Assessment of the Performance of MLP & Hierarchical Clustering Method in the Prediction of Breast Cancer

Jayshree Bhattacharya[1], Pranati Rakshit[2*], Arindam Roy[3], Dharmpal Singh[4], Debolina Dalui[5]

[1, 2,3,4,5]Computer Science and Engineering, JIS College of Engineering, JIS University, Kalyani, India

*Corresponding Author: pranati.rakshit@jiscollege.ac.in, Tel.: +91 9434437490

Abstract— Breast cancer is such a type of cancer which can leads to the death of women .in this type of cancer there is excessive growth of cells of the breast tissue are done. Lump, pain or sore, bloody nipple, and change in size are being visible for the case of breast cancer. Breast cancer took places when the tissues of the breast are not in a normal manner and not divided in a proper way. Those abnormal cells make clotted mass tissues, which therefore that becomes a tumor. There are two types of tumors, one is benign, which are not cancerous and that no becomes the cause of anyone's death .Another type of tumor is Malignant tumor, that is cancerous and very harmful. The second type of tumor can lead to the death. Soft computing techniques are become popular for the case of medical disease diagnosis and also for the prediction. In this paper we combine the Hierarchical Clustering Method and Multilayer Perceptron to predict of the breast cancer .The accuracy of the proposed methods is 81% for the clustering method and 94% for Multilayer Perceptron.

Keywords—Soft Computing, Breast Cancer, Benign, Malignant, Hierarchical Clustering Method, Multilayer Perceptron

I. INTRODUCTION

Breast cancer is one of the most harmful of cancers which can cause the death of a woman. Approximately 4 lakh women are affected by this harmful cancer out of which 50 thousand women are from India. It is very necessary the early detection of breast cancer which can provide the timely treatment for the patient. So, the probability of recovery of this harmful disease can be increased [1]. The toxicity and pain can be reduced if the disease is found in initial stages. An important role is played by soft computing for the case of early prediction of the tumor whether it is Benign or Malignant. In the case of benign tumor, the cells grow in an abnormal manner and made a lump. But it is not spread to any other parts of body and it not becomes the cause of cancer. For the case of malignant tumor, the cancer cells may spread to the others part of body generally those cells break away from actual cancer occupied body parts and then spread to other parts of the body. Those cells can spread by the bloodstream. When these cells get a new area they are being divided and make a new tumor. Breast cancer takes places when cells of the breast are malignant [2]. If the disease can be predicted during the early stage period, tis cancer also can be cured. The identification of the tumor using soft computing techniques which can gives a result with suitable accuracy with the help of actual methods and attributes reference values. The prediction of this method would help doctors, scientists, pharmacists to understand about the nature of tumor. The MLP and the Hierarchical Clustering method both are used here to make the prediction more accurate.

II. RELATEDWORK

Many studies have been done which have focused on breast cancer. These works are applied by using different technique to the given problem and got high classification accuracies. Soft computing technique is become very powerful in medical field. Sometimes these techniques give the predictions with high accuracy. Amjad Khan et al. [4] proposed a soft computing technique, by which the early detection of breast cancer can done. Multilayer Perceptron is an approach which brings the prediction with more accuracy. J.Padmavathi [11] describes an approach to predict the breast cancer using MLP.N.S.Nithiya et.al [10] proposed a paper to enhance the clustering method by surveying on many research papers about the clustering technique used in medical field. Ms.Sneha Badrinath Sanap et.al [13] developed an expert system based on supervisory neural network based learning approach. Here the testing is done with MLP.S.A Mojarad et.al [14] investigated the effectiveness of MLP to predict breast cancer progression using a set of four biomarkers of breast tumour.

III. METHODOLOGY ANDIMPLEMENTATION

In this research work two methods are implemented they are discussed below. We acquired a dataset with 570 instances from kaggle datasets. In this dataset, in the

diagnosis field where malignant is represented with a value of „„,1‴ and benign is represented with „„0‴.The attributes which are present here The radius of tumor, The texture of Tumor, The perimeter of Tumor, The area of Tumor and The smoothness of Tumor along with another column of diagnosis. Factor analysis: At first we have applied Factor analysis of that dataset. Factor analysis is such a type of statistical method which is used to describe the variability among observed and the correlated variables in terms of a potentially lower number of unobserved variables called factors. The information which is gained about the interdependencies between observed variables can be used later to reduce the set of variables in a dataset. Factor analysis is used to reduce a large number of variables into fewer numbers of factors. This technique extracts maximum common variance from all variables and puts them into a common score. Using the values of antecedent item, the correlation of coefficient of the items have been computed and these coefficients have been stored in table 1 termed as correlationmatrix.

Table 1: correlation matrix

1.0000	0.3238	0.9979	0.9874	0.1706
0.3238	1.0000	0.3295	0.3211	-0.0234
0.9979	0.3295	1.0000	0.9865	0.2073
0.9874	0.3211	0.9865	1.0000	0.1770
0.1706	-0.0234	0.2073	0.1770	1.0000

Table 2 : Eigen value

3.1723	0	0	0	0
0	0.0014	0	0	0
0	0	0.0166	0	0
0	0	0	0.7886	0
0	0	0	0	1.0211

Table 3: Eigen vector

A	C	D	E
0.306584	0.158962	0.03045	0.00000729
0.06096	2.6E-05	0.654319	0.28472896
0.308691	0.174808	0.020851	0.00061009
0.304483	0.666019	0.029207	0.00001521
0.019265	0.00011	0.265225	0.71470116

The major factors have been calculated as per the formula ($\sqrt{\text{Eigen value}} \times$ Eigen vector) using the selected Eigen value as furnished in table 2 and Eigen vector as furnished in table 3. The major factors have been furnished in table 4.

Table 4: Cumulative Effect Value of Items

A	0.593584
B	0.977352
C	0.591464
D	0.654076
E	0.992058

Now a relation can be formed by using the cumulative effect value of all the elements to make the total effect value. Total effect value = (0.593584) × A + (0.977352) × B +(0.591464) ×C+(0.654076)×D+(0.992058)×E...Nowbyusingthis relation, a resultant total effect value being formed (due to size of data it is not shown here).

Implementation by Hierarchical Clustering Method:
It is an algorithm that groups similar objects into groups which are called clusters .Basically this is based on in between two nearest clusters. First all the data are setting as a cluster .After doing a few iteration it reaches to the final cluster. The Concept of hierarchical clustering has been used using the Matlab command to form the cluster on the total effect to select the optimal cluster for the formation of knowledge base. The command is furnished as follows: Step 1: X = [total effect value]; Step 2: Using this, the whole data set can be defined as matrix. Then pass it to pdist. Y = pdist(X) Step 3: The pdist is a function which calculates the distance between every object. Z = linkage(Y) The linkage function generates a hierarchical cluster tree, after that the linkage information is returning in a matrix Z. After creating the hierarchical cluster tree we will partition the data into the clusters. Step 4: T = cluster (Z, 2).here _2' is taken because the two type of output i.e. Malignant and Benign.

Implementation by Multilayer Perceptron Method:
The Multilayer Perceptron or MLP is a Feed Forward Artificial Neural Network which is able to generate a set of output after receiving a set of input. An MLP can be distinguished by a number of layers of input nodes which are connected in directed graph between the output and input layers. Back propagation process is used by MLP for training the network. It is a deep learning method. Using MLP we get Confusion matrix shown in fig 1.
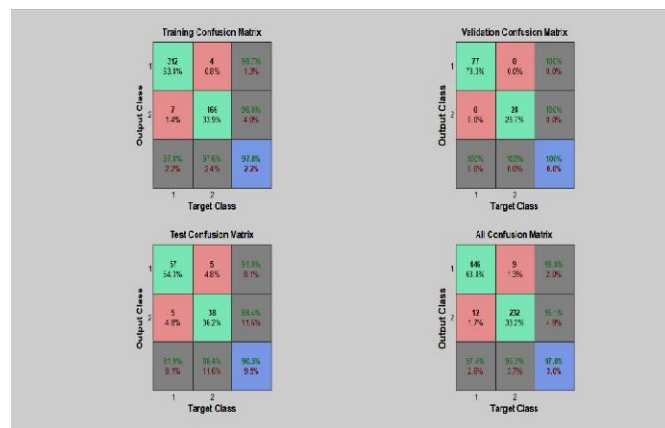

Fig: 1 Confusion Matrix

Here the accuracy is shown by the confusion matrix which is got from Matlab. A Confusion matrix describe you the

accuracy of the classifier. It lists correct classifications as „true positives‟ or „true negatives‟, and incorrect classifications as „false positives‟ or „false negatives‟. It is also possible to derive the sensitivity, specificity, positive predictive value, and others. This is particularly important in diagnostic tests for the presence or absence of a particular condition or disease based on a single test.

IV.   RESULTS AND DISCUSSION

For the case of Hierarchical Clustering Method the result is shown in the table 5.

Table: 5

Total Error	-11158.5
Average Error	19.57632

Here we can clearly see the error is 19.57632 %.so, the accuracy of the prediction is approximate 81%. For the case of MLP the accuracy of the result is 90.5% for test dataset.

V. CONCLUSION AND FUTURESCOPE

In this paper, we have applied two prediction method or breast cancers on two parameters: benign and malignant tumour. It has been observed that Soft Computing methods provide some good solutions to biologically related problem of some medical domain as breast cancer. Here we use the popular MLP method and Hierarchical Clustering Method to get the accuracy of the prediction. We acquired data set of 570 instances. By combining the Hierarchical Clustering Method and Multilayer Perceptron to predict of the breast cancer we get the result with the accuracy of the proposed methods is 81% for the clustering method and 90.5% for Multilayer Perceptron. As the error is not so much this combined method is actually efficient to make the proper prediction of a tumour in future. In this research paper we have just implemented MLP and Hierarchical Clustering method to get the accuracy of prediction. In future k-means and Fuzzy c-means can be implemented on this paper .So, after applying those two techniques on this work we can a again get some value of accuracy on the prediction of atumor whether it is benign or malignant. By comparing those values we will get the best accuracy value of the result for prediction.

REFERENCES

[1] Kowsalya, shanmugapriyaa, ─ A Survey on Diagnosis Methods of Breast Cancer using Mammography‖ International Journal of Engineering Technology Science and Research , ISSN 2394 – 3386 Vol. 02, Issue 4(April2015),pages,100-107.

[2] VikasChaurasia, Saurabh Pal, BB Tiwari ,‖ Prediction of benign and malignant breast cancer using data mining techniques‖ Volume: 12 issue: 2, page(s): 119-126 . Issue published: June 1, 2018.

[3] M. Durairaj , K. Priya ,‖Breast Cancer Prediction Using Soft Computing Techniques – A Survey‖,International Journal of Computer Sciences and Engineering,Vol.-6, Issue-8, Aug 2018 E- ISSN: 2347-2693.

[4] Amjad Khan , Zahid Ahmed Ansari,‖ Early Detection of Breast Cancer using Soft Computing‖, IJSRD - International Journal for Scientific Research & Development| Vol. 5, Issue 10, 2017 | ISSN (online): 2321-0613.

[5] Ibrahim M. El-Hasnony, Hazem M. El-Bakry, Ahmed A. Saleh,‖ Classification of Breast Cancer Using Softcomputing Techniques‖ International Journal of Electronics and Information Engineering, Vol.4, No.1, PP.45-54, Mar.2016.

[6] SoumadipGhosh,SushantaBiswasandDebasreeChanda,─Breast Cancer Detection using a Neuro-fuzzy basedClassification Method‖ Indian Journal of Science and Technology, ISSN (Print): 0974- 6846, Vol 9(14), April 2016, pages, 1-15.

[7] N.S.Nithya,Dr.K.Duraiswamy,P.Gomathy,─ASurveyon Clustering Techniques in Medical Diagnosis‖ International Journal of Computer Science Trends and Technology (IJCST) Vol. 1 Issue2, Nov-Dec 2013.

[8] SL.Aarthy,S.Prabu.,─acomputerizedapproachonbreastcancer detection and classification‖ IIOAB Journal, ISSN: 0976-3104, Vol. 7/Issue 5/2016/157-169.

[9] Elouedi, H., Meliani, W., Elouedi, Z. and Amor, N.B., A hybrid approach based on decision trees and clustering for breast cancer classification. IEEE 6th International Conference of Soft Computing and Pattern Recognition (SoCPaR), 2014, August, pp. 226-231.

[10] N.S.Nithiya,Dr.K.Duraiswamy,P.Gomathy ,‖A Survey On Clustering Techniques In Medical Diagnosis‖,International Journal Of Computer Science Trends And Technology (IJCST)- Volume 1 issue 2,Nov-Dec2013.

[11] J.Padmavathi,‖A Comparative study on Breast Cancer Prediction Using RBF and MLP‖,International Journal of Scientific & Engineering Research, Volume 2, Issue 1, January-2011 1 ISSN 2229-558.

[12] TubaKiyanandTulayYildirim,─BreastCancerDiagnosisUsing Statistical Neural Networks‖, Istanbul University, Journal of Electrical and Electronics Engineering, Year 2004, 4(2), 1149-1153.

[13] Ms.SnehaBadrinathSanap,VijayL.Agrawal─DiagnosisofBreast Cancer using Multilayer Perceptron (MLP)‖,International Journal of Advance Foundation and Research in Computer (IJAFRC) Volume 3, Issue 4, April - 2016. ISSN : 2348 –4853.

[14] S.A Mojarad, S.S.Dlay& W.L. Woo and G.V Sherbet,‖ Breast cancer prediction and cross validation using multilayer perceptron neural networks‖, 2010 7th International Symposium on Communication Systems, Networks and Digital Signal Processing, CSNDSP2010.4.760764.

[15] ShikhaAgrawal,JitendraAgrawal,‖Neural Network Techniques for Cancer Prediction: A Survey‖, 19th International Conference on Knowledge Based and Intelligent Information and Engineering Systems, https://doi.org/10.1016/j.procs.2015.08.234.

Authors Profile

Jayshree Bhattacharya has received her B.tech degree in Computer Science and Engineering Form West Bengal University of Technology in 2018. She is currently pursuing her M.tech degree in Computer Science and Engineering from Maulana Abul Kalam Azad University of Technology.

Dr. Pranati Rakshit is an Assistant Professor in the Department of Computer Science and Engineering of JIS College of Engineering, India. She has more than 19 years of work experience associated with teachin and research. She has completed her Ph.D. degree from Jadavpur University, kolkata, West Bengal, in the field of Pattern Recognition and Medical image analysis. She has completed her Master degree and B.E. Degree in Computer Science & Engineering. She has worked in the field of Data Mining, image Processing, IOT also. She has supervised more than 20M.Tech Projects. She has a good number of research publications. She is a life member of Indian Society of Technical Education.

Arindam Roy is currently pursuing his Post Graduate Degree in the field of Computer Science and Engineering from JIS College of Engineering, Kalyani, Nadia. He did B.tech in Computer Science and Engineering from Maulana Abul Kalam Azad University of Technology, West Bengal (Formerly known as West Bengal University of Technology). He completed his Graduation in the year 2017 from JIS College of Engineering, Kalyani, and Nadia. He has interest on Web Development and Designing. He has experience on Adobe Photoshop. He did a Training on Android on the year2016-17.

Dr.Dharmpal Singh received his Bachelor of Computer Science and Engineering and Master of Computer Science and Engineering from West Bengal University of Technology. He has done his PhD inyear 2015. He has about 12 years of experience in teaching and research. At present, heis with JIS College of Engineering, Kalyani, and West Bengal, India as an Associate Professor and Head of the department. He has published 32 papers in referred journal and conferences index by Scopus, DBLP and Google Scholar and editorial team and senior member of many reputed journal index by SCI, Scopus, DBLP and Google Scholar. He has organized seven national levels Seminar/Workshop, published three patent and has applied for the AICTE Project in year of 2019. He is a member of the Computer Science Teachers Association (CSTA), Computer Society of India (CSI) and also a member of International Association of Computer Science and Information Technology(IACSIT).

Debolina Dalui has completed M.tech in the department of Computer Science and Engineering from a renowned Engineering college of West bengal.