

Prediction of Breast Cancer Detection Using K-Nearest Neighbors, Support Vector Machine and Feed Forward Back Propagation Neural Network

Jalpa J. Patel¹, Dr.Sarman. K.Hadia²

¹Electronics and Communication Engineering Dept., Charusat, Changa, Anand, Gujarat.

²Electronics and Communication Engineering Dept., Charusat, Changa, Anand, Gujarat ,

Abstract— Now a days breast cancer is a most common problem in women. This work related to diagnose cancerous cell for find out cancer detection. Regular checkups are critical for early recognized and treatment of this type of cancer. Now days computer-aided methods for breast cancer analysis allow for another and earlier way of breast cancer analysis. In this paper classification of the images for breast cancer detection is accessible. Classification of the images is accomplished using Feed Forward Back Propagation Neural Network (FFBPNN) , Support Vector Machines(SVM) and k-NN algorithms are used. Classification of the breast cancer images is to improved by using Feed Forward back propagation neural network (BPNN) method. The accuracies of the FFBPNN, SVM and k-NN are 98.8%, 97.7% and 96.8% respectively.

Keywords— K-nearest Neighbors, Support Vector Machine, Artificial Neural Network, Feed Forward Back Propagation Neural Network, and Breast Cancer.

I. INTRODUCTION

The analysis of the breast cancer is detected by biopsy method. Biopsy method is a procedure of laboratory and it is completed by a pathologist to find out the cancer. Pathologist bring samples of tissues from the breast area for biopsy method. There are more than a few techniques for collecting samples of breast tissues. Now the cancer tissues are analyzed in microscope it is called as histopathology images. Pathologist examines the histopathology images and categorizes them as malignant or benign images.

Early diagnosed of breast tumor increases the persistence rate and also increased the treatment choices. Here they are presenting dissimilar pre-processing methods for mammogram images. The main purpose of this type of method is to improve the mammogram image class as well as quality. After that those type of images are ready for the segmentation and feature extraction. Different types of filtering methods are obtained for pre-processing. Filters are used to improve the image quality, eliminate the noise, preserve the boundaries within an image and image enhance and smoothen the image[1], Sertan Kaymak et.al. proposed Breast cancer generally develop on women breast area. They are applying different type of techniques like Back propagation Neural Network and Radial Based Neural Network. Automatic classification of the images is diagnosis. They achieved accuracy for algorithm BPNN and RBNN are 59.0% and 70.4% respectively[2].

II. LITERATURE SURVEY

Yu-Dong Zhand et.al. proposed to find out abnormal mass identification in breast .they are using convolution neural network to improved up to nine layer and achieve accuracy is 94.0%. sensitivity is 93.4 and specificity is 94.6%. they are using open access data set from mini MIAS data base. Also they are using three additional activation function like rectified linear unit, leaky rectified linear unit and parametric rectified unit.[3]

MadhuKumari et. al. proposed breast lesion detection for women. Breast cancer became the major source of mortality between women. Different types of classifier used for Wisconsin breast cancer dataset (WBCD) for proposed experiment. K-NN classifier achieved highest accuracy for predicted data values. The outcome of the method archived accuracy 99.28% for this study.[4]

Yeh et al. proposed an novel technique for lesion detection for breast cancer, now they are applying statistical methods in combined with swarm optimization and to achieve the accuracy of 98.71%. [5].

Polat et al , present LS-SVM classifier technique for detection of breast cancer and using this method they are achieved classification accuracy 98.53% for 10-fold cross validation method.[6].

Akay was introduced a new method for identify breast cancer lesion analysis using support vector machine classification algorithm on the most analytical features and obtain the classification accuracies 99.02% without cross-validation method [7].

Data Set : Here we are using WDBC breast cancer dataset is available in the UCI machine learning repository. This data is collected from the records histories of patients in the USA as maintained in the General Surgery Department of the

Clinical Sciences Center in Madison, Wisconsin. This dataset contains 569 samples with 32 attributes each. We have utilized 70% of the instances for training purposes, 15 % of instances for validation purposes and while the remaining 15% have been used for testing. These type of testing data are applied over the different types of classification methods to check the accuracy of the systems[8].

III. METHODOLOGY

Most of the databases are liable to noisy and unpredictable data because of their origin from various sources. So, we can make data for pre-processed an absolutely needed step for the classification of data set. In pre-processing steps provide noise removed process, image cleaning, image resize and normalized image for fit to classification process. Here we have used three different classification method for find out cancerous and non-cancerous mass. Firstly the dataset is divided into three parts. The first part is used to train the model, second part is used for validate the model and the third part is for testing the models.

Cross validation method is used in SVM and k-NN classifier and find out tuning have also been carried out to check over fitting of data and it gives more correct predicted results. And finally the confusion matrix and prediction accuracy have been compared to select the greatest model for the prediction of breast cancer data.

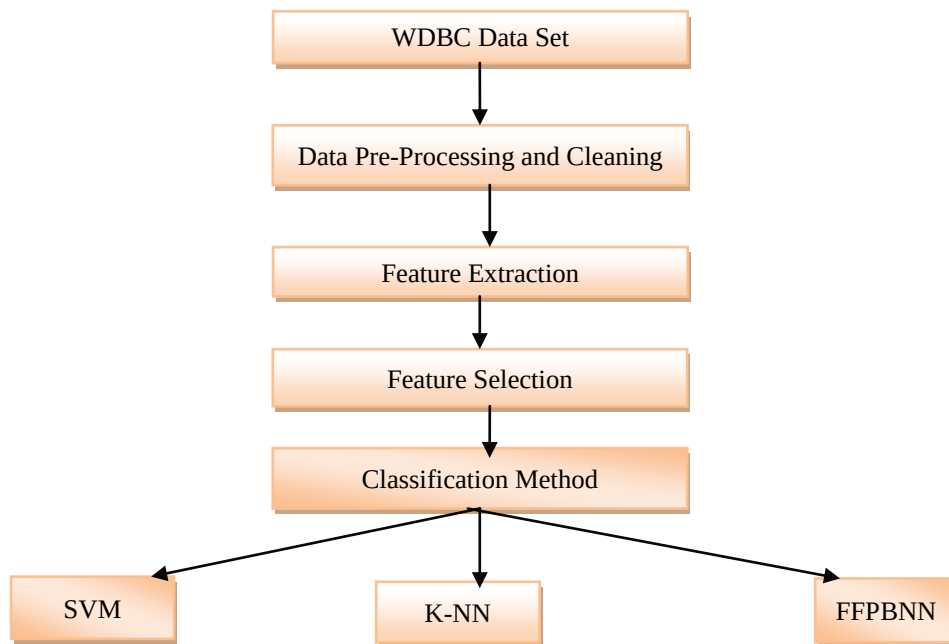


Fig.1 Block Diagram For Proposed Method

A. *K-Nearest Neighbor (KNN) :*

The k-nearest neighbor's algorithm is type of a machine learning algorithms. This algorithm is completely supported by the idea that "object are near each other will also have the same characteristics. So if you know characteristics feature of predicted data for nearest neighbor".

TABLE I
 STEPS FOR K-NEAREST NEIGHBORS ALGORITHM

Sr. No.	Steps
1.	Assign value for K-parameter, which represents the number of the nearest neighbors.
2.	Give the New Sample
3.	Compute the Distance Between Instances and Training Instances
4.	Select the K- Nearest Neighbors of New Sample
5.	Assign the class of the same classes of the K- Nearest Neighbors.

Generally K-NN method is used for pattern recognition to identify particular class based on training data sets.

B. *Support Vector Machine (SVM) :*

Support Vector Machine is also consider as a pattern classifier as well as machine learning method. SVM is a classifier that can be defined by a separating hyperplane. It is also used to separate two or multi class problem. There does not exit any linear separable hyperplane for any data set. Kernel function is used for maximum –margin separate hyperplane.

C. Feed Forward Back Propagation Neural Network (FFBPNN) :

Artificial Neural Network is demanded to be one of the very much powerful classifier for very complex input to output data[9].

It is familiarize the capability of human brain neuron in perceptual acts. It is train the artificial neurons using defined input and according to that calculate the greatest weight of each neuron so that the minimum error can be occur. This type of error is represented using mean square error (MSE).

One of popular Artificial Neural Network algorithm is Feed-Forward Back -propagation Neural Network (FFBPNN) that can regulate the weight of each and every neuron during this training process to minimize the MSE. In this method error is calculated and than according to error update the weight again and again up to reached minimum error. This method is more robust according to researches proofs. So, for this motive, we are using FFBPNN to classify used for mammograms images. The model of FFBPNN selected in this study contains of input layer, hidden layer, and output layer[10, 11].

IV. RESULTS AND DISCUSSION :

In this paper we are comparing study of different types of classification method. we are applying k-NN classifier, Support Vector Machine and Back Propagation Algorithm to this data base after applying pre-processing techniques.

In support vector machine we are using 10-fold cross validation function to classify the data. And based on that we have to classify malign and benign cell.

We have also made use of confusion matrix to define the performance of the above classification models based on a collection of test data whose response were previously known.

TABLE II
CONFUSION MATRIX FOR ALL CLASSIFICATION METHOD

Classification Methods								
K-Nearer Neighbor			Support Vector Machine			Feed Forward Back Propagation Neural Network		
Benign	354(TP)	3(FP)	Benign	355(TP)	2(FP)	Benign	357(TP)	6(FP)
Malignant	15(FN)	197(TN)	Malignant	11(FN)	201(TN)	Malignant	1(FN)	205(TN)

According to confusion matrix Table.2 in K-NN classification method we have to get accuracy 96.8%, in SVM classification we can get 97.7% and using FFBPNN we have achieve maximum accuracy is 98.8%.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} * 100 \text{ -----(i)}$$

$$\text{Specificity} = \frac{TN}{TN+FP} * 100 \text{ -----(ii)}$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} * 100 \text{ -----(iii)}$$

Where TP is true positive, TN is true negative, FP is false positive and FN is false negative . Sensitivity represent the percentage of truly predicted malignant class, specificity represent the percentage of truly predicted benign class and accuracy is percentage of correctly predicted cancer and non-cancer cases.

TABLE IV:
COMPARISON WITH EXISTING METHODS IN TESTING PHASE

	Accuracy (%)	Sensitivity (%)	Specificity(%)
LPSVM [15]	97.14	98.24	95.08
SSVM [15]	96.57	96.58	96.55
PSVM [15]	96	97.36	93.44
St-SVM [15]	94.86	95.65	93.33
Proposed model using K-NN	96.80	96.2	98.5
Proposed model using SVM	97.70	96.0	99.0
Proposed model using FFBPNN	98.80	99.7	97.1

Accuracy Chart

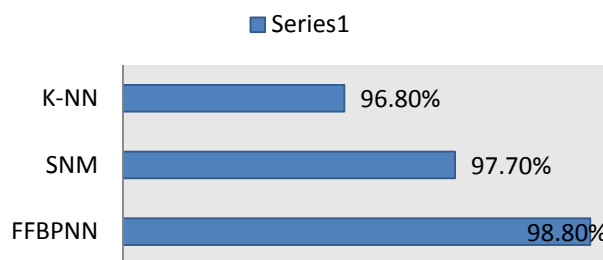


Fig.2 Accuracy Chart for Different Classification Method

V. CONCLUSION :

Based on the above result, it can be conclude that FFBPNN is more accurate in classification with sensitivity of 99.7% and accuracy of 98.8% for benign and malignant lesion as compare with other classification method K-NN achieve sensitivity of 96.2% and accuracy of 96.8%. And with SVM classifier sensitivity of 96.0% and accuracy of 97.7% for benign and malignant lesion.

REFERENCES

- [1] R.Ramani, Dr. N. Suthanthira, S.Valarmathy "The Preprocessing Methods of Mammogram Images for Breast Cancer Detection" I.J.image Graphics and Signal Processing 2013, 5, 47-54.
- [2] SurtanKaymak, Adbulkader, DilberUzun "Breast cancer image classification using artificial neural networks" ICSCCW 2017, 24-25 August 2017, Budapest, Hungary.
- [3] Yu-Dong Zhang, Chichun Pan, Xianqing Chen, Fubin Wang "Abnormal Mass Identification by nine layer convolution neural network with parametric rectified linear unit and rank -based stochastic pooling" Journal of Computer Science 27 (2018) 57-68
- [4] MadhuKumari, Vijendra Singh "Breast Cancer Prediction system" ICCIDS 2018 Procedia Computer Science 132(2018) 371-376
- [5] Yeh WC, Chang WW, Chung YY "A new hybrid approach for mining breast cancer pattern using discrete particle swarm optimization and statistical method. Expert Systems with Applications" 2009 May 1;36(4):8204-11.
- [6] Polat K, Güneş S. "Breast cancer diagnosis using least square support vector machine" Digital Signal Processing. 2007 Jul 1;17(4):694701
- [7] Akay MF. "Support vector machines combined with feature selection for breast cancer diagnosis "2009 Mar 1;36(2):3240-7
- [8] Breast Cancer Statistics. <https://ww5.komen.org/BreastCancer/Statistics.html>.
- [9] M. J. Islam, M. Ahmadi, and M. A. Sid-Ahmed, "An Efficient Automatic Mass Classification Method in Digitized Mammograms Using Artificial Neural Network," International Journal of Artificial Intelligence & Applications (IJAIA), vol. 1, no. 3, pp. 1-13, 2010.].
- [10] Nadeem Tariq "Breast Cancer Detection using Artificial Neural Networks"Journal of Molecular Biomarkers & DiagnosisTariq, J MolBiomarkDiagn 2017
- [11] MellisaPratiwia, Alexandra, JeklinHarefa, SakkaNandaa"Mammograms Classification using Gray-level Co-occurrence Matrix and Radial Basis Function Neural Network"International Conference on Computer Science and Computational Intelligence (ICCCSI-2015)Procedia Computer Science 59 (2015) 83 – 91
- [12] Md. Milon Islam, Hasib Iqbal, Md. RezwanulHaque, and Md. Kamrul Hasan "Prediction of Breast Cancer Using Support Vector Machine and K-Nearest Neighbors"2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC) 21 - 23 Dec 2017, Dhaka, Bangladesh.
- [13] Mugahed A. Al-antari, Mohammed A. Al-masni1 , Sung-Un Park"An Automatic Computer-Aided Diagnosis System for Breast Cancer in Digital Mammograms via Deep Belief Network", Taiwanese Society of Biomedical Engineering 2017
- [14] Aritra Basu1, Rohit Roy and N. Savitha"Performance Analysis of Regression and Classification Models in the Prediction of Breast Cancer" Indian Journal of Science and Technology, Vol 11(3), DOI: 10.17485/ijst/2018/v11i3/119179, January 2018
- [15] Ahmad Taher Azar, Shaimaa Ahmed ,El-Said , "Performance analysis of support vector machines classifiers in breast cancer mammography recognition" Neural Computing and Applications, Springer, Volume 24, Issue 5, pp 1163–1177, April 2014.