EasyChair Preprint

№ 2930

A Computer Vision Based Framework for Detecting Breast Cancer Using Mammography

R. Benazir Begam, K. R. Sreegayathry, S. Srinagavaishnavi and Srinithi Venkatesh

March 11, 2020

A COMPUTER VISION BASED FRAMEWORK FOR DETECTING BREAST CANCER USING MAMMOGRAPHY

Mrs.R.Benazir Begam,
Assistant Professor,
Department of ECE,
Rajalakshmi Engineering College, Chennai,
Tamilnadu
benazirbegam.r@rajalakshmi.edu.in

K.R.Sree Gayathry [1], S.Srinagavaishnavi [2],
Srinithi Venkantesh [3] , Student,
[123]Department of ECE,
[123]Rajalakshmi Engineering College,
Chennai, Tamilnadu
sreegayathry.kr.2016.ece@rajalakshmi.edu.in
srinagavaishnavi.s.2016.ece@rajalakshmi.edu.in
srinithi.venkatesh.2016.ece@rajalakshmi.edu.in

ABSTRACT

Mammography is a specialized medical imaging that uses a low-dose x-ray system to examine the breasts. A mammogram is a mammography exam report that helps in the detection and diagnosis of breast diseases in women at an early stage. This project proposes to classify mammography breast scans into their respective classes and uses attention learning to localize the specific pixels of malignancy using a heat map overlay. The attention learning model is a standard encoder-decoder circuit wherein convolutional neural networks perform the encoding and recurrent neural networks perform the decoding. Convolutional neural networks enable feature extraction from the mammography scans which is thereafter fed into a recurrent neural network that focuses on the region of malignancy based on the weights assigned to the extracted features over a series of iterations during which the weights are continuously adjusted owing to the feedback received from the previous iteration or epoch. Mammography images are equalized, enhanced and augmented before extracting the features and assigning weights to them as a part of the data preprocessing procedures. This procedure would essentially help in tumor localization in case of breast cancers.

Keywords— Back propagation Neural Network (BPNN); Breast Cancer Detection; GLCM; Wavelet transform.

INTRODUCTION

The major health issue that arises these days had led to much advancement in the medical field, yet there are certain diseases which remains incurable. Certain diseases even seem to be more challenging to guess what it is in the earlier stages. Among these many diseases (stroke,

cancer, heart attack, viral hepatitis, chronic liver diseases, coronary artery disease etc.), the death rate due to cancer is becoming more and more each year. Once a tissue is affected by radiation, it takes almost 15 years to turn into a cancerous tissue. Even though we have this time period for perfect curability, many of the patients who have this kind of basic abnormalities don't take it as a big deal since the symptoms are not indicating cancerous growth. Nearly 60% patients with breast cancer are diagnosed in advanced stages.

The growing cancer burden is due to a number of factors, including residence growth and ageing as well as the changing commonness of certain causes of cancer connected to social and economic progress. There are so many cancers, among which Breast cancer is the most important cause of cancer death among women.

II. LITERATURE SURVEY

The literature survey conveys the proposed research work on measures adopted in order to carry out mammography scans classification for the diagnosis of malignancy associated with each scan.

Classification based on multiple association rule (CMAR): This method proposed a neural network classifier. Nodes in input layer are representative of one characteristic from each rule. The number of input nodes were equivalent to the number of characteristics, number of hidden nodes were equivalent to the number of rules and number of output nodes were equivalent to the number of mammography classes. Backpropagation is used for learning the network model with a 10-fold cross validation and sigmoid activation function. The sensitivity and specificity are calculated to plot the ROC curve in order to measure the performance of the classifier. Classifier based on multiple association rule with neural network yields an accuracy of 84.5% albeit in a smaller dataset with huge bias in the training instances available under each class [1].

Two-dimensional discrete wavelet transform classifier: Feature vectors are generated by grey level co-occurrence matrix to all detailed co-efficients from 2D- discrete wavelet transformation of the region of malignancy. Derivation of relevant features are done through f-test and t-test of random sampling. Area under receiver operating characteristic (ROC) curve is better. Accuracy is abnormally high due to absence of measures to avoid overfitting such as a dropout. Assignment of weights to the features representing the region of malignancy or interest & adjustment of these weights from the attained feedback is absent thereby denying the process of attention [2].

Ada boost based multiple support vector machines for recursive feature elimination (SVM-RFE) for mammogram classification: it is a wrapper variant feature selection procedure. Ranking of features is done by the SVM-RFE by calculating information gain during iterative backward feature elimination. In each iteration the SVM-RFE sorts features in working set in order of distinctions between the objective functions and removes the feature with minimal distinction. Ensemble method is used to combine the SVM-RFE with the boosting approach to carry out replication of original dataset by random resampling to gain a higher improvement of this ensemble each replicate is different from one another to attain maximal classification accuracy. Ensemble method of integrating multiple SVM-RFE with AdaBoost performs great on the classification paradigm but a simpler visualization of the region of interest on the scans in an unbiased mammography dataset is missing [3].

Classification of normal and abnormal patterns for diagnosis of breast cancer in digital mammograms in the DDSM dataset performs feature extraction using a grey level co-occurrence matrix (GLCM). This is followed by offering the GLCM as an input to the neural network in order to train the classifier and test its performance on the test data allocated from the DDSM dataset. It produced a classification output between one of the two classes (cancer positive and normalcy). The classification report had a large number of true positives but an even larger number of false negatives which indicated that the classifier had a lot of misclassifications of cancer positive scans as normal ones. Thereby the classifier results were not reliable [4].

III. METHODOLOGY

The objective of the paper is to propose a methodology which could classify the mammography scans into one of the given classes and localize the malignancy region on the test samples using attention learning model

Flow Diagram Of Malignancy Detection

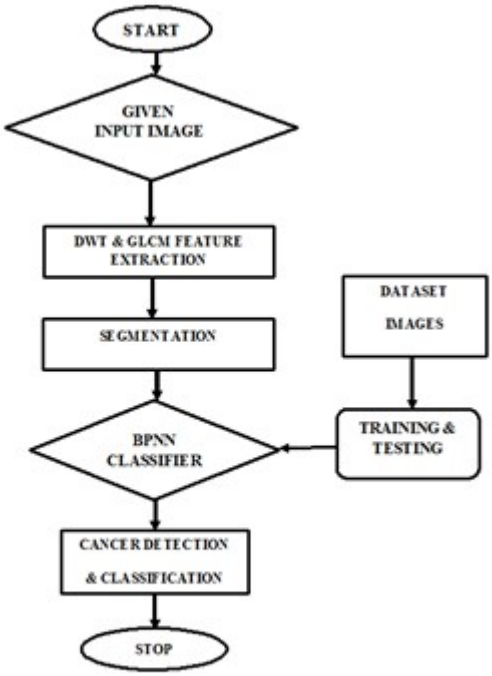


Figure 3.1 Flow Chart of Breast Cancer Detection

Data Acquisition

The data acquisition process was done by this project because there are no publicly available datasets pertaining of mammography dataset. However, dataset images accurately represent the state of cancer affected and normal. This is unlikely in fast moving vehicles are not treated because images are not clearly and with blurring and shakable things. Therefore, hand collected own dataset of images, which will plan on making a public dataset in future. The dataset contains images of traffic dataset across two classes with about 40-50 images, totaling about 140/150 images. Some of them dataset images the figure 3.2

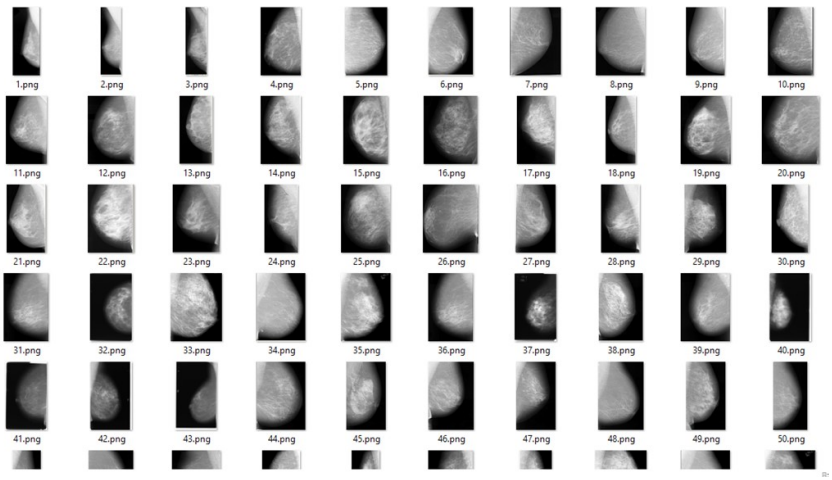


Figure 3.2 Input samples Dataset

Preprocessing

In this medical diagnosis systems, X-rays are used as prediction tool in mammography for the inspection of human breast. These images are recorded as specialized images which are then experiential by radiologists for any possible abnormality and noise removal of input sample image is removed with the help of median filter.

Bilateral filter has Non-linear nature and have fringe protected and noise decreasing flatted types of pictures. It restores the strength of every picture quantified usual strength size of neighbor pixels. This size of pixel represents the Gaussian distribution. Critically the quantities rely on not particularly norm closeness of true pictures, still we use radiometric closeness. This method protects keen portion.

This method provides input-output straight relationship and reshape the picture or reduce the noise. If we proposed for this method, we get un sharp screening. This filter exclusively applies sharp boundary and decrease contrast. The one advantage a Gaussian filter has over a median filter is that it's faster because multiplying and adding is probably faster than sorting.

Feature Extraction

The high information contained sub-band is selected from the DWT, this sub-band is taken as an input for feature extraction.

In Grey-Level Co-occurrence Matrix (GLCM) ellaporate the co-ordinate difference between each intensity presented in different gray levels i and j at specific angle θ. The following properties of GLCM is estimated

1.Power estimating uniformity of grey scale distribution,

2. Entropy measuring randomness;

3.Standard deviation;

4.Mean;

5. Texture feature.

Finally, 5 The characteristics are removed for the given input pictures and the given values are stored in an array for classification purpose.

E. Segmentation

In segmentation method transfer digitalize picture into sequence of pictures, that kind of pictures are collection of pixel that we refer as perfect pictures. This method presented in simpler and vary the presentation of a picture into variable form that represent more understandable and handled very lucid way to analyze. That notable pictures are removed from original pictures.

If we notice that after considering segmented process the blur area that is the abnormality exist in the infected portion will appear clearly. If we consider the depth of subtraction operation, succeeding the blur picture and the threshold is selected. In this pictures enlargement choose threshold to reduce intra class difference of saturated gray image.

K-means clustering is a type of unsupervised learning, which is used when you have unlabeled data (i.e., data without defined categories or groups). The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable *K*. The algorithm works iteratively to assign each data point to one of *K* groups based on the features that are provided. Data points are clustered based on feature similarity. The results of the *K*-means clustering algorithm are:

1. The centroids of the *K* clusters, which can be used to label new data
2. Labels for the training data (each data point is assigned to a single cluster)

Rather than defining groups before looking at the data, clustering allows you to find and analyze the groups that have formed organically. The "Choosing K" section below describes how the number of groups can be determined.

E. Backpropagation Neural Network

BPNN is a learning algorithm used for training the artificial neural network. Mainly, the backpropagation algorithm consists of two stages i.e. - forward pass and backward pass through which the various layers or sections of the network are trained.

The algorithm for BPNN can be given as follows:

1. The first step is to initialize the weights randomly.

2. In second step, an input vector pattern is provided to the network.

3. Check the outputs of the network by directing input signals forward.

4. Calculate $\partial j = (yj-dj)$ for all output neurons, where dj is the desired output of neuron j and yj is its current output: $yj= g(\sum iwijxi) = (1 + e-\sum iwijxi)-1$, assuming a sigmoid activation function.

5. For remaining neurons (from last hidden layer to first), compute $\partial j = \sum kwjkg'(x)\partial k$, where ∂k is the ∂j of the succeeding layer, and $g'(x) = yk(1 -yk)$.

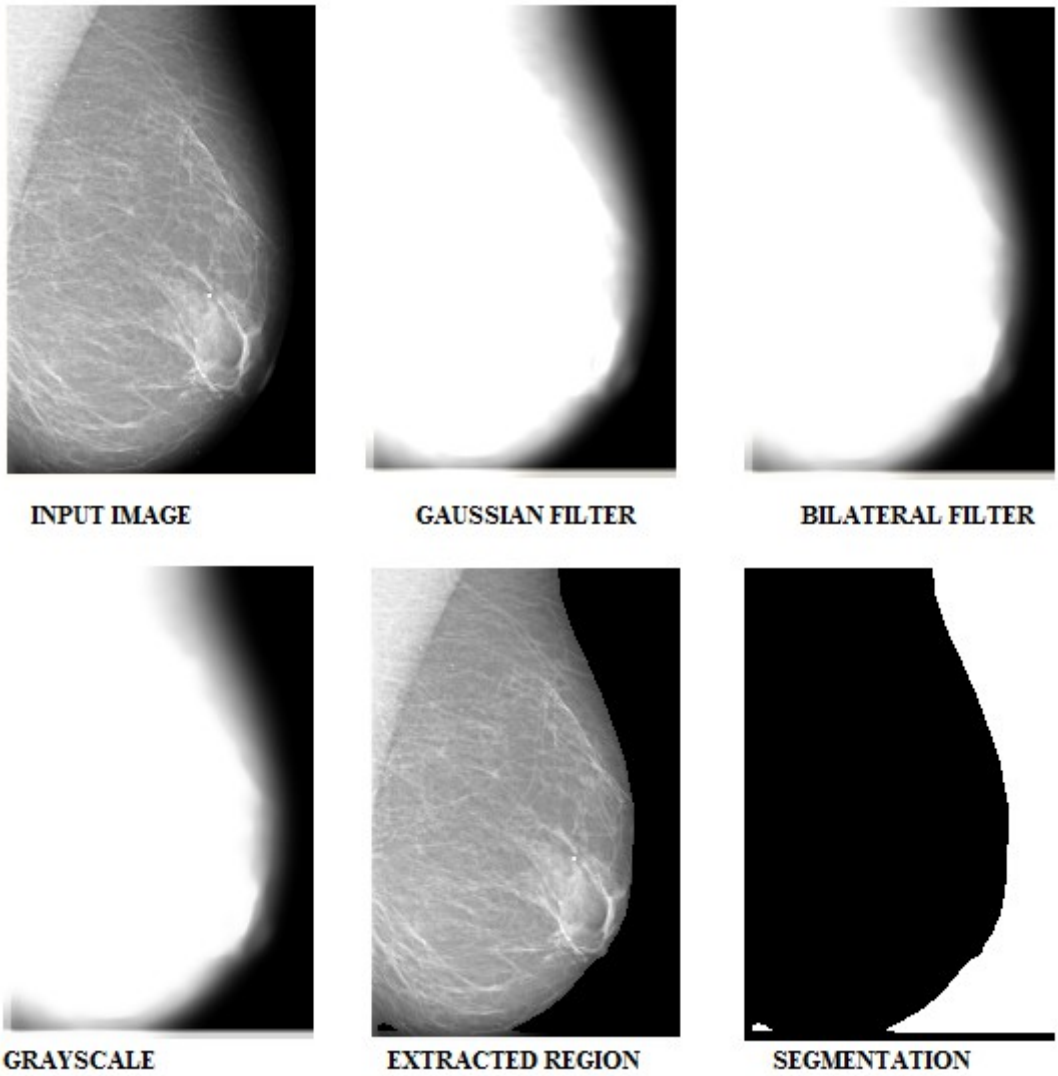
6. Update the weights according to: wij(t + 1) = wij(t) - yiyj(1-yj)∂j , where, wij is a parameter called the learning rate. 7. Go to step 2 for a some number of iterations, or until the error is decreased to a pre specified value.

IV. RESULTS AND DICUSSION

In this paper, detection of breast cancer was done using BPNN algorithm. There are several other algorithms are also available such as ANN, SVM, KNN etc. But as per the accuracy level is concerned, BPNN provides a better result. The average accuracy was improved to 89.77% and high accuracy changed to 96%.

From the above table it is inferred that, using DNN algorithm accuracy and level of sensitivity is increased than another algorithm. Thus, the earlier detection of breast cancer is detected without any physical contact. This method avoids destruction of the part being tested.

This paper uses the wavelet transformation and GLCM feature extraction. These features were useful to distinguish the maximum number of samples accurately. Finally, the simulated results show that used methodologies provides better classifier rate with minimum error rate for all test samples.



V. CONCLUSION

This paper focus on the earlier diagnosis of breast cancer, as the detection bring about the success of about 96% accuracy by the use of BPNN algorithm for classification. By the use of python as a software platform, training and the computational time has been reduced to a greater extent than others. This project ensures the greater level of detection of breast cancer at earlier stage, by which mortality rate of cancer affected person can be reduced and earlier diagnosis would increase the life time of a patient by giving them a right treatment at a right stage.

REFERENCES

[1] Zaiane, Osmar R., Maria-Luiza Antonie, and Alexandru Coman. "Mammography classification by an association rule-based classifier." Proceedings of the Third International Conference on Multimedia Data Mining. Springer-Verlag, 2002.

[2] Beura, Shradhananda, Banshidhar Majhi, and Ratnakar Dash. "Mammogram classification using two-dimensional discrete wavelet transform and gray-level co-occurrence matrix for detection of breast cancer." Neurocomputing 154 (2015): 1-14.

[3] Yoon, Sejong, and Saejoon Kim. "AdaBoost-based multiple SVM-RFE for classification of mammograms in DDSM." BMC medical informatics and decision making 9.1 (2009): S1.

[4] Nithya, R., and B. Santhi. "Classification of normal and abnormal patterns in digital mammograms for diagnosis of breast cancer." International Journal of Computer Applications28.6 (2011): 21-25.

[5] Balleyguier, Corinne, et al. "BIRADS™ classification in mammography." European journal of radiology 61.2 (2007): 192-194.

[7] 1. “Latest Global Cancer Data: Cancer burden rises to 18.1 million new cases and 9.6 million cancer deaths in 2018”, International Agency for Research on Cancer, World Health Organization, 12 September 2018.

[8]. M.M.Mehdy, E.E.Shair and P.Y.Ng, “Artificial Neural Networks in Image Processing for Earlier Detection of Breast Cancer”, Hindawi, Computational and Mathematical Methods in Medicine, Volume 2017, Article ID 2610628.

[9]. Vishnukumar K.Patel, Prof.Syed Uvaid and Prof.A.C.Suthar, “Mammogram of Breast Cancer Detection Based Using Image Enhancement Algorithm”, Internationa Journal of Engineering Technology and Advanced Engineering, Volume 2, Issue 8, August 2012.

[10]. Melanie A. Sutton, “Breast Cancer Detection Using Image Processing Techniques”, IEEE International Conference on Fuzzy System- Febraury 2000.

[11]. Bhagyashri k Yadav, Dr. Prof. M. S. Panse, “Virtual Instrumentation Based Breast Cancer Detection and Classification Using Image- Processing”, International Journal of Research and Scientific Innovation (IJRSI), Volume V, Issue IV, April 2018.

[12]. Moh’d Rasoul A Al-Hadidi, Mohammed Y. Al-Gawagzeh, “Solving Mammography Problems of Breast Cancer Detection Using Artificial Neural Networks and Image Processing Techniques”, Indian Journal of Science and Technology, Vol 5, No.4 (April 2012), ISSN: 094-6846.