

SVM BASED APPROACH OF DETECTION AND CLASSIFICATION OF TUMORS IN MAMMOGRAPHY

M. Lakshmitha
M.E. VLSI Design
Akshaya College of Engineering and
Technology, Coimbatore

A. Abdul Hayum, M.E.,
Assistant Professor, Department of ECE
Akshaya College of Engineering and
Technology, Coimbatore

Abstract: One of the most lethal cancer threatening women all over the world is the breast cancer. It is medically defined as abnormal cell growth in the breasts. It is estimated that by the end of 2020 about 17 lakh women will be diagnosed to have breast cancer. After detected with carcinoma out of 100 women in US only 89 women can survive for at least 5 years. The diagnosis of carcinoma has different procedures evolved from various technologies. But the diagnosis results are not exact all the time and the major drawback reported to be is false-positives. The advancement of image processing and artificial intelligence had led to diagnosis results of cancer masses more accurate and in short time. This project provides a solution for early detection by processing the mammogram images to locate the carcinoma and to specify the stage of cancer. It employs the machine learning approaches to perform the classification after the identification of the location of cancer masses. To classify the cancer stages the unsupervised learning technique named as support vector machines is implemented.

Keywords: Breast cancer, SVM, classification, Expectation-Maximization

I. INTRODUCTION

The abnormal growth of cells forms masses medically defined as a tumour. This tumour may or may not be cancerous. The tumour which does not spread to other parts is called benign and when it is more lethal it is malignant. The symptoms of a tumour arrive very late making it difficult to cure [12]. Thus, doctors always recommend to have periodic screening of breasts to ensure early detection.

Initially the cancers were more unpredictable such that the cause or treatment was nor liable. The screening methods were very

traditional sometimes yielding inaccurate results. Also, the process of diagnosis took long time to end up with results.

Cancers may occur in all body parts and the tests done for screening differs according to the type of cancer. For most of the cancers, doctors may recommend a blood work but when a breast cancer is suspected then the screening is done via imaging and biopsy. Before recommending any tests, the patients will be asked to undergo physical examination. When they feel any tumour then the successive procedures are done. Mammogram is the imaging technique that is widely used. It is an X-ray passed through the breasts to detect the presence of lumps, mostly mammogram can detect small masses before they are fully grown to be detected by a physical test. This is a screening mammogram. When there is something suspicious then a diagnostic mammogram is done.

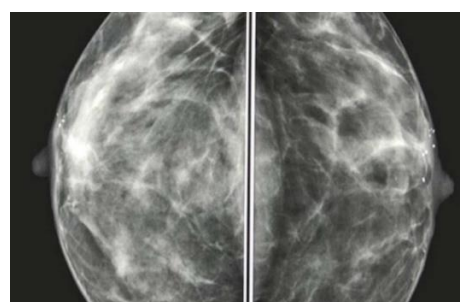


Fig.1.1. Mammogram image

Diagnostic Mammogram examines a portion of breasts in a greater detail whereas screening mammogram examines the whole breasts.

The pathology report describes the type of cancer mentioning some of the attributes such as size of the tumour, rate of growth and other factors that influence the treatment. So far it is the responsibility of the doctors to examine the imaging results and find the tumour location but with the emerging AI technology makes it possible to develop a system that automatically detects and classifies the cancer

as benign or malignant reducing the manual work thus minimising human errors.

II. METHODOLOGY

The detection of cancer from the mammogram images uses many techniques from image processing and machine learning. The method of training a machine by incorporating a large amount of data and making it perform the given task is machine learning. This method involves a series of steps from collecting the required data followed by pre-processing and classification.

a. Dataset:

The dataset is a collection of handwritten digits that are used in image classification, clustering and other processing methods. As far as breasts are concerned, we have the major characteristics of the breast and the tumour as attributes. The attributes we consider here are

- Radius
- Texture
- Perimeter
- Area
- Smoothness
- Compactness
- Concavity
- Concave points
- Fractal dimension

b. Pre- processing:

It is the most important step to be done before working with the dataset. It is the process of finding, removing and replacing the bad or missing data. The pre-processing ensures the data is clear and efficient for further operations as it enhances the quality of the data to run our model.

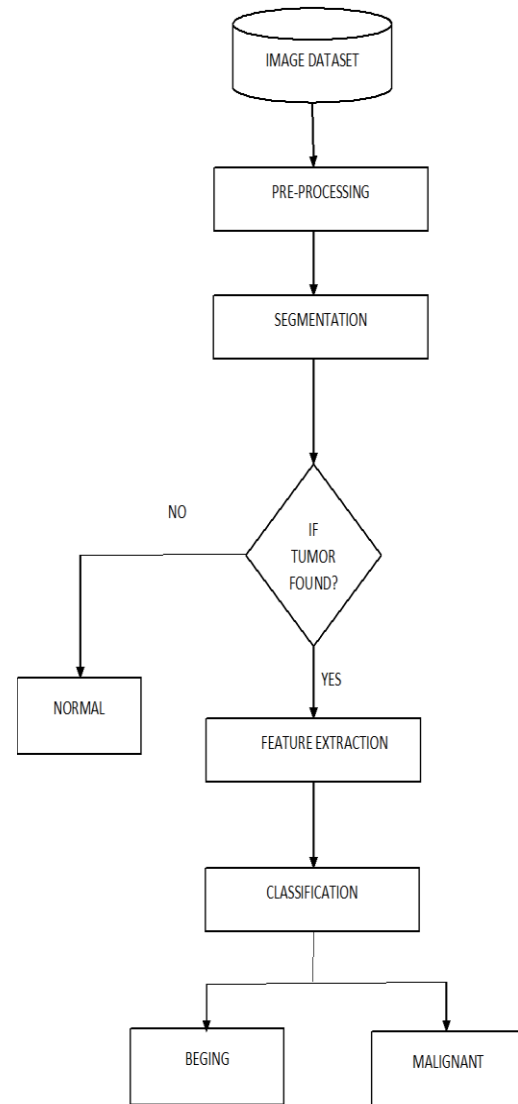


Fig 2.1. Flow chart of breast cancer classifier

c. Image Noise Removal:

In pre-processing section, the input image may be in different size, contains noise and it may be in different colour combination. These parameters need to be modified according to the requirement of the process.

Image noise is most apparent in image regions with low signal level such as shadow regions or under exposed images.

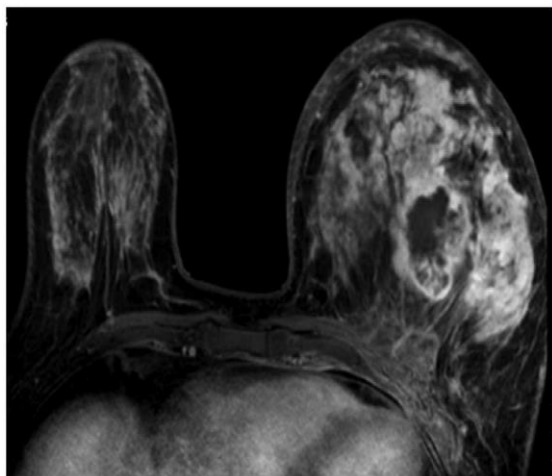


Fig 2.2 Input image

There are so many types of noise like salt – and – pepper noise, film grains etc., All these noises are removed by using filtering algorithms. Among the several filters, Weiner filter is used.

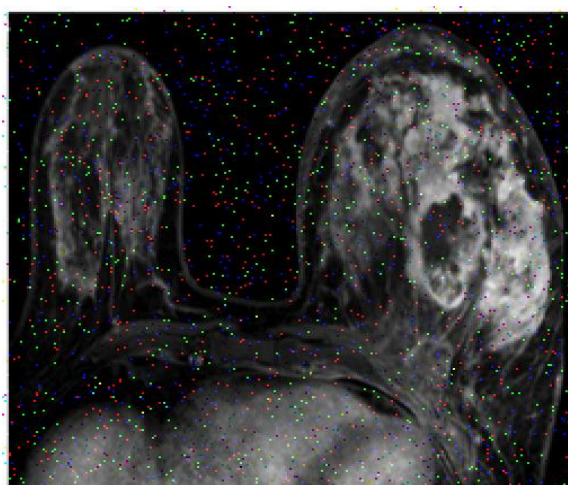


Fig 2.3 Noisy image

In pre-processing module image acquired will be processed for correct output. Pre-processing was done by using some algorithm. Bharat Anusha et.al (2018) proposed that for all images the pre-processing should be done so that the result can be obtained in the better way.

To find out the transformation between two images precisely they should be pre-processed to improve their quality and accuracy of result

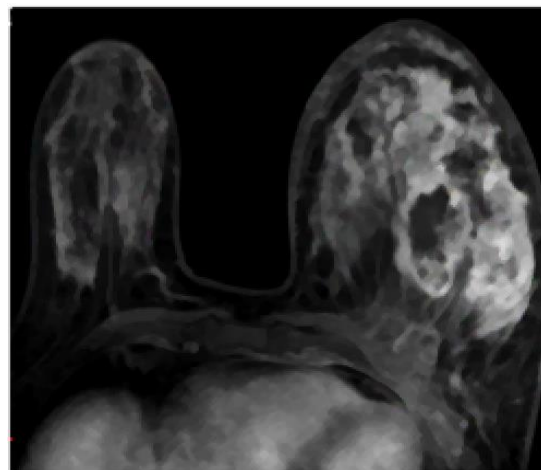


Fig 2.4 Noise Removed Image

d. Segmentation:

Image segmentation is a technique in which the image is broken down into multiple segments also known as image objects or pixels. After the data pre-processing the image is left to undergo segmentation in order to transform the data in a meaningful form. There are different segmentation techniques, here the method of Expectation-Maximization is used

EXPECTATION – MAXIMIZATION:

The EM algorithm is an efficient iterative procedure to compute the Maximum Likelihood (ML) estimate in the presence of missing or hidden data. Each iteration of the EM algorithm consists of two processes: E-step and M-step

Algorithm:

E-Step: Perform probabilistic assignments of each data point to some class based on the current hypothesis for the distributional class parameters – update variables

M-Step: Update the hypothesis for the distributional class parameters based on the new data assignments – update hypothesis

e. Pattern Recognition:

When the image is segmented then the next step is the pattern recognition which is done by two processes, feature extraction and classification.

It is the process of reduction of the image dimensions and to extract some essential features that can be used in classification. Two features are considered statistical and size-based features.

- i. Statistical feature extraction: This is used to derive the boundary of the image based on the features of skew, mean, variance, standard deviation.

- ii. Size based feature extraction: A gray level co-matrix is constructed with minimum 0 and maximum 2. The size-based features are acquired from control, energy, entropy and homogeneity of the GLCM.

III. EXPERIMENTAL RESULTS AND DISCUSSION

Once the features are extracted from the respective images they are used for the classification. Machine learning comes into picture, we use the supervised learning method for classifying the type of cancer.

Cancers are distinguished into two categories as benign and malignant. The mechanism used for this classification is Support Vector Machines. This classifier is preferred because of the need to classify the MRI images into 2 categories.

IV. SUPPORT VECTOR MACHINES

Support Vector Machines (SVMs) are a relatively new supervised classification technique to the land cover mapping community. The training data is distributed in space and when a new data enters this is marked inside the space in the right position where the space is separated by a hyper plane. This plane is the important factor of a SVM classifier. It divides the maximum and minimum data points that are distributed in space.

The data set acquired is grouped into training set and the test set. The hyper plane attempts to separate the classes so that each fall on either side of the plane, and by a specified margin. Ali Amna et.al (2018) discussed about the cost function as there is a specific cost function for this kind of model which adjusts the plane until error is minimized.

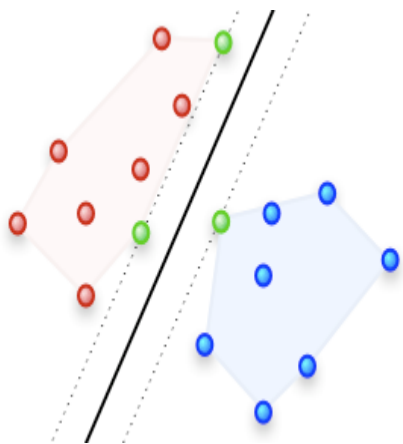


Fig 2.5. Hyper plane separating two classes

The dataset is trained with the provided train data, where the data is divided into two sections. This is done by the function 'svmtrain()'. When the train procedure is complete the Euclidean distance is found with the training and the test set. A matlab function 'cvEuclidist (Trainingset, testset)' is used to find the distance.

The MRI images that are used in classification are pre-processed so that all the images are of same standards and it becomes easier for classification and provides effective results. The images are thus classified and a pop-up message appears showing the stage of tumour such as malignant or benign.

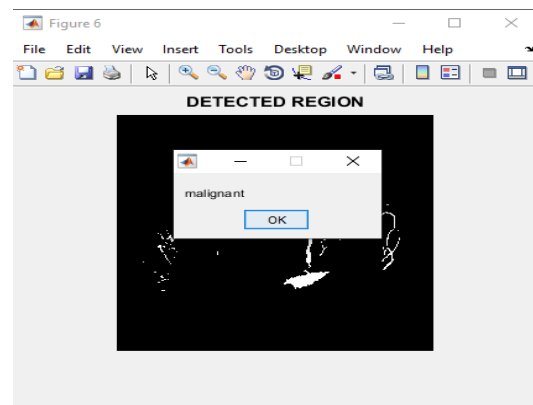


Fig.2.6. Result window

V. CONCLUSION

Breast cancer has serious health concerns among women all over the world. The mortality rate is very high in breast cancer patients. But with suitable and efficient screening, the mortality rate can be reduced by an early detection. The initial step based on gray level information of image and segment the breast cancer. Each tumour region is extracted by the statistical and GLCM feature extraction method. Finally, the SVM classifier is used for classification. Finally, the proposed model seems suited for control parameter setting of the machine learning algorithms in one side and automated breast cancer diagnosis on the other side.

VI. REFERENCES

[1] Wang Chuin-Mu, Xiao-Xing Mai Geng-Cheng Lin, Chio-Tan Kuo, 2018, "Classification for Breast MRI Using Support Vector Machine", *IEEE 8th International Conference on Computer and Information Technology Workshops*.



- [2] Ali Amna, Khan Umer, Ali Tufail, Kim Minkoo, 2010 “Analyzing Potential of SVM based Classifiers for Intelligent and Less Invasive Breast Cancer Prognosis”, *Second International Conference on Computer Engineering and Applications*.
- [3] Kavitha K R, Rajendran Syamili G, Varsha J, 2016, “A Correlation Based SVM-Recursive Multiple Feature Elimination Classifier for Breast Cancer Disease using Microarray “, *Intl. Conference on Advances in Computing, Communications and Informatics (ICACCI), Sept. 21-24, 2016, Jaipur, India*.
- [4] Yang Xiufeng, Hui Peng, Shi Mingrui, August 2013, “SVM with Multiple Kernels based on Manifold Learning for Breast Cancer Diagnosis “, *Proceeding of the IEEE International Conference on Information and Automation Yinchuan, China*.
- [5] Yoon Sejong and Kim Saejoon, 2009, “Multiple SVM-RFE using Boosting for Mammogram Classification” , *International Joint Conference on Computational Sciences and Optimization*.
- [6] Zhou Bin, Xuan Jianhua, Zhao Hongzhi, Gloria J. Chepko, Matthew T. Freedman, Yingyin Zou Kevin, 2007, “Polarization Imaging for Breast Cancer Diagnosis Using Texture Analysis and SVM”, *IEEE/NIH Life Science Systems and Applications Workshop (LISSA 2007)*.
- [7] Jaffar Arfan .M, Ahmed Bilal, Hussain Ayyaz, Naveed Nawazish, Jabeen Fauzia and Anwar M. Mirza, 2009, “Multi domain Features based Classification of Mammogram Images using SVM and MLP“, *Fourth International Conference on Innovative Computing, Information and Control*.
- [8] Khourdifi Youness, Bahaj Mohamed, 2018, “Feature selection with Fast Correlation-Based Filter for Breast cancer prediction and Classification using Machine Learning Algorithms “, *International Symposium on Advanced Electrical and Communication Technologies (ISAECT)*.
- [9] Bharat Anusha, Pooja N, Reddy Anishka R, 2018, “Using Machine Learning algorithms for breast cancer risk prediction and diagnosis”, *IEEE Third International Conference on Circuits, Control, Communication and Computing*.
- [10] Siyabend Turgut, Datekin Mustafa and Ensari Tolga, 2018, “Microarray Breast Cancer Data Classification Using Machine Learning Methods”.
- [11] Burges, A Tutorial on Support Vector Machines for Pattern Recognition: <http://research.microsoft.com/pubs/67119/svmtutorial.pdf>
- [12] World Health organization statistics: www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/