



Review Article

Breast cancer classification using improved fuzzy c means algorithm

Hye-jin Kim*

Kookmin University 77, Jeongneung-ro, Seongbuk-gu, Seoul, Korea

ABSTRACT

Abnormal growth in the breast tissue prompts to the strange cell development in the breast. The researchers typically research for the size of the tumour in a mammogram, because mammograms contain irregular measurements of large scale and smaller scale calcifications. The nearness of these irregular measures of calcium stores in the breast ought to never be ignored as these are indications of early breast malignancy. To decipher this statement in a mammogram precisely, the quality of the pictures ought to be at its incomparable. The proposed research work is conveyed out for examinations of different screening strategies to recognize the unique phases of breast malignancy. In India for every 4 minutes, the women are diagnosed with this disease. And a woman dies with this disease for every 13 minutes. This disease is prominent with the people living in the ruler area while comparing the people in the urban areas. Therefore, it is very important to find and treat this disease as early as possible. The breast tumour region, perimeter and breadth are assessed from mammogram picture databases. The Bits Errors Degree (BER), Highest Indication to Clatter Percentage (PSNR) and Callous Tetragonal Inaccuracy (MSE) values are determined for both Abnormal and normal images. These analyses were used to approve the presence or absence of the disease and to support the evaluation process for finding the disease. This quality assessment is used to understand the reality on Earth for a specific diagnosis that is a specific type of chromatin in a carcinogenic core that may indicate an irregular protein sequence.

Keywords: PSNR, MSE, Malignancy, Cancer, Classification, Fuzzy Means Algorithm.

Received - 25-09-2021, Accepted- 22-12-2021

Correspondence: Hye-jin Kim* ✉ khj5187@kookmin.ac.kr

Department of General Education, Kookmin University 77, Jeongneung-ro, Seongbuk-gu, Seoul, Korea.

INTRODUCTION

This article explains about the importance of Breast cancer [1]. Breast cancer accounts for 28% to 35% of all lady malignancies in all conurbations transversely India. It's very important to find the Breast cancer at earlier stage. All over the country U.S is mostly affected with this disease In India, for every 4 [2] minutes the women are diagnosed with this disease. And a woman dies with this disease for every 13 minutes. This disease is more common with the people living in rural [3] area while comparing the people in urban area. Therefore, it is very important to find and treat this disease as early as possible. According to the affected areas and severeness of the problem [4] this cancer is divided into different stages, with the different stage it is divided into different types, the early-stage cancer is more treatable while comparing the later stages.

Motivation and previous work

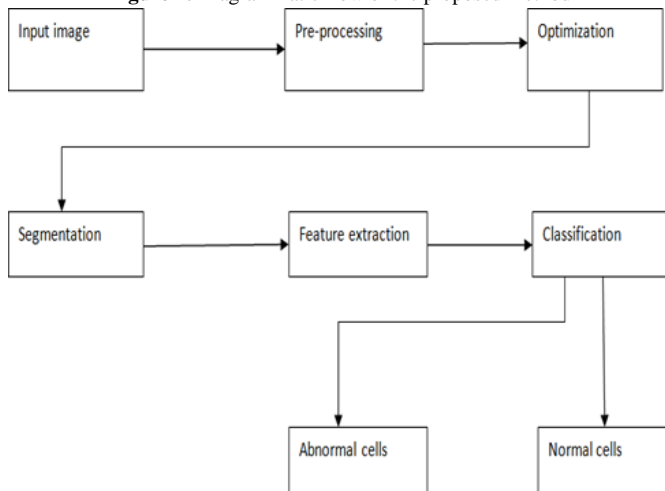
Jiyo S. Athertya et al. (2016) [5] developed an involuntary separation of delineations from CT images using Fuzzy crooks. In this method the automatic initialization of contours has been demonstrated using active contour method. Fuzzy corner gave an

Accuracy of 80% with high Dice coefficient and low Hausdorffs distance. This algorithm is suitable for noisy images also. It might be a daunting task in case of lax tissue image means and it has the complexity in finding of corners of the image. Elisee Ilunga Mbuyamba et al. (2016) [6] proposed an alternative Active Contour Model (ACM) driven by Multi-population Cuckoo Search algorithm. This strategy assists the converging of control points towards the global [7] minimum of the energy function unlike ACM which is often trapped in local minimum. The algorithm has been tested and implemented on MRI images. Three metrics, Jaccard index, Dice coefficient and Hausdorffs distances [8] has been used to assess the results. This method requires lesser iterations, robust and more effective. It takes more computational time for computing the iterations. Agus Pratondo et al. (2016) [9] delivered improved robust edge stop.

Functions (ESFs) for edge based active contour models. Robust Edge Stop functions uses gradient information which fails to stop contour evolution when the image has poor boundaries. In this

method the new ESFs has been used which has gradient information as well as probability scores to classify the mass. This method was evaluated using two quantitative measures namely Jaccard index and Dice coefficient. This method converges faster and gives global contours but it is a complex method. Radha M et al. (2016) [10] proposed an image enhancement technique for breast cancer detection. Mean filters, Median filter, Wiener filter and linear filter are used for pre-processing among these filters median filter provides best results. Image segmentation is performed through thresholding technique and K-Mean's [11] algorithm. The tumour edges are detected using canny edge detection technique [12]. This algorithm results in a higher accuracy. The limitation of this method is the difficulty in finding blurred image edges.

Figure 1: Diagrammatic flow of the proposed method



Proposed work

The proposed method comprises of four steps. In the first step pre-processing is done where the unwanted parts such as labels has been removed. In the second step optimization has been done where the image gets optimized for the further processing methods. the third step segmentation has been done where the exact affected parts can be obtained. segmentation is followed up by the fourth step feature extraction where some special features get extracted and made ready for the classification.

Pre-processing

The main purpose of pre-processing [13] is to improve the image quality in an effective manner. The proposed method consists of few pre-processing steps, in the first step background [14] image should get removed, in the second step pectoral muscle should get removed based on the image orientation, in the third step the image should get enhanced where the quality should get improved without any artifacts. The fig 2 shows the flow diagram of pre-processing method.

Labels and other artifacts removal from the background

To find the tissue boundary of the breast, we do the following; First, we switch from the unsigned portion of the image to the double decade, and then we get the image energy, which is equal

to the second power of the decimal image. The model of image energy is shown in Figure after conversion, once again according to the threshold value we transform that again to the original binary image.

Figure 2: Flow diagram for pre-processing

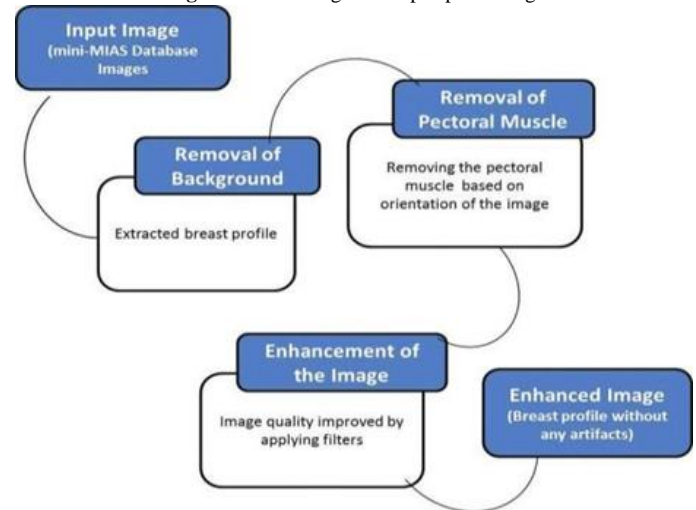
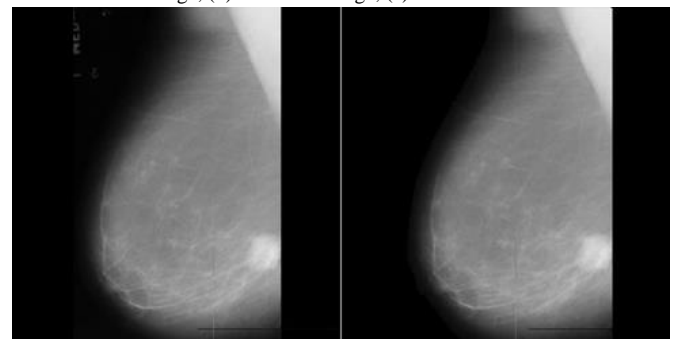


Figure 3: Mdb 209 image: Removal of labels and other artefacts, (a) input image, (b) binarized image, (c) label removed



Background areas will not cover the breast area, it mostly occurs in the right and left of the image. As the background regions are dark and their grey level value will always be close to zero, this grey level does not differ from each other. Therefore, find the new function by fixing an intensity value. That is, here the taken parameters are fixed with the specified value, the range within that fixed value is given a maximum intensity value. The range above and below the fixed specified value is denoted as zero. Thereby the excessive part gets removed. fig 3 shows the output of label removed image.

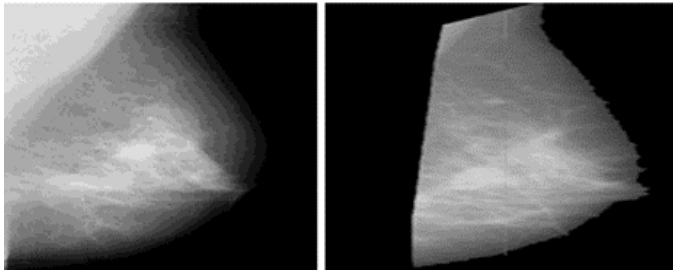
Removing of pectoral muscle

The pectoral muscle is a muscle that comes behind the border of the breast. This is an unwanted portion that basically comes in the mammogram image; therefore, the removal [15] of this region will be helpful for the further segmentation process. There are various methods available to remove that region. In this thesis, we remove that unwanted region by finding the correct border of the breast part. By finding the correct border, it's easy to evacuate the remaining unwanted regions.

In figure 4 the first image is the input image taken from

mini MIAS database; the second image shows the pectoral muscle removed output

Figure 4: Mdb209 image: Removal of pectoral muscle, (a) input image, (b) Pectoral muscle removed output



Enhancement of image

Enhancement is the final step in the pre-processing techniques. During this enhancement, the quality of the image gets improved, which is more important for the next undergoing process. spatial domain and frequency domain are the two basic classification techniques of image enhancement [16] in our proposed method median filtering is used to enhance image [17] equality. Median filtering, enhancement is basically done by calculating the median value of the image pixel value. Algorithm to find the median value is shown below.

Step 1: The pectoral muscle removed image is obtained.

Step 2: If the obtained pixel is noisy, it should undergo a further process.

Step 3: Find the median value of the method shown in fig 3.11 and replace all the noisy pixels using the median value.

Step 4: Shift the window.

Step 5: Repeat step 3 for all the pixel values.

Step 6: Obtain the enhanced image.

Measurement of PSNR Value and MSE Value

The performance of these median filter is calculated using the Peak Signal to Noise Ratio (PSNR) value and MSE Values.

$$(x + a)^n = \sum_{k=0}^n \binom{n}{k} x^k a^{n-k} \quad (1)$$

Peak Signal-to-Noise Ratio (PSNR) is the ratio between the maximum potential value of a signal (power) and the noise distortion that affects the quality of its representation. Since many signals have a very wide dynamic range (the ratio between the largest and smallest values of the convertible size), the PSNR is usually expressed in terms of logarithmic decibels. Improving the image or improving the visual quality of a digital image is subjective [18]. Here the obtained PSNR value is 35.28 which shows that the final enhanced image is better in quality. MSE value is 2.9861 which shows the result has less error and the obtained image quality is very good. In contrast to the proposed technique gives more accurate result. These results suggest that, our current study has convincingly enhanced the quality of the image with better in contrast saying that a system provides a high-quality picture is different for each person. For this reason, it is

necessary to establish quantitative/empirical measures to compare the effects of image enhancement methods on image quality. Using similar test images, we can systematically compare different image enhancement methods. to identify if a particular method produces better results. The metric under investigation is the peak-to-signal-to-noise ratio. If we can show that an algorithm that is similar to the original or a set of instructions can improve the known image of the degenerate, then we can more accurately conclude that this is a better algorithm [19].

Optimization segmentation

Basically, the medical images are not very crystal clear, it contains more noises and picture quality is also very low while comparing other digital images, so segmenting those images directly may lead to poor segmentation, therefore the identification of cancerous cells becomes complex. To reduce those complexities the obtained medical image is first optimized.

Optimization Using Independent Search Krill Herd Technique

In this manuscript, a brilliant algorithm, named krill herd (KH) is used to solve the optimization tasks successfully [6]. The KH algorithm works according to the simulation behaviour of the krill members. The krill members try to preserve a more density which moves automatically due to their mutual effects. For a krill individual, this movement can be defined as:

$$f(x) = a_0 + \sum_{n=1}^{\infty} \left(a_n \cos \frac{n\pi x}{L} + b_n \sin \frac{n\pi x}{L} \right) \quad (2)$$

Nmax is denoted as the maximum speed and on is denoted as the weight of the motion induced in the range [0,1], N old is the last motion induced provide due to the fellow citizen and a target j is the objective route effect providing by the best rill distinct. The measured values of the maximum induced speed are considered as 0.02(MS). The result of the neighbours can be considered as an attractive/repulsive tendency among the individuals for a neighbourhood search. In this study, the result of the neighbours in a krill movement member is strong-minded as surveys:

$$\sin \alpha \pm \sin \beta = 2 \sin \frac{1}{2}(\alpha \pm \beta) \cos \frac{1}{2}(\alpha \mp \beta) \quad (3)$$

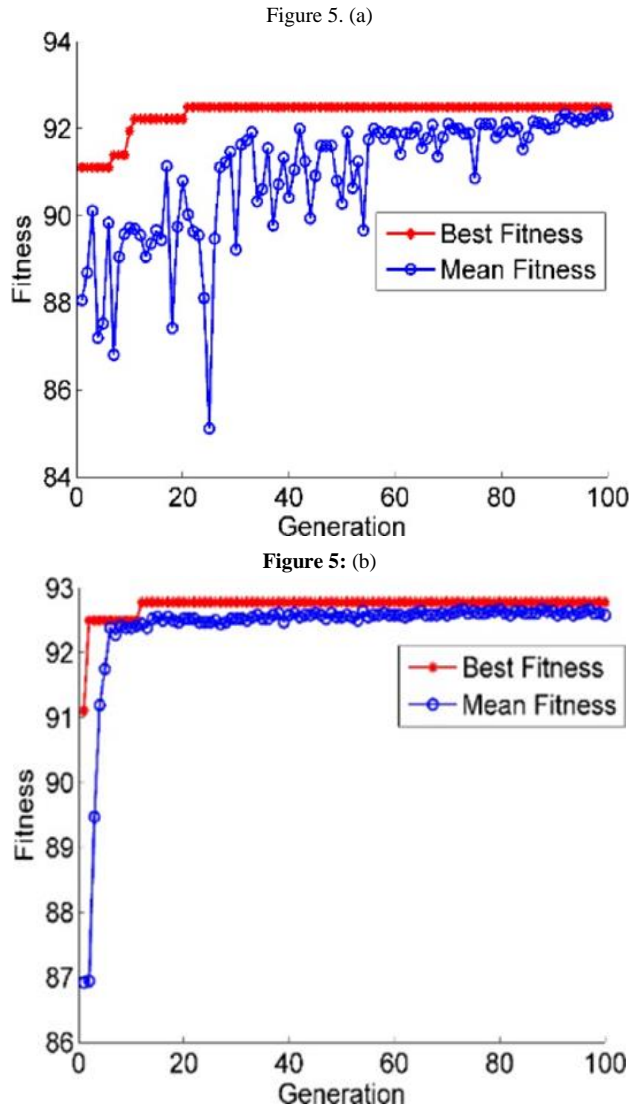
The neighbour's vector might be appealing or appalling since the standardized esteem be negative or positive. The distance for every krill individual can be resolved utilizing diverse techniques. Here, it is resolved by utilizing accompanying equation for every cycle:

$$\cos \alpha + \cos ds(\beta) = 2 \cos \frac{1}{2}(\alpha + \beta) \cos \frac{1}{2}(\alpha - \beta) \quad (4)$$

Where ds, i denotes the distance for ith krill individual and N denotes the total number of krill individuals, the value 5 given in the denominator of the equation is empirically found. By Utilizing the above condition, if the separation of two krill individual is not exactly

the characterized distance, then they are neighbours. The well-known main vector of every krill member is the most reduced fitness member. The krill member with the best fitness on the i th singular krill is considered utilizing the above equation. This dimension drives to the worldwide membership which can be written as, were, C_{best} is the most successful member of the krill movement with the best wellness to the i th krill membership. The C_{best} is characterized as an objective which drives the answer for the worldwide optima, and it ought to be more powerful than other krill individual, such as neighbours.

Figure 5: Fitness curve, (a) fitness curve output of proposed method, (b) fitness curve output of PSO technique



In figure 5, its shown that, the first graph produces the fitness curve output using independent free search krill herd optimization technique. The second graph shows the fitness curve output using PSO technique, from the graph it's clear that IFSKHO technique produces best fitness value while comparing PSO technique.

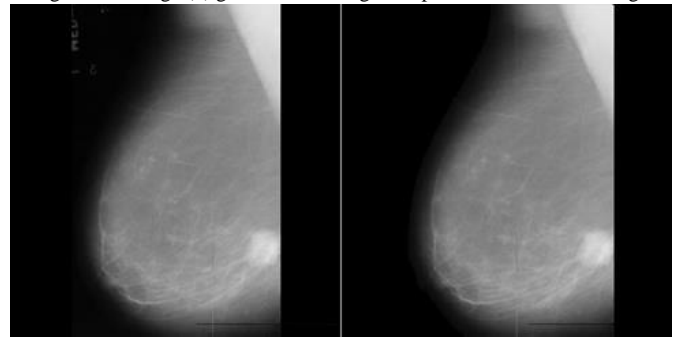
Defuzzification

Defuzzification is done to convert the fuzzy matrix [18] function to the crisp partition. Several methods have been developed

to de- fuzzy the partition matrix U , of which the maximum membership algorithm is one of the important types.

$x_i = \text{argi} \{ \max (u_{ij}) \}$, where $j = 1, 2, \dots, x$ with the above equation, the above method is de-fuzzy and useful for segmentation.

Figure 6: Results: (a) mdb271 original image (b) optimized image, (c) FCM segmented image (d) ground truth image compared with resultant image



Segmentation

This document introduces a new advanced FCM clustering method called Advanced FCM (AFCM) algorithm to generate partitioning. The penalty time is based on the local dependence of an object, which is inspired by the Neighbouring EM (NEM) algorithm and adapted based on FCM reference. The advantage of this algorithm is that it can process all the levels of noise by adjusting the penalty rate. Besides, in this algorithm the membership changes, while the calculations on the central objects are the same as in the standard FCM algorithm. This adaptive FCM is proposed only after understanding that it considers both spatial and feature information together during segmentation. Therefore, it is easy to implement. Experimental results and comparisons of the multiple derivatives of this algorithm in different images show that the proposed algorithm is efficient and powerful

Step 1: Initialize the cluster value C .

Step 2: Adaptive FCM value is made to run using the objective function.

Step 3: Calculate the clustered centre value.

Step 4: If the criteria meet up, repeat until the final clustered value is obtained, if not, then assign the object value to the clustered function.

Step 5: If condition satisfied, then stop, otherwise calculate the population entropy function and again find the centroid value.

Step 6: Evaluate the suitability cost again and replace with the existing rate if the obtained value is higher.

Step 7: Determine the excellent value by using the gold Selection method.

Step 8: Load the ground truth which labelling from an Experienced radiologist.

Step 9: Comparison between the output in step (6) and ground truth for the abnormal region.

Step 10: Perform the FCM iteration again, if criteria match to update

the new cluster centre value, if not then repeat the step 7 again.

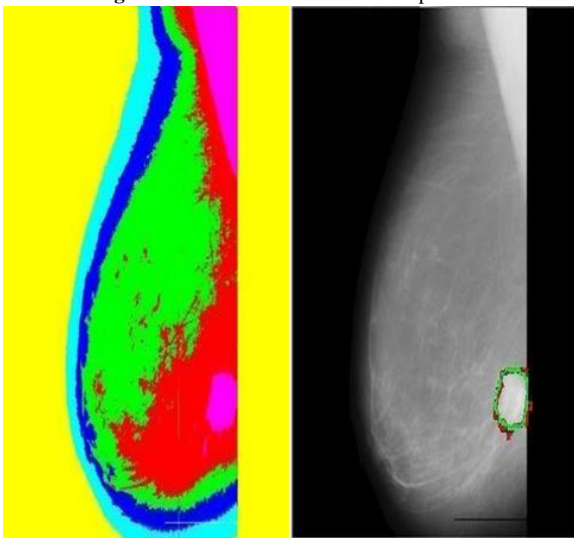
Feature extraction

Feature extraction is needed whenever there is a need of various information, by using this we can sort out the most relevant information needed for classification, so it is considered to be one of the important methods for image processing steps. There exist many algorithms like template matching, histogram matching, DWT, PCA, and many. In this proposed method, we use DWT with Adaptive FCM to extract the complete features which can be useful during the classification procedure. Discrete Wavelet Transform (DWT) is one of the most powerful methods for feature extraction why because it captures both time and frequency information. DWT 2D signal can be obtained by multiplying two 1D functions.

$$\alpha(i, j) = \alpha(i) \alpha(j)$$

This distorts the effort Duplicate as sub-bands with the pair of Strainers namely Low pass filter and High Pass filter. Here the Haar filter is used to decompose the image into sub-bands. After the first phase decomposition the input image is divided into four bands, namely the LL1 represents the rough image. The detailing can be divided into three-phase and the first one is horizontal profile HL, 1 $(i, j) = (i) * (j)$, the second one is vertical profile LH, 2 $(i, j) = (i) * (j)$, and the third one is diagonal details HH input image, 3 $(i, j) = (i) * (j)$. By using this filter in one stage the image is decomposed into four sub-bands (LL, LH, HL, HH) where the three types of detail imaging are obtained over there for each resolution. There are three types of detail images for each Resolution: Horizontal (HL), Vertical (LH), and Diagonal (HH). Activities can be repeated in the Low-Low (LL) band by utilizing the second stage of the homogeneous filter bank. Thus, a generic creates 2D DWT, used in image compression produces the hierarchical structure as shown in Figure.

Figure 7: Structure of wavelet decomposition

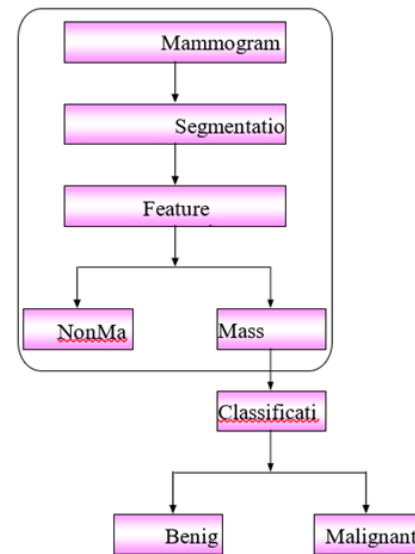


RESULTS AND DISCUSSIONS

Classification is a process used in medical image processing to distinguish benign and malignant tumours cells. Breast

cancer classifies according to different program criteria and serves a different purpose. The major categories are histopathological type, tumour quality, tumour status, and expression of proteins and genes. In this thesis, the Gray Level Co-occurrence Matrix is used to distinguish the cancerous and non-cancerous cells.

Figure 8: Classification of Mammogram



They are the size of the tumour, mobility, their spread in the lymph nodes, and their spread in the other parts. Basically, the work of classifiers is to classify the good tissue and the bad tissue, which is said in other words that classifier has to classify the cancer cells and non-cancerous cells. The classification of cancer cells can be done by identifying the subtypes count in the feature extracted cells. Estrogen Receptor (ER) and Human Epidermal Growth Factor (HER2), basal-like Luminal-A, and Luminal-B are most probably used subtypes as a predictive factor for classifying abnormal cells from the normal cells. Our analyses were led in two sections, to assess these marks (starting now and into the foreseeable future we allude to every single prognostic signature and organic pathways as just marks, except if explicitly recognized) for their capacity. Breast cancers are classified into known sub-atomic or subtypes (basal-like, HER2-enhanced, luminal-A and luminal-B), ER-status (ER+ and ER-). Separate great and awful malady forecast in patients. The Figure shows the various sub-types occurrence in the different stages of cancer, and the methods used currently to treat the cancer cells. AR stands for Androgen Receptor, BCL is B-cell Lymphoma, CK is Cytokeratin, EGRF is Epidermal Growth Factor Receptor, ER Estrogen Receptor, ERBB2 is Erythroblastic Leukaemia Oncogene homolog, PARP is poly (ADB-ribose) Polymerases, and PR is Progesterone Receptor. Appraisals of mammography sensitivity can run from 75% to 90% with specificity from 90% to 95%. The positive prescient estimation of mammography for bosom malignant growth ranges from 20% in ladies under age 50 to 60% to 80% in

ladies age 50-69. And the negative prescient estimation can be range from 90 to 95% in the age > 40 years women.

Figure 9: Performance analysis of mammogram image
Performance Analysis of Mammogram Images

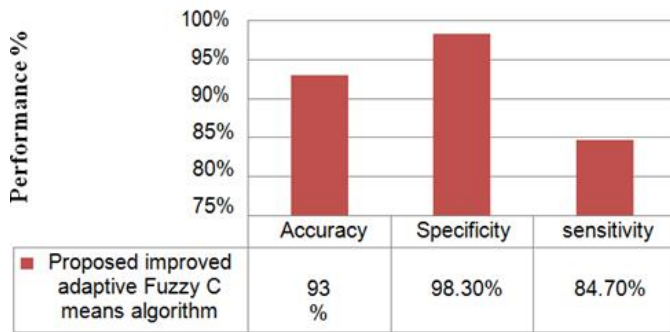
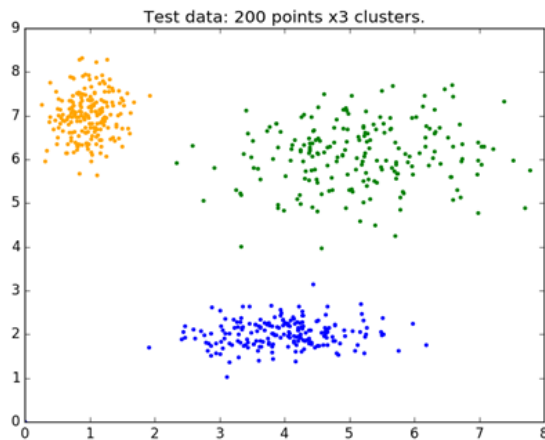


Table 1: The values of reliability ratio with the diagnostic procedure

Parameters	Diagnostic Procedures
	Mammography
Sensitivity (%)	52.4
Specificity (%)	66.7
Positive Predictive value (%)	40.7
Negative Predictive Value (%)	76.2

Figure 10: Showing the Training Process of the Fuzzy C Means Algorithm



CONCLUSION

We developed a new methodology for automatic mass detection to aid in the manual identification of masses in mammography images. The suggested project begins by identifying the tumour inside the specified image region of interest using a fuzzy c-means technique, and then verifies the image features (i.e. texture) produced from the FCM input data using the GLCM feature texture to help in the segmentation process. As can be observed, the suggested methodology's results were more closely matched with those from the Mini-MIAS database, demonstrating that the suggested technique is capable of properly and automatically extracting masses from ROI.

SPECIAL THANKS

This work was supported by G Sai Chandrasekhar and N. Thirupathi Rao (Department of CSE, Vignans Institute of Information Technology, Visakhapatnam, AP, India.)

REFERENCES

- Bhattacharyya D, Reddy BD, Kumari NMJ, & Rao NT, 2021. Comprehensive analysis on comparison of machine learning and deep learning applications on cardiac arrest, *J Med P'ceutical Allied Sci*, 10(4), 3125-3131 doi:10.22270/jmpas,V10I4,1395.
- ST Acton and DP Mukherjee, 2000. "Scale space classification using area Mukherjee, "Scale Trans Image Process, vol 9, no 4, pp 623-635.
- Eali Stephen Neal Joshua, Debnath Bhattacharyya, Midhun Chakkravarthy, Yung-Cheol Byun, 2021. "3D CNN with Visual Insights for Early Detection of Lung Cancer Using Gradient-Weighted Class Activation", *J Healthc Eng*, vol Article ID 6695518, 11 pages, doi:10.1155/2021/6695518.
- Chandra Sekhar P, Thirupathi Rao N, Bhattacharyya D & Kim T, 2021. Segmentation of natural images with k-means and hierarchical algorithm based on mixture of pearson distributions, *J Sci Ind Res*, 80(8), 707-715, Retrieved from scopus.
- Joshua ESN, Battacharyya D, Doppala BP, & Chakkravarthy M, 2022. Extensive statistical analysis on novel coronavirus, Towards worldwide health using apache spark, doi:10.1007/978-3-030-72752-9_8 Retrieved from scopus.
- Aziz Maiandra, Bhagirathi Halalled, 2015. "Threshold based segmentation technique for mass detection in mammography", *Vol 5, jack 472-478*.
- Joshua ESN, Chakkravarthy M, Bhattacharyya D, 2020. An extensive review on lung cancer detection using machine learning techniques, a systematic study *Revue d'Intelligence Artificielle*, Vol 34, No 3, pp 351-359, doi:10.18280/ria, 340314.
- Ramani, Ramani, Salamat, "A Survey of current image segmentation techniques for detection of Breast cancer", vol. 2, no. 5, pp.1124-1129, Oct. 2012.
- Dale JR, Loew WK, Raccoon D, Tupac AE & Putti TC, 2008. Automatic breast cancer grading of histopathological images", in *Proc 30th Anu, Int Conf, IEEE Eng Med Biol Soc, (EMBS)*, pp 3052-3055.
- Demir C & Yenter B, 2005. „Automated cancer diagnosis based on histopathological images, a systematic survey“, Dept Compute Sci, Rensselaer Polytech, Inst Troy, NY, USA, Tech Rep, TR-05-09.
- Elstone CW, Ellis IO & Pinder SE, 1999. Pathological prognostic factors in breast cancer“, *Critical Reviews in Oncology/Haematology*, vol 31, no 3, pp 209- 223.
- Frierson HF Jr, 1995. „Interobserver reproducibility of the Nottingham modification of the bloom and Richardson histologic grading scheme for infiltrating ductal carcinoma“, *Amer J Clin Pathos*, vol 103, no 2, pp 195-198.
- Gene's tie C, 1997. „Comparison of the prognostic value of Scarf- Bloom-Richardson and Nottingham histological grades in a series of 825 cases of breast cancer, Major importance of the mitotic count as a component of both grading systems“, *Anticancer Res*, vol 18, no 1B, pp 571-576.
- Held M, 2010. „Cell Cognition Time-resolved phenotype annotation in high-throughput live cell imaging“, *Nature Methods*, vol 7, no 9, pp 747-754.
- Hong-Ze Li, Sen Guo, Chun-jibe Li & Jing-qi Sun, 2013. "A hybrid annual power load forecasting model based on generalized regression neural network with fruit fly optimization algorithm", *Knowledge- Based Systems*, vol 37, pp 378-387.
- Kumari NMJ, & Krishna KK, 2018. Prognosis of Diseases

- Using Machine Learning Algorithms, a Survey in 2018 International Conference on Current Trends towards Converging Technologies, ICCTCT, pp 1-9, IEEE.
17. Bhattacharyya D, Kumari NMJ, Joshua ESN, & Rao NT, 2020. Advanced Empirical Studies on Group Governance of the Novel Corona Virus, MERS, SARS and EBOLA, a systematic study, Int J Cur Res Rev, Vol, 12(18), 35.
 18. Joshua ESN, Chakkravarthy M, Bhattacharyya D, 2020. An extensive review on lung cancer detection using machine learning techniques, a systematic study Revue d'Intelligence Artificielle, Vol 34, No 3, pp 351-359, doi:10 18280/ria, 340314.
 19. Eali Stephen, Neal Joshua, Debnath Bhattacharyya, Midhun Chakkravarthy, Yung-Cheol Byun, 2021. "3D CNN with Visual Insights for Early Detection of Lung Cancer Using Gradient-Weighted Class Activation", J Healthc Eng, vol Article ID 6695518, 11 pages, doi:10 1155/2021/6695518.

How to cite this article

Hye-jin Kim, 2022. PSNR, MSE, Malignancy, Cancer, Classification, Fuzzy Means Algorithm. J. Med. P'ceutical Allied Sci. V 11 - I 1, P- 4093 - 4099. doi: 10.55522/jmpas.V11I1.1431.