International Journal of Engineering & Technology, 7 (3.34) (2018) 807-809

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Building a Prediction Model to Predict the Breast Cancer using ANN'S Kernel Based Method

Naganandini.G[1], Vishwanth.R.H[2]

Research Scholar, School of Computing and IT, REVA University, Bengaluru
Professor, School of Computing and IT, REVA University, Bengaluru
*Corresponding Author Email: [1]*nandini.geneflux@gmail.com* , [2]*vishwanath.rh@reva.edu.in*

Abstract

The most vulnerable disease is Breast Cancer. Many different methods and processes were identified for the early detection and for the remedy of Breast Cancer .Till date only two genes have been identified and which accounts for genetic characteristics to a large extent. A particular variant has been identified as a major hetrozygous component which amounts as a major component for the breast cancer. The rigorous working on Human Genome project has proved that the family history plays a major role in detecting Breast cancer and also helps in gaining knowledge about genetic variations which depicts as a high risk factor among all the cancer types. The overall observation has concluded that the risk of this disease is mostly among the women community who has a family history. The ultimate aim is to classify the genes that are most significant and non-significant among all the genes present in the breast cancer tissue at the early stages using Naïve's Bayesian's and c5.0 algorithm and hence build a predictor model to predict breast cancer using ANN's Kernal based method.

Keywords: *Mutations, oncology,neural networks,machine learning, kernel based*

1. Introduction

Breast cancer has been in the top list as the most commonly analysed disease in the women community across the world. Remarkable progress has been made in analyzing and treating breast cancer. The rate of survival of women seen a increase from 1980s , however overall breast cancer deaths have now begun to decrease. The five yearcomparative rate of survival for all women with breast cancer is eighty five percent; the five year rate of survival for women with this deadly disease breast cancer is now ninety six percent. Significant changes in the breast cancer genes contributes for only a part of hereditary breast cancer. This implies that some genes contributes to cancer caused by hereditary factors too. The segregation of most unique and non-unique genes among all the genes present in the breast cancer tissue at the early stage , and predict the breast cancer using neural network's Kernel based method.

2. Background

Currently, research focus of in the field of identifying the breast cancer genes . The samples are collected and epidemiology studies are being carried out . Investigations for both familial and non-familial for BRCA1 genes is under progress.

3. Literature Survey

Qing Ping et al [1] has identified many symptom patterns for symptom clusters from the survivors of breast cancer which was got from both public media and secondary data with the help of improvised K-medoid clustering. A lot of survey were done and finally more than 50,000 available messages were collected via social media through Practo.com and a lot of questionnaires were posed , around 650 were gathered as a part of a survey in research . The symptoms of data collected from the public media were comparatively less when compared to that of the primary data from the research scholars which made the task of partitioning easier. The performance of clustering were improved by the transformation of K-medoid clustering by reassigning thE points in the cluster .The assignment of these symptoms to other clusters and the reduction of ASW on the whole avoided the problem of local optima trapping. On the whole ASW helped in improvised presentation of the final clustering. The aggregation of all the symptoms of common symptoms from various sources was grouped into various clusters. The clustering results segregates symptom clusters that are common among social media data and clinical data by some major symptoms. The gathered study data worked cooperatively to achieve the integrative results.

Ritu Chauhan et al [2] proposes the efficiency of combination of Hierarchical Agglomerative Clustering and K-means for the detection of number of clusters ,for the increase in the quality of clusters. Partitioning of n different objects and assigning them to k clusters in which every object gets inside any of the clusters framed by the nearest mean. Clustering makes all the tasks easier. A clear distinction between the different clusters is made maximum through this method. The ultimate clusters k with a lot of distance difference could be determined only through the data sets. Reducing the distance between the clusters within the system or reducing squared error function was the final goal.

Dechang Chen et al [3] proposed predictions accurately about the survival rates of patients suffering from cancer to diagnose patients for further process and treatment. Prediction of survival rate was determined by three factors (tumor, lymph and metastasis) system .Due to lot of disadvantages of this system a

new system came into picture. Nowadays due to the presence of huge data sets there was a need for a very powerful prediction systems. Luckily this was done using the latest machine learning and data that were partitioned were done PAM . A combination of dissimilarity measure along with hierarchical clustering is identified to find the different clusters. The overall approach finally adds and hence predicts accurately the rate of the breast cancer hit patients survivors.

S M Halawani et al [4] firmly asserted and made comparisons between different probabilistic algorithms and hence concluded that probabilistic clustering algorithms proved better in their performance where every data points were put under a particular cluster due to invariable distance measures.

Charles Edeki et al [5] proposes that a huge amount of efforts were made by computer to store ,manage ,process and analyze biological databases to design and implement . The primary focus was to predict the usage of different statistical tools such as R and Python and many other open source softwares to mine the data and predict their visualization and the extent of survivability of patients . The other primary goal was to develop and design a new algorithm to support different experts in their respective field. Different data mining and machine learning techniques were used to predict the survival rate of different breast cancer hit patients.

Zakaria Suliman zubi et al [6] performed few data mining techniques to identify different characteristics and to segregate each gene group for identification of cancer.

Labeed K Abdulgafoor et al [7] proposed the image analysis, feature extraction, time series to cluster and thenuse algorithm to be used for intensity based segmentation.

Sahar A.Mokhtar et al [8] have segregated classification models into different types so that the prediction of different breast masses were identified as severe .and could be done efficiently using different Machine Learning techniques .

Rajashree Dash et al [9] concluded that a hybridized K-means algorithm describes the dimensionality reduction stages through PCA, and places each point to the clusters.

The grouping algorithms performs better when compared to all other grouping algorithms arranged into one cluster. Only very few machine learning techniques were applied for the analysis of breast cancer data sets., each of the algorithm excelled each other in such a way that it was declared as the useful algorithm and they were used for the prediction of survival rate of cancer hit patients. Further PCA analysis can be performed to visualize all the information ,project the similarities , differences and contributions of each variables its dependencies and helps in detection of patterns.

4. Implementation

It could be applied when the mean of a cluster were defined and users were supposed to specify k. K-means were not suitable for discovering clusters of very different size. C5.0 proves better performance than Naïve Bayesian Classifier for Breast Cancer disease hit patients to get better results with accuracy, low error rate and performance. The performance ofC5.0 is at high level compared with other classifiers. Implementation using K-mediod clustering where each data object is considered for clustering and hence produces more accurate clusters and hence applying Artificial Neural Network's Kernel based method.

5. Methodology or Approach

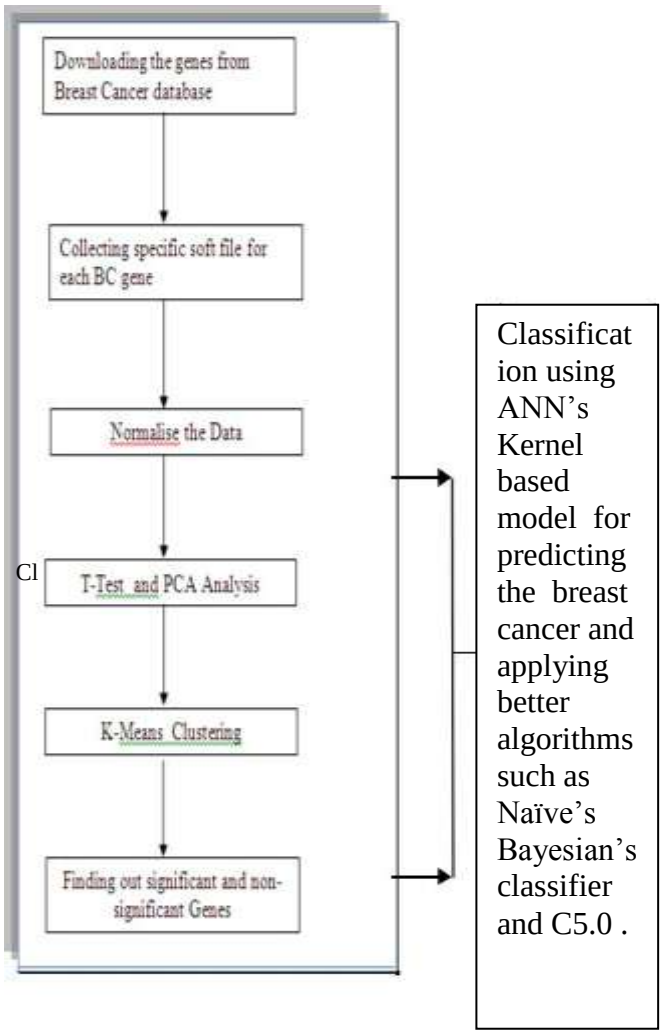


FIG 6.0.1

6. Relevant Work Done

plotfit()

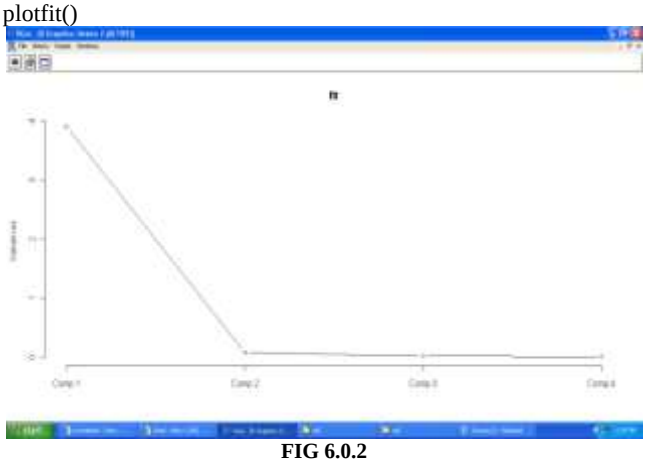


FIG 6.0.2

Biplot(fit)

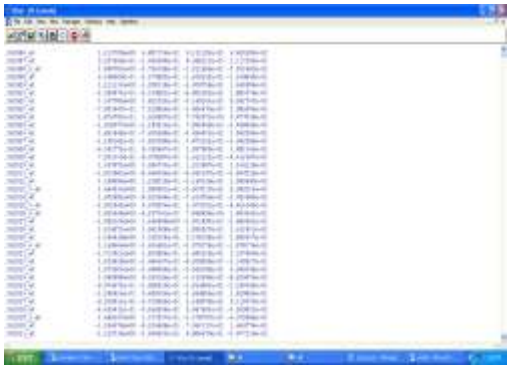


FIG 6.0.3
Aggregate()

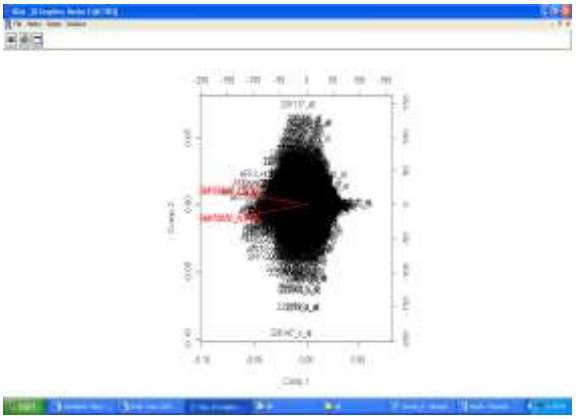


FIG 6.0.4
Clusplot()

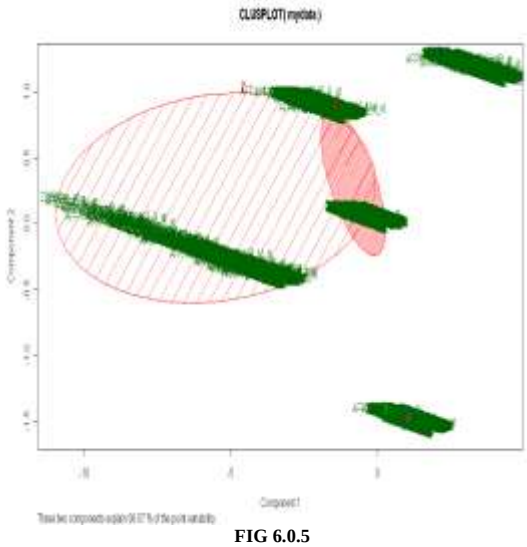


FIG 6.0.5
Plot(fit)

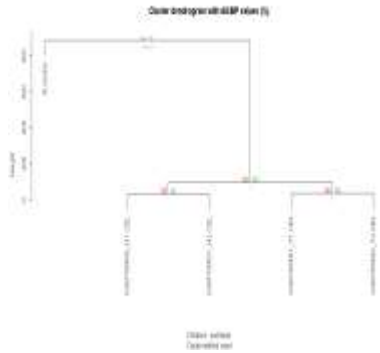
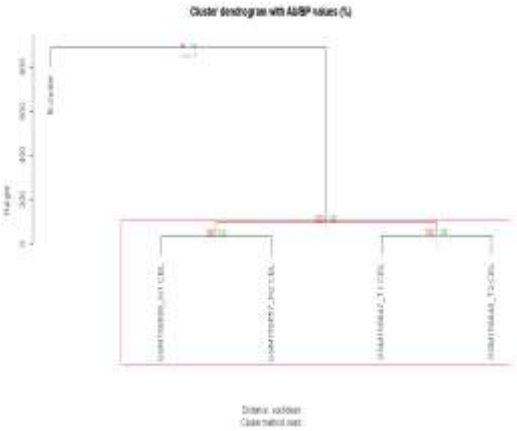


FIG 6.0.6

Pvrect()



7. Expected Outcome

Classification of significant and non significant breast cancer genes at the early stage using ANN's Kernel Based model and applying k-Means and k-Mediods clustering and hence building a predictor model to predict the Breast cancer. Applying Naïve's Bayesian and C5.0 algorithm to improvise the accuracy of segregation.

References

[1] Qing Ping, Christopher C. Yang, Sarah A. Marshall, Nancy E. Avis, and Edward H. Ip "Breast Cancer Symptom Clusters Derived From Social Media and Research Study Data Using Improved K-Medoid Clustering" IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS, VOL. 3, NO. 2, JUNE 2016.

[2] Ritu Chauhan , Harleen Kaur and

[3] Mafshar Alam "Data clustering method for Discovering Clusters in

[4] spatialcancerdatabases"International Journal of Computer Applications (0975-8887) Volume 10-No.6, November 2010.

[5] Dechang Chen, Kai Xing, Donald Henson, Li Sheng, Arnold M. Schwartz, and Xiuzhen Cheng2 "Developing Prognostic Systems of Cancer Patients by Ensemble Clustering" Hindawi publishing corporation, Journal of Biomedicine and Biotechnology Volume 2009,Article Id 632786.

[6] S M Halawan ,M Alhaddad and A Ahamad"A study of digital mammograms by using Clustering algorithms" Journal of Scientific & Industrial Research Vol. 71, September 2012, pp. 594-600. Number of Symptoms Risk Store Medium Low No yes No No International Research Journal of Engineering and Technology (IRJET)e-ISSN: 2395-0056 Volume: 02 Issue: 08 | Nov-2015.www.irjet.net p-ISSN: 2395-0072 © 2015, IRJET ISO 9001:2008 Certified Journal Page 1182

[7] Charles Edeki and Edmand "Comparative Study of Data Mining and Statistical Learning Techniques for PredictionofCancer Survivability" Mediterranean journal of Social Sciences Vol 3November 2012, ISSN: 2039-9340.

[8] Zakaria Suliman and zubi "Improves Treatment Programs of Breast Cancer using Data Mining Techniques" Journal of Software Engineering and Applications, February 2014, 7, 69-77.

[9] Labeed K Abdulgafoor and Aji George "Detection of Tumor usingModified K-Means Algorithm and SVM" International Journal ofComputer Applications (0975 – 8887) National Conference on Recent Trends in Computer Applications NCRTCA 2013.

[10] Sahar and Ala M Elsayad "Predicting the Severity of Breast Masses with Data Mining Methods" International Journal of Computer Science Issues, Vol. 10, Issues 2, No 2, March 2013 ISSN (Print):1694-0814| ISSN (Online):1694-0784 www.IJCSI.org.

[11] Rajashree Dash , Debahuti Mishra , Amiya Kumar Rath , MiluAcharya "A hybridized K-means clustering approach for high dimensional dataset" International Journal of Engineering, Science and Technology Vol. 2, No. 2, 2010, pp. 59-66.