TUMOUR GROWTH STAGE CLASSIFICATION FOR BREAST CANCER USING BACK PROPAGATION ALGORITHM

G.Sam Singh[1], R.Selva Priya[2], K.Sabana Asmin[3], S.Muthu Rajashree[4], R.Selva Lakshmi[5]

[1]M. Tech , Assistant Professor, Dept. of Electronics and Communication Engineering , Dr.Sivanthi Aditanar college of Engineering., Tiruchendur, India

[2,3,4,5] B.E, Student, Dept. of Electronics and Communication Engineering, Dr.Sivanthi Aditanar college of Engineering., Tiruchendur, India

Abstract

Breast cancer is the leading cause of death in women all over the world and can be diagnosed by various tests. Among that Mammogram is a goal standard technique. Identification of early signs of breast cancer requires high-quality images and skilled mammography interpretation. Over the years the help of technology such as data mining and machine learning can substantially improve the diagnosis accuracy. However, standard and accurate classification is still poses a great challenge for researchers. Difficulties are often found to find the sets of features that provide adequate distinctiveness required for classifying normal and malignant one. Therefore, the main aim is to propose a system for diagnosis, prognosis and prediction of breast tumours using Artificial Neural Network models based on Back Propagation Neural Network algorithm.

Index Terms: Adaptive Median, Back Propagation, Feature Extraction, K-Means, Support vector Machine.

I. INTRODUCTION

Cancer is not a single disease but a wide range of different diseases of which there were hundreds of types. Thus, breast cancer refers to the erratic growth of cells that originate in the breast tissues and nodules. . The group of these cells formed as a lump is called tumours. Tumors can be malignant or benign. Malignant tumors penetrate deep into nodes and destroy healthy among women between 40 and 60. It rarely occurs in men. Early diagnosis and sustained treatment were the biggest hurdles and main indication of person survival. To provide a better assistant for robust and reliable classification of breast tumour , Back Propagation algorithm can be a powerful tool for such distributed diagnosis. This provides an accurate result, whether the tumour is benign or malignant.

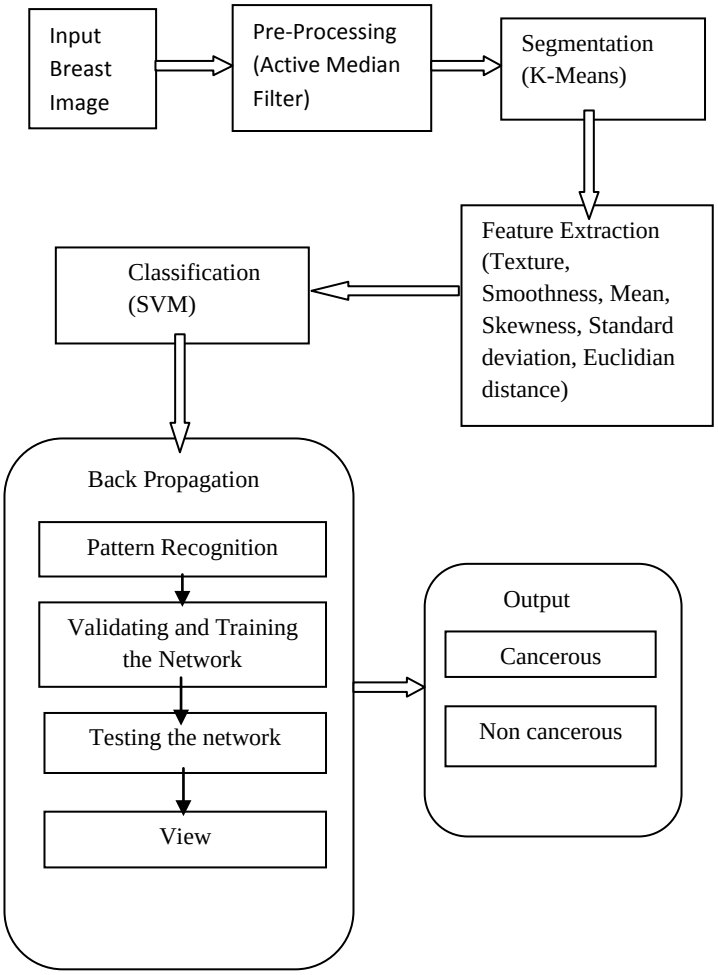
II. OBJECTIVE

- To detect the tumour using Ultrasound screened images
- To classify the growth stage of breast tumour using a Back Propagation algorithm.

III. PROPOSED SYSTEM

The proposed method comprises four stages: pre-processing, segmentation, feature extraction and classification stage. The pre-processing stage involves techniques that include noise removal, enhancement and segmentation process is done by K-means clustering. The feature extraction stage entails extraction of features like entropy, standard deviation, texture, Euclidean Distance, variance, tumour area, shape, concavity etc. Lastly, the classification stage involves the use of artificial neural network by back propagation to classify breast tissue into normal and abnormal.

IV. BLOCK DIAGRAM:



V. METHODOLOGY

STEP 1: Input raw Breast Ultrasound screened images.

STEP 2: Pre-process using Adaptive Median Filter

STEP 3: Segment using K-Means Clustering on the ROI.

STEP 4: Extract features like entropy, Standard Deviation, Euclidean Distance, Variance, Tumour area, Shape etc.

STEP 5: Classification of cancer stages using Back propagation Algorithm.

VI. PRE-PROCESSING:

The primary tasks of pre-processing step is obtaining visually informative images and ease the subsequent image processing and automated evaluation steps[10]. In this paper, the main focus is on adaptive median filter.

A. Adaptive Median Filtering

The adaptive median filtering is an advanced method compared with standard median filtering. The Adaptive Median Filter performs spatial processing and it will determine which pixels in an image have been affected by noise. It compares each pixel in the image to its surrounding neighbour pixels. The size of the neighbourhood is adjustable[1]. A pixel that is different from a majority of its neighbours, as well as being not structurally aligned with those pixels to which it is similar, is labelled as impulse noise. These noise pixels are then replaced by the median pixel value of the pixels in the neighbourhood that have passed the noise labelling test.

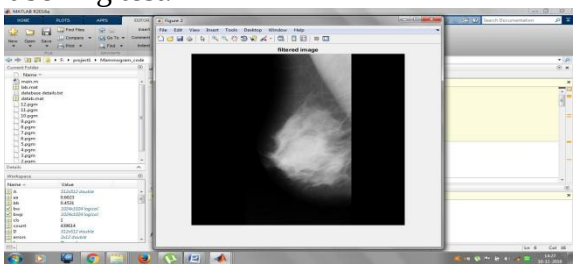


Fig 1: Filtered image output using adaptive median filter

VII. SEGMENTATION:

Image segmentation is a technique for dividing and discrediting an image into a number of segments in order to modify the depiction of an image into more significant and easier one to examine[4]. During the evaluation of mammogram images, it is necessary to delete background from breast profile.

A. K-means clustering

Clustering is a method to divide a set of data into specific number of groups. K-Means clustering is a popular method.. In K-Means clustering, It classifies a given set of data into k number of disjoint cluster. The algorithm consists of two separate phases. It calculates the k centroid in first phase and in the second phase it takes each point to the cluster which has nearest centroid from the respective data point. There are different methods to define the distance of the nearest centroid and one of the most used methods is Euclidean distance[9]. Then it recalculatethe new centroid of each cluster and based on that a new Euclidean distance is calculated between each center and each data point . It assigns the points in the cluster which have minimum Euclidean distance. Each cluster is defined by its member objects

and centroid of the image. The centroid for each cluster is defined as the point to which the sum of distances from all the objects in that cluster is minimized. So *K* - means is an iterative algorithm in which it minimizes the sum of distances from each object to its cluster centroid, over allclusters[6].

The algorithm for *k*-means clustering is following as:

1. Initialize number of cluster k andcentre for the image.
2. Calculate the Euclidean distance *d*, between the centre and each pixel of an image using the relation givenbelow.

$$d=p(x,y)-c_k$$

3. Assign all pixels of an image to the nearest centre based on distance*d*.
4. Recalculate new position of the centre after each pixel is assigned.
5. Repeat the process until it satisfies the requirement of given tolerance or errorvalue.
6. Reshape the clustered pixels intoimage.

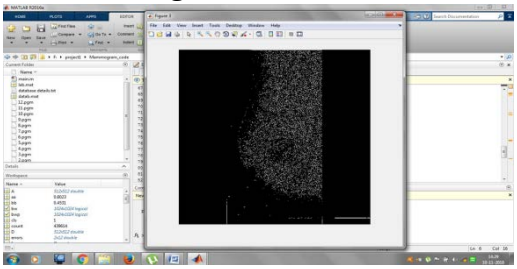


Fig 2: K-Means output

VIII. FEATURE EXTRACTION

Feature extraction a kind of dimensionality decrease that productively speaks to intriguing parts of a picture as a conservative component vector. This methodology is valuable when picture sizes are huge and a decreased component portrayal is required to rapidly total undertakings[7]. Feature identification, include extraction, and coordinating are regularly consolidated to take care of normal PC vision issues.

A. Mean:

Mean is described by adding all pixel values of an image divided by total no of pixels in an image.

$$M=\frac{1}{m*n}\sum_{x=0}^{m-1}\sum_{y=0}^{n-1}f(x,y)$$

B. Standard deviation:

It is the measure of probability distribution of an observed population and can serve as a measure of homogeneity.

$$SD=\sqrt{\frac{1}{m*n}\sum_{x=0}^{m-1}\sum_{y=0}^{n-1}f(x,y)}$$

C. Skewness:

Skewness can be quantified to define the extent to which a distribution differs from normal distribution.

$$S_k(x)=\frac{1}{m*n}\frac{\sum(f(x,y)-M)}{SD^3}$$

D. Euclidian distance:

The Euclidian distance is the straight line distance between two pixels.

$$D(a,b)=\sqrt{(bx-ax)^2+(by-ay)^2}$$

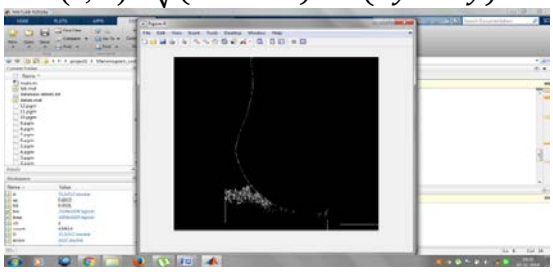


Fig 3: Feature extracted image output

IX. CLASSIFICATION:

The features or features subset are used by classifiers to classify images into normal and abnormal. The classification technique consists of the training phase and the testing phase. The training phase involves the analysis of features present in a known data by processing which precedes classification. The testing phase validate the input and provide the required output.

A. Support vector machine classifier (SVM):

SVM is a method which works accurately and efficiently with high dimensionality feature spaces. The original idea of the SVM is to construct a hyper plane as the decision surface in such a way that the margin of separation between the positive and negative samples is maximized in an appropriate feature space [2].Kernel functions are employed to perform the nonlinear mapping, which computes the inner product matrix, the so-called kernel matrix, For kernel-based methods, the kernel matrix acts as a bottleneck. All information are extracted from the Kernel matrix.

B. Back propagation algorithm

It is a supervised multi layer perception based learning method and generalization of delta rule. They can be used to find a pattern within the dataset to recognize the cancer automatically [9]. They works on self-learn logic principle as they require a database of inputs with desired output, making up a training set and update their weights accordingly. The confusion matrix determine how well the neural network fit into the dataset[3].

This algorithm performs four different operations

- Feed Forward computation
- Back Propagation to the output layer
- Back Propagation to the hidden layer
- Weight update

The training procedure for Back Propagation algorithm includes Feed forward and Back Propagation in two parts. During feed forward procedure feature values from the input layer propagates to the hidden layer and finally activation values of output node are generated. During Back Propagation procedure, the errors of the output node can be obtained through computing the difference between actual result and expected result[5]. Then the error signal from the output layer propagates backwards through the network and updates the each node's weight. The algorithm is stopped when the value of error function is sufficiently small. This is the rough and basic formula of Back Propagation algorithm.

X. CONCLUSION:

In this various levels of pre-processing techniques were studied and the stages of Breast images which includes detection, pre-processing and classified using Back Propagation Algorithm, Also to check its malignancy using MATLAB v2017. The values were also formulated from the various parameters in detection of malignant tumour. It clearly demonstrates the effectiveness of Back Propagation algorithm in the classification of breast tumours. Also this paper has the reference on the parameters of processing the tumour. Further work can be incorporated by using confusion matrix to find the accuracy level in detection of a malignant tumour.

REFERENCES

[1] Kesari Vermaa , Bikesh Kumar Singhb , A.S. Thokec Zelst, "An Enhancement in Adaptive Median filter for Edge Preservation", Procedia Computer Science 48 , Elseiver 2015.

[2] V.Vijaya Lakshmi,G.Krishnaveni," Performance Assessment by using SVM and ANN for Breast cancer Mammography image classification ",International Journal of Engineering Technology Science and Research IJETSR,Volume 4, Issue 9 September 2017

[3] Kalpana Kaushik,Anil Arora, "Breast Cancer Diagnosis using Artificial Neural Network"International Journal of Latest Trends in Engineering and Technology (IJLTET) , Vol 7 issue 2 July 2016.

[4] R.C. Gonzalez , Woods R.E ,Digital Image Processing, third ed , Pretice Hall, Englewood cliffs,NJ,2008.

[5] Paulin et al F,"Classification of Breast cancer by comparing Back propagation training algorithms" International journal on Computer science and Engineering, Vol. 3 No. 1 Jan 2011.

[6] S. M. Aqil Burney,Humera Tariq, K-Means Cluster Analysis for Image Segmentation,International Journal of Computer Applications (0975 – 8887) Volume 96– No.4, June 2014 .

[7] Prannoy Giri and Saravanakumar," Breast cancer detection using image processing techniques", Orient J.Comp. Science &Tech, An International open Free access Journal, Vol 10,391-399,2017.

[8] Pankaj Kumar Sharma ,"Implementation of Cancer Detection Based on Back Propagation" Vihar Univ, International Journal of Converging Technologies and Management(UCTM), Vol 3,Issue 1,2016.

[9] I.Salama, Gouda , M. B. Abdelhalim, and Magdy Abd-elghany Zeid."Breast Cancer Diagnosis on Three Different Datasets using Multiclassifiers",Breast Cancer (WDBC) 32.569 (2012).

[10]R.Ramani, N. Suthanthira Vanitha, and S. Valarmathy. "The PreProcessing Techniques for Breast Cancer Detection in Mammography Images."International Journal of Image, Graphics & Signal Processing 5.5 (2013).