



International Journal of Applied Business and Economic Research

ISSN : 0972-7302

available at <http://www.serialsjournals.com>

© Serials Publications Pvt. Ltd.

Volume 15 • Number 22 • 2017

An Overview of Predictive Models in Breast Cancer Diagnosis

Durgalakshmi B¹ and Vijayakumar V¹

¹Department of SCSE, VIT University, Chennai. Email: durgalakshmi.b2013@vit.ac.in

ABSTRACT

Cancer is a major threat to the human community all around the globe. It is a destructive disease, which deadly affects many people and has influenced the numerous lives each and every day. One of the dominant types of cancer cells is the ones which affect the breast region of women and is termed as breast cancer; it is the leading cause for death in women worldwide. This cancer is a threat to all developing and as well as developed countries in the world in the most recent years. The early finding and forecast of malignancy and benign stages helps in identifying the survival of the patients. The techniques that are available in machine learning which incorporates various approaches like probabilistic, statistical and optimization techniques. This paper briefs the various predictive models here based on different ML systems that deals with essential information elements and samples in cancer research.

Keywords: Breast cancer, WDBC.

1. INTRODUCTION

Breast cancer is one among the most commonly heard types of cancer that predominantly affects the women and may be categorized as modifiable and non-modifiable. Modifiable factors are can be controlled and cured by habits and environmental issues. Non-modifiable factors are comes from gender and family history [1]. Based on the recent analysis, 1 on 28 women in India is possibility of prone to get breast cancer so the early detection techniques on breast cancer diagnosis are still lacking in the accurate prediction of cancer. Moreover, unaware of this disease seriousness, proactive measures and treatment may increase the survival risks.

As per the scientific researchers proposed the automatic detection of breast cancer, Data mining is another scientific field that encloses several techniques and algorithms to analyse the hidden facts and knowledge from the complex data sources. In data mining, Classification techniques plays a role as a predictor of end result by defining the class labels of a dataset. Hence, the classification techniques contributes in predicting the disease and its exact stages in medical domain for building knowledge prediction model.

A performance based summary of the popular machine learning algorithms and classification techniques has been drafted out in this paper as to classify the research problems that need to be targeted for near future. Moreover, this paper highlight the comparative study between classification algorithms that exactly predicts the malignant and benign stages of breast cancer.

1.1. Machine learning techniques

The tendency on the use of ML techniques is increasing with the increase in suggesting patient tailored medication, disease prognosis and prediction. In such cases the features taken into account are the prognostic features and predictive features in the field of cancer therapy. These features are not dependant on any particular kind of treatment. On using ML technique, the main things to note are the type of method, the type of data to be used. Much work is done on methods to choose most informative features that can be used later in the classification in Machine Learning.

For a prognostic procedure the factors that are taken into consideration are gene expression profiles, parameters related to clinical observation and factors from patient's family history. On aiming the cancer prediction the tasks to be performed are prediction of vulnerability (probability of developing), reappearance (probability of developing the same type of cancer again after treatment) and endurance (prediction of survival in terms of both disease oriented and overall progress oriented). The expectation of disease outcome generally alludes to the instances of (i) lifespan probability, (ii) survivability, (iii) advancement in disease and (iv) treatment affectability

2. RELATED WORK

A prediction on the survivability rate of patients with an indication of the breast cancer has been explored using data mining techniques in *Bellaachia et. al.* This kind of data is used for the Surveillance, Epidemiology, and End Result exploration for public usage. The total no. of cancer data used for the pre-processing stage is 151, 886 records, which taken from the above-mentioned database in 16 various fields. Three data mining techniques have been investigated by the authors: neural network with back propogation, the C4.5 version of the decision tree algorithms and the very popular Naive Bayes algorithm, Finally they conclude out of three algorithms applied, the C4.5 algorithm has been better than other two techniques.

Orlando Anunciação et. al., investigate the implication of decision trees for diagnosis and identification of breast cancer. The dataset used for this work is prepared by the department of genetics, Faculty of Medical Sciences Universidade Nova de Lisboa with hundred and sixty four controls and ninety four patient data's is used implied in WEKA , a machine learning tool. To confirm the association origin statistical permutation-based test was used. Finally, a conclusion from the cancer patient data held a statement that there was evidence of only one control stage for the thirteen different case with a Fisher Exact Test value of 9.7×10^{-6} along with a p-value of 0.017. Hence the analysis using a decision tree was found to be helpful in analysing the cancer data, the statistical associations between the data was kept intact due to the correlation between the leaf nodes in the decision tree.

Sumbaly et. al., proposed a computer-aided diagnosis system for detection of the cancerous cells in the breast region of women by analysing the association rules and by employing neural network. The reduction of the highly voluminous data related to the breast cancer database has been experimented based on the

application of association rules for intelligent classification using a neural network. The proposed technique performance is compared with existing model. A Wisconsin breast cancer database which composes of 9 attributes and 699 cases has been utilized for this purpose. Finally, the proposed innovative model has been stated to be effectual for diagnosing diseases even for other medical applications.

Ali Raad et. al., proposed a study on the development of an innovative method for breast cancer differentiation amongst several possible classes of breast cancer. They have utilized the Wisconsin diagnostic dataset and the prognostic breast cancer dataset for feature selection and for the implementation of neural network techniques in order to classify the data with the effectual experimentation on multi-layer perceptrons and its radial basis function. Totally nine features has been selected as a representation and is fed into the input layer to the neural network. Finally, input features are primely classified into two main classes of cancer cells (benign and malignant). The proposed technique shows 97 % positive rate of classification using radial basis function network.

Sudhir. D et. al., proposed innovative method to disease identification using supervised classifier called support vector machine. It is only for classification purpose, not for function approximation. Using kernel patron function implementing the classifier. The data's used in this work totally 683. In a further set of data of 117 data's is set up using a neural network. Using Support Vector Machine and artificial neural network, the identification the cancerous cell which actually cause the breast cancer are comparably detected in a precisely efficient way with a higher accuracy than by the professional people. According to the statement from the people in the medicinal domain there is a possibility of 85% of breast cancer detection in general and this detection rate has been effectively made to reach around 97% with the employment of the Support Vector machine recognition. This high rate of accuracy can be taken into consideration by the Doctor's to make decisions on the requirement of Biopsy and will help in the reduction of unwanted biopsies which induce a trauma in the patients.

Thein et. al., have proposed an approach for differentiating the classes of breast cancer with the help of the neural network. To achieve this author have fulfilled the local optima problem of neural network differential evolution algorithm for conclusive the optimal value or near optimal value for artificial neural network parameters. To avoid the issue longer iteration in training time and lower classification time is needed, the differential evolution algorithm further collaborates with island based model. It increases the overall accuracy and takes very reduce training time by making an analysis between two different migration topologies.

The applications of various types of statistical approaches for the purpose of identifying the presence of breast cancer and to enhance the automated breast cancer diagnosis system has been discussed by *Tuba kiyari et. al.*, 2004. A WBCD database has been used by the researcher to evaluate the performance of the multilayer perceptron. Neural networks with several variations has been explored such as the one with Radial basis function (RBF), one with the general regression (GRNN), one with probabilistic theory (PNN) and one with a multi layer perceptron (MLP) for the experimentation on classification performance and the comparative results are as mentioned: 96.18% for Radial Basis Function (RBF), 98.8% for GRNN, 97% for PNN, and 95.74% for MLP. Hence it is clear from the results that neural networks provide a better way to structure the cancer data and can be applied to develop a system to diagnose breast cancer.

Senturk et. al., have analyzed the various performance of seven classification prediction models such as the Artificial Neural Networks (ANN), the Support Vector Machines (SVM), the Naive Bayes (NB), the Decision Trees (DT), the Logistic Regression (LR), the Discriminant Analysis (DA), and the K-nearest neighbor (KNN) for the early automatic identification and detection of breast cancer through Rapid Miner Tool. The authors have concluded that the supervised classifier support vector machine algorithm has outperformed than all other six techniques in the prediction of the existence and non-existence of the disease with 96%.

Ronak Sumbaly et. al., Cancer is of an immense concern throughout all the contries in the world. It is a dangerous disease, which is destroys several lives in many cases and has shattered the life of their loved one's life too. It affects the next generation of lives too in many cases as it has the ability to be passed across the genes in a hereditary manner. Breast cancer has been accounted to be the second prime type of cancer causing huge amounts of deaths. Women have been found to be more susceptible to this deathly disease in all regions of the developed and also the developing countries in the world in the last few years. Around 40,000 women die in every year due to this disease, Everyday in the world for every 13 minute's one woman lost their life from this disease. The only way in which we can save our mankind especially women who are susceptible to this disease is developing some advanced and accurate detection system of breast cancer. Authors present an automated breast cancer identification system which shows their sincere efforts in being a support towards breast cancer detection with the help of decision tree based data mining technique. The classification of breast cancer can be pinned down to two categories such as benign (this type cancerous type of lacks an ability to invade the neighboring tissue and is comparatively less harmful than the other type) and malignant (this type of cancerous cells has an ability to invade neighboring tissue and is more harmful) breast tumors. The intent of this paper is to portray the several vicious effects of cancer along with a general note on its types, risk factors, symptoms which implicitly support the need for the application of the various data mining approaches for the development of a efficient and accurate identification and diagnosis system against the cancer cells.

3. CONCLUSION

This article holds a survey on the performance of diverse machine learning algorithms such as the Support Vector Machine(SVM), decision trees has been studied. In the past, Many researchers have applied majorly classification algorithm to predict the predicting cancer, mainly breast cancer. In the last years, the major focus is on the areas of progressing in the development of predictive models by unsupervised and supervised ML methods which aims to forecast exact disease outcomes. from the analysis of previous results, it is clearly shows that there is a need of integrating the heterogeneous data, which is applied on various machine learning techniques like feature selection and classification in cancer research.

References

- Williams, Kehinde, Peter Adebayo Idowu, Jeremiah Ademola Balogun, and Adeniran Ishola Oluwaranti. "Breast cancer risk prediction using data mining classification techniques." *Transactions on Networks and Communications*, 3(2), 01-11, 2015.
- Bellaachia, Erhan Guven, Predicting Breast cancer survivability using data mining techniques, Department of computer science, George Washington University. *International Journal of Computer Science and Information Technologies*, Vol. 4 (6) , 2013, 1023-1028

- Orlando Anunciação, Bruno C. Gomes, Susana Vinga, Arlindo L.Oliveira, José Rueff. A Data Mining Approach for the Detection of High-Risk Breast Cancer Groups, *Advances in bioinformatics*, Vol74, pp 43-51.
- Sumbaly, Ronak, N. Vishnusri, and S. Jeyalatha. "Diagnosis of Breast Cancer using Decision Tree Data Mining Technique." *International Journal of Computer Applications*, 98(10), 16-24, 2014.
- Ali Raad*, Ali Kalakech**, Mohammad Ayache, Breast cancer classification using neural network approach ;MLP & RBF, 13th International Arab Conference on Information Technology , ACIT'2012, Dec.10-13, ISSN.182-0857
- Sudhir D. Sawarkar, Ghatol, Neural network aided Breast Cancer Detection & diagnosis using Support Vector Machine, *Proceeding of 7th WSEAS International conference on Neural Network*, June 12-14, 2006, (pp 158-163).
- Thein, HtetThazin Tike, and Khin Mo Mo Tun. "An Approach for Breast Cancer Diagnosis Classification Using Neural Network." *Advanced Computing*, 6(1), 1-10, 2015.
- Tuba kiyan, Tulay Yildirim "Breast cancer diagnosis using statistical neural networks", *Journal of Electrical and Electronic Engineering*. International XII. Turkish Symposium on Artificial Intelligence and Neural Networks – TAINN 2003
- Senturk, ZehraKarapinar, and Resul Kara. "Breast Cancer Diagnosis via Data Mining: Performance Analysis of Seven different algorithms." *Computer Science & Engineering*, 4(1), 35-46, 2014.
- Ronak Sumbaly, N. Vishnusri, Jeyalatha Diagnosis of Breast Cancer using Decision Tree Data Mining Technique, *IJCA*, Volume 98, No.10, July 2014.

