BREAST CANCER DETECTION USING IMAGE PROCESSING AND MACHINE LEARNING

Mrs. K.Maheswari, Associate professor, Dept of ECE, Sanskrithi School of Engineering

T.S.Afreen
Dept of ECE
Sanskrithi school of engineering
Puttaparthi 515134

P.Seema Kousar
Dept of ECE,
Sanskrithi school of engineering,
Puttaparthi 515134.

B. Vijay Lakshmi
Dept of ECE
Sanskrithi school of engineering
Puttaparthi 515134

K.Pavitra
dept of ECE,
Sanskrithi school of engineering,
Puttaparthi 515134.

Abstract: — — Nowadays breast cancer is the frequent type of cancer in women which leads to death. The mammography and ultrasound are the common ways to the breast cancer . Our paper describes Machine Learning for identification of breast cancer using mammography imagesUltrasound and Elastography are the combined imaging techniques to separate benign and malignant breast lesions.support vector machine is a classifier which is used to classification of combined B-mode andElastography image.. Our project helps the physician to detect the breast cancer earlier.

Keywords: Breast Cancer, Elastography, Image Processing, B-mode (ultra sound), SVM

Introduction:

Breast cancer constitutes a significant threat on women health and is considered the second leading cause of their death. Breast cancer is a result of an abnormal behavior in the functionality of the normal breast cells. Therefore, breast cells tend to grow uncontrollably forming a tumor which can be felt as a lump in the breast.The earlier detection of breast cancer leads to decreasing the number of deaths by applying relevant prescription.. In general, palpation, ultrasound and mammography are the most common ways of diagnosis. However,Nowadays Elastography and ultrasound techniques are playing a major role for the detection of breast cancer.

The Computer-aided diagnosis is a tool using a merge of ultrasound(B-mode) and elastography images gives a high quality of output rather than other digitalimaging techniques due to its exact value of classification . Machine learninggains a data by using mathematical and statistical models.

Machine learning finds an important role in biomedical applications in which accuracy of measurements is a crucial factor.Consequently machine learning algorithms can helps todetect the breast cancer at its initial stage. Machine learning tools can establish mostpredicative features from the noisy and complex datasets.

Whereas, for exceptional accuracy the false negative and false positive should be reduced.

Literature Survey:

B.M.Gayathri , C.P.Sumathi and T.Santhanam [1] cancer detection of breast Machine Learning Algorithm –ASurvey Machine Learning Algorithms are developed to reduce the time taken for the process of diagnosis and to decrease the death .This paper encapsulates the survey on breast cancer diagnosis with

various machinelearning algorithms and techniques, which causes the improvement of accuracy and predictingthe cancer at earlier stage. We know The number of papers that areimplemented to diagnose the breast cancer by the help of these survey paper

S. Punitha, S. Ravi and M. Anousouya Devi [2] “Breast Cancer Detection in Digital Mammograms using Segmentation Techniques” Segmentation and Edge Detection algorithms Mammography is the adequate and method for the early diagnosis of the breast cancers through screening and accurate detection of masses, microcalcifications and architectural distortions. The breast cancer detection accuracy and efficiency can be increased by applying various image analysis techniques on digital mammograms on the dense regions of the breasts helping the radiologists to identify suspicious regions preventing unwanted biopsies and traumatic treatments. This paper focuses on the various image analysis techniques such as segmentation and edge detection algorithms for the detection breast abnormalities and compares its advantages and disadvantages.

Ashmithakhaleel khan &Naufal p [3] Automatic lesion detection based on Wavelet usingMean Filter, Adaptive Mean filter, Weiner Filter. In thispaper, the authors recommended a new method for the segmentation process that helps to detect the lesions.In this method Pre-processing and DWT of image can be done before segmentation.

Preprocessing plays a vital role in medical images to remove the unwanted noise and for increasing the intensity of image. Discrete Wavelet Transformation that is used for segmentation because it contains most accurate information of the input image.

Amandeep Singh, Amanpreetkaur [4] Breast tumor detection using segmentation technique from CT scan Crop Segmentation, Noise Reduction, Edge Detection, Global Thresholding .They presented the imaging technique is used to detect the information of the tissue, the bio medical images having the capability to help the physician in detecting the disease which is caused by abnormal growth of the cells. The evolving software and algorithms is used to analyze the images also helps the physician in their works. The clue and rigid is auto extracting of the small modules or tumor from the bio medical imagethat is identified at starting stage that gives the information of the early cancer. This work merges the image threshold, edge detection and segmentation helps in detection of cancer.

METHODOLOGY :

The implemented classification approach consists of three main consecutive stages. Firstly, the dataset is extracted by using image processing algorithms then data pre-processing procedures to the dataset are applied. Finally, machine learning techniques are used for classification.

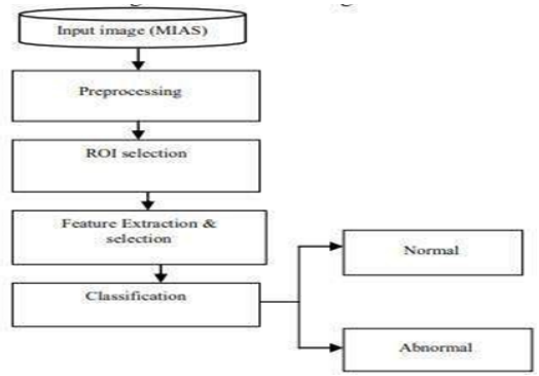


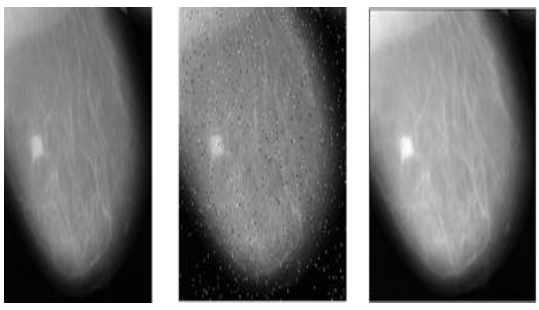
Image Acquisition:-

In UK research group, the organization named Mammography Image Analysis Society (MIAS) which have fashioned a digital mammography database. The images are in Gray scale file format (PGM – Portable Gray Map). The padded/clipped original MIAS Database (digitized at 50 micron pixel edge)

has been taken because its every image contains 1024 pixels x 1024 pixels known as the mini-MIAS database. The databasewhich weusedcalled as mini-MIAS database. as it contains complete information about abnormalities of each mammographic image.

Pre-Processing:

For early detection of breast cancer Digital mammogram is a better technique model. The high computational capabilities are required to process the image . To enhance the image the Computer image processing techniques will be applied . Thisproject paperdemonstrates about pre-processing is main step in the mammogram process because of its poor captured mammogram image quality. To correct and adjust the mammogram image for further study and processing the pre-processing technique is used. In pre-processing different types of filters and techniques are available. To improve image quality, and to remove the noise, to get the preserves edges within an image, to enhance and smoothen the image the filters are used. The experimental results summarizes that the noise removal can be done by using adaptive median filter. Because the adaptive median filter in noise removal process. and it estimates perfect PSNR values. The digital mammogram fig (a) is taken to process the adaptive median filter method to remove noise as show in fig (b) we get a high intensity output after filtering .The filtered image as shown in fig (c).



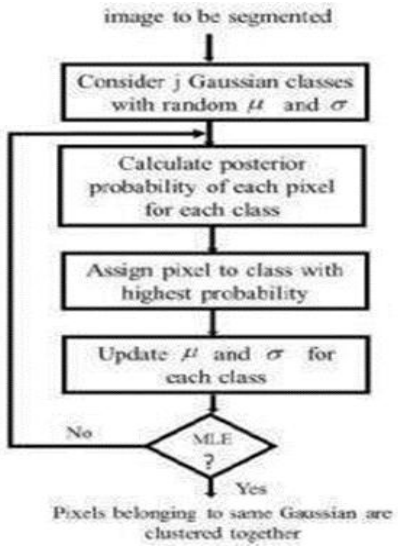
(a) (b) (c)

Fig: a. digital mammogram ,Fig: b. noisy imageFig: c. filtered output image.

ROI Selection:

In this stage, the output from pre-processing stage is used as input to find the ROI which is done using K means based GMM segmentation technique. K-means is iterative unsupervised clustering algorithm.Each cluster canidentify by its own centre point K-means finds a local minimum of the cost functionand converges.Euclidean distance metric is used as Dissimilarity measure to find distance between pixel and centroid of each class.

This is soft clustering algorithm. Each cluster is considered as a generative model with mean and variance. Mixture models are to estimate the parameters of probability distribution like mean and variance. The basic block diagram of K- means based GMM is shown in figure.



Feature Extraction :

Many features have been extracted from the mammograph. Extraction of feature plays important role in identifying disease.Feature extraction is used to reduce the redundant data from original image. In this output of the previous stage is passed to standard deviation filter. As the standard deviation has capability of measuring the variations, andedge

sharpening, the brightness level of edge image value can be changed by large image value.

Width to Height Ratio:

It is the ratio between the width (W) as the minor axis andthe height (H) as the major axis. The difference of Elastogram and theB-mode is calculated by using this equation.

$$D = |(WH)B - (WH)E|$$

• Area difference: The area difference is determined by the number of pixels in elastography of the tumor and no of pixels in b-mode images of the tumor.

$$A = |NB - NE|$$

Feature selection :

Based wrapper based method Feature selection byGA and Optimization of feature selection is done by Wrapper Method/ Hill Climbing, A random search technique, selecting the similar features from the available feature set by sing wrapper method. In the proposed algorithm genetic algorithm is integrated as a wrapper model.

The evaluated population can be tested in each generation with the help of termination algorithm. I The crossover, mutation and fitness computation steps are repeated when the break down occurs.

Table 1: Properties of Texture and co-occurrence matrix

Energy = $\sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1} p^2(i,j)$	Contrast = $\sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1}(i,j)^2 p(i,j$
Correlation = $\sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1}\frac{(i-\mu)(j-\mu)p(i,}{\sigma_i\sigma_j}$	Homogeneity = $\sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1}\frac{p(i,j)}{1+\lvert i-j\rvert}$
Mean: $\mu_i = \frac{1}{N}\sum_{j-1}^{N} f_{ij}$	SD: σ_i = $\left(\frac{1}{N}\sum_{N}^{1}(f_{ij}-\mu_i)^2\right)^{\frac{1}{2}}$
Skewness: γ_i = $\left(\frac{1}{N}\sum_{N}^{1}(f_{ij}-\mu_i)^3\right)^{\frac{1}{3}}$	Entropy: ε_i = $-\sum_{j-1}^{N}(f_{ij}-\log f_{ij})$

MACHINE LEARNING:

A probabilistic neural network (PNN) is a neural network, used for classification and pattern recognition problems. In PNN, The probability density function of each class can be estimated by using a non- parametric function and parzen window.

A PNN is an implemented by a kernel discriminant analysis. The kernel discriminant analysis is a statistical algorithm. The operations are performed in probability neural network with four layers.

Pattern layer, input layer, output layer, summation layer.

A PNN is an implementation of a statistical algorithm called kernel discriminant analysis in which the operations are organized into a multilayered feedforward network with four layers. Input layer, pattern layer, summation layer , output layer.

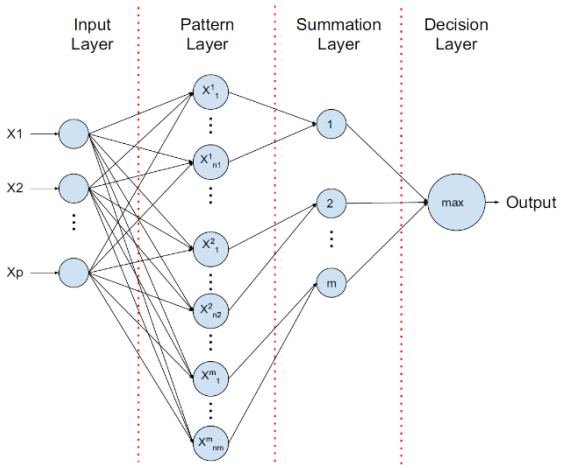
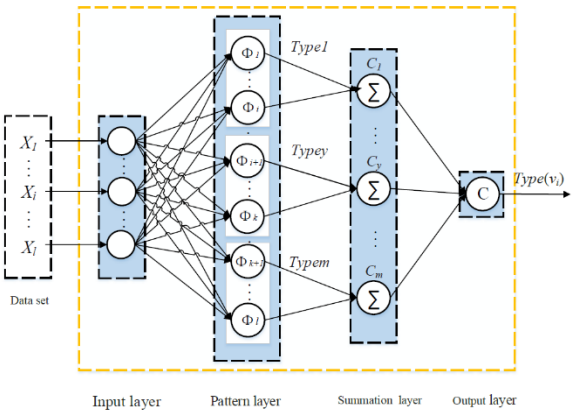


Fig : PNN Architecture



$$\Phi = \frac{1}{(2\pi)^{\frac{d}{2}}\sigma^d} e^{-\frac{(X-x_{ij})^T(X-x_{ij})}{2\sigma^2}}$$ The Euclidean distance between the feature vector of training sample X and radial center x_{ij}

$$v_i = \frac{\sum_{j=1}^{L}\Phi_{ij}}{L}$$ Averaging L patterns of class i

$$Type(v_i) = arg\max(v_i)$$ Select the class that gives maximum output in the summation layer

Fig : working of PNN classifier

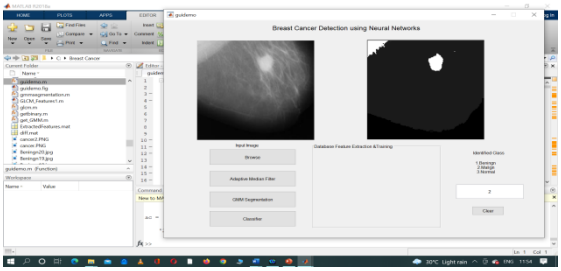


Fig b. output image for malign case cancer

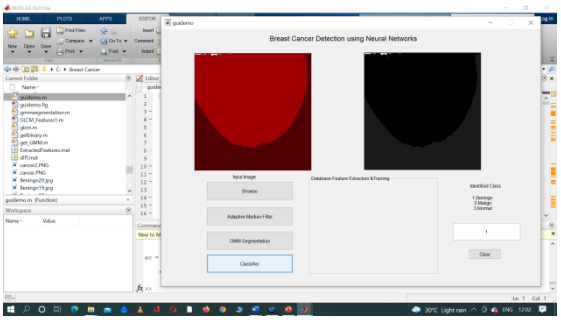


Fig c. output image of normal stage cancer

IV.SIMULATION RESULTS:

The decreasing number of mortality in patients can be possible by detecting a cancer at initial stage. Furthermore, the detection of cancer at an earlier stage allows treatment with breastconservation and hence a better quality of life. In these project, the proposed algorithm istested on the real life problem.Hence it is found that

neural networks improves the identification of breast cancer classification . Here are the outputimages of Breast cancer detection in different stages.

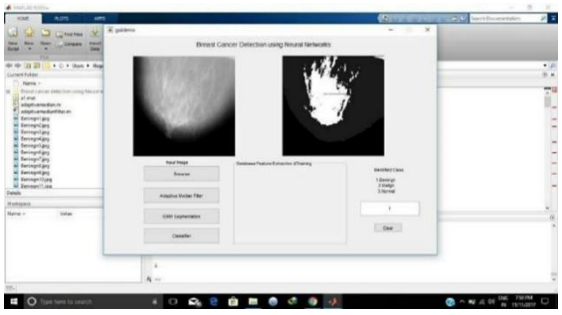


Fig a. output image of benign case cancer

V.CONCLUSION:-

The proposed algorithm is tested on a real life problem for the diagnosis of breast cancer.The study of this object is to create an effectivetool for building neural models to help us making a proper classification of variousclasses of breast cancer. Neural network will be approach drove by the learning algorithm works as well ,in terms of accuracy.

Using this model, an automated classification of various types of breast cancer was performed by avoiding the question of the expert concerning the recognition of cancer required, improving the identification of breast cancer classification. Through the result analysis, it was found that the computing time is reduced and the solutions quality is improved significantly by this proposed model.

ACKNOWLEDGMENT:

I would like to express my deep gratitude to Mrs. Maheshwari, Associate professor of Sanskrithi school of engineering Puttaparthi, my research supervisors, for their patient

guidance, enthusiastic encouragement and useful critiques of this research work.

REFERENCES:-

1. Survey by Indian cancer society, Indian Cancer Society (2018)
2. Padmanabhan, S., Sundarajan, R.: Enhanced Accuracy of Breast Cancer Detection in Digital Mammograms using wavelet analysis. IEEE Trans. Imag. Proc. (2012).
3. Spandana, P., Rao, K.M.M., Jwalasrikala,
4. J.: Novel Image Processing Techniques for Early Detection of Breast Cancer In Matlab and Lab View Implementation. IEEE Point of- Care Healthcare Technologies (PHT), Bangalore, India, pp. 16-18(2013)
5. H. Alto, R. M. Rangayyan and J. L. Desautels, Content-based retrieval and analysis of mammographic masses. Journal of Electronic Imaging, Vol 14, No.2,:023016–023016, 2005
6. Y. Tao, S. Lo, M.T. Freedman and J. Xuan, A preliminary study of content based mammographic masses retrieval. In Medical Imaging, 65141Z. International Society for Optics and Photonics, 2007.
7. B. Zheng, A. Lu, L. A. Hardesty, J. H. Sumkin, C. M. Hakim, M. A. Ganott and D. A. Gur, A method to improve visual similarity of breast masses for an interactive computer-aided diagnosis environment. Medical Physics, Vol 33, No.1, pp: 111–117, 2006.
8. C. Wei, Y. Li and P. Huang, Mammogram retrieval through machine learning within bi- rads standards. Journal of Biomedical Informatics, Vol 44, No.4, pp: 607–614, 2011.
9. Narvaez, G. Diaz and E. Romero, Multi view information fusion for automatic bi rads description of mammographic masses. In SPIE Medical Imaging, 79630A. International Society for Optics and Photonics, 2011 .
10. J. Liu, S. Zhang, W. Liu, X. Zhang and D. Metaxas, Scalable mammogram retrieval using anchor graph hashing. In 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI), pp; 898– 901, 2014.
11. W. Liu, J. Wang, S. Kumar and S. Chang, Hashing with graphs. In Proceedings of the 28th international conference on machine learning (ICML-11), pp; 1–8, 2011 .
12. G. Hong and N. Ashok, Breast cancer diagnosis using genetic programming