Research Journal of Applied Sciences 9 (2): 110-118, 2014
ISSN: 1815-932X

Classification of Feature Extracted, Selected and Segmented Mammogram Image Using Hybrid Algorithm-Monkey Search Optimization (MSO) and Support Vector Machine (SVM)

[1]S. Kanimozhi Suguna and [2]S. Uma Maheswari
[1]Department of Computer Applications,
Anna University Regional Centre-Coimbatore,
641 047 Coimbatore, Tamil Nadu, India
[2]Department of Electronics and Communications Engineering,
Coimbatore Institute of Technology, 641 014 Coimbatore, Tamil Nadu, India

Abstract: Classification is most important for analyzing the mammograms of breast cancer. This study proposes a different approach based on Metaheuristic Algorithm is presented for classifying the mammogram image. The foraging behavior of monkey is optimized as Monkey Search Optimization (MSO) which is the subset of the metaheuristic algorithm. Feature extracted image is given as input for the process of classification. To solve complex problems by cooperation the behaviors are considered. Several algorithms based on population-based metaheuristic algorithms were introduced in the literatures to solve different problems like optimization problems. This is the new technique proposed for classifying the mammogram images. Results are presented based on simulation made with the implementation in MATLAB which is tested on the images of MIAS database.

Key words: MSO, climb, watch and jump, cooperation, somersault, stochastic perturbation mechanism, termination

INTRODUCTION

Breast cancer is where cancerous (Malignant) cells are found in the breast tissue. Radiograph of the breast tissue is called mammogram. Getting a mammogram is an effective way to detect breast cancer in its early stages. Breast cancer is the second leading cancer next to Cervical cancer.

X-rays are used for diagnosing breast cancer through Mammographic screening. Breast cancer is most common in women than men worldwide. According to estimates of lifetime risk will be the US National Cancer Institute about 13.2% of women in the US will develop breast cancer which is the same as saying one in eight people.

In 2009, about 562,340 Americans died of cancer, >1,500 people a day. About 1,479,350 new cancer cases were diagnosed in 2009. In the United Sates, cancer is the second most common cause of death and accounts for nearly one of every four deaths. The chance of developing invasive breast cancer at some time in a woman's life is about one in eight (12%).

X-ray mammography is the most common investigation technique used by radiologists in the screening and diagnosis of breast cancer they could help the radiologists in the interpretation of the mammograms and could be useful for an accurate diagnosis. To perform a semi-automated tracking of the breast cancer, it is necessary to detect the presence or absence of lesions from the mammograms.

LITERATURE REVIEW

The preprocessing technique is also known as background suppression. Before getting into the presence or absence of lesions preprocessing is done to remove the pectoral region as it is not in need of the whole image (Ramirez-Villegas *et al*., 2009). In most conditions any of the filter is used as by Thangavel *et al*. (2005) and Kumar *et al*. (2010). Next to the preprocessing is selecting the Region of Interest or Feature Extraction. It focuses on the relevant feature extraction and possible removal of irrelevant features (Roselin *et al*., 2011). The advantage of doing feature extraction is to compress the input space for classification which reduces the overhead of the classifier. The classification would help the doctors to judge the status of the patient and provide treatment accordingly.

Corresponding Author: S. Kanimozhi Suguna, Department of Computer Applications,
Anna University Regional Centre-Coimbatore, 641 047 Coimbatore, Tamil Nadu, India

Overview of metaheuristic algorithms: This study explains a new algorithm called Monkey Search Optimization for extracting the feature in mammogram image. This MSO is a Nature-Inspired Evolutionary algorithm.

OVERVIEW OF THE PROBLEM

In this study, researchers focus on classification which plays a major role in mammogram image. The process of extracting the feature is the last stage in preprocessing of mammogram images (Abdallah *et al*., 2008) is given as input for classification. Earlier to classification process involves three major processes such as: removing unwanted region, removing the pectoral muscle region and selecting the region of interest (Feature Extraction). As the three processes are completed, it reduces the area to be classified.

Comparison is made based on the behaviors of three different monkey species such as Howler, Spider and Squirrel Monkeys (Amato *et al*., 2007). The behaviors of monkeys differ with one another based on the food type, movement, searching for food, resting, watching and doing other activities. Types of food are leaf, fruits and insects. Howler monkey depends on leaves, Spider monkey focuses on leaves and fruits whereas Squirrel monkeys have both fruits and insects. Mammogram image is considered as forest. Forest will have both edible and non-edible foods. This study presumes the region of interest part as the look edible food for monkeys which is available in abundant. Hence, the feature is selected and extracted.

BASIC OF THE MSO

MSO focuses on the foraging behavior of monkey. Monkeys go in search of foods available in the branches of the trees. Each branch is considered as feasible solution. Selection of the branch is based on the probability mechanism. The process of MSO is organized in three processes such as climb, watch-jump and somersault processes. Climb process is further classified as climb up and climb down along with either long or short step climb process (Mucherino and Seref, 2007; Mucherino *et al*., 2013; Ramirez-Villegas *et al*., 2009; Wang *et al*., 2010; Zhao and Tang, 2008).

Climb process: Let assume the concept by having the decision variable vector $C_i = (c_1, c_2, c_3, ..., c_n)^T$ and the objective function for minimization as f(C). $\Delta C = (\Delta c_1, \Delta c_2, \Delta c_3, ..., \Delta c_n)^T$ is a randomly generated vector. The pseudo-gradient of function f(C) at the point C can be expressed as $(f_1'(C), f_2'(C), ..., f_n'(C))^T$ where:

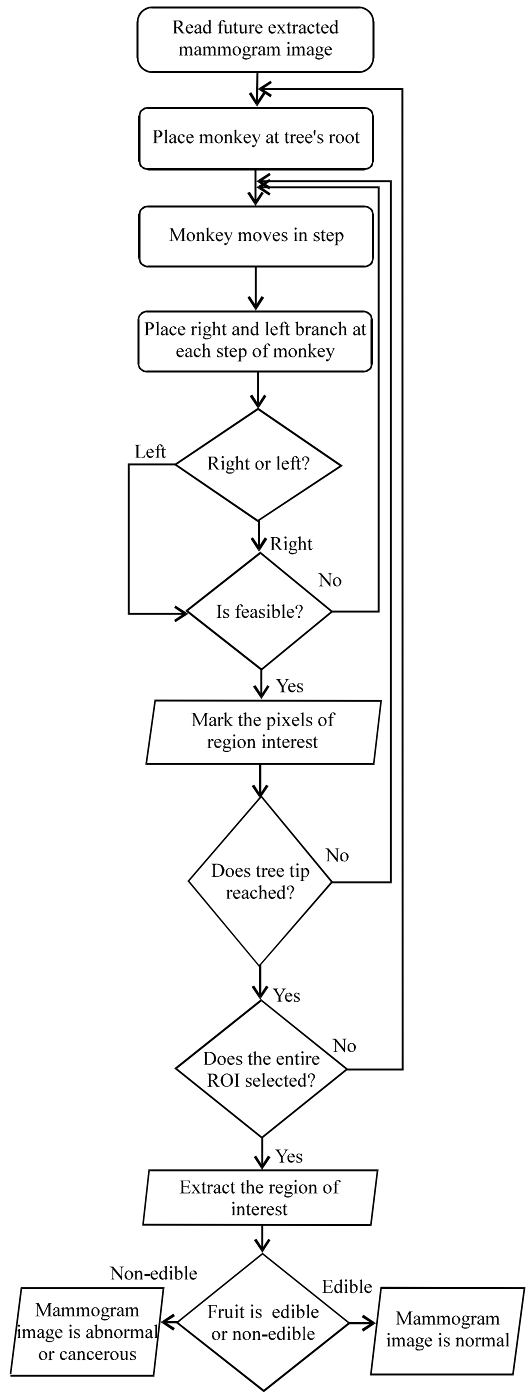

Fig. 1: Flowchart for classifying mammogram image using MSO

$$f_j'(C) = \frac{f(C+\Delta C)-f(C-\Delta C)}{2\Delta c_j} j \in \{1, 2, ..., n\} \quad (1)$$

As the generations of ΔC are repeated, local optimum is found by the decrease in the objective function f(C) based on the sign caused as a result of slow replacement of C with C+ΔC or C-ΔC.

Watch-jump process: This process is made once the monkey reaches the tip of the tree. Since, the monkey has to move on to find availability of any other feasible solution in the forest. Besides, searching for food the monkey has to be aware of the presence of enemies. If there is any feasible solution then it is replaced with the current solution (Mucherino and Seref, 2007; Mucherino *et al*., 2013; Ramirez-Villegas *et al*., 2009; Wang *et al*., 2010; Zhao and Tang, 2008).

Cooperation process: This process focuses on the cooperation among monkeys. Hence, the monkeys at better solution will communicate with monkeys at poor solution and make them to climb to the better of the area (Wang *et al*., 2010).

Somersault process: In this phase monkeys, will move from one position to another position to find the availability of new solution (Wang *et al*., 2010) in the given forest area. Figure 1 depicts the MSO for removing or segmenting the pectoral muscle region in mammogram image.

Algorithm for MSO:

```
1   Set parameters M, C_i, ΔC_i
2   Randomly generate monkey population M
3       for i = 1 to M
4           initialize monkey's position at tree's root
5           C_i = (c_1, c_2, c_3,...c_n)
6           if change in position
7                   Then ΔC = (Δc_1, Δc_2, Δc_3,...Δc_n)
8             end if
9       end for
10      for each monkey i do f(C+ΔC)
11          generate right and left branches for each step
12          using probability mechanism choose either right left branch
13          if the distance is small
14                  Then choose ΔC_i = (0, ..., 0, ΔC_{i,j} 0,...0)^T
15          else if distance is large
16                    then choose ΔC_i = (Δc_{i,1}, Δc_{i,2}, Δc_{i,3},...Δc_{i,n})
17          end if
18          end if
19          if the branch has a fruit
20                 mark as feasible solution C_i
21          else
22          move next step (repeat from step 1)
23                 end if
24                     continue the process form step 10 until the monkey
                   reaches the tree tip
25          end for
26          while tree tip is reached by monkey do
27                 for each monkey i do f(C-ΔC)
28                        move to next tree
29                        repeat the process from step 4
30                            if local feasible solution (tree with edible fruits)
                          is obtained
31                              then cooperate the monkeys, C_i"
32                                     consider the edible fruit as the normal
                                 mammogram image
33                    else
34                         the fruit is considered as non-edible
35                      end if
36              end for
37          end while
38          while local optimum is obtained
39          move on to new domain u_j in search of better solution
40               Find C_i'''
41               Update with monkeys initial position C_i
42               Move all monkeys to C_i'''
43          end while
44          if all the trees with fruits are identified and marked
45                  then mark it either as normal or cancerous and
               terminate the process
46          end if
```

IMPLEMENTATION OF MSO TO THE PROBLEM

Given the details of the image, removal of pectoral muscle region in the mammogram image problem can be defined energy of monkeys is represented as:

$$\|CE\|_2 = \sqrt{\int_{-\infty}^{\infty} f(C) e^{-ihw} dr} \tag{2}$$

Where:

$$f(c) = (hw)_{ij} \tag{3}$$

where, hw is the energy required by each monkey for moving from one place to another. Total energy is calculated based on:

$$T(c) = \|CE\|_2 (hw)_{ij} = \|CE\|_2 \sum_{i=1, j=1}^{M} (hw)_{ij} \tag{4}$$

With the following conditions:

$$CI_m \, CI_v = CI_a - CI_d \tag{5}$$

$$-CI_l^M \le CI_l \le CI_l^M, l \in L \cup L' \tag{6}$$

$$0 \le c_j \le c_j^M, j \in \{1, 2, ..., n\} \tag{7}$$

Solution representation: The solution for the problem is represented in the equation:

$$C_i = (c_{i,1}, c_{i,2}, c_{i,3}, ..., c_{i,n})^T$$

Objective function: The objective function is modified as follows:

$$\min f(r) = \begin{cases} CE \sum_{i=1, j=1}^{M} (hw)_{ij} + G_0 \sum_{l \in L \cup L'} \max\{0, |CI_l| - CI_l^M\} \\ G_1 \end{cases}$$

Generate initial population for M monkeys:

$$C_i = (c_1, c_2, c_3, ..., c_n)^T$$

While (t<MaxGeneration) or (segmentation completed):
Where:

- CI_m = Image matrix
- CI_v = Pixel value
- CE = Total energy
- CI_a = Active pixels in the image
- G_0 = Gravitational force when monkey jumps from one tree to another
- G_1 = Penalty if monkey is unreachable
- RI_l^M = Maximum of how many pixels affected in the image
- CI_l = Original image
- c_j = Number of monkeys in correct position
- c_j^M = Maximum number of monkeys
- L = Set of existing path for movement
- L' = Set of new path for movement

Climb process: Initialize random position for M monkeys as $C_i = (c_{i,1}, c_{i,2}, c_{i,3}, ..., c_{i,n})^T$ evaluate its current position to reach the border for which select either of large-step or small-step climb process:

- The large-step climb process: Generate:

  (i) $$\Delta C_i = (\Delta c_{i,1}, \Delta c_{i,2}, \Delta c_{i,3}, ..., \Delta c_{i,n})^T$$

  Interval $[-c_L, c_L]$ where c_L is the length of large-step climb process.

  (ii) Calculate:

  $$f(C_i+\Delta C_i), f(C_i-\Delta C_i)$$

  (iii) If:

  $$f(C_i+\Delta C_i)<f(C_i-\Delta C_i) \text{ and } f(C_i+\Delta C_i)<f(C_i)$$

  Then:

  $$C_i = C_i+\Delta C_i$$

  Else if:

  $$f(C_i+\Delta C_i)<f(C_i-\Delta C_i)$$

  And:

  $$f(C_i-\Delta C_i)<f(C_i)$$

  Then:

  $$C_i = C_i-\Delta C_i$$

  (iv) Repeat i to iii of large-step climb until $N_{C,L}$ has been reached

- The small-step climb process: (Generate)

  (i) $$\Delta C_i = (0,...0, \Delta c_{i,j}, 0,...0)^T$$

where, $j \in \{1, 2, ..., n\}$ and $\Delta c_{i,j}$ is a non-zero integer with an interval of $[-c_S, c_S]$ where c_S is the length of small-step climb process.
(ii) and (iii) steps for small-step climb process is similar to the (ii) and (iii) steps in large-step climb process.
(iv) Repeat (i) to (iii) of small-step climb process until has been reached.

Watch-jump process: Check whether there is higher position when compared to current position based on the eyesight b:

(i) If higher position is available then generate an integer $c'_{i,j}$ for an interval of $[c_{i,j}-b, c_{i,j}+b]$ randomly.

Where:

$$j \in \{1, 2, ..., n\}$$

Let:

$$C'_i = (c'_{i,1}, c'_{i,2}, ..., c'_{i,n})^T$$

(ii) If:

$$f(C'_i) < f(C_i)$$

Let:

$$C_i = C'_i$$

(iii) Repeat i to ii until maximum allowable number N_w has been reached.

Co-operation process: Assume optimal solution in one iteration as:

$$C^* = (c^*_1, c^*_2, ..., c^*_n)^T$$

Initial position is:

$$C_i = (c_{i,1}, c_{i,2}, c_{i,3}, ..., c_{i,n})^T$$

(i) Generate real number β in [0, 1] randomly.
(ii) Calculate:

$$c'''_{i,1} = \text{round}(\beta c^*_j + (1-\beta)c_{i,j}, j \in 1, 2, ..., n)$$

Let:

$$C_i = (c_{i,1}, c_{i,2}, ..., c_{i,n})^T$$

(iii) Set $C_i = C_i''$ and repeat the climb process.

Somersault process:
(i) Generate the real number α in [c, d].
(ii) Calculate:

$$u_j = \frac{1}{M}\sum_{i=1}^{M} c_{i,j}, j \in \{1, 2, ..., n\}$$

Where:

$$\hat{U} = (u_1, u_2, ..., u_n)^T$$

as pivot of Somersault process.
(iii) For:

$$\forall i \in \{1, 2, ..., M\}, \forall j \in \{1, 2, ..., n\}$$

Calculate:

$$C'''_{i,j} = c_{i,j} + \text{round}\left(\left|u_j - c_{i,j}\right|\right)^T$$

Let:

$$C'''_i = \left(c'''_{i,1}, c'''_{i,2}, ..., c'''_{i,n}\right)^T$$

(iv) Set $C_i = C_i'''$ and repeat climb process. If $c_{i,j}'''$ is new position>upper limit c_j^M.
Then:

$$c'''_{i,j} = c_j^M$$

Else if $c'''_{i,j}<0$
Then:

$$c'''_{i,j} = 0$$

Stochastic perturbation mechanism: When $c_{i,j}$ is same for all monkeys then $u_j = c_{i,j}$ and hence $c'''_{i,j} = c_{i,j}$ for monkey k, k∈{1, 2, ..., M}. Let:

$$c_{k,j} = e$$

where, e is a uniformly distributed integer from is $[0, C_i^M]$. Hence, the global optimization (Serra *et al.*, 1997) is obtained.

EXPERIMENTAL RESULTS

MSO algorithm is carried out for MIAS database of 322 images for removing the pectoral muscle region in the mammogram image. Results of few image samples after the removal of pectoral region and the feature selected and extracted region of interest are shown in Fig. 2-11. From

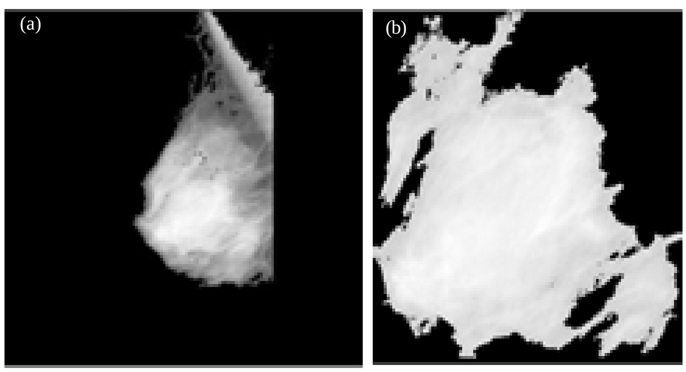

Fig. 2: a) Pectoral region removed; b) region of interest for the dataset 1

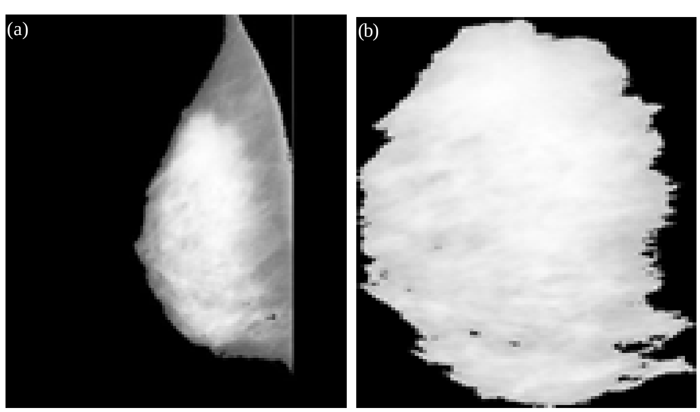

Fig. 3: a) Pectoral region removed; b) region of interest for the dataset 2

the figures it is observed that some figures have region of interest whereas the others don't have the region of interest (Fig. 5, 6 and 8). The dataset with the region of interest is classified as the abnormal mammogram dataset whereas the others (Fig. 5, 6 and 8) are normal. Table 1 gives the details regarding the image dataset. Figure 12-17 made for the Table 1 shows the variations in the values of the datasets.

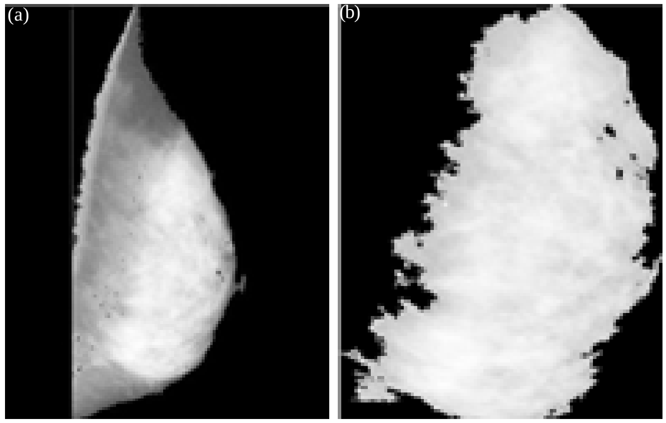

Fig. 4: a) Pectoral region removed; b) region of interest for the dataset3

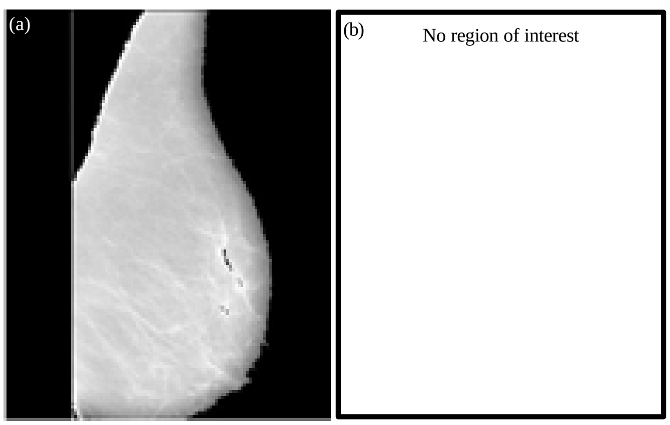

Fig. 5: a) Pectoral region removed; b) region of interest for the dataset4

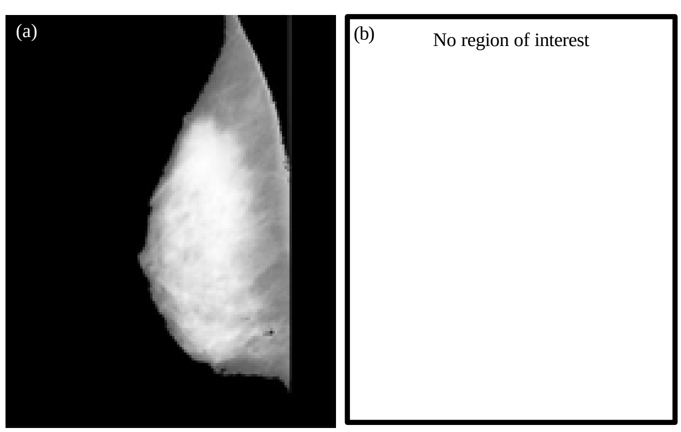

Fig. 6: a) Pectoral region removed; b) region of interest for the dataset5

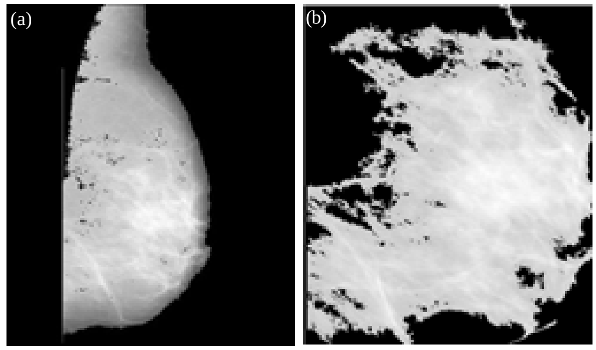

Fig. 7: a) Pectoral region removed; b) region of interest for the dataset6

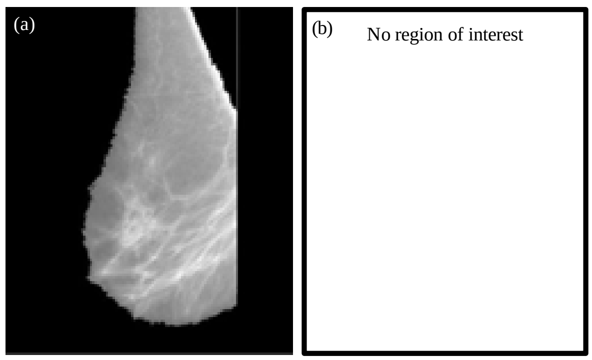

Fig. 8: a) Pectoral region removed; b) region of interest for the dataset7

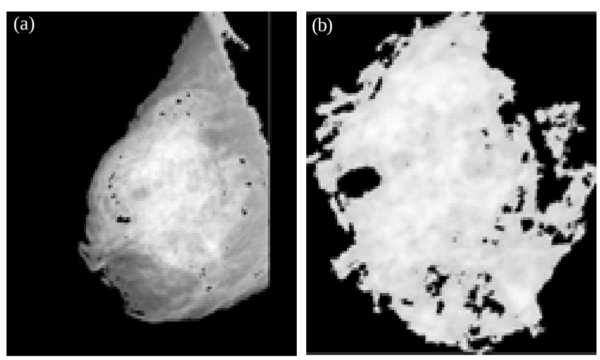

Fig. 9: a) Pectoral region removed; b) region of interest for the dataset8

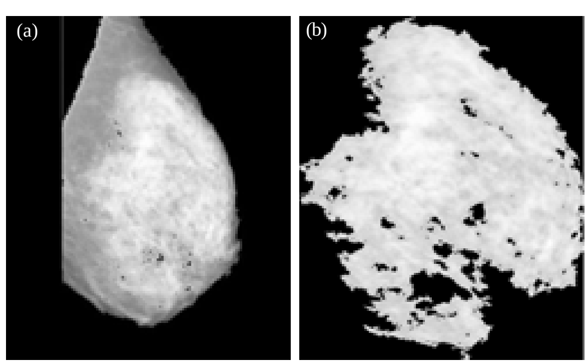

Fig. 10: a) Pectoral region removed; b) region of interest for the dataset9

Table 1: Details of the feature extracted image

Dataset	Mean	Entropy	Skewness	Kurtosis	Height	Width	Energy	CPU Time
1	110.4654	0.9873	-0.104	4.465	288	322	52527	154.6445
2	143.4284	0.8913	-1.0937	14.1107	580	294	120003	140.8140
3	119.0332	0.9797	NaN	NaN	562	386	126815	166.0427
4	0	0	0	0	0	0	0	181.8056
5	0	0	0	0	0	0	0	122.2423
6	117.7582	0.9587	NaN	NaN	464	434	124743	229.1243
7	0	0	0	0	0	0	0	134.6789
8	106.9031	0.9925	-0.2115	7.2043	452	378	94195	168.2245
9	108.8603	0.9908	-0.0763	5.6188	656	470	171699	150.3154
10	0	0	0	0	0	0	0	131.6287

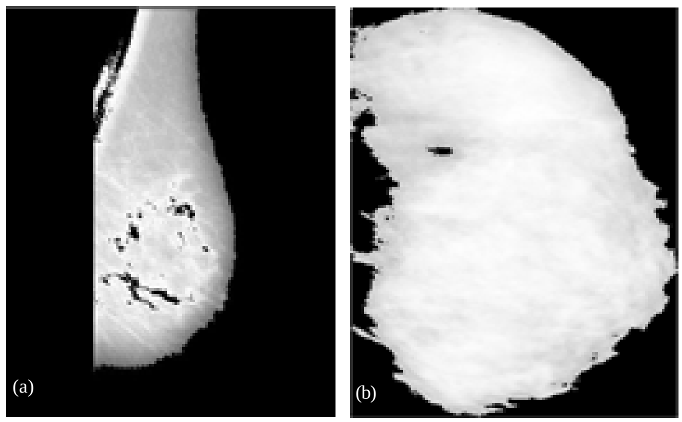

Fig. 11: a) Pectoral region removed; b) region of interest for the dataset10

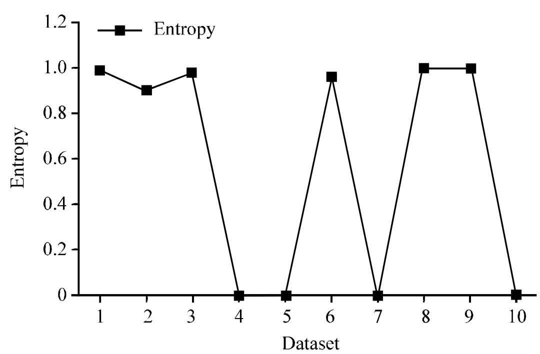

Fig. 12: Variation in entropy for the dataset

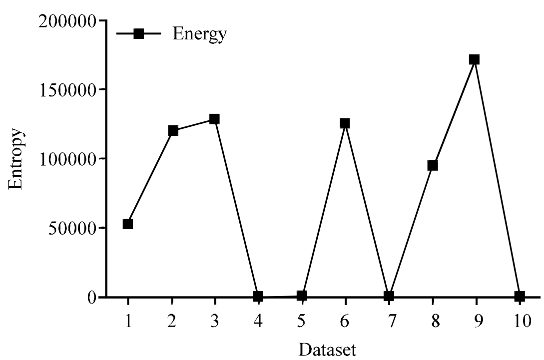

Fig. 13: Variation in energy for the dataset

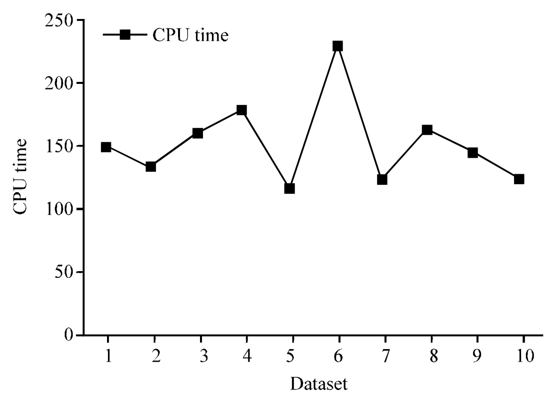

Fig. 14: Variation in CPU time for the dataset

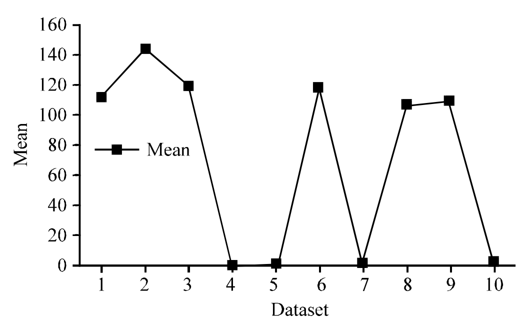

Fig. 15: Variation in mean for the dataset

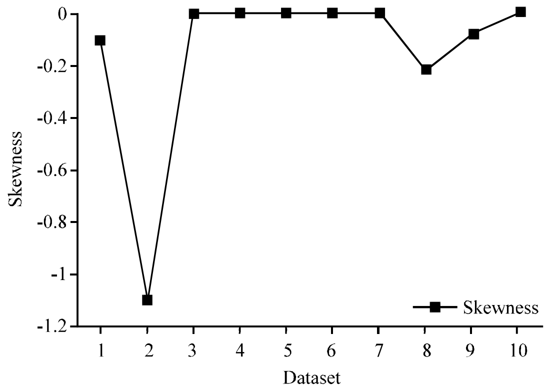

Fig. 16: Variation in skewness for the dataset

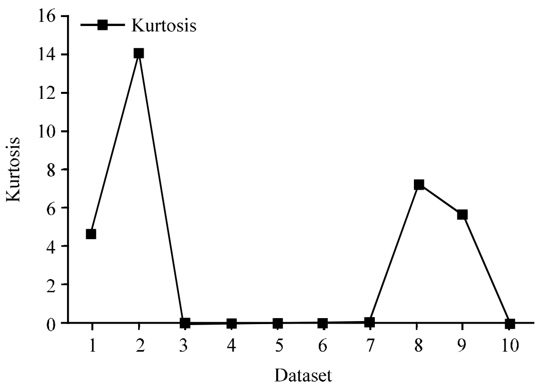

Fig. 17: Variation in kurtosis for the dataset

CONCLUSION

This study proposes a Monkey Search algorithm for performing feature extraction in the mammogram image. As a future enhancement, implementation of the algorithm for classification of the region of interest is in progress.

REFERENCES

Abdallah, M.H., A.A. AbuBaker, R.S. Qahwaji and M.H. Saleh, 2008. Efficient technique to detect the region of interests in mammogram images. J. Comput. Sci., 4: 652-662.

Amato, K.R., D.D. Onen, S.L. Emel and H.C. May, 2007. Comparison of foraging behavior between howler monkeys, spider monkeys and squirrel monkeys. Ecology, 9: 28-31.

Kumar, S.V., M.N. Lazarus and C. Nagaraju, 2010. A novel method for the detection of microcalcifications based on multi-scale morphological gradient watershed segmentation algorithm. Int. J. Eng. Sci. Technol., 2: 2616-2622.

Mucherino, A. and O. Seref, 2007. Monkey Search: A Novel Metaheuristic Search for Global Optimization. In: Data Mining, Systems Analysis and Optimization in Biomedicine, Seref, O., O.E. Kundakcioglu and P.M. Pardalos (Eds.). American Institute of Physics, Melville, New York, ISBN-13: 9780735404670, pp: 162-173.

Mucherino, A., O. Seref and P.M. Pardalos, 2013. Simulating protein conformations through global optimization. http://arxiv.org/pdf/0811.3094.pdf.

Ramirez-Villegas, J.F., E. Lam-Espinosa and D.F. Ramirez-Moreno, 2009. Microcalcification detection in mammograms using difference of gaussians filters and a hybrid feedforward-kohonen neural network. Proceedings of the 12th Brazilian Symposium on Computer Graphics and Image Processing, October 11-15, 2009, Rio De Janiero, Brazil, pp: 186-193.

Roselin, R., K. Thangavel and C. Velayutham, 2011. Fuzzy-rough feature selection for mammogram classification. J. Electron. Sci. Technol., 9: 124-132.

Serra, P., A.F. Stanton and K. Sabre, 1997. Pivot method for global optimization. Phys. Rev. E, 55: 1162-1165.

Thangavel, K., M. Karnan and A. Pethalakshmi, 2005. Performance analysis of rough reduct algorithms in mammogram. Int. J. Global Vision Image Proc., 5: 13-21.

Wang, J., Y. Yu, Y. Zeng and W. Luan, 2010. Discrete monkey algorithm and its application in transmission network expansion planning. Proceedings of the Power and Energy Society General Meeting, July 25-29, 2010, Minneapolis, MN., USA., pp: 1-5.

Zhao, R. and W. Tang, 2008. Monkey algorithm for global numerical optimization. J. Uncertain Syst., 2: 165-176.