Mammogram Image Segmentation using K-means and Optimized Grasshopper Algorithm

Bhawna Utreja[1*], Reecha Sharma[2], Amit Wason[3]
[1,2]Department of Electronics and Communication Engineering, Punjabi University Patiala, India
[3]Ambala College of Engineering and Applied Research, Devsthali, Ambala, India

Abstract – Breast Cancer is serious widespread disease has high death rates among women. Early detection of BC is crucial in improving patient's health. Computer Aided Diagnosis system have been emerged for improving diagnosis accuracy which would avoid needless biopsy and not miss the treatment time. This requires an algorithm that is used for finding breast lesion region i.e. Region of Interest with larger accuracy. However, precise image segmentation has important role in finding lesion region and a challenging problem due to various artifacts. In this paper, clustering based segmentation K-means has been hybridized with grasshopper optimization algorithm for finding Region of Interest. Experimental results show that proposed method has highest accuracy of 91.51% as compared to k-means hybridized with Particle Swarm Optimization and Fruitfly Optimization Algorithm.

Keywords: Computer Aided Diagnosis, Grasshopper Optimization Algorithm, Particle Swarm Optimization, Fruitfly Optimization Algorithm, Region of Interest.

1. Introduction

Women all around the globe are mostly affected by Breast Cancer (BC) and it is the second most rank in causing deaths after lung cancer. Early detection through screening techniques has proven to be the key factor for reducing deaths [1]. Mammography, Ultrasound and Magnetic Resonance Imaging are the most common screening methods used for detecting BC. But the images obtained from these screening methods have shortcomings such as poor resolution, low contrast and blurred images due to noise and acoustic shadowing. This makes false interpretation and the need of unnecessary biopsy occurs which is time consuming and painful procedure. Radiologists predict that a mammography has accuracy over than 90% in its early stages, may fail to see 10-15% of BC. Computer Aided Diagnosis (CAD) systems have emerged for improving diagnosis accuracy of early BC which would prevent unnecessary biopsy and not miss the treatment time [2]. CAD system is generally used by the radiologists as a second opinion for detection of BC [3]. Mammogram image segmentation plays important role in finding breast lesion region i.e. Region of Interest (ROI) helpful in diagnosis of BC [4] [5]. In this paper, automated segmentation K-means hybrid with Grasshopper Optimization Algorithm (GOA) is proposed for locating ROI. The proposed technique consists of two steps: (a) Pre-processing (b) Segmentation

This paper is organized in the following sections: Section 2 presents the research background related to BC and segmentation techniques used. Section 3 presents the research methodology in which dataset, pre-processing, and segmentation has been done. Section 4 illustrates the experimental results and its discussion. Section 5 reports the conclusion.

2. Related Work

Chowdhary et al. [6] proposed novel on integration of intuitionistic possibilistic fuzzy c-mean (IPFCM) clustering System with possibilistic c-mean algorithm (PFCM). This clustering approach take advantages of PFCM that reduces effect of noise, conquer the problem of coincident cluster and less sensitivity to outlier. Experimental results show that proposed technique has tested on MIAS database and achieved high efficiency.

Punitha et al. [7] developed an automated system for early detection of breast masses which help radiologist for correct diagnosis. Gaussian filtering was used at preprocessing stage and then ROI extracted by using Dragon Fly optimized region growing technique. After segmentation, GLCM and GLRLM methods were used for texture feature extraction and after that fed to Feed-forward Neural network trained with Levenberg Marquardt back propagation algorithm. The performance of proposed segmentation was measured by parameter Jaccard Index to be 90%. The proposed system was applied on DDSM database and has overall accuracy of 98%.

Parvathavarthini et al. [8] proposed Intuitionistic fuzzy clustering with optimization algorithm crow search approach with neighbourhood attraction for finding ROI. Optimization technique crow search approach efficiently find optimal value of global centroid and ROI i.e. breast masses based on intensity levels. The proposed method was compared with PSO-IFCM-NA and outperforms better in terms of segmentation of regions. This method gave better results in terms of Jaccard index and Dice index value in excess of 96% and 98% that help radiologist in selecting the disease affected area for right treatment of patient.

Arjmand et al. [9] proposed a novel segmentation method i.e. combination of k-means and cukoo search optimization for finding masses in MRI breast images. This method implemented on RIDER dataset and outperforms than existing methods i.e. FCM and K-means respectively.

Punithavathi et al. [10] proposed an improved Gray Level Co-occurrence Matrix (GLCM) feature-based extraction method for detection of breast lesion. This improved GLCM method extracted texture as well as tamura features mined from segmentation method i.e. Optimized Kernel Fuzzy Clustering Algorithm (OKFCA). Tamura features were also called texture directionality; extracted in three steps namely calculation of Gradient, obtaining histogram direction gradient and lastly sharpness of the histogram. This method trained the features that will be used for further processing and proved to be best in texture features extraction for MIAS database.

Guo et al. [11] presented automated segmentation of pectoral muscle region in blurred mammogram images. This method identified boundary as well as entire shape of pectoral muscle. However, pectoral muscle and gland tissue was similar in intensity and texture, therefore trained deep neural network was used to differentiate for identification of boundary. The boundary with high confidence can be found out by uniformity of predictions from multiple converged models. Generative adversarial network (GAN) was used for predicting entire shape of pectoral muscle. Experimental results show that this method estimate boundary of pectoral muscle for blurred boundaries and improved segmentation performance.

Sha et al. [12] proposed automatic method based on deep learning and GOA for locating ROI for mammogram images. Median filter has been used for noise reduction followed by convolutional neural network used with GOA for finding optimal affected area. Extracted features are based on geometric, textures and statistical features and selection of relevant extracted feature has been done by using GOA. The prosed method has been implemented on MIAS and DDSM dataset and perform better as compared to other methods.

Ali et al. [13] developed a fully automated pectoral muscle segmentation algorithm for MLO views in mammogram images. First step consists of noise removal and sharpening of images. In second step fully convolutional neural network improved with residual connections applied for segmentation. This algorithm achieved better results as compared to U-Net-based architecture on datasets MIAS, INBREAST, and DDSM. Finally, connected component analysis used to remove the wrongly predicted pixels and canny edge detection applied for finding actual boundary of pectoral muscle.

Fang et al. [14] proposed a multilayer perceptron neural network based on whale optimization algorithm (WOA) for BC detection. This method simulated in MIAS and DDSM datasets and has better performance in terms of accuracy etc.

Navneet et al. [15] presented a new approach DLHO by integrating Harris Hawk Optimization (HHO) with dimension learning hunting (DLH) strategy for BC detection. Its aim is to remove weakness of HHO such as disparity in exploration and exploitation, early convergence etc. Proposed approach has been analysed on different datasets and outcomes are in favour of it.

3. Proposed Method

The Proposed methodology consists of four different stages: Image acquisition, image preprocessing, segmentation and optimization techniques as discussed here under:

3.1 Pre processing

It is difficult for analysing images obtained from dataset as there is noise present in mammogram images. Therefore, Preprocessing has been used for removing the noise, as well as for undesired label and any discrepancies in the image's intensity [17]. Two filters namely; median and gaussian filter are used for decreasing noise and traces present in images as well as for maintaining sharp edges and original features of image. Afterwards, Intensity normalization and CLAHE technique has been used for improving contrast of image.

The flow chart of proposed methodology is shown in Fig 1.

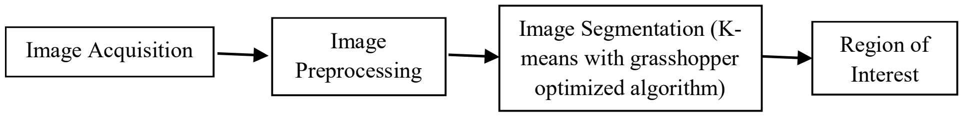

Fig 1. Proposed Methodology for Breast Cancer Segmentation

3.2 Segmentation

Third step being, Segmentation is process of find the ROI i.e. lesion region from mammogram image. In mammogram images, pectoral muscle in top of image has same intensity as that of breast tissue. Pectoral tissue is not a part of breast tissue and therefore excluded from the analysis. Therefore, it is necessary for correctly identify the lesion i.e. ROI for diagnosing breast cancer. There are many techniques used for segmentation such as threshold, graph based, active contour based etc. This suggested breast lesion segmentation has two phases; k-means technique for localizing precise affected region of breast and further k-means optimized with grasshopper algorithm for obtaining maximum accuracy.

3.2.1 K-means Segmentation

K-means algorithm is unsupervised learning technique used for partitioning image into number of clusters [18]. Steps of k-means algorithm are as follows:

1. Pick K points as cluster centres randomly in object space.
2. Allocate each pixel in image to nearest cluster having minimum distance to the cluster centre.
3. Recalculate cluster centre by taking mean of all pixels in cluster.
4. Repeat steps 2 and 3 until cluster centres no longer change or no pixel change cluster.

This algorithm aims at minimizing an objective function V, a squared error function given by:

$$V = \sum_{i=1}^{k} \sum_{j=1}^{l} \left\| x_i - c_j \right\|^2 \tag{1}$$

Where $\left\| x_i - c_j \right\|^2$ shows distance between a data point xi (pixel) and cluster center cj. This technique automatically gives the information of pixels for recognizing cluster in the form of breast lesion region.

3.2.2 K-means with optimized grasshopper algorithm

GOA is swarm-based optimization method that simulates grasshopper's mimicking behaviour and their social interaction in nature. The position of individually grasshopper in the swarm signifies a probable solution of a specified optimization problem. Position of each grasshopper is built on three forces: S_i social interaction between it and the other grasshoppers, G_i gravity force on it and the A_i wind advection [19].
Mathematical model used for simulating swarming behaviour of grasshoppers is given by [2]:
Position of ith grasshopper is given as:

$$X_i = S_i + G_i + A_i \tag{2}$$

Where S_i is social interaction, G_i is gravity force on ith grasshopper and A_i is wind advection.
Random behaviour of grasshopper is given by:

$$X_i = r_1 S_i + r_2 G_i + r_3 A_i \tag{3}$$

Where r_1, r_2, r_3 are random numbers in [0, 1].
Social interaction force S_i is calculated as:

$$S_i = \sum_{j=1, i\neq j}^{N} S\left(d_{ij}\right) \underset{d_{ij}}{\rightarrow} \tag{4}$$

Where d_{ij} distance between *i*th and *j*th grasshopper is:

$$d_{ij} = \left|x_j - x_i\right| \tag{5}$$

And $\underset{d_{ij}}{\rightarrow}$ is unit vector from i_{th} to j_{th} grasshopper is:

$$\underset{d_{ij}}{\rightarrow} = \frac{x_j - x_i}{d_{ij}} \tag{6}$$

Function S defines social forces is evaluated as

$$S(r) = fe^{\frac{-r}{l}} - e^{-r} \tag{7}$$

Where *f* is intensity of attraction and *l* is attractive length scale. The distance between grasshoppers is normalized to [1, 4].

Gravity force on i_{th} grasshopper is evaluated as:

$$G_i = -g \underset{e_g}{\rightarrow} \tag{8}$$

Where g is gravitational force and $\underset{e_g}{\rightarrow}$ is unity vector close to center of earth.

Wind advection component is evaluated as:

$$A_i = u \underset{e_w}{\rightarrow} \tag{9}$$

Where u is constant drift and $\underset{e_w}{\rightarrow}$ is unity vector in direction of wind.

Equation can be re-written by putting all components as:

$$X_i = \sum_{j=1, i\neq j}^{N} s\left(\left|x_j - x_i\right|\right) \frac{x_j - x_i}{d_{ij}} - g \underset{e_g}{\rightarrow} + u \underset{e_w}{\rightarrow} \tag{10}$$

Where N is number of grasshoppers.

A modified solution of equation (2) is given by:

$$X_i^d = c\left\{\sum_{j=1, j\neq i}^{N} c\, \frac{ub_d - lb_d}{2}\, s\left(\left|X_j^d - X_i^d\right|\right) \frac{X_j - X_i}{d_{ij}}\right\} + \underset{T_d}{\rightarrow} \tag{11}$$

The general pseudocode for the GOA algorithm is as follows:

The general pseudo code for GOA

1. Initializing: Generate the initial swarm such as highest c_{max} and lowest c_{min} reducing factor, and the number of simulation rounds.
2. Evaluate the function considering the agents in the swarm
3. Best_Sol (BS) = the best solution (agent)
4. For (d=1: maximum number of simulations)
   - Normalize the distance between grasshoppers in the interval [1, 4].
   - Update the position using the equation 11.
   - Apply the relevant constraints
   - Update the BS if there is any better solution.
5. $T_{itr} = itr + 1$
6. End for
7. Return BS

As shown in Fig. 2, input image is passed to the system aided design where it pre-processed using the morphological operators that is shown in (b) part of the image. After the application of k-means with GOA, (c) is generated and to find the best suitable patches from each region, masking has been done and as a result (d) and finally segmented image (e) are obtained.

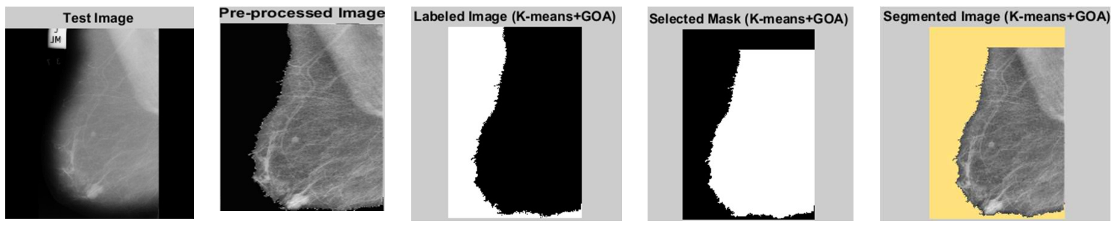

Fig. 2 Images attained in various steps of proposed mammogram segmentation technique (a) Input Image (b) Pre-processed Image (c) Region after applying k-means with GOA (d) Binary Image Mask using K-means with GOA (e) Segmented Image.

4. Results and Discussion

In this section, investigation of different performance metrics has been done for finding the strength of proposed approach. The performance of proposed algorithm is evaluated through various performance metrics such as accuracy, specificity, sensitivity, dice, Jaccard index etc.

Accuracy (A) can be assessed by calculating the closeness of proposed algorithm segmented result and ground truth results using Eq. (12). Sensitivity (α) is percentage of data points correctly segmented estimated by Eq. (13) and specificity (β) find percentage of negative data points correctly removed given by Eq. (14). Dice (D) measures how far the spatial overlap exists between two images is given by Eq. (15). A lower value indicates less overlapping while a value closer to one indicates perfect agreement. Jaccard Index (JI) measures similarity between segmentation and ground truth results by using Eq. (16). Matthews Correlation Coefficient (MCC) is important decisive factor for measuring the performance of proposed algorithm estimated using Eq. (17).

$$A = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

$$\alpha = \frac{TP}{TP + FN} \tag{13}$$

$$\beta = \frac{TN}{TN + FP} \tag{14}$$

$$D = \frac{2 * TP}{2 * TP + FP + FN} \tag{15}$$

$$JI = \frac{TP}{FP + TP + FN} \tag{16}$$

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{17}$$

Where TP is True Positive pixels correctly classified; TN is True Negative means normal pixels correctly unclassified as tumour pixels; FP is False Positive means normal image pixels wrongly classified as tumour pixels; and FN is False Negative means tumour pixels wrongly unclassified as tumour pixels respectively.

Table 1 represents the comparative results of various parameters of k-means with different optimization algorithms i.e., Particle Swarm Optimization (PSO), Fruit Fly Optimization (FOA) and Grasshopper Optimization Algorithms (GOA).

The average value of Specificity of the proposed method is more but less sensitivity as compare to other algorithms.

Table 1 Results of proposed technique using MIAS dataset

Parameters	**Results of pre-processed sample input image 1**				**Results of pre-processed sample input image 2**			
	K-mean	K-mean & PSO	K-mean & FOA	Proposed method (k-mean & GOA)	K-mean	K-mean & PSO	K-mean & FOA	Proposed method (k-mean & GOA)
Accuracy	89.77%	87.88%	88.51%	91.51%	94.07%	93.42%	93.77%	95.26%
Specificity	0.78	0.74	0.75	0.84	0.90	0.89	0.90	0.93
Sensitivity	0.99	0.99	0.99	0.97	1.00	0.99	0.99	0.98
F-measure	0.913	0.899	0.904	0.925	0.927	0.920	0.924	0.940
Precision	0.84	0.82	0.83	0.88	0.86	0.85	0.86	0.90
MCC	0.81	0.77	0.79	0.83	0.88	0.87	0.88	0.90
Dice	0.91	0.90	0.90	0.93	0.93	0.92	0.92	0.94
JI	0.84	0.82	0.82	0.86	0.86	0.85	0.86	0.89
Time Complexity (seconds)	2.69	13.79	5.03	48.55	0.91	3.08	0.73	4.83

From the Table 1, it is clear that the realized values of the dice and jacquard Index are better for GOA showing the quality of breast region i.e. Region of Interest (ROI). The execution time for proposed technique is more but has high accuracy as compare to other methods.

Fig. 3 and Fig. 4 represent the comparative analysis of different optimization techniques such as PSO, FOA and GOA with K-means segmentation technique for accuracy and MCC.

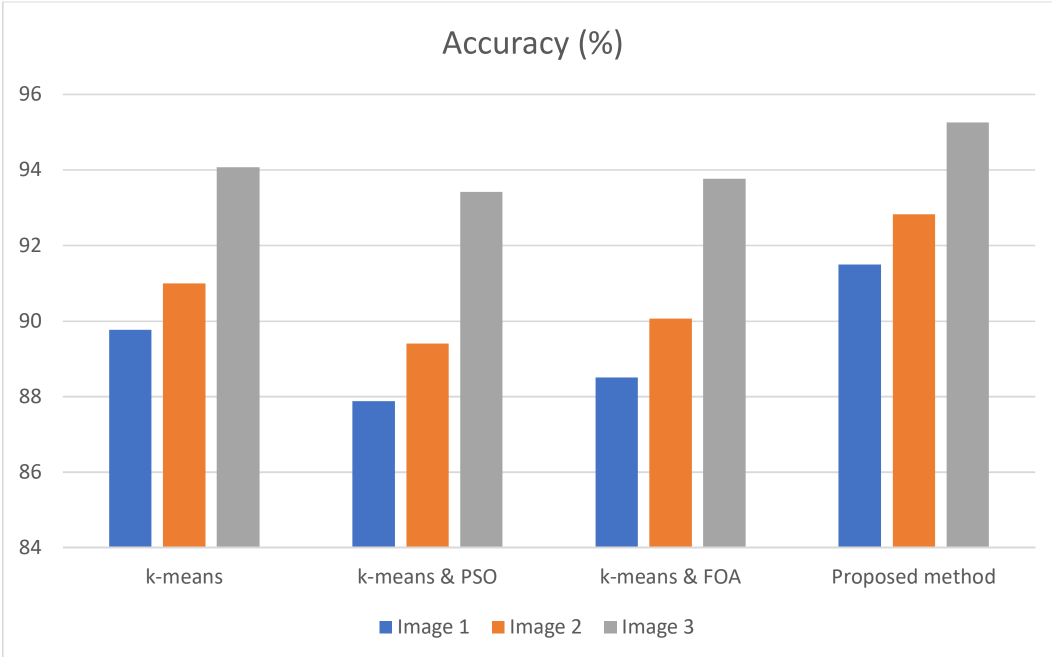

Fig. 3 Comparative Analysis for segmentation Accuracy

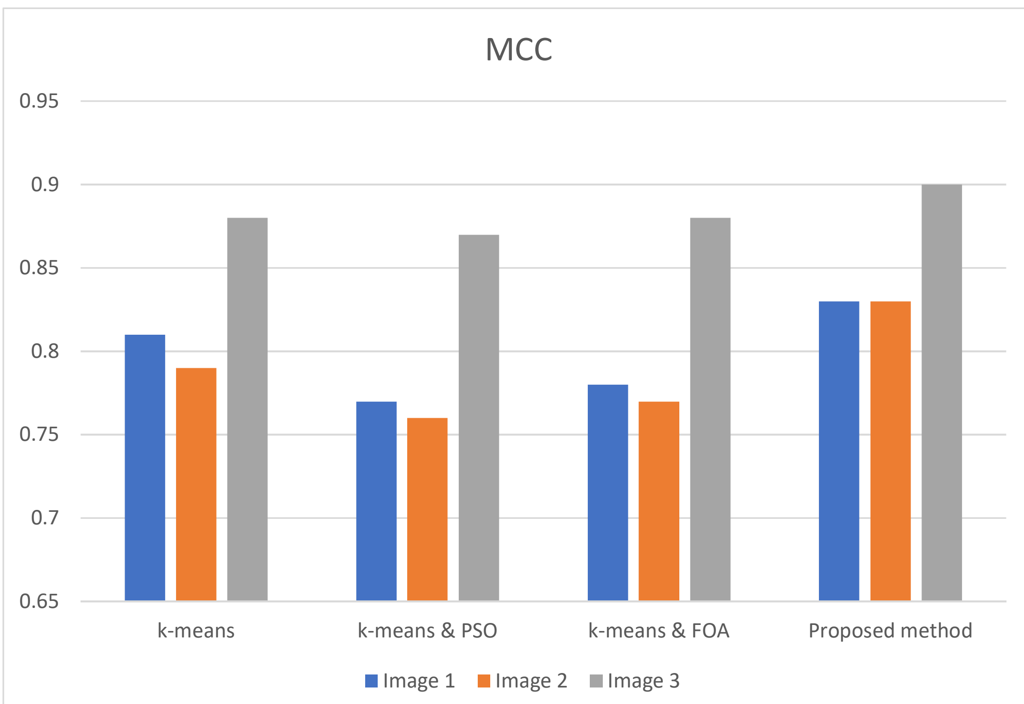

Fig. 4 Comparative Analysis of MCC with different techniques

5. Conclusion

In this study, an automated method k-means hybridize with grasshopper optimization algorithm is presented for mammogram image segmentation. The proposed method has been implemented on MIAS datasets comprises of two stages: pre-processing and segmentation. Pre-processing has been done for decreasing noise and improving quality of original dataset's image. The presented method has been compared with k-means, K-means hybridize with FOA and PSO respectively. From results it is clear that GOA performs better than other algorithms. The presented method GOA has accuracy 91.51% as compared to PSO 87.88% and FOA 88.51% respectively.

References

[1] Blanks, RG, Wallis, MG and Given-Wilson, "Observer variability in cancer detection during routine repeat (incident) mammographic screening in a study of two versus one view mammography", Journal of Medical Screening, Vol. 6, no. 3, pp. 152-158, 1999.

[2] Cheng, HD, Cai, X, Chen, X, Hu, L and Lou, X, "Computer aided detection and classification of microcalcifications in mammograms: a survey", Pattern Recognition, Vol. 36, no. 12, pp. 2967-2991, 2003.

[3] Eltoukhy, MM, Faye, I & Samir BB, "A statistical based feature extraction method for breast cancer diagnosis in digital mammogram using multiresolution representation", Computers in Biology and Medicine, Vol. 42, pp. 123–128, 2012.

[4] Ichikawa, T, Matsubara, T, Hara, T, Fujita, H, Endo, T & Iwase, "Automated detection methods for architectural distortion areas on mammograms based on morphological processing and surface analysis", in SPIE Medical Imaging 2004: Image processing, Proceedings of SPIE, conference, ed. JM. Fitzpatrick and M. Sonka, San Diego, CA, pp. 920-925, 2004.

[5] J. Virmani and R. Agarwal, "Effect of despeckle filtering on classification of breast tumors using ultrasound images", Biocybernetics and Biomedical Engineering, 2019.

[6] Chowdhary, C. L., & Acharjya, D. P., "Segmentation of mammograms using a novel intuitionistic possibilistic fuzzy c-mean clustering algorithm", In Nature Inspired Computing (pp. 75-82). Springer, Singapore, 2018.

[7] Punitha, S., Amuthan, A., & Joseph, K. S., "Benign and malignant breast cancer segmentation using optimized region growing technique", Future Computing and Informatics Journal, Vol. 3, No. 2, pp. 348-358, 2018.

[8] Parvathavarthini, S., Karthikeyani Visalakshi, N., & Shanthi, S., "Breast cancer detection using crow search optimization based Intuitionistic fuzzy clustering with neighborhood attraction. Asian Pacific journal of cancer prevention: APJCP, Vol. 20, No.1, pp. 157, 2019.

[9] Arjmand, Amir, Saeed Meshgini, Reza Afrouzian, and Ali Farzamnia. "Breast tumor segmentation using K-means clustering and cuckoo search optimization." In 2019 9th International conference on computer and knowledge engineering (ICCKE), pp. 305-308. IEEE, 2019.

[10] Punithavathi, V., & Devakumari, D., "Detection of Breast Lesion Using Improved GLCM Feature Based Extraction in Mammogram Images", Studies in Indian Place Names, Vol. 40, No. 70, pp. 358-366, 2020.

[11] Guo, Y., Zhao, W., Li, S., Zhang, Y., & Lu, Y., "Automatic segmentation of the pectoral muscle based on boundary identification and shape prediction", Physics in Medicine & Biology, Vol. 65, No. 4, pp. 045016, 2020.

[12] Sha, Zijun, Lin Hu, and Babak Daneshvar Rouyendegh, "Deep learning and optimization algorithms for automatic breast cancer detection", International Journal of Imaging Systems and Technology, Vol. 30, No. 2, pp. 495-506, 2020.

[13] Ali, M. J., Raza, B., Shahid, A. R., Mahmood, F., Yousuf, M. A., Dar, A. H., & Iqbal, U., "Enhancing breast pectoral muscle segmentation performance by using skip connections in fully convolutional network", International Journal of Imaging Systems and Technology, Vol. 30, No. 4, pp. 1108-1118, 2020.

[14] Fang, Hong, Hongyu Fan, Shan Lin, Zhang Qing, and Fatima Rashid Sheykhahmad, "Automatic breast cancer detection based on optimized neural network using whale optimization algorithm", International Journal of Imaging Systems and Technology, Vol. 31, No. 1, pp. 425-438, 2021.

[15] Kaur, Navneet, Lakhwinder Kaur, and Sikander Singh Cheema, "An enhanced version of Harris Hawks optimization by dimension learning-based hunting for breast cancer detection", Scientific reports, Vol. 11, No. 1, 2021.

[16] https://www.kaggle.com/datasets/ambarish/breakhis.

[17] K. D Marcomini, A. A. Carneiro, and H. Schiabel, " Application of artificial neural network models in segmentation and classification of nodules in breast ultrasound digital images." International journal of biomedical imaging, 2016.

[18] Vivona, Letizia, Donato Cascio, Francesco Fauci, and Giuseppe Raso, "Fuzzy technique for microcalcifications clustering in digital mammograms", BMC Medical Imaging, Vol 14, no. 1, pp. 23, 2014.

[19] Mirjalili, Seyedeh Zahra, Seyedali Mirjalili, Shahrzad Saremi, Hossam Faris, and Ibrahim Aljarah, "Grasshopper optimization algorithm for multi-objective optimization problems", Applied Intelligence, Vol 48, no. 4, pp. 805-820, 2018.