DATA MINING CLASSIFICATION TECHNIQUES APPLIED FOR BREAST CANCER DIAGNOSIS AND PROGNOSIS

SHELLY GUPTA

AIM & ACT, Banasthali University, Student M.Tech. (CS),
Banasthali, India

DHARMINDER KUMAR

Dean, Faculty of Engineering and Technology, GJUS&T
Hisar, India

ANAND SHARMA

Department of CSE, GJUS&T, Project Fellow,
Hisar, India

Abstract

Breast Cancer Diagnosis and Prognosis are two medical applications pose a great challenge to the researchers. The use of machine learning and data mining techniques has revolutionized the whole process of breast cancer Diagnosis and Prognosis. Breast Cancer Diagnosis distinguishes benign from malignant breast lumps and Breast Cancer Prognosis predicts when Breast Cancer is likely to recur in patients that have had their cancers excised. Thus, these two problems are mainly in the scope of the classification problems. This study paper summarizes various review and technical articles on breast cancer diagnosis and prognosis. In this paper we present an overview of the current research being carried out using the data mining techniques to enhance the breast cancer diagnosis and prognosis.

Keywords: Breast cancer; Diagnosis; Prognosis; Data Mining; Classification.

1. Introduction

Breast cancer has become the leading cause of death in women in developed countries. The most effective way to reduce breast cancer deaths is detect it earlier. Early diagnosis requires an accurate and reliable diagnosis procedure that allows physicians to distinguish benign breast tumors from malignant ones without going for surgical biopsy. The objective of these predictions is to assign patients to either a "*benign*" group that is non-cancerous or a "*malignant*" group that is cancerous. The prognosis problem is the long-term outlook for the disease for patients whose cancer has been surgically removed. In this problem a patient is classified as a 'recur' if the disease is observed at some subsequent time to tumor excision and a patient for whom cancer has not recurred and may never recur. The objective of these predictions is to handle cases for which cancer has not recurred (censored data) as well as case for which cancer has recurred at a specific time. Thus, breast cancer diagnostic and prognostic problems are mainly in the scope of the widely discussed classification problems. These problems have attracted many researchers in computational intelligence, data mining, and statistics fields.

Cancer research is generally clinical and/or biological in nature, data driven statistical research has become a common complement. Predicting the outcome of a disease is one of the most interesting and challenging tasks where to develop data mining applications. As the use of computers powered with automated tools, large volumes of medical data are being collected and made available to the medical research groups. As a result, Knowledge Discovery in Databases (KDD), which includes data mining techniques, has become a popular research tool for medical researchers to identify and exploit patterns and relationships among large number of variables, and made them able to predict the outcome of a disease using the historical cases stored within datasets.

The objective of this study is to summarise various review and technical articles on diagnosis and prognosis of breast cancer. It gives an overview of the current research being carried out on various breast cancer datasets using the data mining techniques to enhance the breast cancer diagnosis and prognosis.

2. Breast Cancer: An Overview

Breast cancer is the most common cancer disease among women, excluding non-melanoma skin cancers. The information about the tumor from certain examinations and diagnostic tests are gathered using staging to determine how widespread the cancer is. The stage of a cancer is one of the most important factors in selecting treatment options, and it uses the *Tumour, Nodes and Metastasis* (TNM) system. When a patient's *T*, *N*, and *M* categories have been determined then this information is combined in a process known as stage grouping to

determine a woman's disease stage. This is expressed as from Stage 0 (the least advanced stage) to Stage *IV* (the most advanced stage)[1]. Breast cancer is a malignant tumor, grew from cells of the breast. Hence, cancer of breast tissue is called breast cancer. Worldwide, it is the most common form of cancer in females that is affecting approximately 10% of all women at some stage of their life in the Western world. Although significant efforts are made to achieve early detection and effective treatment but scientists do not know the exact causes of most breast cancer, they do know some of the risk factors (i.e. ageing, genetic risk factors, family history, menstrual periods, not having children, obesity)that increase the likelihood of developing breast cancer in females. The survival analysis that is the study of time to an event of interest, such as disease occurrence or death also gives the physicians and the patient better information with which to plan treatment and may eliminate the need for a prognostic surgical procedure.

3. Knowledge Discovery and Data Mining

This section provides an introduction to knowledge discovery and data mining. We list the various analysis tasks that can be goals of a discovery process and lists methods and research areas that are promising in solving these analysis tasks.

3.1. *The Knowledge Discovery Process*

The terms Knowledge Discovery in Databases (KDD) and Data Mining are often used interchangeably. KDD is the process of turning the low-level data into high-level knowledge. Hence, KDD refers to the nontrivial extraction of implicit, previously unknown and potentially useful information from data in databases. While data mining and KDD are often treated as equivalent words but in real data mining is an important step in the KDD process. The following fig. 1 shows data mining as a step in an iterative knowledge discovery process.
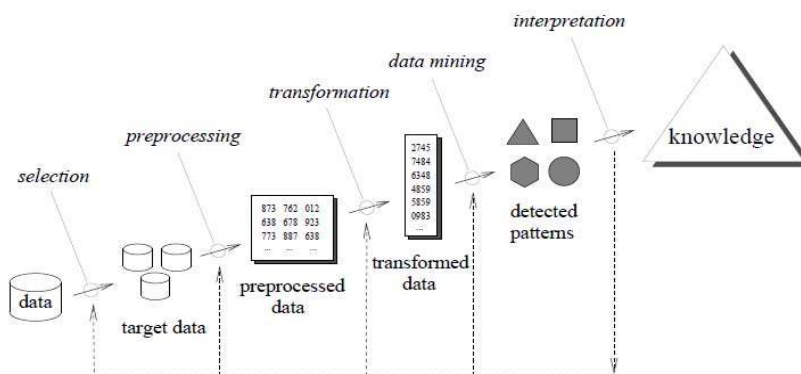

Fig. 1. Steps in KDD

The Knowledge Discovery in Databases process comprises of a few steps leading from raw data collections to some form of new knowledge[2]. The iterative process consists of the following steps:

(1) *Data cleaning:* also known as data cleansing it is a phase in which noise data and irrelevant data are removed from the collection.

(2) *Data integration*: at this stage, multiple data sources, often heterogeneous, may be combined in a common source.

(3) *Data selection:* at this step, the data relevant to the analysis is decided on and retrieved from the data collection.

(4) *Data transformation:* also known as data consolidation, it is a phase in which the selected data is transformed into forms appropriate for the mining procedure.

(5) *Data mining:* it is the crucial step in which clever techniques are applied to extract patterns potentially useful.

(6) *Pattern evaluation:* this step, strictly interesting patterns representing knowledge are identified based on given measures.

(7) *Knowledge representation:* is the final phase in which the discovered knowledge is visually represented to the user. In this step visualization techniques are used to help users understand and interpret the data mining results.

3.2. *Data Mining Process*

In the KDD process, the data mining methods are for extracting patterns from data. The patterns that can be discovered depend upon the data mining tasks applied. Generally, there are two types of data mining tasks: *descriptive data mining tasks* that describe the general properties of the existing data, and *predictive data mining tasks* that attempt to do predictions based on available data. Data mining can be done on data which are in quantitative, textual, or multimedia forms.

   Data mining applications can use different kind of parameters to examine the data. They include association (patterns where one event is connected to another event), sequence or path analysis (patterns where one event leads to another event), classification (identification of new patterns with predefined targets) and clustering (grouping of identical or similar objects).Data mining involves some of the following key steps[3]-

(1) *Problem definition:* The first step is to identify goals. Based on the defined goal, the correct series of tools can be applied to the data to build the corresponding behavioural model.

(2) *Data exploration:* If the quality of data is not suitable for an accurate model then recommendations on future data collection and storage strategies can be made at this. For analysis, all data needs to be consolidated so that it can be treated consistently.

(3) *Data preparation:* The purpose of this step is to clean and transform the data so that missing and invalid values are treated and all known valid values are made consistent for more robust analysis.
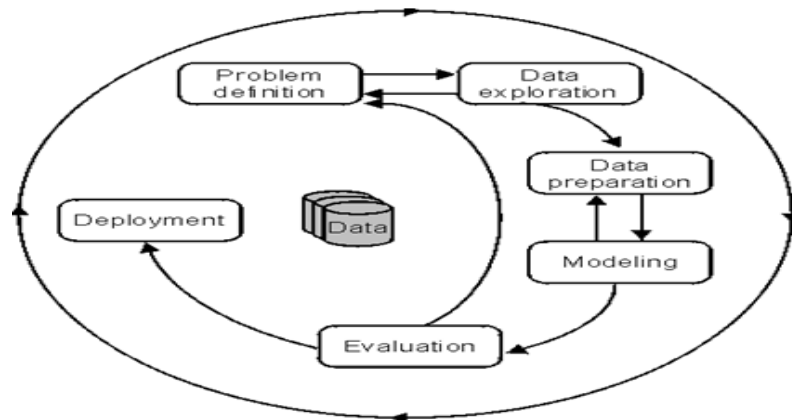

Fig.2. Data Mining Process Representation

(4) *Modelling:* Based on the data and the desired outcomes, a data mining algorithm or combination of algorithms is selected for analysis. These algorithms include classical techniques such as statistics, neighbourhoods and clustering but also next generation techniques such as decision trees, networks and rule based algorithms. The specific algorithm is selected based on the particular objective to be achieved and the quality of the data to be analysed.

(5) *Evaluation and Deployment:* Based on the results of the data mining algorithms, an analysis is conducted to determine key conclusions from the analysis and create a series of recommendations for consideration.

4. Data Mining Classification Methods

The data mining consists of various methods. Different methods serve different purposes, each method offering its own advantages and disadvantages. However, most data mining methods commonly used for this review are of classification category as the applied prediction techniques assign patients to either a "*benign*" group that is non- cancerous or a "*malignant*" group that is cancerous and generate rules for the same. Hence, the breast cancer diagnostic problems are basically in the scope of the widely discussed classification problems.

In data mining, classification is one of the most important task. It maps the data in to predefined targets. It is a supervised learning as targets are predefined. The aim of the classification is to build a classifier based on some cases with some attributes to describe the objects or one attribute to describe the group of the objects. Then, the classifier is used to predict the group attributes of new cases from the domain based on the values of other attributes. The commonly used methods for data mining classification tasks can be classified into the following groups[4].

4.1. *Decision Trees (DT's)*

A decision tree is a tree where each non-terminal node represents a test or decision on the considered data item. Choice of a certain branch depends upon the outcome of the test. To classify a particular data item, we start at the root node and follow the assertions down until we reach a terminal node (or leaf). A decision is made when a terminal node is approached. Decision trees can also be interpreted as a special form of a rule set, characterized by their hierarchical organization of rules.

4.2. *Support Vector Machine(SVM)*

Support vector machine (SVM) is an algorithm that attempts to find a linear separator (hyper-plane) between the data points of two classes in multidimensional space. SVMs are well suited to dealing with interactions among features and redundant features.

4.3. *Genetic Algorithms (GAs) / Evolutionary Programming (EP)*

Genetic algorithms and evolutionary programming are algorithmic optimization strategies that are inspired by the principles observed in natural evolution. Of a collection of potential problem solutions that compete with each other, the best solutions are selected and combined with each other. In doing so, one expects that the overall goodness of the solution set will become better and better, similar to the process of evolution of a population of organisms. Genetic algorithms and evolutionary programming are used in data mining to formulate hypotheses about dependencies between variables, in the form of association rules or some other internal formalism.

4.4. *Fuzzy Sets*

Fuzzy sets form a key methodology for representing and processing uncertainty. Uncertainty arises in many forms in today's databases: imprecision, non-specificity, inconsistency, vagueness, etc. Fuzzy sets exploit uncertainty in an attempt to make system complexity manageable. As such, fuzzy sets constitute a powerful approach to deal not only with incomplete, noisy or imprecise data, but may also be helpful in developing uncertain models of the data that provide smarter and smoother performance than traditional systems.

4.5. *Neural Networks*

Neural networks (NN) are those systems modeled based on the human brain working. As the human brain consists of millions of neurons that are interconnected by synapses, a neural network is a set of connected input/output units in which each connection has a weight associated with it. The network learns in the learning phase by adjusting the weights so as to be able to predict the correct class label of the input.

4.6. *Rough Sets*

A rough set is determined by a lower and upper bound of a set. Every member of the lower bound is a certain member of the set. Every non-member of the upper bound is a certain non-member of the set. The upper bound of a rough set is the union between the lower bound and the so-called boundary region. A member of the

boundary region is possibly (but not certainly) a member of the set. Therefore, rough sets may be viewed as with a three-valued membership function (yes, no, perhaps). Rough sets are a mathematical concept dealing with uncertainty in data. They are usually combined with other methods such as rule induction, classification, or clustering methods.

5. Data Mining Classification Techniques for Breast Cancer Daignosis

Clinical diagnosis of breast cancer helps in predicting the malignant cases. A lump felt during the examination roughly give clues as to the size of tumour and its texture. The various common methods used for breast cancer diagnosis are Mammography, Biopsy, Positron Emission Tomography and Magnetic Resonance Imaging. The results obtained from these methods are used to recognise the patterns which are aiming to help the doctors for classifying the malignant and benign cases. There are various data mining techniques, statistical methods and machine learning algorithms that are applied for this purpose. This section consists of the review of various technical and review articles on data mining techniques applied in breast cancer diagnosis.

In [5] A. Soltani Sarvestani, A. A. Safavi, N.M. Parandeh and M.Salehi provided a comparison among the capabilities of various neural networks such as Multilayer Perceptron (MLP), Self Organizing Map(SOM), Radial Basis Function (RBF) and Probabilistic Neural Network(PNN) which are used to classify WBC and NHBCD data. The performance of these neural network structures was investigated for breast cancer diagnosis problem. RBF and PNN were proved as the best classifiers in the training set. But the PNN gave the best classification accuracy when the test set is considered. This work showed that statistical neural networks can be effectively used for breast cancer diagnosis as by applying several neural network structures a diagnostic system was constructed that performed quite well.

In [6] Orlando Anunciacao, Bruno C. Gomes, Susana Vinga, Jorge Gaspar, Arlindo L.Oliveira and Jose Rueff explored the applicability of decision trees for detection of high-risk breast cancer groups over the dataset produced by Department of Genetics of faculty of Medical Sciences of Universidade Nova de Lisboa with 164 controls and 94 cases in WEKA machine learning tool. To statistically validate the association found, permutation tests were used. They found a high-risk breast cancer group composed of 13 cases and only 1 control, with a Fisher Exact Test(for validation) value of $9.7*10^{-6}$ and a p-value of 0.017. These results showed that it is possible to find statistically significant associations with breast cancer by deriving a decision tree and selecting the best leaf.

In [7] Dr. Medhat Mohamed Ahmed Abdelaal and Muhamed Farouq investgated the capability of the classification SVM with Tree Boost and Tree Forest in analyzing the DDSM dataset for the extraction of the mammographic mass features along with age that discriminates true and false cases. Here, SVM techniques show promising results for increasing diagnostic accuracy of classifying the cases witnessed by the largest area under the ROC curve comparable to values for tree boost and tree forest.

In [8] Wei-pin Chang, Der-Ming and Liou explored that the genetic algorithm model yielded better results than other data mining models for the analysis of the data of breast cancer patients in terms of the overall accuracy of the patient classification, the expression and complexity of the classification rule. The artificial neural network, decision tree, logistic regression, and genetic algorithm were used for the comparative studies and the accuracy and positive predictive value of each algorithm were used as the evaluation indicators. WBC database was incorporated for the data analysis followed by the 10-fold cross-validation. The results showed that the genetic algorithm described in the study was able to produce accurate results in the classification of breast cancer data and the classification rule identified was more acceptable and comprehensible.

In [9] K. Rajiv Gandhi, Marcus Karnan and S. Kannan in their paper constructed classification rules using the Particle Swarm Optimization Algorithm for breast cancer datasets. In this study to cope with heavy computational efforts, the problem of feature subset selection as a pre-processing step was used which learns fuzzy rules bases using GA implementing the Pittsburgh approach. It was used to produce a smaller fuzzy rule bases system with higher accuracy . The resulted datasets after feature selection were used for classification using particle swarm optimization algorithm. The rules developed were with rate of accuracy defining the underlying attributes effectively.

In [10] J. Padmavati performed a comparative study on WBC dataset for breast cancer prediction using RBF and MLP along with logistic regression. Logistic regression was performed using logistic regression in SPSS package and MLP and RBF were constructed using MATLAB. It was observed that neural networks took slightly higher time than logistic regression but the sensitivity and specificity of both neural network models had

a better predictive power over logistic regression. When comparing RBF and MLP neural network models , it was found that RBF had good predictive capabilities and also time taken by RBF was less than MLP .

In [11] Chul-Heui Lee, Soen-Hak Soc and Sang-Chul Choi in their study proposed a new classification method based on the hierarchical granulation structure using the rough set theory. The hierarchical granulation structure was adopted to find the classification rules effectively. The classification rules had minimal attributes and the knowledge reduction was accomplished by using the upper and lower approximations of rough sets. A simulation was performed on WBC dataset to show the effectiveness of the proposed method. The simulation result showed that the proposed classification method generated minimal classification rules and made the analysis of information system easy.

In [12] Aboul Ella Hassanien, and Jafar M.H.Ali in their paper presented a rough set method for generating classification rules from a set of observed 360 samples of theWBC data. The attributes were selected, normalized and then the rough set dependency rules were generated directly from the real value attribute vector. Then the rough set reduction technique was applied to find all reducts of the data which contains the minimal subset of attributes that are associated with a class label for classification. They showed that the total number of generated rules was reduced from 472 to 30 rules after applying the proposed simplification algorithm. They also made a comparison between the obtained results of rough sets with the well known ID3 decision tree and concluded rough sets showed higher accuracy and generated more compact rules.

In[13] Sudhir D. Sawarkar et al applied SVM and ANN on the WBC data .The results of SVM and ANN prediction models were found comparatively more accurate than the human being. The 97% high accuracy of these prediction models can be used to take decision to avoid biopsy.

In [14] Sepehr M. H. Jamarani et al presented an approach for early breast cancer diagnosis by applying combination of ANN and multiwavelet based sub band image decomposition. The proposed approach was tested using the MIAS mammographic databases and images collected from local hospitals. The best performance was achieved by BiGHM2 multiwavelet with areas ranging around 0.96 under ROC curve. The proposed approach could assist the radiologists in mammogram analysis and diagnostic decision making.

6. Data Mining Classification Techniques for Breast Cancer Prognosis

Once a patient is diagnosed with breast cancer, the malignant lump must be excised. During this procedure physicians must determine the prognosis of the disease. This is the prediction of the expected flow of the disease. Prognosis is important because the type and intensity of the medications are based on it. Prognosis problem is also called as "analysis of *survival* or *lifetime* data". It poses a more difficult problem than that of diagnosis since the data is *censored*. That is, there are only a few cases where we have an observed recurrence of the disease. In this case, we can classify the patient as *recur* and we know the *time to recur* (TTR). On the other hand, we do not observe recurrence in most patients. For these, there is no real point at which we can consider the patient a non recurrent case. So, the data is considered censored since we do not know the time of recurrence. For such patients, all known is only the time of their last check-up. We call this the disease-free survival time (DFS). Prognosis helps in establishing a treatment plan by predicting the outcome of a disease.[15] There are three predictive foci of cancer prognosis: 1) prediction of cancer susceptibility (risk assessment), 2) prediction of cancer recurrence and 3) prediction of cancer survivability. The most widely accepted prognostic factor for breast cancer is the American Joint Commission on Cancer (AJCC) staging system based on the TNM system (T, tumor; N, node; M, metastasis)[16] and survival is considered as any incidence of breast cancer where the person is still living from the date of diagnosis. The objective of prognostic predictions is to handle cases for which cancer has not recurred (censored data) as well as case for which cancer has recurred at a specific time. Thus, breast cancer prognostic problems are mainly in the scope of the widely discussed classification problems. This section consists of the review of various technical and review articles on data mining techniques applied in breast cancer prognosis.

C4.5 is a well known decision tree induction learning technique which has been used by Abdelghani Bellaachia and Erhan Gauven [17] along with two other techniques i.e. Naïve Bayes and Back-Propagated Neural Network. They presented an analysis of the prediction of survivability rate of breast cancer patients using above data mining techniques and used the new version of the SEER Breast Cancer Data. The preprocessed data set consists of 151,886 records, which have all the available 16 fields from the SEER database. They have adopted a different approach in the pre-classification process by including three fields: STR(Sarvival Time Recode), VSR(Vital Status Recode), and COD(Cause Of Death) and used the Weka toolkit to experiment with these three data mining algorithms. Several experiments were conducted using these algorithms. The achieved

prediction performances are comparable to existing techniques. However, they found out that model generated by C4.5 algorithm for the given data has a much better performance than the other two techniques.

In [18] M. Lundin et al has applied ANN on 951 instances dataset of Turku University Central Hospital and City Hospital of Turku to evaluate the accuracy of neural networks in predicting 5, 10 and 15 years breast cancer specific survival. The values of ROC curve for 5 years was evaluated as 0.909,for 10 years 0.086 and for 15 years 0.883, these values were used as a measure of accuracy of the prediction model. They compared 82/300 false prediction of logistic regression with 49/300 of ANN for survival estimation and found ANN predicted survival with higher accuracy.

In [19] W. Nick Street applied ANN classification to Wisconsin Prognostic Breast Cancer and SEER datasets for the analysis of survival. He developed a novel encoding as good and poor prognosis of censored data in an ANN architecture to provide a framework for prognostic prediction.

In [20] Chih-Lin Chi et al used the Street's ANN model for Breast Cancer Prognosis on WPBC data and Love data. In their research they used recurrence at five years as a cut point to define the level of risk. The applied models successfully predicted recurrence probability and separated patients with good(>5 yrs) and bad(<5 yrs) prognoses.

In [21] Delen et al compared ANN, decision tree and logistic regression techniques for breast cancer survival analysis. They used the SEER data's twenty variables in the prediction models. The decision tree with 93.6% accuracy and ANN with 91.2% were found more superior to logistic regression with 89.2% accuracy.

In [22] Jong Pill Choi et al compared the performance of an Artifical Neural Network, a Bayesian Network and a Hybrid Network used to predict breast cancer prognosis. The hybrid Network combined both ANN and Bayesian Network. The Nine variables of SEER data which were clinically accepted were used as inputs for the networks. The accuracy of ANN(88.8%) and Hybrid Network(87.2%) were very similar and they both outperformed the Bayesian Network. They found the proposed Hybrid model can also be useful to take decisions.

In [23] Muhammad Umer Khan et al investigated a hybrid scheme based on fuzzy decision trees on SEER data, they performed experiments using different combinations of number of decision tree rules, types of fuzzy membership functions and inference techniques. They compared the performance of each for cancer prognosis and found hybrid fuzzy decision tree classification is more robust and balanced than the independently applied crisp classification.

In [24] Harry B. Burke et al compared the TNM staging system's predictive accuracy with that of ANN for 5years survival of patients. They made the comparison over three different datasets these are SEER data, PCE data and PCE colorectal dataset. In all cases they found ANNs more accurate than the TNM staging system.

7. Conclusion

This paper provides a study of various technical and review papers on breast cancer diagnosis and prognosis problems and explores that data mining techniques offer great promise to uncover patterns hidden in the data that can help the clinicians in decision making. From the above study it is observed that the accuracy for the diagnosis analysis of various applied data mining classification techniques is highly acceptable and can help the medical professionals in decision making for early diagnosis and to avoid biopsy. The prognostic problem is mainly analysed under ANNs and its accuracy came higher in comparison to other classification techniques applied for the same. But more efficient models can also be provided for prognosis problem like by inheriting the best features of defined models. In both cases we can say that the best model can be obtained after building several different types of models, or by trying different technologies and algorithms.

Acknowledgment

References

[1] El-Sebakhy A. Emad, Faisal Abed Kanaan, Helmy T., Azzedin F. and Al-Suhaim F., "Evaluation of breast cancer tumor classification with unconstrained functional networks classifier," *Computer Systems and Applications, IEEE International Conference*, 2006, pp. 281 – 287.
[2] Osmar R. Zaïane, *Principles of Knowledge Discovery in Databases*. [Online]. Available: *webdocs.cs.ualberta.ca/~zaiane/courses/cmput690/notes/Chapter1/ch1.pdf*
[3] The Data Mining Process. [Online]. Available: http://publib.boulder.ibm.com/infocenter/db2luw/v9/index.jsp?topic=/com.ibm.im.easy.doc/c_dm_process.html.

[4] Han J. and Kamber M., *Data Mining: Concepts and Techniques*, 2nd ed., San Francisco, Morgan Kauffmann Publishers, 2001

[5] Sarvestan Soltani A. , Safavi A. A., Parandeh M. N. and Salehi M., "Predicting Breast Cancer Survivability using data mining techniques," *Software Technology and Engineering (ICSTE), 2nd International Conference*, 2010, vol.2, pp.227-231.

[6] Anunciacao Orlando, Gomes C. Bruno, Vinga Susana, Gaspar Jorge, Oliveira L. Arlindo and Rueff Jose, "A Data Mining approach for detection of high-risk Breast Cancer groups," *Advances in Soft Computing*, vol. 74, pp. 43-51, 2010.

[7] Abdelaal Ahmed Mohamed Medhat and Farouq Wael Muhamed, "Using data mining for assessing diagnosis of breast cnacer," *in Proc. International multiconfrence on computer science and information Technology*, 2010, pp. 11-17.

[8] Chang Pin Wei and Liou Ming Der, "Comparision of three Data Mining techniques with Ginetic Algorithm in analysis of Breast Cancer data".[Online]. Available:http://www.ym.edu.tw/~dmliou/Paper/compar_threedata.pdf

[9] Gandhi Rajiv K., Karnan Marcus and Kannan S., "Classification rule construction using particle swarm optimization algorithm for breast cancer datasets," *Signal Acquisition and Processing. ICSAP, International Conference*, 2010, pp. 233 – 237.

[10] Padmavati J., "A Comparative study on Breast Cancer Prediction Using RBF and MLP," *International Journal of Scientific & Engineering Research*, vol. 2, Jan. 2011.

[11] Lee Heui Chul, Seo Hak Seon and Choi Chul Sang, "Rule discovery using hierarchial classification structure with rough sets," *IFSA World Congress and 20th NAFIPS International Conference*, 2001, vol.1 , pp. 447-452.

[12] Hassanien Ella Aboul and Ali H.M. Jafar, "Rough set approach for generation of classsification rules of Breast cnacer data," *Journal Informatica*, 2004, vol. 15, pp. 23–38.

[13] Sudhir D., Ghatol Ashok A., Pande Amol P., "Neural Network aided Breast Cancer Detection and Diagnosis",*7th WSEAS International Conference on Neural Networks*, 2006.

[14] Jamarani S. M. h., Behnam H. and Rezairad G. A., "Multiwavelet Based Neural Network for Breast Cancer Diagnosis", *GVIP 05 Conference*, 2005, pp. 19-21.

[15] Pantel Patrick **,** *Breast Cancer Diagnosis and Prognosis.*[Online]. Available: http://citeseer.nj.nec.com/pantel98breast.html.

[16] Choi J.P., Han T.H. and Park R.W., " A Hybrid Bayesian Network Model for Predicting Breast Cancer Prognosis", *J Korean Soc Med Inform*, 2009, pp. 49-57.

[17] Bellaachia Abdelghani and Erhan Guven, "Predicting Breast Cancer Survivability using Data Mining Techniques," *Ninth Workshop on Mining Scientific and Engineering Datasets in conjunction with the Sixth SIAM International Conference on Data Mining,*" 2006.

[18] Lundin M., Lundin J., BurkeB.H.,Toikkanen S., Pylkkänen L. and Joensuu H. , "Artificial Neural Networks Applied to Survival Prediction in Breast Cancer", *Oncology International Journal for Cancer Resaerch and Treatment,* vol. 57, 1999.

[19] Street W.N., "A Neural Network Model for Prognostic Prediction", *Fifteenth International Conference on Machine Learning, Madison, Wisconsin, Morgan Kaufmann*, 1998.

[20] Chi C.L., Street W.H. and Wolberg W.H., "Application of Artifical Neural Network- based Survival Analysis on Two Breast Cancer Datasets", *Annual Symposium Proceedings / AMIA Symposium*, 2007.

[21] Delen Dursun , Walker Glenn and Kadam Amit , "Predicting breast cancer survivability: a comparison of three data mining methods," *Artificial Intelligence in Medicine* ,vol. 34, pp. 113-127 , June 2005.

[22] Choi J.P., Han T.H. and Park R.W., " A Hybrid Bayesian Network Model for Predicting Breast Cancer Prognosis", *J Korean Soc Med Inform*, 2009, pp. 49-57.

[23] Khan M.U., Choi J.P., Shin H. and Kim M, "Predicting Breast Cancer Survivability Using Fuzzy Decision Trees for Personalized Healthcare", *Conf Proc IEEE Eng Med Biol Soc.*, 2008, pp. 48-51.

[24] Burke H. B. Et al , "Artificial Neural Networks Improve the Accuracy of Cancer Survival Prediction", *Cancer,* 1997, vol.79, pp.857-862.