IDENTIFICATION OF TUMOUR USING K-MEANS ALGORITHM

G.Durgadevi[1]- Research Scholar, Dr.Himanshu Shekhar[2]- Professor

[1,2]*Department of Electronics and Communication Engineering, Hindustan University, Chennai, India*

ABSTRACT

Breast cancer *is cancer that develops from breast tissue. Signs of breast cancer may include a lump in the breast, a change in breast shape, dimpling of the skin, fluid coming from the nipple, or a red scaly patch of skin. In those with distant spread of the disease, there may be bone pain, swollen lymph nodes, shortness of breath, or yellow skin. The uncontrolled division of cells is termed as cancer. It is a highly heterogeneous disease and western women commonly witness this. Mammography is used to diagnose breast cancer. Sometimes the mammography results and additional tests may be performed in special circumstances.This basic test mode helps in identifying breast cancer at early stage and this early stage detection would support in recovering more number of women from this serious disease. Medical centres depute highly skilled radiologists & were given responsibility of analysing this mammography results but still human errors are inevitable. An error frequency ratio is high when radiologists exhausted in their analysis task and leads variations in either observations i.e., internal or external observation. Also, quality of the image plays vital role in Mammographic sensitivity and leads variation. In this research work, the automation process using k-means clustering algorithm is attempted in diagnosis of breast cancer. The algorithm yielded a quality analysis process for breast cancer images.*

Keywords: *Mammogram, Breast cancer, k-means algorithm*

---

1. INTRODUCTION

Unlimited multiplication of a specific group of cells in a particular area of the human body referred as CANCER. A lump or mass of an additional tissue will be formed on a group of divided cells which splits quickly. These lump or masses are identified as Tumours. Cancer cells are known as malignant tumours. Breast cells were basis for forming malignant tumour, known as Breast Cancer. Clusters of micro calcifications, architectural distortions and masses are notable & caution able signs. The growth ratio of breast cancer reported very high in present years. At the same time, survival rate also increased potentially over past, it's majorly due to improved efficiency in diagnosis and treatments [1].

Screening of breast cancer primarily takes an anatomic approach through X-ray mammography which requires the breast tumour to be developed to a stage where it is significantly denser than healthier tissue. As a consequence, mammography misses 5%–15% of non-palpable breast lesions that are not sufficiently denser than healthy tissue [2], [3]. In addition, increased density is not always tied to the presence of cancer: dense lesions of tissue that are further investigated via biopsy are often found to be benign [2]. Instead of relying on density changes, cancer can also be detected by using early molecular signatures.

In 2011 United States, the American Cancer Society had come out with their pre-analysis report that nearly 230,480 fresh cases of invasive breast cancer and nearly 57,650 fresh cases of non-invasive breast cancer would be in treatment for breast cancer. 39,520 women would die out of the total affected cases. Mammography, a famous and well-known process in diagnosing Breast Cancer – which uses low-dose X-rays, high-contrast and high-resolution detectors and the X-ray system designed exclusively to image the breasts. In Breast Cancer Screening and diagnosis, it's understood that Mammography serves the purpose of application. Mammography is of two types, 1. Screen Film Mammography (SFM) – film screen is acting as an end recording device & 2. Full-Field Digital Mammography (FFDM) – digital detectors acting as an end recording media. In Image Processing and further grading support, the FFDM produced digital images has more advantages rather traditional film screen [4].

Digital Mammogram is one of the important methods to identify the Breast Cancer at an early stage at some extend. The advantages of digital mammography include the lack of ionizing radiation, its non-invasiveness, the relatively compact instrumentation, and its cost-effectiveness. As Siddiqui et al.[5] mentioned, Mammography is very effective and the results were highly reliable in identifying breast cancer and it's proven, a minimum number of radiologists were tasked in interpreting and diagnosis of Mammograms which is more by and large from population screening. It was mentioned in the report by Wroblewska et al. [6] that there is always a risk missing breast cancer cases, involved in mammographic image observance because unusual identifications were embedded and hidden by variance in structures of breast tissue.

2. RELATED WORK

2.a. Studies on different techniques

Abou-Chadi et al. [7], taken support of neural networks in identifying candidate circumscribed lesions in digitized mammograms. Back Propagation algorithm is used in training neural networks. The process of neural is majorly differentiates the histogram of cancerous tissue and the normal tissue.Brake et al. [8] noted in his studies how digital mammograms used in single and multi- layer detection of masses. In mammograms, scaling plays a vital role in automated process of detecting masses, it's mainly because of the possible range of masses can have.

The work carried out, it was experimented that if detection of masses canbe done in single scale or might be suitable to use the result at various levels of scaling in multi-scale scheme.Chan et al [9], has done research and introduced a computerized mode of detecting micro-calcification in digital mammograms. This mode works on variation in image in which the signal suppressed image is subtracted from a signal enhanced image, which removes the structured background in the mammogram. For extracting micro-calcification signals, global and local threshold values based techniques are used.

Karssemeijer [10] has done his studies and developed a data based calculate method for detecting of micro-calcifications in digital mammograms. Bayesian image analysis is base for the statistical models and general framework.Nakayama et al. [11] took support of filter bank in identifying linear and nodular patterns. The sub images were generated with the elements of a Hessian matrix at all resolution level with support from filter bank. The small and eigenvalues were calculated and a new filter bank resulted with three properties, follows, 1. Nodular patterns can be enhanced with various sizes, 2. Various sizes can be enhanced in both nodular and linear patterns and 3. By removing these patterns, an original image canbe re-build. In mammograms, filter bank is used in enhancing micro-calcifications.

Yu et al. [12] has given in proposal that two steps of CAD system for the automatic clustered micro-calcifications detection. In first step, wavelet and gray level statistical properties used in potential micro-calcification pixels segment and establish them into objects of potential individual micro-calcification. In second step, 31 statistical properties were used to check these potential objects. Enough support taken from Neural Networks too [14]. The outcome results were promising but not guaranteed; it's due to training set usage in testing.

3. ALGORITHM USED

3.aK-means algorithm

A set D = {**xi** | *i*= 1, . . . ,*N*}, where **xi** denotes the i^{th} data point.

Set of d-dimensional vectors. The process initiated with k points chosen from the initial k cluster data or "centroids". The initial value taken by using sampling at random on dataset, fixing it as the clustering solution, a small data subset or unsettled global mean of k times data.

Repeat this algorithm process till convergence,

Step 1: Assigning Data from set D

Every data point from set D is assigned to its closest centroid, with ties arbitrarily broken. Data partitioning is resulted.

Step 2: "means" Relocation
Every cluster representative data is replaced to the centre (mean) of all the data points assigned to it. The replacement is to the expectations (weighted mean) of the data partitions taking place if the data points reached the probability measure (weights).

Euclidean distance is the default measure of closeness, during this scenario, non-negative cost function applies always in Equation 1,

$$\sum (\text{argmin} \| X_i - C_j \|^2) \quad (1)$$

4. SYSTEM ARCHITECTURE

Mammogram result is taken as an input and given to pre-processing phase for filtering the data. In low-level image processing, pre-processing becomes an inevitable problem. By using various filtering techniques, noise presented in the image can be filtered out. The gray level of an image reduces while high pass filter passes the changes to a low pass filter. It means, the value smoothens and sharp edges were removed frequently, while applying low pass filter. The Median Filter is the best of low pass filters. The filter considers an image of area 3x3, 5x5, 7x7, etc., an element array is resulted by taking all the pixel values. The median value of an array is calculated and resulted by ordering element array. A famous sorting technique, Bubble sort is used in this element array sorting in an Ascending order, which returns a median value from the middle elements of the sorted array. The set, the median values of the array elements calculated for all the pixels, were resulted to an output image array [13]. The complete image array is arrived by repeating the Median Filter process.

End of pre-processing phase, all the processed data is fed into first classification algorithm (i.e. k-means algorithm). With the help of k-means algorithm, processed data can be converted into specified clustered data.

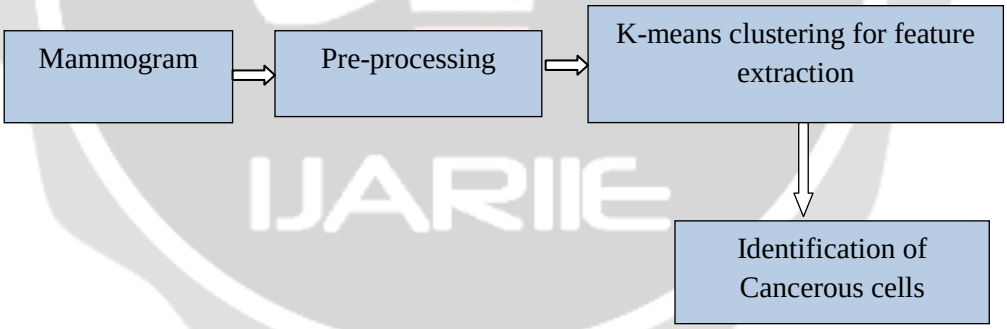

Figure 1. System Architecture for Breast Cancer Identification

5. RESULTS AND DISCUSSION

The identification of breast cancer from the MRI images is made automatic using K-means clustering and wavelet transform. The data base consisting of both normal and abnormal images of the breast are used for this analysis. The human perception at many times may lead to erroneous diagnosis. Variation in diagnosis may produce adverse effect on the patients. Hence to improve the accuracy this system is made automatic using machine vision algorithms.

5.1 MRI Data base collection

The MRI images both consisting of normal and abnormal breast images are collected from a well known MRI scan centre. An opinion from the Radiologist is obtained before implementing the Image processing algorithms for

identification. The data base nearly consists of 51 images corresponding to three categories namely normal and abnormal (primary and secondary stages) respectively.

5.2 Pre-processing

The surface plot analysis is done to find out the variation in the intensity values for normal and abnormal breast images as shown in Figure 2 and 3 respectively. The future work for this stage will include noise removal and edge detection. Presently, the images if corrupted with noise are removed using a suitable filtering method. The edges are identified so as to delineate the normal tissue of the breast from the abnormal region.

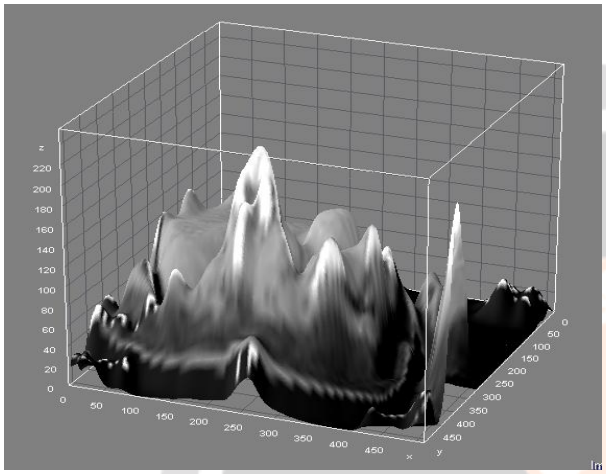

Figure 2. Histogram image of normal breast

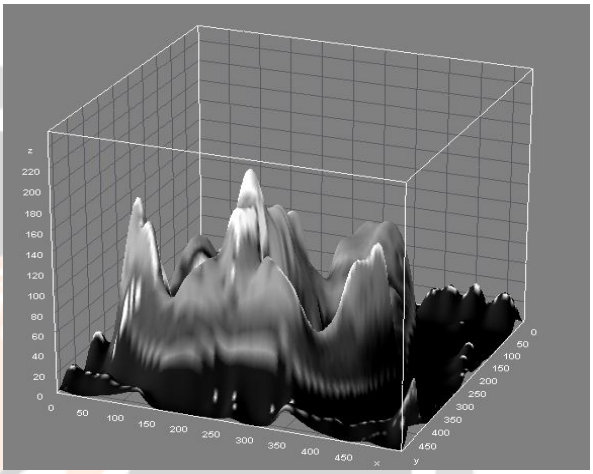

Figure 3. Histogram image of abnormal breast

6. Outputs for K-means clustering algorithm

The main idea is to define k centers, one for each cluster. These centers should be placed by trial and error basis because different location causes different results. So, the better choice is to place them as far as possible away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest center. When no point is left out, the first step is completed and an early grouping is done. At this point it is necessary to re-calculate 'k' new centroids for the clusters resulting from the previous step. After these k new centroids are generated, a new binding has to be done between the same data set points and the nearest new center. A loop has been generated. As a result of this loop it is noticed that the 'k' centers change their location step by step until no more changes are done for moving the centers. The suitable values are k=3 & 5. The presence of tumours on the breasts is evident from the outputs in Figure 4.

Figure 4. Output for the cancerous breast image using K means algorithm

7. CONCLUSION AND FUTURE ENHANCEMENT

A framework for identification of breast cancer from the mammographic images has been proposed. From the study of available literature, it is found that the classification to the problem of mammographic image analysis is rare. Hence it is strongly believed that the proposed system's performance can be scaled up and further enhanced by framing new function that is more adaptable to mammograms.

REFERENCES

[1] K. Polat and S. Genes, "Breast cancer diagnosis using least square support vector machine," Digit. Signal Process., vol. 17, no. 4, pp. 694–701, Jul. 2007.

[2] R. Manoharan et al., "Raman spectroscopy and fluorescence photon migration for breast cancer diagnosis and imaging," Photochem.Photobio., vol. 67, no. 1, pp. 15–22, Jan. 1998.

[3] R. T. Osteen, J. L. Connolly, M. E. Costanza, J. R. Harris, and D. F. Hayes, "Cancer of the breast," in Cancer Manual, 9th ed. New York: Am. Cancer Soc., 1996, pp. 320–339.

[4] [29]M. Siddiqui, M. Anand, P. Mehrotra, R. Sarangi, N. Mathur, Biomonitoring of organochlorines in women with benign and malignant breast disease, Environmental Research 98 (2) (2005) 250–257.

[5] A. Wroblewska, P. Boninski, A. Przelaskowski, M. Kazubek, Segmentation and feature extraction for reliable classification of microcalcifications in digital mammograms, Opto-Electronics Review 11 (3) (2003) 227–235.

[6] Guido M. te Brake and NicoKarssemeijer "Single and MultiscaleDetection of Masses in Digital Mammograms" IEEE transactions onmedical imaging, vol. 18, no. 7, July 1999.

[7] NohaYoussry, Fatma E.Z. Abou-Chadi, Alaa M. El-Sayad, "A neuralnetwork approach for mass detection in digitized mammograms,"ACBME, 2002.

[8] H. P. Chan, K. Doi, C. J. Vyborny, K. L. Lam, and R. A. Schmidt,"Computer-aided detection of microcalcifications in mammogramsmethodology and preliminary clinical study," Investigative Radiol.,vol. 23, pp. 664–671, 1988.

[9] N. Karssemeijer, "Recognition of clustered microcalcificationsusinga random field mode, biomedical image processing and biomedicalvisualization," Proc. SPIE, vol. 1905, pp. 776–786, 1993.

[10]Ryohei Nakayama and Yoshikazu Uchiyama "Development of NewFilter Bank for Detection of Nodular Patterns and Linear Patterns inMedical Images" Systems and Computers in Japan, Vol. 36, No. 13,2005.

[11] Songyang Yu and Ling Guan, "A CAD system for the automaticdetection of clustered microcalcifications in digitized mammogram 3films," IEEE Trans. Med. Imag., vol. 19, pp. 115-126, February 2000.

[12] R.C. Gonzalez, R.E. Woods, "Digital Image processing", Pretice Hall. 2007E. D. Pisano, C. Gatsonis, R. E. Hendrick, M. J. Yaffe, J. K. Baum, S. Acharyya, and J. B. Cormack, "Diagnostic performance of digital versus film mammography for breast-cancer screening," New England J. Med., vol. 353, no. 17, pp. 1773–1783, Oct. 2005.

[13] T. C. S. S. Andre and R. M. Rangayyan, "Classification of tumors and masses in mammograms using neural networks with shape and texture features," in Proc. 25th Ann. Int. Conf. IEEE EMBS, vol. 3. Sep. 2003, pp. 2261–2264.

[14] Sujatha, K. (2012) Flame Monitoring in power station boilers using image processing, ICTACT Journal on Image and Video Processing, Dr.M.G.R Educational & Research Institute, Indexed in IET Inspec.