RESEARCH ARTICLE

Breast Cancer Detection using Crow Search Optimization based Intuitionistic Fuzzy Clustering with Neighborhood Attraction

Parvathavarthini S[1]*, Karthikeyani Visalakshi N[2], Shanthi S[3]

Abstract

  Objective: Generally, medical images contain lots of noise that may lead to uncertainty in diagnosing the abnormalities. Computer aided diagnosis systems offer a support to the radiologists in identifying the disease affected area. In mammographic images, some normal tissues may appear to be similar to masses and it is tedious to differentiate them. Therefore, this paper presents a novel framework for the detection of mammographic masses that leads to early diagnosis of breast cancer. **Methods:** This work proposes a Crow search optimization based Intuitionistic fuzzy clustering approach with neighborhood attraction (CrSA-IFCM-NA) for identifying the region of interest. First order moments were extracted from preprocessed images. These features were given as input to the Intuitionistic fuzzy clustering algorithm. Instead of randomly selecting the initial centroids, crow search optimization technique is applied to choose the best initial centroid and the masses are separated. Experiments are conducted over the images taken from the Mammographic Image Analysis Society (mini-MIAS) database. **Results:** CrSA-IFCM-NA effectively separated the masses from mammogram images and proved to have good results in terms of cluster validity indices indicating the clear segmentation of the regions. **Conclusion:** The experimental results show that the accuracy of the proposed method proves to be encouraging for detection of masses. Thus, it provides a better assistance to the radiologists in diagnosing breast cancer at an early stage.

Keywords: Image segmentation- neighborhood attraction- intuitionistic fuzzy C-means clustering- mammogram images

Introduction

  Cancer has been a huge threat to human life. It is the uncontrolled growth of cells that can spread over the organs of the body. Among all types of cancers, breast cancer is the most fatal disease that has been a great menace to middle-aged women. But the consoling fact is that there are several possibilities for early detection and diagnosis of breast cancer that improve the long term survival rate. Computer Aided Diagnosis (CAD) plays a vital role in the detection, examination and follow-up procedures of a patient.

  The challenge before radiologists is that there are huge volumes of images that are tedious to handle manually. This leads to the difficulty in interpretation of the results. Several imaging modalities like Computed Tomography, Magnetic Resonance Imaging, Ultrasound and Mammography are used for early screening of cancer. Therefore, an automatic detection of breast masses in the digital mammogram has become the need of the hour. They support radiologists in identifying the disease. Quantification of breast features is the essential task to be done by a mammography based CAD system.

Sometimes, the abnormalities are hidden within the thick breast tissue and there may be overlapping of benign and malignant masses. These problems highlight medical imaging analysis as an open issue in the research world. The availability of an automated tool facilitates the radiologist to concentrate on these suspicious areas. This work provides the tool to identify masses.

  Image segmentation attempts to partition an image into numerous non-overlapping areas, each with the homogeneous properties. There are several methods for performing segmentation such as histogram based methods, region growing, thresholding, clustering and template matching. This work employs Intuitionistic fuzzy clustering for segmenting the mammogram images to identify the region of interest.

  Intuitionistic Fuzzy Sets (IFS) are a subset of fuzzy sets. Fuzzy sets are proposed by Zadeh (1965) and are represented as follows

$$fs = \{< x, \mu_{fs}(x) >| \ x \in X\}$$

  where μfs: X → [0, 1] denotes the membership and vfs: X → [0, 1] denotes the non-membership and $v_{fs}(x) = 1 - \mu_{fs}(x)$.

*[1]Department of Computer Technology, [3]Department of Computer Applications, Kongu Engineering College, Perundurai, [2]Department of Computer Science, Government Arts and Science College, Kangeyam, Tamilnadu, India. *For Correspondence: varthinis@gmail.com*

Intuitionistic fuzzy set (Atanassov, 2003) is defined as a triplet that deals with membership, non- membership and a powerful third parameter called indeterminacy or hesitancy for estimating the uncertainty. An Intuitionistic fuzzy set is given as

$$ifs = \{< x, \mu_{ifs}(x), \nu_{ifs}(x) >| x \in X\}$$

and indeterminacy can be computed as

$$\pi_{ifs}(x) = 1 - \mu_{ifs}(x) - \nu_{ifs}(x)$$

such that $0 < \mu_{ifs}(x) + \nu_{ifs}(x) < 1$ where π_{ifs} is the hesitancy value.

The vagueness and ambiguities present in the medical images make them suitable for representing them as an IFS. Intuitionistic fuzzy clustering has been the favorite choice of researchers who work in the field of image processing. There are only a handful of works in the area of hybrid intuitionistic fuzzy clustering algorithms. However, the application of intuitionistic fuzzy C-Means clustering to images has several notable references in the literature.

Vlachos and Sergiadis (2005) considered the intensity values of the pixels to represent them as an IFS. Jawahar and Ray (1996) modeled an IFS as a fuzzy histogram by representing the inherent quantization error as the hesitancy degree. Pelekis et al., (2007) modified FCM to cluster intuitionistic fuzzy data and discriminated the images from historical and art databases into different clusters.

Tumor/Haemorrhage detection using IFCM is done by Chaira and Anand (2011). Histogram thresholding is done to remove the noisy pixels and edge detection is performed to identify the tumor region. Chaira (2011) used Yager intuitionistic fuzzy generator and modified the objective function of IFCM to incorporate hesitancy values also for detecting tumors in CT brain images. Chaira (2010) also proposed another method of IFCM using Sugeno intuitionistic fuzzy generator to extract color regions from remote sensing images. Chaira (2014) explored Type II fuzzy set to enhance medical images by constructing Hamacher T conorm to get a clear image without noise. Huang et al., (2015) improved Chaira's multiobjective criterion function to incorporate neighborhood attraction information and hybridized it with GA to enhance the performance of the segmentation method. Son et al., (2012) introduced intuitionistic possibilistic fuzzy geographic weighted clustering and proved its efficiency by executing it over real datasets. Ananthi et al., (2016) segmented gray scale images using entropy and the value that minimizes entropy is taken as the threshold to segment the image. Binu (2015) worked on several hybrid approaches like PSO, genetic algorithm and cuckoo search algorithms to find that PSO suits well for large scale datasets. Balasubramaniam and Ananthi (2016) segmented crop images to find the nutrition deficiency and the membership matrix shows the deficiency regions that occur due to lack of minerals like zinc, nitrogen etc.

Kuo et al., (2018) developed hybrid kernel intuitionistic fuzzy C-Means algorithm for the purpose of analyzing clusters. The kernel IFCM algorithm is combined with Particle Swarm Optimization, Genetic Algorithm and Artificial Bee Colony optimization techniques and the results are tested over six benchmark datasets.

Zhao et al., (2018) introduced a multiobjective evolutionary IFCM with multiple image spatial information to perform segmentation of synthetic Berkeley and magnetic resonance images. Cordeiro et al., (2016) proposed a semi-supervised growcut algorithm that utilizes fuzzy Gaussian membership functions to segment and classify the region of interest from mammograms.

Pavan et al., (2017) estimated breast density using post-processed digital mammograms and utilized an optimized FCM for classifying fibroglandular tissue in mammograms. Dhahbi et al., (2015) deals with feature extraction using the curvelet transforms and used k-nearest neighbor to classify the tumors as malignant or benign. Shi et al., (2018) estimated skin-air boundary using a gradient weight map and detected pectoral region unsupervised pixelwise labeling and used texture filter to detect calcifications. Shanthi and Bhaskaran (2013) used discrete wavelet transform and multiscale surrounding region dependence matrix computation to detect breast cancer from benchmark and real datasets.

Yang (2017) proposed progressive support pixel correlation statistical method to segment medical images. Anter and Hassenian (2018) segmented liver tumor from CT images using a fast FCM and PSO to enhance optimize cluster results. Keller et al., (2011) proposed an adaptive FCM by utilizing the tissue properties and linear discriminant analysis to segment parenchymal tissue in digital mammography. Kontos and Maragoudakis (2013) used GA to reduce the features and incorporated a hybrid boosting and genetic subsampling to segment ROI from mammogram images.

Possibilistic FCM is used by Vega-Corona et al., (2011) to find more or less homogeneous regions in the image. Mean and standard deviation features are extracted from the image and fed as input to the neural network to identify micro-calcifications. Shanthi and Baskaran (2011) extracted the region of interest that are susceptible masses from a mammographic image and finally classified the mass as benign or malignant. Parvathavarthini et al., (2017) combined cuckoo search with IFCM and clustered benchmark datasets from UCI. Cluster indices are computed for analyzing the results and proved to be efficient. An application of Intuitionistic fuzzy PSO to medical datasets is proposed by Parvathavarthini et al., (2017). Parvathavarthini et al., (2018) hybridized FCM and IFCM with crow search optimization resulting in very low error rates for benchmark datasets.

Materials and Methods

Intuitionistic Fuzzy C-Means Clustering with Neighborhood Attraction (IFCM-NA)

Intuitionistic Fuzzy C-Means (IFCM) algorithms are proposed by various researchers. Shen et al., (2005) proposed FCM based on neighborhood attraction and it is utilized by Huang et al., (2015) to modify the IFCM algorithm proposed by Chaira.

Pseudocode for IFCM Clustering with Neighborhood Attraction

Read the image
Apply preprocessing to remove background and pectoral region
Convert data into IFS representation
Extract first order moments from the image
Find the appropriate lambda value that maximizes entropy
Select initial seeds randomly
Repeat

Calculate membership using

$$U_{ij} = \frac{1}{\sum_{r=1}^{C}\left(\dfrac{dis(A_j,V_i)}{dis(A_j,V_r)}\right)^{\frac{2}{m-1}}}, 1 \le i \le C, 1 \le j \le n, m = 2$$

Compute neighborhood tuning matrix

$$NbU_{ij} = U_{ij}^{*} / \left(U_{ij}^{*} + \sum_{k \in S} U_{jk}^{*} \right)$$

Estimate hesitancy using $u_{ik}^{*} = u_{ik} + \pi_{ik}$

Calculate $J_2 = \sum_{i=1}^{C} \pi_i^{*} e^{1-\pi_i^{*}}$

where $\pi_i^{*} = \dfrac{1}{L}\sum_{j=1}^{n}\pi_{ij}, j \in [1,L]$

Update centroids
Compute fitness using

$$J = \sum_{i=1}^{C}\sum_{j=1}^{n} NbU_{ij}^{m} \| A_j - V_i \| + \sum_{i=1}^{C} \pi_i^{*} e^{1-\pi_i^{*}}$$

Until Convergence is reached

Crow Search Optimization algorithm
There are a vast variety of nature inspired metaheuristic algorithms that can simulate the activities of living beings from the nature. Generally, real world problems are nonlinear in nature. Optimization techniques try to find the optimal solution from a set of feasible solutions. In case of clustering problems, the objective function is to be minimized subject to certain constraints. The main issue with the clustering algorithms is that they tend to find only local optimal results. The role of an optimization algorithm is to discover global optimal solutions.

Crow search algorithm was proposed by Askarzadeh (2016) by observing the behavior of crows. Crows are intelligent birds that can memorize faces and the places where they store food. A flock of crows exhibit some similarities in their behavior pattern. They follow each other to acquire best food sources. The crows search the environment (search space) for a feasible solution (each position in the environment) and the best food source is the global solution. The quality of the food source serves as the fitness function for the algorithm.

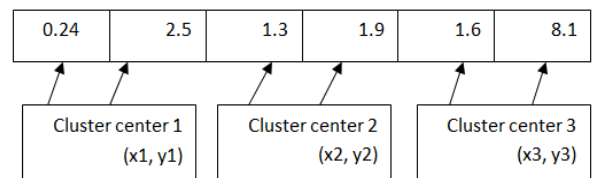An optimization algorithm is characterized by two

Figure 1. Encoding of a Crow in the Initial Population

main factors namely intensification and diversion. The parameter named Awareness Probability (AP) controls these factors and achieves a balance between these two. Intensification denotes the regions that are thoroughly searched so as to find better solutions. The unexplored regions that must be visited in the search space are given by diversification.

Solution Encoding
In order to represent the dataset as a crow, encoding is to be done. The value of each particle is encoded into a string sequence constructed by a set of real values. If there are n data points and they are to be grouped into C clusters, the cluster centers can be combined as a string to represent every single crow. If there are d dimensions in the data, the length of each crow is C×d words. The initial population is randomly generated and it represents the vector of different cluster centers. For example, let d=2, C=3, then the string of a crow is given by {(0.24, 2.5), (1.3, 1.9), (1.6, 8.1)}. This can be pictorially represented as in Figure 1.

Let N be the population size and $x^{i,iter}$ denotes the position of the crow i at the iteration iter. The crows are capable of remembering their hiding places and the best position attained by crow i so far is given by $m^{i,iter}$. The pseudocode for crow search algorithm is given below.

Pseudocode for Crow Search Algorithm

Initialize flock size N, number of iterations maxiter, flight length and awareness probability
Initialize a random position for all the crows in population
Initialize the memory same as that of position as crows do not have any experience currently
Estimate the fitness function
While iter < maxiter
for i=1:N
Generate a new position by choosing crow j and its follower crow i
If crow j is not aware of follower crow i

$$x^{i,iter+1} = x^{i,iter} + r_i . fl^{i,iter} . (m^{i,iter} - x^{i,iter})$$

Else
    A random position is chosen to fool crow i
End for
Check feasibility of new positions
Estimate the fitness function of new positions
If fitness of new position > fitness of memorized position
    Update the memory $m^{i,iter+1} = x^{i,iter+1}$
End while

Crow Search based Intuitionistic Fuzzy C-Means Clustering with Neighborhood Attraction (CrSA-IFCM-NA)

This is a hybrid approach that facilitates the extraction of ROI efficiently. The methodology utilizes the most exclusive property of IFS that is the ability to deal with uncertainty by means of hesitancy factor. The competence of the algorithm further increases with the implementation of neighborhood attraction that is dependent on the relative location and features of neighboring pixels and this improves the performance of the segmentation process. In addition to this, the optimization technique serves as the best way to reach global optimal results and to obtain best initial centroids. The workflow of the proposed method is shown in Figure 2.

The image obtained from the database is preprocessed so as to remove the unwanted background portion and the pectoral region is also removed. After removing the necessary artifacts, the image is converted into intuitionistic fuzzy representation. As a result, each pixel in the image is now represented as a triplet with membership, non-membership and hesitancy factors.

In order to perform this conversion, Yager method of intuitionistic fuzzification based on fuzzy complement is employed. This is denoted by

$$\mu_A(d_j; \lambda) = 1 - (1 - \mu_A(d_j))^{\lambda} \qquad (1)$$

$$v_A(d_j; \lambda) = 1 - (1 - \mu_A(d_j))^{\lambda(\lambda+1)} \qquad (2)$$

where $\lambda \in [0,1]$ and hesitancy is found by summing up membership and non-membership values and subtracting the result from one. To find the appropriate value for lambda, entropy is calculated for all lambda values in the range 0 to 1. The value that maximizes the entropy is taken for each image.

The formula for finding entropy is given by

$$IFE = \frac{1}{U \times V} \sum_{i=0}^{L-1} \sum_{j=0}^{P-1} \frac{2\mu_i(d_j)v_i(d_j) + \pi_i^2(d_j)}{\pi_i^2(d_j) + \mu_i^2(d_j) + v_i^2(d_j)} \quad (3)$$

After the fuzzification process is over, the features are extracted from the image. A 5x5 window is moved over the image matrix to obtain the features. Statistical features that include first-order moments like mean, median, mode, standard deviation and kurtosis are obtained from the image and passed on to the clustering algorithm as input.

Segmentation of the image is performed using the proposed clustering method and various regions in the images are clustered based on the intensity values of the pixels. The population size (no. of crows), number of clusters and the maximum number of iterations are initialized. The two problem specific parameters like flight length and awareness probability are assigned. In this work, the initial centroid values are taken as crows, every position of the crow is considered as a feasible solution and the objective function of the clustering algorithm determines the fitness of the centroids.

Initially, the position and memory matrices take same random values as the crows do not have any experience and they are assumed to hide their food at the initial positions. Generally, the membership value of IFCM clustering is determined by the distance measure or similarity measure. The difference in the intensities of a pixel and the cluster center is considered for finding the similarity and this method has less resistance to noise. Therefore the neighborhood attraction method proposed by Shen is followed so that every pixel tries to attract its neighbor towards its own cluster. Two factors that determine neighborhood attraction are the pixel intensities or feature attraction and the spatial position of the neighbors or distance attraction. Figure 3 shows the definition of neighborhood structure.

The neighborhood attraction is computed as

$$d^2(A_j - V_i) = \|A_j - V_i\|^2 (1 - \alpha H_{ij} - \xi F_{ij}) \qquad (4)$$

Where H_{ij} is the feature attraction and F_{ij} is the distance attraction factor. The values of α and ξ range in between 0 and 1. The feature attraction is calculated using the following formula

$$H_{ij} = \frac{\sum_{k=1}^{S} u_{ik} g_{jk}}{\sum_{k=1}^{S} g_{jk}} \qquad (5)$$

Where g_{jk} is the difference between the intensity levels of current pixel j and its neighboring pixel k, u_{ik} is the membership of neighboring pixel k to the ith cluster and S is the number of neighborhood pixels considered for tuning.

Based on the membership values, the neighborhood tuning membership matrix NbU_{ij} is obtained as

$$NbU_{ij} = U_{ij}^* / \left(U_{ij}^* + \sum_{k \in S} U_{jk}^* \right) \qquad (6)$$

Where the number of neighboring pixels S may be set to 4 or 8. The non-membership and hesitancy values are calculated and the new membership value is calculated as

$$u_{ik}^* = u_{ik} + \pi_{ik} \qquad (7)$$

Using the hesitancy obtained for every iteration, second part of objective function is computed as:

$$J_2 = \sum_{i=1}^{C} \pi_i^* e^{1-\pi_i^*} \qquad (8)$$

where $$\pi_i^* = \frac{1}{L} \sum_{j=1}^{n} \pi_{ij}, j \in [1, L] \qquad (9)$$

The final objective function of CrSA-IFCM-NA is given by

$$J = \sum_{i=1}^{C} \sum_{j=1}^{n} NbU_{ij}^m \| A_j - V_i \| + \sum_{i=1}^{C} \pi_i^* e^{1-\pi_i^*} \qquad (10)$$

Thus the fitness of the crows is evaluated. Now, choose a crow i as the follower of crow j. The fitness of both the crows is compared with the awareness probability and if it is high, a new position is generated using (11), meaning that, the follower is clever enough to reach hiding position of crow j.

$$pos^{A,t+1} = pos^{A,t} + r_i . fl^{A,t} . (mem^{B,t} - pos^{A,t}) \text{ If } r_i \geq AP^{B,t} \quad (11)$$

If the awareness probability is low, then crow j is aware of its follower and thus chooses a random position to fool the follower. The feasibility of the new position is then checked and position is updated only if it is feasible. Otherwise, no change to the position is made. The fitness of the new position of crows is evaluated again. If the quality of the new position is better than the earlier position, the memory of crows is updated using

$$mem^{A,t+1} = pos^{A,t+1} \text{ if } f(pos^{i,t+1}) \text{ is better than } f(mem^{i,t}) \quad (12)$$

where f(.) denotes the objective function value.

Similarly, the position and memory of all crows are updated until the maximum number of iterations is reached. The resulting best initial centroids are those which minimize the fitness function to a greater extent.

The cluster centroids are updated as

$$V_i = \frac{\sum_{j=1}^{n} U_{ij}^* A_j}{\sum_{j=1}^{n} U_{ij}^*} \qquad \text{where } 1 \leq i \leq C \quad (13)$$

The IFCM algorithm with neighborhood attraction is run till the maximum number of iterations is reached. The resulting segmented portions are taken and the region of interest is extracted to identify mammographic mass. The resulting images seem to assist the radiologist for easy diagnosis.

Results

The mammogram images used for investigation are taken from the mini MIAS database which consists of 322 images with the size of 1,024 x 1,024 pixels. The dataset details given are: Reference number, character of background tissue, class of abnormality present, the severity of the abnormality, x and y coordinates of the center of abnormality and the approximate radius of a circle enclosing the abnormality. The character of the background tissues considered for experiment is either fatty or fatty glandular. The severity of abnormality is benign for one image and six images contain malignancy.

The main objective is to segment the region of interest from the mammogram image using a novel hybrid clustering technique. The methodology is applied to all 322 images. In order to consider all kinds of masses and abnormalities, the sample images are chosen such that they belong to all categories like well-defined / circumscribed masses, spiculated masses, ill-defined masses and asymmetrical images. All the 322 images are considered to find the Jaccard index and Dice index. The segmented images shown in Figure 4 are chosen such that they cover samples from various characters of background tissues and different categories of abnormalities. The parameters used for experiments are given in Table 1. The images are segmented and the results are shown in Figure 4 with the region of interest encircled. Original images, their preprocessed versions and ROI extracted images are shown in Figure 4. Segmentation accuracy is proved by means of visual inspection by radiologists or by comparing the segmented area with ground truth results.

In order to effectively validate the clusters found by the clustering algorithm, the cluster validity indices are computed. Internal indices like Silhouette measure, DB index are calculated to evaluate the quality of the resulting clusters. Dice index compares the segmented image with the ground truth images. A higher value facilitates better results in terms of Silhouette index and Dice index whereas a lower value of DB index is preferable.

Silhouette coefficient (Rousseeuw, 1987) is calculated as:

$$SC = \frac{1}{N} \sum_{i=1}^{N} s(x) \quad (14)$$

s(x) is determined as follows

$$s(x) = \frac{b(x) - a(x)}{\max\{a(x), b(x)\}} \quad (15)$$

where a(x) denotes compactness and b(x) indicates separation.

Table 1. Parameters for CrSA-IFCM-NA

Parameter	Value
No. of crows	10
Awareness Probability	0.1
Flight length	2
Lambda	0 to 1 (based on the value that maximizes entropy)
Max Iterations	100

Table 2. Comparison of CrSA-IFCM-NA and PSO-IFCM-NA for Various Cluster Indices

Image	Silhouette Index		DB Index		Dice Index	
	CrSA-IFCM-NA	PSO-IFCM-NA	CrSA-IFCM-NA	PSO-IFCM-NA	CrSA-IFCM-NA	PSO-IFCM-NA
MDB028	.95.23	92.41	9.84	11.34	97.12	96.73
MDB091	96.83	95.72	10.76	12.17	94.19	94.65
MDB092	96.11	94.66	13.42	13.56	97.28	97.03
MDB184	93.64	93.05	15.7	16.5	89.39	89.21
MDB265	98.32	97.8	6.34	7.29	99.36	98.76
MDB267	95.51	95.45	12.58	12.89	96.94	96.45
MDB271	96.03	93.07	9.46	9.84	98.27	95.78
1R	91.29	91.32	9.11	10.08	94.98	94.87
2R	93.27	91.92	13.9	13.96	93.18	93.06

DB index is defined as the ratio of within cluster and between cluster distances.

The formula for DB Index is given by

$$\frac{1}{k}\sum_{i=1}^{k} max_j \frac{s(C_i)+s(C_j)}{d_c(C_i,C_j)} \qquad (16)$$

Where k is the number of clusters, s(C) is the average distance among the instances in cluster C,

dc(C_i ,C_j) measures the distance between the centers of C_i and C_j.

Dice index (Dice, 1945) measures how far the spatial overlap exists between two binary images. A lower value indicates less overlapping while a value closer to one indicates perfect agreement.

$$D(A,B) = \frac{2|A\cap B|}{|A|+|B|} \qquad (17)$$

The Jaccard index measures the similarity and diversity between the segmentation results and the ground truth values. It is also known as Intersection over Union and is computed as

$$J(A,B) = \frac{|A\cap B|}{|A\cup B|} \qquad (18)$$

Real Dataset

Even though the algorithm performs well on the public database, it is necessary to evaluate the algorithm against the real time dataset to prove its efficiency. A set of real images is obtained from the radiologist and the algorithm is executed over them. Out of 15 images received, 12 cases are malignant and 3 cases are benign. The images originally have a size of 2,020 x 2,708 pixels.

After applying the proposed methodology to these images, the sample results are shown in Figure 5. The region of interest is marked in the original image itself.

We have implemented the PSO based IFCM algorithm with neighborhood attraction (PSO-IFCM-NA) for MIAS dataset. The results of crow search based IFCM algorithm with neighborhood attraction (CrSA-IFCM-NA) is compared with PSO-IFCM-NA and the results are shown in Table 2.

Discussion

To make a complete evaluation of the performance of the proposed method, we compare our results with the state-of-the-art methods in the literature. However, it is very hard to make a direct comparison due to application
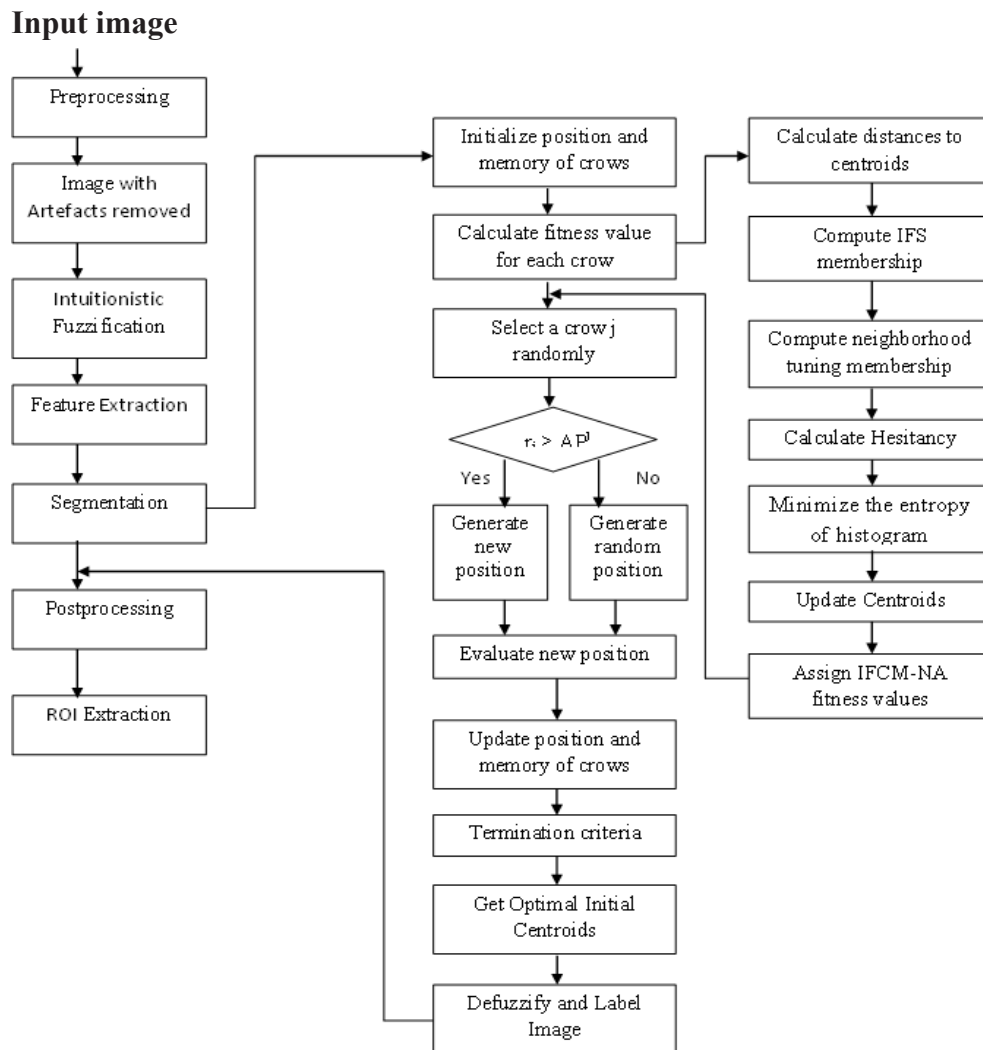

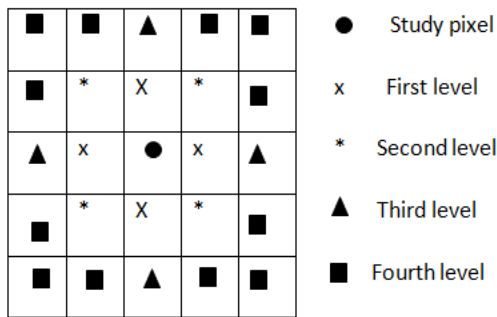Figure 2. Workflow of CrSA-IFCM-NA

Figure 3. Definition of Neighborhood Structure

of algorithms over different datasets (e.g. MIAS, DDSM, inBreast, private datasets obtained) and the number of images considered for evaluation by various authors. Also, the variations in the types of evaluation metrics (Visual assessment by radiologists, Jaccard Index, Dice Index, accuracy, sensitivity) are diverging. Thus, we
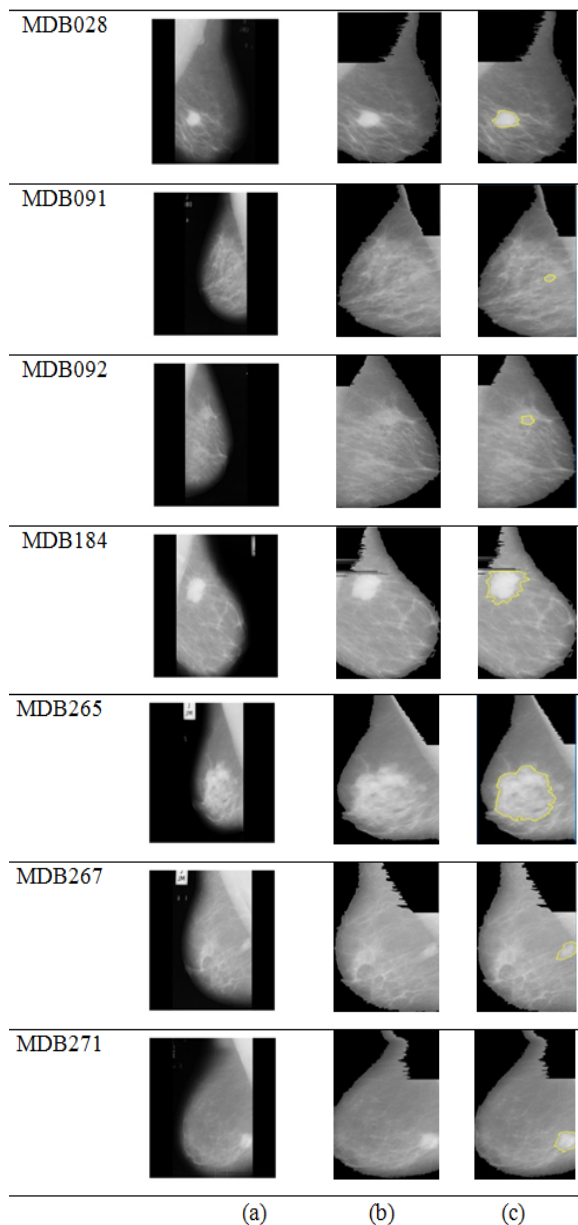

(a)Original image   (b) ROI encircled in original image
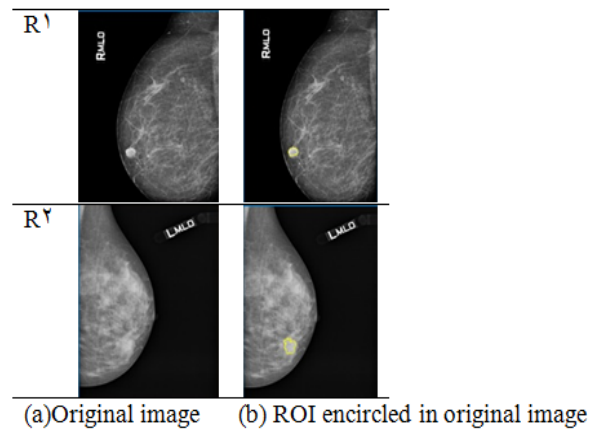
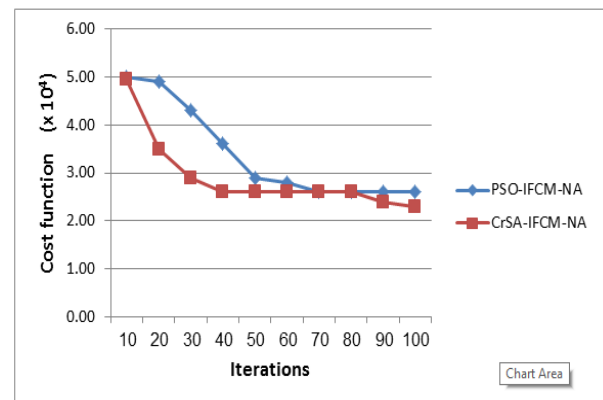Figure 5. Results of Rreal Dataset

Figure 6. Comparison of Convergence Graph of Fitness Function

made a comparison against studies that used the MIAS database (Rampun et al., 2017; Shi et al., 2018) only and have similar metrics. All the 322 images are considered for finding the Jaccard and Dice index.

Many papers in the literature (Binu, 2015) have chosen the population of 10 as this is the global convergence operator commonly preferred by several optimization algorithms. The images themselves are large three-dimensional datasets. The crow Search algorithm is repeated for 100 iterations. Therefore, a population of ten crows undergoes 100 iterations which is a complex task when the number of crows is increased. Apart from these 100 iterations for crows, the entire CrSA-IFCM-NA algorithm is executed for 30 runs and the average values are taken for comparison. If the number of crows is increased, then the number of iterations and runs proportionately increase, leading to the increase in time complexity.

Table 2 shows that the proposed method outperforms

Figure 4. (a) Original Image (b) Preprocessed Image (c) Image with ROI Encircled

Table 3. Comparison with Other State-of-the-Art Techniques

		Indices	Our Method	
Peng Shi (2018)	Jaccard	94.89 ± 6.77	Jaccard	eciD
	Dice	96.40 ± 5.28	96.45±5.27	98.21±3.97
Andrik Rampun (2017)	Jaccard	94.9 ± 6.7		
	Dice	97.9 ± 5.2		

PSO for almost all the sample set of images taken. mdb265 produces the highest value in case of silhouette measure, Dice index and also the lowest value in case of DB index. The results of real datasets show a little less accuracy due to the fact that there may be poor illumination or noise in the figures obtained. Even though the ROI is extracted for the image mdb184, it does not constitute a clear boundary which is indicated by a low level of value for the silhouette and DB indices. Real datasets also exhibit a good performance with the indices. A clear picture of the separation of ROI can be seen in both the real dataset and the mini MIAS images.

From Table 3, it is evident that the mean values of all 322 images in 30 runs are obtained and it is found that CrSA-IFCM-NA produces the higher results in terms of both Jaccard and Dice indices. Our method gives a higher value of over 96% for Jaccard index and over 98% for Dice index.

Figure 6 shows the convergence of the proposed method in comparison to PSO-IFCM-NA. The average values of every ten Iteration are depicted in the graph and it can be seen that the proposed method converges and reaches global minima faster than PSO-IFCM-NA.

The advantages of the proposed methods over FCM is that our method is highly robust to noisy images and also the inherent noise can be well represented using the indeterminacy or hesitancy value of the Intuitionistic Fuzzy Set. The CrSA-IFCM-NA algorithm performs an efficient detection of masses from mammogram images. It recommends the radiologist and assist him in selecting the disease affected area so that a patient can be sent for the right kind of treatment. The optimal global centroid for the IFCM-NA is fixed by means of crow search optimization technique. The extracted features are also given as input to the clustering algorithm to segment the image into various parts. The region of interest is separated based on the intensity levels for further diagnosis.

Acknowledgements

References

Ananthi VP, Balasubramaniam P, Lim CP (2014). Segmentation of gray scale image based on intuitionistic fuzzy sets constructed from several membership functions. *Pattern Recogn*, **47**, 3870-80.

Anter AM, Hassenian AE (2018). Computational intelligence optimization approach based on particle swarm optimizer and neutrosophic set for abdominal CT liver tumor segmentation. *J Comput Sci-Neth*, **25**, 376-7.

Askarzadeh A (2016). A novel metaheuristic method for solving constrained engineering optimization problems: crow search algorithm. *Comput Struct*, **169**, 1-12.

Atanassov KT (2003). Intuitionistic fuzzy sets: past, present and future. *In EUSFLAT Conf*, pp 9-12.

Balasubramaniam P, Ananthi VP (2016), Segmentation of nutrient deficiency in incomplete crop images using intuitionistic fuzzy C-means clustering algorithm. *Nonlinear Dynam*, **83**, 849-6.

Binu D (2015), Cluster analysis using optimization algorithms with newly designed objective functions. *Exp Syst Appl*, **42**, 5848-9.

Chaira T (2010). Intuitionistic fuzzy set approach for color region extraction. *J Sci Ind Res India*, **69**, 426-2.

Chaira T (2011). A novel intuitionistic fuzzy C means clustering algorithm and its application to medical images. *Appl Soft Comput*, **11**, 1711-7.

Chaira T, Anand S (2011). A novel intuitionistic fuzzy approach for tumour/hemorrhage detection in medical images. *J Sci Ind Res India*, **70**, 427-4.

Chaira T (2014). An improved medical image enhancement scheme using Type II fuzzy set. *Appl Soft Comput*, **25**, 293-8.

Cordeiro FR, Santos WP, Silva-Filho AG (2016). A semi-supervised fuzzy GrowCut algorithm to segment and classify regions of interest of mammographic images. *Expert Syst Appl*, **65**, 116-6.

Dhahbi S, Barhoumi W, Zagrouba E (2015). Breast cancer diagnosis in digitized mammograms using curvelet moments. *Comput Biol Med*, **64**, 79-90.

Dice LR (1945). Measures of the amount of ecologic association between species. *Ecology*, **26**, 297-2.

Huang CW, Lin KP, Wu MC, et al (2015). Intuitionistic fuzzy c-means clustering algorithm with neighborhood attraction in segmenting medical image. *Soft Comput*, **19**, 459-70.

Jawahar CV, Ray AK (1996). Incorporation of gray-level imprecision in representation and processing of digital images. *Pattern Recogn Lett*, **17**, 541-6.

Keller B, Nathan D, Wang Y, et al (2011). Adaptive multi-cluster fuzzy C-means segmentation of breast parenchymal tissue in digital mammography. In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp 562-9.

Kontos K, Maragoudakis M (2013). Breast cancer detection in mammogram medical images with data mining techniques. In IFIP International Conference on Artificial Intelligence Applications and Innovations, pp 336-7.

Kuo RJ, Lin TC, Zulvia FE, Tsai CY (2018). A hybrid metaheuristic and kernel intuitionistic fuzzy c-means algorithm for cluster analysis. *Appl Soft Comput*, **67**, 299-8.

Parvathavarthini S, Visalakshi NK, Shanthi S, Mohan JM (2018). Crow search optimization based fuzzy C-means clustering for optimal centroid initialization. *Taga J Graphic Technol*, **14**, 3034-5.

Parvathavarthini S, Karthikeyani NK, Shanthi S, Lakshmi K (2017). Crow-search-based intuitionistic fuzzy C-means clustering algorithm. Developments and Trends in Intelligent Technologies and Smart Systems, pp129.

Parvathavarthini S, Visalakshi NK, Shanthi S, Lakshmi K (2017). An application of pso-based intuitionistic fuzzy clustering to medical datasets. *ICTACT J Soft Comput*, **8**, 1531-8.

Pavan ALM, Vacavant A, Trindade AP, de Pina DR (2017). Fibroglandular tissue quantification in mammography by optimized fuzzy C-means with variable compactness. *IRBM*, **38**, 228-3.

Pelekis N, Iakovidis DK, Kotsifakos EE, Karanikas H, Kopanakis I (2007). Intuitionistic fuzzy clustering to information retrieval from cultural databases. In 22nd European Conf. on Operational Research, EURO XXII, Prague.

Rampun A, Morrow PJ, Scotney, Winder J (2017). Fully automated breast boundary and pectoral muscle segmentation in mammograms. *Artif Intell Med*, **79**, 28-1.

Rousseeuw PJ (1987), Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math*, **20**, 53-5.

Shanthi S, Bhaskaran VM (2011). Intuitionistic fuzzy C-means and decision tree approach for breast cancer detection and classification. *Eur J Sci Res*, **66**, 345-1.

Shanthi S, Bhaskaran VM (2013). A novel approach for detecting and classifying breast cancer in mammogram images. *Int J Intell Info Technol (IJIIT)*, **9**, 21-9.

Shen S, Sandham W, Granat M, Sterr A (2005). MRI fuzzy segmentation of brain tissue using neighborhood attraction with neural-network optimization. *IEEE T Inf Technol B*, **9**, 459-7.

Shi P, Zhong J, Rampun A, Wang H (2018). A hierarchical pipeline for breast boundary segmentation and calcification detection in mammograms. *Comput Biol Med*, **96**, 178-8.

Son LH, Cuong BC, Lanzi PL, Thong NT (2012). A novel intuitionistic fuzzy clustering method for geo-demographic analysis. *Expert Syst Appl*, **39**, 9848-9.

Vega-Corona A, Quintanilla-Domínguez J, Ojeda-Magana B et al (2011). Microcalcifications detection using PFCM and ANN. In Mexican Conference on Pattern Recognition. Springer, Berlin, Heidelberg, pp 260-8.

Vlachos IK, Sergiadis GD (2005). Towards intuitionistic fuzzy image processing. In Computational Intelligence for Modelling, Control and Automation, 2005 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, International Conference on, 1, pp 2-7.

Yang SC (2017). A robust approach for subject segmentation of medical Images: Illustration with mammograms and breast magnetic resonance images. *Comput Electr Eng*, **62**, 151-5.

Zadeh LA (1965). Fuzzy sets. *Info Control*, **8**, 333-8.

Zhao F, Liu H, Fan J, et al (2018). Intuitionistic fuzzy set approach to multi-objective evolutionary clustering with multiple spatial information for image segmentation. *Neurocomputing*, **312**, 296-9.