Breast Cancer Detection In Data Mining: A Review

S. Yuvarani and Dr. C. Jothi Venkateswaran

Assistant Professor, Department of Computer Application,
Thanthai Hans Roever College, Perambalur-621212
Assistant Professor& Head, Department of Computer Science,
Presidency College, Chennai-600 005

ABSTRACT

The purpose of this paper was during the literature review of journals and publications in the field of data mining in computer science and engineering. The research focused on more recent publications in breast cancer. In this paper we have discussed various evaluation and technological articles on breast cancer analysis and prediction. Also we focus on current research being carried out using the data mining algorithms and techniques, to enhance the breast cancer identification and diagnosis.

Keywords: Breast Cancer; Data mining Techniques; literature review;

I. INTRODUCTION

Data mining is an interdisciplinary field and is fast reputation because of exploring Database technology, Information Science, Machine learning and Neural networks along with the Statistical techniques. Though data mining algorithms are not applied on the medical data by common people but the knowledge obtained can be very useful for them if shared with in an comprehensible form.

Cancer is a common term used to describe a group of over a hundred diseases that occur when malignant forms of abnormal cell growth develop in one or more body organs. There are three ways that cancer spreads in the body. Cancer can spread through tissue, the lymph system, and the blood: Tissue, The cancer spreads from where it began by growing into nearby areas. Lymph system, the cancer spreads from where it began by getting into the lymph system. The cancer travels through the lymph vessels to other parts of the body. Blood, The cancer spreads from where it began by getting into the blood. The cancer travels through the blood vessels to other parts of the body.

II.   REVIEW OF LITERATURE

There is large number of papers about applying Data Mining techniques for survivability analysis. Several studies have been reported that they have focused on the importance of technique in the field of Breast Cancer. These studies have applied different approaches to data mining and given problem. Here are some examples:

2.1 In year 2004, authors Dursun Delen et.al.[1] have given their study:

The comparative study of multiple prediction models for breast cancer survivability using a large dataset along with a 10-fold cross-validation provided us with an insight into the relative prediction ability of different data mining methods. Using sensitivity analysis on neural network models provided us with the prioritized importance of the prognostic factors.

2.2 In year 2012, author Shweta Kharya[2] have been given their study:

Breast cancer is one of the leading cancers for women in developed countries including India. It is the second most common cause of cancer death in women. The high incidence of breast cancer in women has increased significantly in the last years. To discussed various data mining approaches that have been utilized for breast cancer diagnosis and prognosis. Breast Cancer Diagnosis is distinguishing of benign from malignant breast lumps and Breast Cancer Prognosis predicts when Breast Cancer is to recur in patients that have had their cancers excised.

2.3 In year 2012, authors K.Rajesh et.al.[3] have given their measurement:

Classify SEER breast cancer data into the groups of "Carcinoma in situ" and "Malignant potential" using C4.5 algorithm. And used a training set of a random sample of 500 records and then applied the classification rule set obtained to the full breast cancer dataset. To obtain an accuracy of ~94% in the training phase and an accuracy of ~93% in the testing phase. To compared the performance of C4.5 algorithm with other classification techniques.

2.4 In year 2013, authors Shiv Shakti Shrivastava et.al.[4] have given their measurement:

Review in the use of data mining in breast cancer. To observed that neural network and decision approach mostly used by various researchers to create a predictive model and decision rules from the breast cancer data. Most of them they have done the comparative study of algorithm to take breast cancer data. To conduct an experimental work and find various if….then rules from decision tree which is represent and used J48 classifier of WEKA which is an extension form of ID3 algorithm of decision tree.

2.5 In year 2013, authors G.Ravi et.al.[5] have given their experiment:

A comparison among are the different Data mining classifiers on the database of breast cancer Wisconsin Breast Cancer (WBC), by using classification accuracy. Aims to establish an accurate was classification model for Breast cancer prediction, in order to make full use of the invaluable information in clinical data, especially which

is usually ignored by most of the existing methods when they aim for high prediction accuracies. They have done experiments on WBC data. The dataset is divided into training set with 499 and test set with 200 patients. In this experiment, to compare six classification techniques in Weka software and comparison results show that Support Vector Machine (SVM) has higher prediction accuracy than those methods. Different methods for breast cancer detection are explored and their accuracies are compared. With these results, we infer that the SVM are more suitable in handling the classification problem of breast cancer prediction, and we recommend the use of these approaches in similar classification problems.

2.6 In year 2014, authors Vikas Chaurasia et.al.[6] have given their experiment:
A diagnosis system for detecting breast cancer based on RepTree, RBF Network and Simple Logistic. In test stage, 10-fold cross validation method was applied to the University Medical Centre, Institute of Oncology, Ljubljana, Yugoslavia database to evaluate the proposed system performances. The correct classification rate of proposed system is 74.5%. To demonstrated that the Simple Logistic can be used for reducing the dimension of feature space and proposed Rep Tree and RBF Network model can be used to obtain fast automatic diagnostic systems for other diseases.

2.7 In year 2014, authors Vikas Chaurasia et.al.[7] have given their experiment:
Three popular data mining methods: Sequential Minimal Optimization (SMO), IBK, BF Tree. An important challenge in data mining and machine learning areas is to build precise and computationally efficient classifiers for Medical applications. The performance of SMO shows the high level compare with other classifiers. Hence SMO shows the concrete results with Breast Cancer disease of patient records. Therefore SMO classifier is suggested for diagnosis of Breast Cancer disease based classification to get better results with accuracy, low error rate and performance. And also shows that the most important attributes for breast cancer survivals are Uniformity of Cell Size, Uniformity of Cell Shape, Bare Nuclei, Bland Chromatin, Single Epithelial Cell Size, Normal Nucleoli, Marginal Adhesion, Clump Thickness and Mitoses. These attributes were found using three tests for the assessment of input variables: Chi-square test, Info Gain test and Gain Ratio test.

2.8 In year 2014, author Hamid Karim Khani Zand.[8] have been given their survey:
The classification of Breast Cancer data can be useful to predict the outcome of some diseases or discover the genetic behavior of tumors. To present a comparative survey on data mining techniques in the diagnosis and prediction of breast cancer and also an analysis of the prediction of survivability rate of breast cancer patients.

III. CONCLUSION
This paper clearly shows that the groundwork results are gifted for the function of the data mining methods into the survivability forecast problem in breast cancer. In the presence of all this challenges also data mining algorithms have good results in

different methodologies of Breast cancer, but more work is required to be done in managing different study and reviews of Data mining Technologies. In data mining Technology is used in novelty of different approaches are linear regression, support vector machine, principal component analysis and various classification and clustering algorithms.

REFERENCES

1.   Dursun Delen, Glenn Walker, Amit Kadam "Predicting breast cancer survivability: a comparison of three data mining methods" Artificial Intelligence in Medicine-2004.

2.   Shweta Kharya "Using Data Mining Techniques for Diagnosis and Prognosis of Cancer Disease" International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.2, No.2, April 2012.

3.   K.Rajesh, Dr. Sheila Anand "Analysis of SEER Dataset for Breast Cancer Diagnosis using C4.5 Classification Algorithm" International Journal of Advanced Research in Computer and Communication Engineering Vol. 1, Issue 2, April 2012.

4.   Shiv Shakti Shrivastava , Anjali Sant , Ramesh Prasad Aharwal "An Overview on Data Mining Approach on Breast Cancer data" International Journal of Advanced Computer Research, Volume-3 Number-4 Issue-13 December-2013

5.   G. Ravi, Dr. G. A. Ramachandra, K.Nagamani "An Efficient Prediction of Breast Cancer Data using Data Mining Techniques" International Journal of Innovations in Engineering and Technology (IJIET) Vol. 2 Issue 4 August 2013

6.   Vikas Chaurasia , Saurabh Pal "Data Mining Techniques: To Predict and Resolve Breast Cancer Survivability" International Journal of Computer Science and Mobile Computing, Vol.3 Issue.1, January-2014, pg. 10-22

7.   Vikas Chaurasia , Saurabh Pal "A Novel Approach for Breast Cancer Detection using Data Mining Techniques" International Journal of Innovative Research in Computer and Communication Engineering, Vol. 2, Issue 1, January 2014.

8.   Hamid Karim Khani Zand "a comparative survey on data mining techniques for breast cancer diagnosis and prediction" Indian Journal of Fundamental and Applied Life Sciences, Vol.5 (S1), pp. 4330-4339, 2015