



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

Efficient Diagnosis of Breast Cancer Patients

Muktai Itankar, Vaishnavi Bandewar, Nikita Desarda, Antara Darbeshwar,

B.E Students, Department of Computer Science and Engineering, Vishwakarma Institute of Information Technology, Savitribai Phule Pune University, Pune, India

ABSTRACT: The term Breast Cancer is referred to as a malignant tumor which develops from cells in the breast. Breast Cancer is the second cause of deaths in women. Even though breast cancer is rare case in men still a few cases have been reported in men as well. In today's world the technological advancements in cancer treatments has lead to different new innovations which has increased the survival rates. The objective of this thesis is to develop an intelligent decision support application for diagnosis of breast cancer using Case-Based Reasoning (CBR) algorithm to predict class of cancer for patients. CBR is a algorithm which uses past cases for solving new problems. This application provides recommendation of hospitals. In this research paper we also proposed a breast cancer classification model based on comparison of case based reasoning algorithm and other machine learning algorithms to classify the breast cancer tumor as malignant or benign.

KEY WORDS: Breast Cancer detection , Case Based Reasoning (CBR), Machine Learning, Recommendation system.

I. INTRODUCTION

Breast Cancer is a chronic disease that requires permanent control diagnosis for death risk prevention. According to American Cancer Society breast cancer is second cause of deaths in women. In 2017, it has been estimated that about 30% of new cancer cases diagnosed are breast cancers. The chances for successful treatment depends mainly on the stage of cancerous tumors and it becomes higher when the tumors are in the early stages of development. Breast Cancer causing factors are family history, race, age. Breast cancer diagnosis, classification and prediction techniques have been widely researched are in past decades in the world of medical informatics. Many articles had been published which tries to classify breast cancer dataset using different techniques such as support vector machine, Bayesian classifiers, fuzzy logic, neural networks and decision trees. This paper presents an intelligent decision support application for diagnosis of breast cancer using Case-based Reasoning algorithm to predict the class of cancer for patients. This paper [3] also demonstrates the comparison of different classification techniques like case based reasoning (CBR) and Support Vector Machine. The data utilized in this research is the breast cancer data. It holds a total of 699 data and a dimension of 699 rows and 11 columns. The attributes details are listed in table 1.

Table 1 Wisconsin database attributes

	Attribute	Domain
1	Sample code number	1-10
2	Clump Thickness	1-10
3	Uniformity of Cell size	1-10
4	Uniformity of Cell shape	1-10
5	Marginal Adhension	1-10
6	Single Epithelial Cell size	1-10
7	Bare Nuclei	1-10



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

8	Bland Chromatin	1-10
9	Normal Nucleoli	1-10
10	Mitoses	1-10
11	Class	2 for Benign & 4 for Malignant

A. Literature survey

This paper [4] demonstrates the comparison of different classification techniques like Case Based Reasoning and other machine learning algorithms. Case-Based Reasoning is an algorithm which uses past cases for solving new problems [5]. Case-Based Reasoning includes four main phases Retrieve, Revise, Reuse. Retrieve means to find out the most similar case. Reuse means to use the previous cases and try to solve the problems. Revise means to update the given solution if necessary. Retain means to consider the new solution as a part of new case. Support Vector Machine (SVM) is the other algorithm which is used for detection of cancer. It is considered as a powerful method that has been used in many different applications. Support vector machine algorithm is also known as support vector networks. The underlying idea of Support vector machine algorithm is the linear separability or the hyper plane classifier. K-Nearest Neighbor (KNN) is yet another algorithm which was used to detect the stage of cancer[6]. This algorithm classifies cases based on the similarities found among them. This algorithm is considered to be popular for pattern recognition. It is a Lazy learning algorithm which is used for regression and classification. An object is classified by a majority of its neighbors. K must always be a positive integer. The neighbors are selected from a set of objects for which the correct classification is known.

Table 2 Comparison of existing techniques

Author	Technique	Algorithm	Pros	Cons
Seema Singh, Sunita Saini, Mandeep Singh	Artificial Neural Network	Adaptive Resonance Theory	The accuracy of clustering of ARNN model is better than other models.	In a developed software system, it would be desirable to improve the accuracy, precision and recall.
Ali Raad, Ali Kalakech, mohammad Ayache	Artificial Neural network	Back Propagation algorithm	The supervised RBF neural approach drove by the learning algorithm GBP works well, in terms of accuracy, efficiency	A detailed comparison between multi layer networks MLP and RBF showed that the model constructed from RBF neural Network is much more efficient than other model based on the MLP.



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

Gouda Salman, M.B.	Data Mining	SMO, J48, MLP	The experimental results in Winconsin dataset shows the fusion between MLP and J48 classifiers with features selection is superior to the other classifiers.	Their accuracy can be increased.
S.Muthuselvan, Dr.K.Soma Sundaram , Dr.Prabasheela	Data Mining	J48	J48 algorithm has accuracy of 86.36%.	The accuracy can be increased.

B. Methodology

The BCDA (Breast Cancer Diagnosis Application) has been implemented. The BCDA has been developed using the android studio platform. Android 4.3 has been used. The implementation stage involves designing and storing the database, designing the GUI (Graphical User Interface), connecting the database and application, and coding the main algorithm. The database has been created in SQLite database engine. The source code is written partially in python language and partially in Java language. The dataset of Wisconsin Breast Cancer has been used for developing the database. The CBR algorithm has developed. The dataset consists of 699 samples. Each sample has nine attributes. The attributes have values ranging from 1-10. The attributes selected are uniformity of cell size, clump thickness, uniformity of cell shape, mitoses, marginal adhesion, single epithelial cell size, normal nucleoli, bare nuclei and bland chromatin. This dataset is retrieved from UCI machine learning and stored in SQLite database. The application takes input as the values of the nine attributes. It then connects with the database, retrieves the data from the database and calculates local and global similarity of the new case with each previous case stored in the database. Based on the similarity, the matched case is found. The result is predicted with the help of the matched case. The result is then displayed on the application.

C. Conclusion

A framework has been proposed in this work to assist the early detection of breast tumors [3]. The classification of class of tumors either benign or malignant group can be done using this algorithm and this algorithm has a great potential to be implemented in diagnosis of breast cancer.

REFERENCES

- [1] Breast Cancer Organization, U.S. Breast Cancer Statistics, March 2017. http://www.breastcancer.org/symptoms/understand_bc/statistics.
- [2] A. Aamodt and E. Plaza, "Case-Based Reasoning: Foundational Issues, Methodological Variation and System Approaches," AICOM, vol. 7 no. 1, 1994, pp. 39-58.
- [3] Kathija, Shajun Nisha, "Breast Cancer Data Classification Using SVM and Naïve Bayes Techniques", International Journal of Innovative Research in Computer and Communication Engineering- Vol. 4, Issue 12, December 2016.
- [4] <https://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/wdbc.data>.
- [5] Breast Cancer, <<http://patient.info/health/breast-cancer-leaflet>>.
- [6] U.S. Cancer Statistics Working Group. United States Cancer Statistics: 1999-2008 Incidence and Mortality Web-based Report. Atlanta (GA): Department of Health and Human Services, Centers for Disease Control and Prevention, and National Cancer Institute; 2012