Automatic Diagnosis of Breast Cancer using Thermographic Color Analysis and SVM Classifier

Asmita T. Wakankar [1], G. R. Suresh [2]

[1] Sathyabama University, Chennai, India
asmita_wakankar@yahoo.com
[2] Easwari Engineering College, Chennai, India
sureshgr@rediffmail.com

Abstract. Breast cancer is the commonly found cancer in women. Studies show that the detection at the earliest can bring down the mortality rate. Infrared Breast thermography uses the temperature changes in breast to arrive at diagnosis. Due to increased cell activity, the tumor and the surrounding areas has higher temperature emitting higher infrared radiations. These radiations are captured by thermal camera and indicated in pseudo colored image. Each colour of thermogram is related to specific range of temperature. The breast thermogram interpretation is primarily based on colour analysis and asymmetry analysis of thermograms visually and subjectively. This study presents analysis of breast thermograms based on segmentation of region of interest which is extracted as hot region followed by colour analysis. The area and contours of the hottest regions in the breast images are used to indicate abnormalities. These features are further given to ANN classifier for automated analysis. The results are compared with doctor's diagnosis to confirm that infra-red thermography is a reliable diagnostic tool in breast cancer identification.

Keywords: Breast cancer, Thermography, Segmentation, Level Set, ANN

1 Introduction

Breast cancer is a type of cancer caused by breast tissue either the inner lining of milk ducts or the lobules that supply the ducts with milk.[1] Breast cancer is caused by combination of multiple factors like inheritance, tissue composition, carcinogens, immunity levels, hormones etc. Currently, the most common methods for detecting the breast diseases are Mammography, Doppler Ultrasonography, Magnetic Resonance Imaging (MRI), Computed Tomography Laser Mammography (CTLM), Positron Emission Mammography (PEM).[2] However theses imaging techniques only provide the anatomical structure information of tumor lacking functional information. Infrared Thermography is functional imaging technique that can detect cancerous tissue indicating cancer infection, inflammation, surface lesions and more. [2] All objects in the universe emit infrared radiations which is a function of their tempera-

J.M. Corchado Rodriguez et al. (eds.), *Intelligent Systems Technologies and Applications 2016*, Advances in Intelligent Systems and Computing 530,
DOI 10.1007/978-3-319-47952-1_2

ture. [4]The objects with higher temperature emit more intense and shorter wavelength infrared radiations. Infrared cameras are used to detect this radiation, which is in the range of 0.9-14 μm and produce image of that radiation, called Thermograms. In case of cancer, once a normal cell begins to transform, it's DNA is changed to allow for the onset of uncoordinated growth. To maintain the rapid growth of these precancerous and cancerous cells, nutrients are supplied by the cells by discharging the chemicals. [3]This keeps existing blood vessels open, awaken inactive ones and generate new ones. This process is called as 'Neoangiogenesis'. [5]This chemical and blood vessel activity in both pre-cancerous tissue and the developing cancer area is always higher. This in turn increases local tissue surface temperatures leading to detection by infrared imaging. These thermal signs indicate a pre-cancer stage that can't be detected by physical examination, or other types of structural imaging technique. Thermography is radiation free, painless and non-invasive technique. It is the best option for screening of young women, pregnant women or women with fibrocystic, large, dense breasts or women with metallic implants.

2 Literature Review

Though extensive work has taken place in the area of thermographic analysis for breast cancer detection, considering the scope of this paper, we have focussed only on the work that uses segmentation techniques like K means, Fuzzy C means and Level Set method. N. Golestani et al 2014 suggested use of K means, Fuzzy C means and Level set method for detection of hottest region in breast thermogram. Abnormality is detected by measuring shape, size and borders of hot spot in an image. [9] Lavanya A. used particle swarm optimization for threshold level determination and K means clustering algorithm to detect hottest region in IR thermogram. Using fractal measures the cancerous and non cancerous cases are identified from segmented region.[14] Authors have implemented K means and Fuzzy C means algorithms earlier for finding the hot spot in breast thermogram.[2] S.S.Suganthi et al implemented segmentation algorithm using anisotropic diffusion filter which smoothens intra region preserving sharp boundaries. The ROI are segmented by level set function based on improved edge information.[15] Prabha S. et al employed Reaction diffusion based level set method to segment the breast using edge map as stopping boundary generated by total variation diffusion filter. Wavelet based structural texture features are calculated for asymmetry analysis. The results are compared against ground truth. [16] Srinivasan s. et al worked on the segmentation of front breast tissues from breast thermograms using modified phase based distance regularized level set method. Local phase information is used as edge indicator for level set function. Region based statistics and overlap measures are computed to compare and validate the segmented ROI against ground truth. [17]

3 System Block Diagram

Early methods of breast thermogram interpretation were solely based on subjective criteria. [6] The doctors look for changes in colour and vascular patterns in the thermograms to detect the abnormality. As the results depend upon experience and skill set of an individual, the variability between diagnoses is very high. To generate more accurate, consistent and automated diagnosis, the proposed method recommends the use of colour analysis as a first stage. Further, few statistical parameters like entropy, skewness and kurtosis are calculated to arrive at abnormality detection. The system block diagram is as shown in the Fig.1

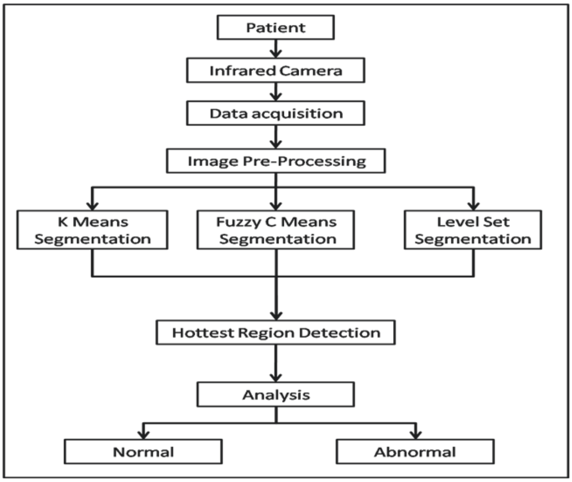

Fig. 1. System Block Diagram

The thermograms are captured at Breast Cancer Clinic using FLIR E30 Infrared camera considering the Pre-Themographic imaging instructions. [2] Total 27 samples are taken including normal and abnormal cases. Few more thermograms are taken from the online thermogram database, Ann Arbor Thermography, (http://aathermography.com) Some image preprocessing steps like background removal and resizing are carried out to remove the undesired body portion. [6] Three methods are used for segmentation of hot spot namely K-Means Clustering, Fuzzy C Means, and Level set method. The statistical parameters indicating abnormality are calculated which are further given to Support Vector Machine (SVM) for classification purpose and finally the results are compared with doctor's diagnosis.

4 Segmentation

The colour bar present on the thermograms gives idea of the temperature distribution indicating the coolest part as blue; intermediate temperature as yellow and red and the warmest part of the image as white. [2] Colour segmentation of the thermal images using clustering algorithms is done to find hottest regions. There are several techniques that can be applied to colour image segmentation. In this study, three techniques namely K-means clustering, fuzzy C means (FCM) clustering and level set are used for colour segmentation of breast infrared images. K means and Fuzzy C means techniques are based on the least square errors while the level set method is based on partial differential equations. K-means clustering computes the distances between the inputs and centers, and assigns inputs to the nearest center. This method is easy to implement, relatively efficient and computationally faster algorithm. But, the clusters are sensitive to initial assignment of centroids. The algorithm has a limitation for clusters of different size and density. [7] FCM clustering algorithm assigns membership to each data point corresponding to each cluster centre on the basis of distance between the cluster centre and the data point.[8] The advantages of FCM are it gives best result for overlapped data set and comparatively better than k-means algorithm. The disadvantages are cluster sensitivity to initial assignment of centroids and long computational time. [9] In Level set method, numerical computations involving curves and surfaces can be performed on a fixed Cartesian grid without having to parameterize these objects. [10] The advantages of Level Set Method are easy initialization, computational efficiency and suitability for Medical Image Segmentation. Disadvantages of Level Set Method are initial placement of the contour, embedding of the object and gaps in the boundaries.

5 Feature Extraction and Classification

Once the segmentation of thermograms is done to find the hottest region, the statistical features like skewness, kurtosis, entropy and area are calculated. The result of the extracted features on segmented thermograms is tabulated along with the graphical representation indicating the better performance of Level Set method.

Classification algorithms categorize the data based on the various properties of images. They work in two stages: training and testing. In the training phase, specific image properties are separated and using these, training class, is prepared. During the testing phase, these feature-space categories are used to classify images. There are different algorithms for image classifications. In this work, Support Vector Machine (SVM) is used for image classification. [11] SVM classifies data by finding the best hyperplane that separates all data points in two distinct groups. SVM method is effective in high dimensional spaces and is memory effective. SVM has a limitation when the number of features is much greater than the number of samples. [12]

5.1 Evaluation of Classifier(SVM) Performance

To evaluate the classifier performance, three parameters namely sensitivity, specificity and accuracy are calculated. Confusion matrix helps to visualize the outcome of an algorithm.

1. Sensitivity / true positive rate:

In case of medical diagnosis, sensitivity is the proportion of people who have the disease and are tested positive. Mathematically, this can be expressed as:

$$\text{sensitivity} = \frac{\text{number of true positives}}{\text{number of true positives} + \text{number of false negatives}}$$

2. Specificity / true negative rate:

Specificity is the proportion of people who are healthy and are correctly diagnosed. Mathematically, this can also be written as:

$$\text{specificity} = \frac{\text{number of true negatives}}{\text{number of true negatives} + \text{number of false positives}}$$

3. Accuracy / Error Rate:

It is the percent of correct classifications. It is the proportion of the total number of predictions that are correct. [29]

$$\text{accuracy} = \frac{\text{number of true positives} + \text{number of true negatives}}{\text{number of true positives} + \text{false positives} + \text{false negatives} + \text{true negatives}}$$

6 Results & Discussion:

In this work, a total of 34 thermograms (27 captured images and 7 online images) were studied. Thermograms of various cases like normal subject, inflammatory cancer patient and patient with fibrocystic change are used to test the system. By using three segmentation techniques K means, FCM and level set method, the hottest region for each case is identified. [14] The results of three different approaches are compared. Following session gives the details of the segmentation, feature extraction and analysis. A comparison of segmentation results between k-means, fuzzy c-means, and level set for a Normal case is shown in figure 2. The segmented region by K means and FCM are yellow and green and is symmetric for both the breasts. The hottest spot window is blank showing that it is a normal case.

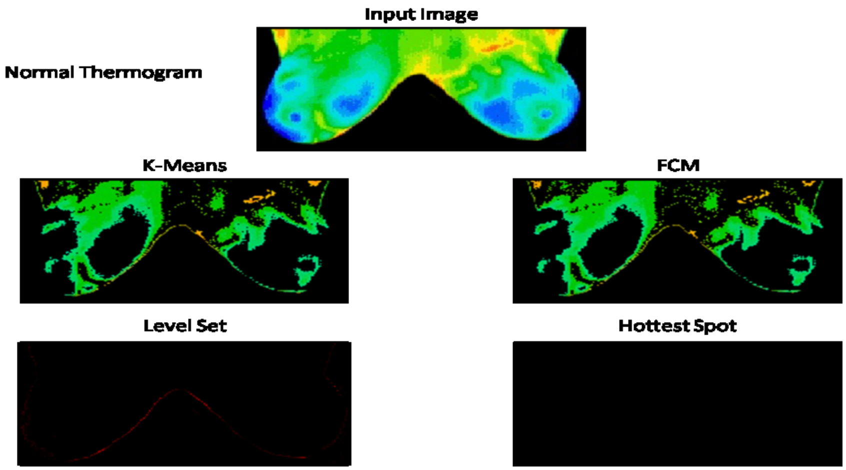

Fig. 2. Hot Region Segmentation of Normal Thermogram

Figure 3 shows segmentation results of all the three methods, for a fibrocystic case. The hottest regions shown by red and orange colour are separated for both the breasts. The area of hottest region is more in the left breast as compared to right breast indicating asymmetry. Also, the Hot spot window shows the presence of small white patches. Thus we can say that there is an abnormality in the left breast.

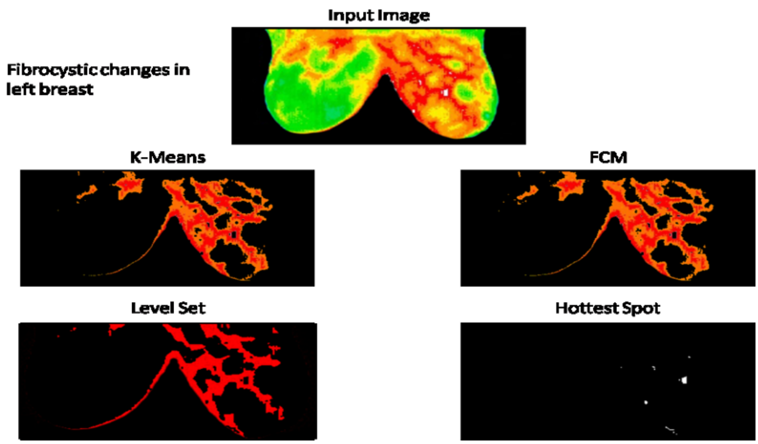

Fig. 3. Hot Region Segmentation of Fibrocystic Changes in Breast Thermogram

Segmentation results for an inflammatory breast cancer case are shown in Figure 4. The hottest spot window shows the presence of white patch in the right breast. So we can notice the asymmetry and elevated temperature for right breast indicating the abnormality.

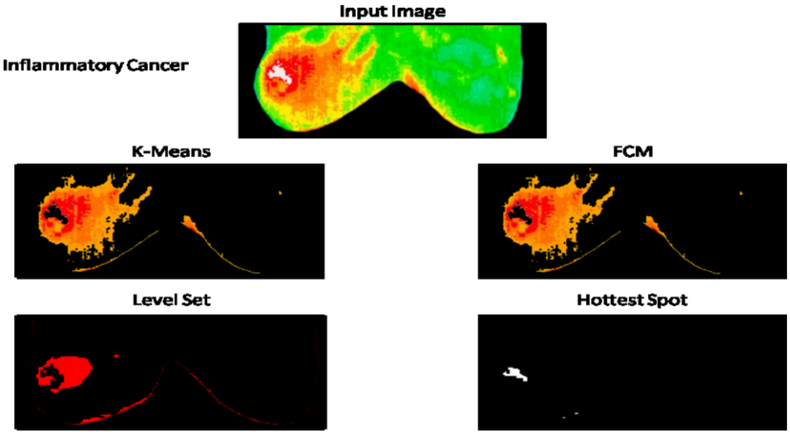

Fig. 4. Hot Region Segmentation of Inflammatory Cancer Patient Thermogram

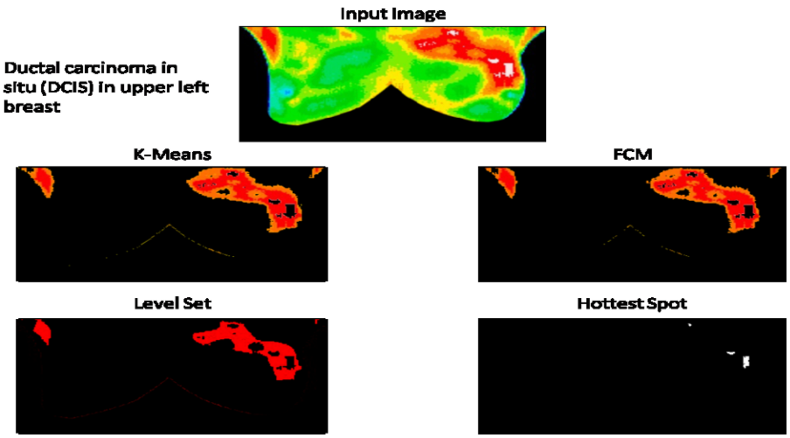

Fig. 5. Hot Region Segmentation of DCIS Patient Thermogram

Figure 5 shows a comparison of segmentation results between k means, fuzzy c means, and level set, for a DCIS in left breast cancer case.

To compare the performance of segmentation algorithms, area of hottest region segmented by three methods is calculated. The results are compared with the actual area of abnormality in thermograms, calculated using the doctors diagnosis. Table 1 shows the Quantitative Parameter Evaluation for 27 patients.

Table 1. Quantitative Parameter Evaluation

Sr. No.	% of Hottest region by K-means	% of Hottest region by FCM	% of Hottest region by Level Set	Actual % of Hottest region	Acceptable Range +	Acceptable Range -
1	2.64	4.29	4.71	3	3.3	2.7
2	10.4	10.39	3.61	3.2	3.52	2.88
3	21.28	21.54	7.58	3.32	3.652	2.988
4	14.93	14.93	6.63	4	4.4	3.6
5	10.92	10.91	3.95	4	4.4	3.6
6	5.55	4.48	4.4	4.1	4.51	3.69
7	5.22	5.81	4.9	4.5	4.95	4.05
8	23.29	23.28	5.69	5.5	6.05	4.95
9	14.47	14.47	6.17	5.8	6.38	5.22
10	9.98	9.97	5.48	6	6.6	5.4
11	9.95	9.71	9.36	8.6	9.46	7.74
12	9.91	9.91	9.08	8.5	9.35	7.65
13	10.08	9.43	9.6	8.8	9.68	7.92
14	11.22	11.29	9.72	9.1	10.01	8.19
15	16.2	17.34	11.61	10.5	11.55	9.45
16	12.02	12.01	11.54	10.8	11.88	9.72
17	12.5	12.5	12.49	11.5	12.65	10.35
18	19.35	19.33	10.18	11.5	12.65	10.35
19	12.5	12.5	12.47	11.5	12.65	10.35
20	37.19	37.12	15.48	15	16.5	13.5
21	21.93	21.27	21.24	21	23.1	18.9
22	21.93	21.27	21.24	21	23.1	18.9
23	23.54	23.12	23	22	24.2	19.8
24	27.51	27.47	23.41	22	24.2	19.8
25	36.89	36.85	30.71	29.5	32.45	26.55
26	36.89	36.85	30.9	29.5	32.45	26.55
27	35.8	35.55	34.54	32	35.2	28.8

A graph of segmented area by different methods versus actual area is plotted. Figure 6 shows that the results of level set method are superior and thus the method is more efficient in finding the hottest region. So feature extraction is performed on thermograms segmented by Level Set algorithm.

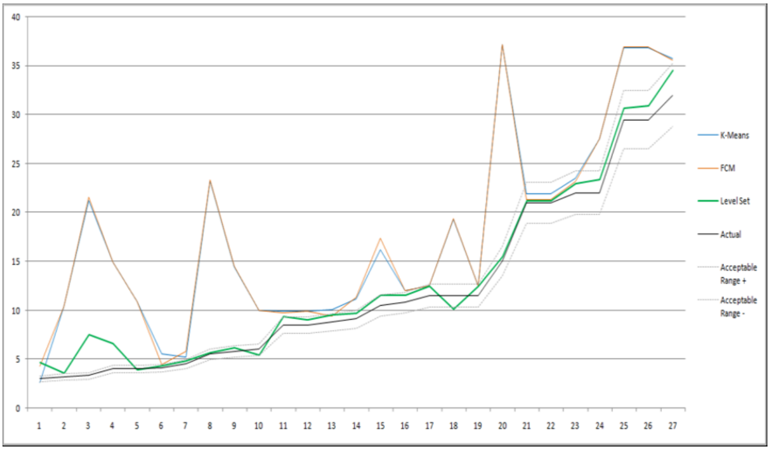

Fig. 6. Comparison of Segmentation Algorithms

A set of features like Skewness, Kurtosis, Entropy, Size and Area of hottest region is calculated. HOS based feature extraction includes the feature Entropy which describes the temperature distribution in the ROI. The Statistical Parameter Evaluation of normal thermograms for 7 volunteers is shown in table 2. It is observed that the parameters percentage, size of the hottest region and entropy are having lesser value for the Normal Thermograms and the parameters skewness and kurtosis are having higher value.

Table 2. Statistical Parameter Evaluation for Normal Thermograms

Sr no	Percentage	Size	Skewness	Kurtosis	Entropy
1	0.3533	89	26.6588	910.3917	0.2679
2	0.4769	639	29.5497	947.6723	0.1489
3	1.2917	452	15.1652	231.9787	0.1403
4	1.5567	471	15.855	265.4904	0.2634
5	10.08	721	13.7705	196.4618	0.199
6	1.972	2456	12.424	156.5434	0.1918
7	2.3006	873	12.6953	167.1503	0.1965

Table 3 gives the details for abnormal thermograms. It is observed that the parameters percentage, size of the hottest region and entropy are having higher value for the Abnormal Thermograms and the parameters skewness and kurtosis are having lesser value.

Table 3. Statistical Parameter Evaluation for Abnormal Thermograms

Sr no	Percentage	Size	Skewness	Kurtosis	Entropy
1	3.0837	1480	10.2861	108.5159	0.2402
2	3.1958	2731	9.82557	98.4661	0.2351
3	3.8906	2430	8.9108	81.1057	0.2579
4	4.3172	1362	8.6429	76.9598	0.3322
5	5.9626	5211	6.97	49.7647	0.371
6	6.3851	2363	6.7477	46.7892	0.3669
7	7.4077	5136	6.2694	40.5905	0.4428

These features are fed to SVM classifier which categorizes the thermograms into Normal and Abnormal cases. Total 34 cases are taken as case study, in which 10 cases are used for training of classifier and 24 cases are used for testing purpose. Out of 24 cases, 18 were abnormal and 6 were normal cases. The system correctly identified 16 abnormal cases and all 6 normal cases. The above values are shown in the form confusion matrix as in table 4. The two nonconclusive cases were online images and had issues related to image capturing.

Table 4. Confusion Matrix

	Disease Present	**Disease Absent**
Test Positive	TP=16	FP=0
Test Negative	FN=2	TN=6

To evaluate the classifier performance, the following three parameters Sensitivity, Specificity and Accuracy are calculated. Table 5 shows these values.

Table 5. Classifier Performance

Sr.No.	Parameters	Value (%)
1.	Sensitivity	88.8
2.	Specificity	100
3.	Accuracy	91.6

7 Conclusion

Infra-red Thermography in medicine is a non-invasive and a non-ionizing method for understanding the internal system, especially for a cancerous tumor, as it relates temperature to physiological parameters such as metabolic rate. [6] This work presented an approach that deals with Automatic Analysis of Breast Thermograms using colour segmentation and feature extraction for abnormality detection. The hottest region of breast thermograms was detected using three image segmentation techniques: k-means, fuzzy c-means (FCM) and level set. Experimental results have shown that the level set method has a better performance as against other methods as it could highlight hottest region more precisely. Some useful features like skewness, kurtosis, area and entropy were extracted from them and given to SVM for classification into normal and abnormal cases. The classifier performed well with the accuracy of 91 %. Proposed method based on Colour Analysis proves to be an effective way for classification of Breast Thermograms helping medical professionals to save their time without sacrificing accuracy. An extension of developed methodology is to quantify degradation of cancer in terms of percentage.

References

1. Ganesh Sharma, Rahul Dave, Jyostna Sanadya, Piyush Sharma and K.K.Sharma: Various Types and management of Breast Cancer: An Overview, Journal of Advanced Pharmaceutical Technology and Research, pp 109-126,1(2) (2010)
2. Shahari Sheetal, Asmita Wakankar: Color Analysis of thermograms for Breast Cancer Detection: Proceedings IEEE International Conference on Industrial Instrumentation and Control (2015)
3. Harold H. Szu, Charles Hsu, Philip Hoekstra, Jerry Beeney: Biomedical Wellness Standoff Screening by Unsupervised Learning: Proceedings SPIE 7343, Independent Component Analyses, Wavelets, Neural Networks, Biosystems and Nanoengineering VII, 734319 (2009)
4. Hossein Zadeh, Imran Kazerouni, Javad Haddaadnia: Diagnosis of Breast Cancer and Clustering Technique using Thermal Indicators Exposed to Infrared Images, Journal of American Science, Vol 7(9), pp 281-288 (2011)
5. Pragati Kapoor, Seema Patni, Dr. S.V.A.V Prasad: Image Segmentation and Asymmetry Analysis of Breast Thermograms for Tumor Detection, International Journal of Computer Applications, Volume 50, No 9 (2012)
6. Asmita Wakankar, G. R. Suresh, Akshata Ghugare: Automatic Diagnosis of Breast Abnormality Using Digital IR Camera, IEEE International Conference on Electronic Systems, Signal Processing and Computing Technologies (2014)
7. Saad N.H.,N.Mohonma Sabri,A.F.Hasa,Azuwa Ali, Hariyanti Mahd Saleh: Defect Segmentation of Semiconductor Wafer Image Using K Means Clustering, Applied Mechanics & Materials, Vol 815, pp 374-379 (2015)
8. Anubha, R.B. Dubey: A Review on MRI Image Segmentation Techniques, International Journal of Advanced Research in Electronics and Communication Engineering, Vol 4, Issue 5 (2015)
9. N. Golestani, M. Tavakol, E.Y.K.Ng: Level set Method for Segmentation of Infrared Breast Thermograms, EXCLI Journal, pp 241-251(2014)

10. Chuming Li, Rui Huang et al: A Level Set Method for Image Segmentation in the Presence of Intensity Inhomogenitis with Application to MRI, IEEE Transactions in Image Processing, Vol 20, No 7 (2011)
11. U. Rajendra Acharya, E.Y.K.Ng, Jen Hong Tan, S. Vinitha Sree: Thermography based Breast Cancer Detection Using Texture Features and Support Vector Machine, Journal of Medical Systems, Vol 36, pp 1503-1510 (2012)
12. Kieran Jay Edwards, Mahamed Gaber: Adopted Data Mining Methods, Studies in Big Data, pp 31-42 (2014)
13. Nijad Al Najdawi, Mariam Biltawi, Sara Tedmori: Mammogram Image Visual Enhancement, mass segmentation & Classification, Applied Soft Computing, Vol 35, pp 175-185 (2015)
14. A. Lavanya: An Approach to Identify Lesion in Infrared Breast Thermography Images using Segmentation and Fractal Analysis, International Journal of Biomedical Engineering and Technology, Vol 19, No 3 (2015)
15. S.S.Suganthi, S.Ramakrishnan: Anisotropic Diffusion Filter Based Edge Enhancement for Segmentation of Breast Thermogram Using Level sets, Elsevier, Biomedical Signal Processing and Control, Vol 10 (2014)
16. Prabha S, Anandh K.R, Sujatha C.M, S.Ramakrishnan: Total Variation Based Edge Enhancement for Levelset Segmentation and Asymmetry Analysis in Breast Thermograms, IEEE, pp 6438- 6441 (2014)
17. S.S.Srinivasan, R. Swaminathan: Segmentation of Breast Tissues in Infrared Images Using Modified Phase Based Level Sets, Biomedical Informatics and Technology, Vol 404, pp 161-174 (2014)