International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 4, Issue 9
September 2017

Performance Assessment by using SVM and ANN for Breast cancer Mammography image classification

I.V.Vijaya Lakshmi[1]
Research Scholar
Sri Venkateswara University, Tirupathi
G.Krishnaveni[2]
Research Scholar, SPMVV, Tirupathi

ABSTRACT

Classification of mammography is very significant research going on because it is popular method to discover breast cancer at a premature stage. The purpose of this paper is to propose ways to automate the process of mammography classification using the feature extraction algorithms and then classified using machine learning algorithms. One of the most active areas of research in supervised machine learning is to study methods for constructing good ensembles of learners. In this paper, the breast cancer data were classified by Support Vector Machine (SVM) and Artificial Neural Network (ANN) classifiers to finalize the classification process. ANN outperforms as supervised learning algorithm for classification problems. The images are collected using mias data set. The objective is to propose good combination of classification algorithms, which gives high-quality results in the classification of mammograms.

KEYWORDS

Mammogram, support vector machine, ANN, classification

OBJECTIVES

The primary objective of this paper is to propose a system for analysis of Mammogram incorporating novel classifier for classification of tumours. The proposed approach is a hybrid approach using Support Vector Machine (SVM) and Artificial Neural Networks (ANN) classifier. The objectives of this paper can be listed as below.

1) Improving the diagnostic and classification accuracy of Nodule type tumours.

2) Improving the average classification accuracy and the same time reducing the variance in the classification performance of SVM and ANN.

3) Creating a data base specific with images to enhance the classification accuracy.

INTRODUCTION

The chances of occurring terrible diseases have increased a lot in this 21st century. Timely detection of these diseases can have variation between life and death. Breast cancer is the main leading cause of death for women (1). Breast cancer tumors up to 30% even though caught early, will metastasize to other organs in the body, such as lungs, brains, bones or livers. For the detection of breast cancer, various techniques are used in mammography is the most promising technique and used by radiologist frequently. Mammogram images are usually of low contrast and noisy. In breast mammography, bright regions represent cancer. There are several features in mammography that help physicians to detect abnormalities in early stage, and these features can be directly extracted by image processing methods.

Cancer is the most wide spread disease around the world. Among this breast cancer is the most common and harmful disease among women. Breast cancer is the uncontrolled growth of breast tissues. It begins in the form of “cysts” in breast tissues. Cysts if huge in number or size can lead to breast cancer. Also micro calcification in breasts can show a possibility of cancer in breasts. Cancer tumour can be classified in two ways- benign (not cancerous) or malignant (is potentially cancer). The word “breast cancer” equals to

malignant tumours that are formed in cells of breast. Mammography is the most effective, low cost, contemporary option of premature detection and highly sensitive technique for detecting small lesions resulting in at least a 30% reduction in breast cancer deaths. Mammograms are used as screening tool to detect early breast cancer in women experiencing no discharge.

For the diagnosis and treatment of cancer, precise prediction of tumors is critically important. Among the existing techniques, supervised machine learning methods are the most popular in cancer diagnosis. Mammography is the best method for early detection of breast cancer. However, mammogram interpretation constitutes a challenge for the specialist. There are evidences that many of the cancers detected at screening mammograms would be retrospectively visible, but were not identified by the radiologist at the moment of the mammogram analysis. Approximately 10% to 30% of breast lesions are missed during routine screenings because of limitations that are specific to human observers (1). With the advances in digital image processing, imaging pattern recognitions and utilization of artificial intelligence, radiologists have the opportunity to improve their diagnosis with the assistance of computer system On the other hand, in a classification task, the radiologist and the computer will both analyze the same lesion in order to evaluate its malignancy probability. In case of disagreement, the computer analysis will be a challenge for the radiologist.

BACKGROUND WORK

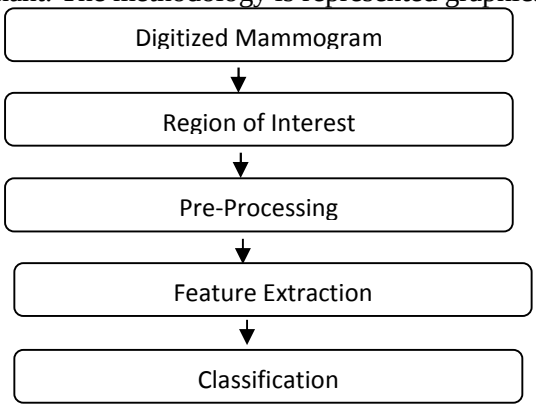
The necessity of analyzing a high number of images to detect a small number of Positive cases, the complex radiographic structure of the breast, the parenchyma density that may obscure a lesion, Positioning error or inappropriate mammography technique, the location of a lesion outside the field of view, subtle characteristics of malignancy in association with radiologist fatigue or view, or distraction contribute to false-negative interpretations of a mammogram. Besides fatigue, distraction and poor experience of the radiologist, the absence of previous imaging studies for comparison and the lack of supplementary mammographic views may lead to misinterpretation.

An effective technique for screening breast cancer is mammography. CAD gives high benefits to Radiologist with Automated breast cancer classification (2). CAD systems aid in classifying lesions as benign or malignant.

In mammography, double reading has shown to be highly beneficially, reducing the number of false-negative results by 5% to 15%, improving the rates of breast cancer detection. In spite of its proven diagnostic benefits, double reading is not always feasible because of logistic and financial issues affecting different institutions.

METHODOLOGY

The mammograms are digitized to find out regions of interest to find the lesions (3). The preprocessing is to be completed to enhance the image to find tumors. The specific features are extracted to further classify the tumors into benign or malignant. The methodology is represented graphically as (figure.1)



International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 4, Issue 9
September 2017

PRE-PROCESSING

Pre-processing is an analysis and manipulation of digitized image to enhance the image and remove unwanted objects from the image. Pre-processing is required to overcome this problem and make efficient feature extraction of images as possible. A lot of pre-processing is done to bring out the required feature from the image (4). In this case background removal and contrast enhancement is done in each image.

1. *Background removal:* Sometimes the image contains certain portions which are not requires for processing. These components in the image may give false or incorrect values. Thus background removal is done for removing the unnecessary components in the image.

2. *Contrast Enhancement*: Image enhancement techniques have been widely used in many applications of image processing where the subjective quality of images is important for human interpretation.

REGION OF INTEREST (ROI)

The ROI is detected from preprocessed mammogram to extract specific regions. These regions contain massive lesions. Each detected ROIs is then labeled at true positive ROI (TP ROI) or false positive ROI (FP ROI). The ratio of number of TPs to number of masses is called sensitivity.

FEATURE EXTRACTION STAGE

The original mammograms in MIAS are very big size (1024 × 1024 pixels) and almost 50% of the whole image comprised of the background with a lot of noise. Preprocessing is an important issue in low level image processing. Preprocessing stage is required to remove the areas that are not related to the detection region (pectoral muscle and any artifact labels that can be applied on mammogram). Therefore cropping process was performed manually to eliminate the noise and background information and Regions of Interest (ROI) image of size 256 × 256 is cropped from the original image.

PROPOSED ALGORITHMS

I. SUPPORT VECTOR MACHINES

Support vector machine, SVM [5] is based on statistical learning theory. It can be treated as a powerful, robust and sophisticated supervised machine learning approach. When the data is not linearly separable, the algorithm works by mapping the input space to higher dimensional feature space and constructs a hyper plane, which splits class members from non-members.

The ‘margin’ separates the two classes as a hyper plane on either side. Maximizing the margins and thus creating the largest possible distance between the separating hyper plane and the samples on either side, is proven to reduce an upper bound on the expected generalization error.

It is initially intended for binary classifying; employ administered culture to find the optimal separating hyper plane between both the groups of data. Having found a plane, it can then forecast the organization of an unlabeled instance by asking on which part of the unraveling plane for the examples. It acts as a linear classifier in a high dimensional characteristic space originated by a ledge of the unique input space, the resultant classifier has in common nonlinear in the given input space and it achieved very good simplification performance by exploiting the boundary between the classes. Support Vector Machine (SVM) is used for both classification and regression challenges. However, it is mostly used in classification problems and a popular machine learning algorithm. Each item is plotted as a point in n-dimensional space (where n is number of features) with the value of a particular coordinate. In this algorithm Support Vectors are the co-ordinates of individual observation. SVM is best applied for segregating the data in to two classes (hyper-plane/ line).

SVM Algorithm

Step 1: Load Dataset

Step 2: Classify Features (Attributes) based on class labels

Step 3: Estimate Candidate Support Value

While (instances! =null)

Do

Step 4: Support Value=Similarity between each instance in the attribute

Find Total Error Value

Step 5: If any instance < 0

Estimate Decision value = Support Value\Total Error

Repeat for all points until it will empty

End If

II. ARTIFICIAL NEURAL NETWORKS

Artificial neural networks are widely used of its efficiency in classification. The neural networks are made up of neurons; these neurons are connected to form input layer, hidden layer and output layer (6). The inputs feed to the network is propagated through the network to obtain an output. Training data is used to train the network using learning algorithms like back propagation. ANN diagnosis can be done by training and testing. The Feed Forward neural network is used for diagnosing breast cancer by comparing the hidden layers.

ANN is performed to distinguish the segmented objects as either micro calcifications or non micro calcifications. The accuracy of the ANN is tested by having a set of labeled test images for determination of true positive (TP) and false positive (FP) detection rates. It is a self-organizing ANN which runs a supervised learning algorithm. The ANN approach shows a promising result to detect micro calcifications or masses. ANN executes the tasks by teaching the system to do definite tasks. The ANN considers patient demographic factor and mammographic result as input for mammography interpretation. It also aids to estimate the corresponding breast cancer risk to further suggest for biopsy decision.

There are two different ways of using ANNs to aid in mammography interpretation. The first approach is to apply the classifier directly to the region of interest (ROI) image data. As a second approach, ANNs can also learn from the features extracted from the pre processed image signals. These automatically extracted features can be used as input to feed the CADx models. In addition to morphological features extracted from mammography images, texture features were also used to feed ANNs in classifying malignant and benign micro calcifications.

RESULTS

Several algorithms of classification have been projected of classification and assessed results of breast cancer disease. They are used to evaluate the missed lesions. The accuracy and efficiency of all the algorithms identified Benign and Malignant as two kinds of tumors and are classified properly from the MIAS data set with the accuracy, Sensitivity, Specificity.Confusion matrix (7) is a matrix which represents a relation between actual values and predicted values. The Confusion matrix is used to evaluate the performance of classification problem as two-class classification problem.

Table 1: Confusion Matrix

Actual Class	Predicted Class	
	POSITIVE	NEGATIVE
POSITIVE	TP	FN
NEGATIVE	FP	TN

TN= True Negative; values which are correct but are not predicted

TP= True positive; Values which are correct and predicted

FP= False Positive; values which are incorrect but are predicted

FN= False Negative; values which are incorrect and not predicted

The measures which are commonly used to evaluate the performance of the proposed method are accuracy, sensitivity, specificity and precision. These are calculated from confusion matrix using formulae.

A. *Accuracy:* Classification accuracy (8) is the percentage of instances that are correctly classified by the model. It is calculated as the sum of correct classification divided by the total number of samples. It is given by the formula:

Accuracy (%) = (TN + TP) / (TN + FN + FP + TP)

B. *Recall:* It is the measure of the ability of a classification model to select instances of certain class from the dataset. It is the proportion of actual positive which are predicted positive.

Recall = TP/TP+FN

C. *Specificity:* This is a measure that is commonly used in two class problems where the focus is on a particular class. It is the proportion of the negative class that was predicted negative and it is also known as the true negative rate.

Specificity = TN/TN+FN

Mammographic Image Analysis Society (MIAS) database is used as reference dataset for evaluation the results in this database. The total number of cases available in MIAS database is as in table 2. MIAS Database consist two experimental datasets on the same images. First dataset are divided into normal and abnormal cases. Second dataset are divided into benign, malignant and normal cases.

Table 2: MIAS data

Cases	**Total images**	**Benign**	**Malignant**
Micro calcifications	25	12	13
Masses	81	49	32
Architectural distortion	19	9	10
Lesions	15	6	9

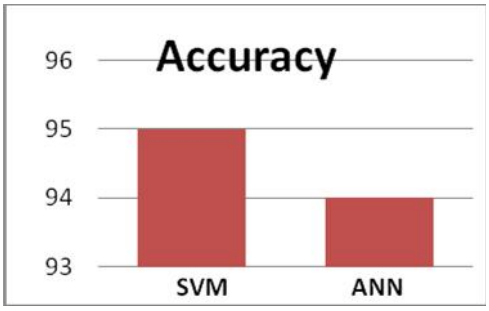
All the 64 benign images and 51 malignant images, totally 115 abnormal images are considered for this study irrespective of the type of abnormalities. From the database, 70 normal images are randomly chosen and considered for performance evaluation.

The results are tabulated in table 3. The classification accuracy is graphically represented in Fig 2.

Table 3: Performance of ANN and SVM algorithms

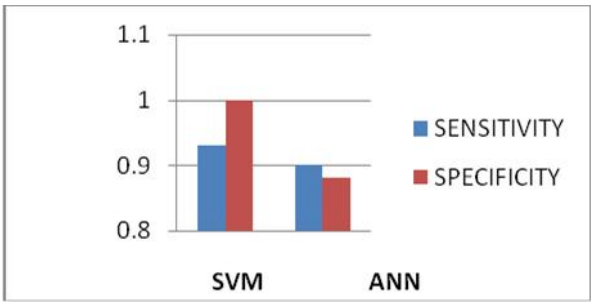
Classifiers	ANN	SVM
Accuracy	93	95
Recall/sensitivity/TPR	0.90	0.93
Specificity(TNR)	0.88	1.00

Fig 2: Performance Evaluation of SVM and ANN in terms of accuracy



From the above SVM classification accuracy is 95%. The precision and recall are also calculated and represented graphically in Fig 3.

Fig 3: Precision and Recall of SVM and ANN



CONCLUSION

The ultimate goal of this paper is to show an efficient classification technique to detect the presence of the tumor cells in breasts and to give an early prediction of breast cancer so that many woman lives could be saved as it was a major public problem. In this approach, the MIAS dataset is used to apply the technique whether the given data is benign or malignant. This prediction gives the maximum accuracy of 95%. The work is to demonstrate the ability of SVM and ANN classifier as second opinion to medical experts reducing the need of unnecessary expenditure. The methodology is applied to digitized mammograms to identify lesions so that prediction time can be reduced and the treatment for the patients can be given as early as possible. The high accuracy achieved by this can aid radiologists in taking correct decisions.

REFERENCES

1. Mahjabeen mirza beg , monika jain ," An analysis of the method employed for breast cancer diagnosis"
2. Khalid basher , anju Sharma , "A Review paper on classification on mammography" , International journal of engineering trends and technology (ijett)- volume 14 number 4 August 2014
3. Bh.nagarajasri, prof.m.padmavathamma, "A novel diagnostic computer aided medical tool for breast cancer based on Neuro fuzzy logic", International journal of scientific & engineering research, volume 4, Issue 12, December 2013
4. C Hashem B. Jehlol, Zainab khyioon Abdalrdha "classification of Mammography Image Using Machine Learning Classifiers and Texture Features" International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163 Issue 9, Volume 2 (September 2015)
5. Kathija,shajun nisha, "Breast cancer data classification using svm and naïve bayes techniques", International journal of innovative research in computer and communication engineering , volume 4, issue 12 , december 2016
6. Ritika bewal, aneecia ghosh, apoorva chaudhary, "Detection of breast cancer using neural networks"
7. R.Ramani, Dr.N.Suthanthira Vanitha "Computer A ided Detection of Tumours in Mammograms" *I.J. Image, Graphics and Signal Processing*, 2014, Published Online March 2014 in MECS DOI:10.5815 /ijigsp.2014.04.07

8. S.Krishnaveni,R.Bhanumathi, T.Pugazharasan "Study of Mammogram Microcalcification to aid tumour detection using Naive Bayes Classifier" International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering Vol. 3, Issue 3, March 2014, ISSN :2320-3765
9. Monika hedawoo , abhinandan jaisawal, nishita mehta, "Comparison of data mining algorithms for mammogram classification", International journal of electronics and data communication, ISSN :2320-2084 , volume 4, issue 7, july2016.
10. Kanchanamani M, VaralakshmiPerumal "Performance evaluation and comparative analysis of various machine learning techniques for diagnosis of breast cancer" *Biomedical Research 2016; 27 (3): 623-631 ISSN 0970-938X*