[6] Hariz, M., Adnan, M., Husain, W., & Rashid, N. A. (2012). Data mining for medical systems: a review. In *Proceedings International Conference on Advances in Computer and Information Technology-ACIT*.

[7] Rafe, V., & Farhoud, R. H. (2013). A survey on data mining approaches in medicine. *International Research Journal of Applied and Basic Science*, *4*(1), 196-202.

[8] Y. Cho, C. Chin, K. Wang, "Based on fuzzy linear discriminant analysis for breast cancer mammography analysis", Proceeding of the IEEE/TAAI, Chung-Li, Taiwan, Noc. 2011 (doi:10.1109/TAAI.2011.18).

[9] Cho, Y. S., Chin, C. L., & Wang, K. C. (2011, November). Based on fuzzy linear discriminant analysis for breast cancer mammography analysis. In *2011 International Conference on Technologies and Applications of Artificial Intelligence* (pp. 57-61). IEEE. (doi:10.1148/radiol.2532090494).

[10] Ferlay, J., Shin, H. R., Bray, F., Forman, D., Mathers, C., & Parkin, D. M. (2010). Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *International journal of cancer*, *127*(12), 2893-2917. (doi:10.1002/ijc. 25516.).

[11] Nasr, S., & Mahmoodian, H. (2019). Insulin drug regulation by general type 2 fuzzy controller with alpha plane. *Journal of Intelligent Procedures in Electrical Technology*, *10*(37), 39-48.

[12] Momeny, M., Gharravi, S., & Hourali, F. (2019). Reducing the impact of SYN flood attacks by improving the accuracy of the PSO algorithm by adaptive effective filters. *Journal of Intelligent Procedures in Electrical Technology*, *10*(37), 51-57.

[13] Shahgholian, G., & Movahedi, A. (2011). Modeling and controller design using ANFIS method for non-linear liquid level system. *International Journal of Information and Electronics Engineering*, *1*(3), 271-275.

[14] Behzadfar, N., Firoozabadi, S. M. P., & Badie, K. (2016). Low-complexity discriminative feature selection from eeg before and after short-term memory task. *Clinical EEG and neuroscience*, *47*(4), 291-297.

[15] Karaboga, D., & Basturk, B. (2007). A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. *Journal of global optimization*, *39*(3), 459-471.

[16] Shahgholian, G., Fazeli-Nejad, S., Moazzami, M., Mahdavian, M., Azadeh, M., Janghorbani, M., & Farazpey, S. (2016, June). Power system oscillations damping by optimal coordinated design between PSS and STATCOM using PSO and ABC algorithms. In *2016 13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)* (pp. 1-6). IEEE.

[17] Dorigo, M., & Stützle, T. (2003). The ant colony optimization metaheuristic: Algorithms, applications, and advances. In *Handbook of metaheuristics* (pp. 250-285). Springer, Boston, MA.

[18] Lotfi Forushani, M., Karimi, B., & Shahgholian, G. (2012). Optimal PID controller tuning for multivariable aircraft longitudinal autopilot based on particle swarm optimization algorithm. *Journal of Intelligent Procedures in Electrical Technology*, *3*(9), 41-50.

[19] Yang, X. S. (2010). A new metaheuristic bat-inspired algorithm. In *Nature inspired cooperative strategies for optimization (NICSO 2010)* (pp. 65-74). Springer, Berlin, Heidelberg.

[20] Gandomi, A. H., & Alavi, A. H. (2012). Krill herd: a new bio-inspired optimization algorithm. *Communications in nonlinear science and numerical simulation*, *17*(12), 4831-4845.

[21] Kao, T. J., Boverman, G., Kim, B. S., Isaacson, D., Saulnier, G. J., Newell, J. C., ... & Kopans, D. B. (2008). Regional admittivity spectra with tomosynthesis images for breast cancer detection: preliminary patient study. *IEEE transactions on medical imaging*, *27*(12), 1762-1768.

[22] Tan, T., Platel, B., Mus, R., Tabar, L., Mann, R. M., & Karssemeijer, N. (2013). Computer-aided detection of cancer in automated 3-D breast ultrasound. *IEEE transactions on medical imaging*, *32*(9), 1698-1706.

[23] Zhou, J., Wu, J., & Chen, H. (2018, November). Immune Multi-Population Firefly Algorithm and its Application in Multimodal Function Optimization. In *2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)* (pp. 319-322). IEEE.

[24] Yang, X. S., & Deb, S. (2009, December). Cuckoo search via Lévy flights. In *2009 World congress on nature & biologically inspired computing (NaBIC)* (pp. 210-214). Ieee.

[25] Dekhici, L., Redjem, R., Belkadi, K., & El Mhamedi, A. (2019). Discretization of the firefly algorithm for home care. *Canadian Journal of Electrical and Computer Engineering*, *42*(1), 20-26.

[26] Yang, X. S. (2009, October). Firefly algorithms for multimodal optimization. In *International symposium on stochastic algorithms* (pp. 169-178). Springer, Berlin, Heidelberg.

[27] Thawkar, S., & Ingolikar, R. (2018). Classification of masses in digital mammograms using firefly based optimization. *International Journal of Image, Graphics and Signal Processing*, *11*(2), 25.

[28] Yang, X. S. (2010). Firefly algorithm, stochastic test functions and design optimisation. *International journal of bio-inspired computation*, *2*(2), 78-84.

[29] Yousefi, M. R. (2017). Magnetic induction tomography: a review of process and medical tomography systems. *Journal of Intelligent Procedures in Electrical Technology*, *8*(31), 33-50.

[30] Kretz, T., Müller, K. R., Schaeffter, T., & Elster, C. (2020). Mammography image quality assurance using deep learning. *IEEE Transactions on Biomedical Engineering*, *67*(12), 3317-3326.

[31] Andreadis, I. I., Spyrou, G. M., & Nikita, K. S. (2014). A CAD $ _ {\bf x} $ Scheme for Mammography Empowered With Topological Information From Clustered Microcalcifications' Atlases. *IEEE journal of biomedical and health informatics*, *19*(1), 166-173.

[32] Milovic, B., & Milovic, M. (2012). Prediction and decision making in health care using data mining. *International Journal of Public Health Science (IJPHS)*, *1*(12), 126-136, 2012.

[33] De Sampaio, W. B., Silva, A. C., de Paiva, A. C., & Gattass, M. (2015). Detection of masses in mammograms with adaption to breast density using genetic algorithm, phylogenetic trees, LBP and SVM. *Expert Systems with Applications*, *42*(22), 8911-8928.

[34] Sampaio, W. B., Diniz, E. M., Silva, A. C., De Paiva, A. C., & Gattass, M. (2011). Detection of masses in mammogram images using CNN, geostatistic functions and SVM. *Computers in biology and medicine*, *41*(8), 653-664.

[35] Nunes, A. P., Silva, A. C., & de Paiva, A. C. (2009, July). Detection of masses in mammographic images using simpson's diversity index in circular regions and svm. In *International Workshop on Machine Learning and Data Mining in Pattern Recognition* (pp. 540-553). Springer, Berlin, Heidelberg.

[36] Ebrahimi, S. M. S., Pourghassem, H., & Keshavarz, A. (2014). Segmentation of melanoma and other pigmented skin lesions in dermoscopic images using fusion of threshoding methods based on reinforcement algorithm. *Journal of Intelligent Procedures in Electrical Technology 4*(16), 37-48.

[37] Javadi, H. R., & Poorghasem, H. (2017). Skin lesion classification from dermoscopy images using color and shape features. *Journal of Intelligent Procedures in Electrical Technology*, *8*(29), 33-40.