ISSN: 2349-9141

Available online at http://www.ijirr.com

International Journal of Information Research and Review
Vol. 2, Issue, 04, pp. 660-663, April, 2015

OPEN ACCESS JOURNAL

Full Length Research Paper

An Approach for Identifying Cancer Using Support Vector Machine Algorithm

***Shobana Angeline, A and Ms. Shahar Banu, S.**

*Department of Computer Applications, B S Abdur Rahman University, India

***Corresponding Author**

Received 22nd March 2015; Accepted 30th April 2015

Abstract

Breast cancer is one of the major causes of death in women when compared to all other cancers. Breast cancer has become the most hazardous types of cancer among women in the world. Early detection of breast cancer is essential in reducing life losses. This paper presents an approach for identifying breast cancer using classification algorithm. This project comprises of three modules user module which user can enter the details once they have enter the details an unique id will be generated to each user .The user will authenticate based on their credentials. The patient enters his patient id; a message box will be prompting them to view the results after the process complete. The second module is classification module in which individual observations are analyzed into set properties. These properties may variously to user to user. Based on that training dataset the test data will assign to the specific category. Third module is prediction module in which the result will be predicted at the last stage with the help of SVM. We compare classification techniques in Waikato Environment for Knowledge Analysis (weka) software and comparison results with the help of Support Vector Machine (SVM) The SVM will predict the result with based on the risk score once the result is predicted it is stored, so that the user can view the result. This project is implemented in java as the front end and mysql as the back end. This project aims to implement an effective prediction on breast cancer using classification algorithm .with the help of it the user can know the cancer status. From this project we infer that the SVM are more suitable in handling the classification problem of breast cancer prediction

Keywords: Breast Cancer. Classification. Decision Tree, SVM and Weka.

To cite this paper: Shobana Angeline, A and Ms. Shahar Banu, S. 2015. An approach for identifying cancer using support vector machine algorithm. *International Journal of Information Research and Review. Vol. 2, Issue, 04, pp. 660-663.*

INTRODUCTION

Breast cancer is one of the major causes of death in women when compared to all other cancers. Cancer is a type of diseases that causes the cells of the body to change its characteristics and cause abnormal growth of cells. Most types of cancer cells eventually become a mass called tumor. The occurrence of breast cancer is increasing globally. It is a major health problem and represents a significant worry for many women. Early detection of breast cancer is essential in reducing life losses. However earlier treatment requires the ability to detect breast cancer in early stages. Early diagnosis requires an accurate and reliable diagnosis procedure that allows physicians to distinguish benign breast tumors from malignant ones. The International Journal of Innovations in Engineering and Technology (IJIET) automatic diagnosis of breast cancer is an important, real-world medical problem. Thus, finding an accurate and effective diagnosis method is very important.

Related works

G. Ravi Kumar (Ravi Kumar *et al*., 2013) Different methods for breast cancer detection are explored and their accuracies are compared. Various data mining techniques are implementer to compare the accuracies. Samar Al-Qarzaie, Sara Al-Odhaibi, Bedoor Al-Saeed, and Dr. Mohammed Al-Hagery (Syed Shajahaan *et al*., 2013) This paper presents an approach for using data mining techniques (decision tree) for the early prediction of breast cancer disease, The drawback is that work cannot be intended to find another data sources to use it in the test and prediction. Abdelghani Bellaachia, Erhan Guven (American Cancer Society, 2005-2006) this paper has outlined, discussed and resolved the issues, algorithms, and techniques for the problem of breast cancer survivability prediction in SEER database. This study clearly shows that the preliminary results are promising for the application of the data mining methods into the survivability prediction problem in medical databases but our analysis does not include records with missing data.

Proposed system

In this paper I have apply classification on the data set, we have to assign the patients to different groups based on their survival time (which is available in number of date). This task is really important. Consider the case that we divided the data

into two different groups; the first group contains the patients who died before three months, and the second group contains the data for the patients who died after three months. The low error Rate and performance of classification techniques .In this project we are implementing support vector machine **(**SVM). SVM is an algorithm that attempts to find a linear separator (hyper-plane) between the data points of two classes in multidimensional space

Classification Techniques

Building accurate and efficient classifiers for large databases is one of the essential tasks of data mining and machine learning research. Building effective classification systems is one of the central tasks of data mining. Many different types of classification techniques have been proposed in literature that includes Decision Trees, SVM etc.

Decision Tree

Decision tree models are commonly used in data mining to examine data and induce the tree and its rules that will be used to make predictions (10). The prediction could be to predict categorical values (classification trees) when instances are to be placed in categories or classes. Decision tree is a classifier in the form of a tree structure where each node is either a leaf node, indicating the value of the target attribute or class of the examples, or a decision node, specifying some test to be carried out on a single attribute-value, with one branch and sub-tree for each possible outcome of the test.

A decision tree can be used to classify an example by starting at the root of the tree and moving through it until a leaf node is reached, which provides the classification of the instance.

Support Vector Machine (SVM)

SVMs are a set of related supervised learning methods that analyze data and recognize patterns, used for Classification and regression analysis. SVM is an algorithm that attempts to find a linear separator (hyper-plane) between the data points of two classes in multidimensional space. SVM represents a learning technique which follows principles of statistical learning theory. Generally, the main idea of SVM comes from binary classification, namely to find a hyper plane as a segmentation of the two classes to minimize the classification error. The SVM finds the hyper plane using support vectors (training tuples) and margins (support vectors). The Sequential Minimal Optimization (SMO) algorithm is a simple and fast method for training a SVM.

RESULT AND DISCUSSION

The user will authenticate based on their credentials. The patient enters his patient id a message box will be prompting them to view the results after the process complete. Once they enter the details, all details are entered into the local database. If the patient gives his patient id and presses submit button his result will appear In next module the individual observations are analyzed into set properties.

PatientID	Name	Age	Gender	MaritalStatus	NoofChildren	LivinArea	Education	OccupationalHazards	Sm
1	A. Swathi	56	Female	Married	3	Rural	School	None	No
2	Aarthi Sridhar	27	Female	Married	1	Rural	College	None	No
3	Abdul	15	Male	Unmarried	0	Urban	School	None	No
4	Abdullah	51	Male	Married	0	Urban	School	None	No
5	Adebaa Fathima.S.M	22	Female	Unmarried	0	Urban	College	None	No
6	Chandramathi	61	Female	Widowed	3	Urban	School	None	No
7	Abhijith Chandran	65	Male	Married	1	Rural	Uneducated	None	No
8	Abinaya.S	76	Female	Widowed	0	Urban	School	Radiation exposure	No
9	Abiramavalli.R	73	Female	Married	2	Rural	Uneducated	None	No
10	Abirami.I	39	Female	Married	2	Urban	Uneducated	None	No
11	Adhithyan.R	64	Male	Married	1	Urban	College	None	Ye
12	Afroze Sultana	56	Female	Married	0	Rural	School	None	No
13	Agila.G	19	Female	Unmarried	0	Rural	School	None	No
14	Aishwarya . L	49	Female	Married	1	Urban	School	Sunlight exposure	No
15	Ajay Kumar.R	64	Male	Married	0	Urban	School	None	No
16	Ajithkumar	5	Male	Unmarried	0	Urban	Uneducated	None	No
17	Akshitha . M	50	Female	Married	0	Rural	Uneducated	None	No
18	Amala	59	Female	Married	1	Rural	Uneducated	Chemical Exposure	No
19	Ammu	43	Female	Married	2	Urban	School	None	No
20	Amutha.N	13	Female	Unmarried	0	Rural	School	None	No
21	Anamika	41	Female	Married	0	Rural	School	None	No
22	Ananth	20	Male	Unmarried	0	Urban	School	None	No
23	Anantha Lakshmi.R	31	Female	Married	1	Urban	School	None	No
24	Anbu	18	Male	Unmarried	0	Urban	School	None	No
25	Andrew	9	Male	Unmarried	0	Urban	School	None	No

Database: medical Table: noncancer

Figure 1 Dataset Table

Figure 1. Homepage

Figure 2. User Information page

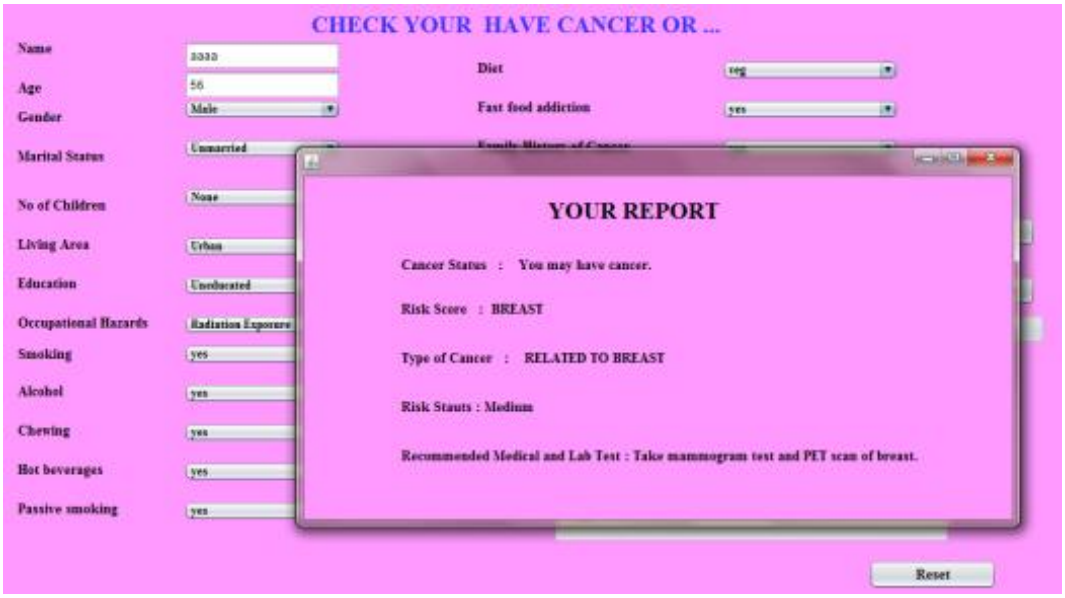

Figure 3. Result page

These properties may variously be categorical by habits type like passive smoking, alcohol, chewing, other classifiers work. Based on that training dataset the test data will assign to the specific category. At last we will compare the result with the decision tree algorithm. The result will be predicted at the last stage with the help of SVM. The SVM will predict the result with based on the risk score once the result is predicted it is stored, so that the user can view the result. SVM performs well on data sets that have many attributes, even if there are very few cases on which to train the model. There is no upper limit on the number of attributes; the only constraints are those imposed by hardware. Traditional neural nets do not perform well under these circumstances. The patient can see the result with the help of the id given to them. They can go the home page in that old user option is given so by clicking the option and enter the id result will be displayed

Algorithm

Step 1: start the process
Step 2: Training set: a number of expression profiles with known labels which represent the true population.
Step 3: Learning/Training: finding a decision rule which explains the training set well.
Step 4: Generalization ability: how does the decision rule learned from the training set generalize to new specimen.
Step 4: stop the process

Conclusion

This research is concerned with the study and analysis of Data Mining and Data classification algorithms, analyzing the existing methods for Predicting breast cancer Disease and to design and develop an efficient and effective method for predicting cancer disease. In this research SVM Classifier techniques has been used. The compactness and connectedness for complementary measures are used and it is found that the efficiency and effectiveness of the method for predicting cancer Disease is better in support vector machine (svm).

REFERENCES

American Cancer Society. Breast Cancer Facts & Figures 2005-2006. Atlanta: American Cancer Society, Inc. (http://www.cancer.org/).

Bellachia, A. and E. Guvan, 2006 .“Predicting breast cancer survivability using data mining techniques”, Scientific Data Mining Workshop, in conjunction with the 2006 SIAM Conference on Data Mining.

Endo, A., T. Shibata and H. Tanaka 2008, Comparison of seven algorithms to predict breast cancer survival, Biomedical Soft Computing and Human Sciences, vol.13, pp.11-16.

Ravi Kumar, G. 2013. An Efficient Prediction of Breast Cancer Data using Data Mining Techniques, *International Journal of Innovations in Engineering and Technology*

Syed Shajahaan, S. S. Shansi, V. Manor Chita, 2013. Application of Data Mining Techniques to Model Breast Cancer Data, International Journal of Emerging Technology and Advanced Engineering

****