

Predictive Data Analytics approaches in Clinical Informatics: A review

Shweta Kharya¹ SunitaSoni² Tripti Swarnkar³

^{1,2}Bhilai Institute of Technology, ³Siksha 'O' Anusandhan (Deemed to be University)

shweta.bitdurg@gmail.com,sunitasoni74@gmail.com,triptiswarnakar@soauniversity.ac.in

Abstract

“Quality care is the right care,at the right time,every time” is a well known adage in the health care industry. To facilitate quality care there is a dire need to create data driven approaches to augment existing decision making processes and also out sized quantity of data which are generated in health care domain is growing with a fast pace. Hence, it becomes important to find the relevant information from the explosive amount of data for effective diagnosis and prognosis. Research community give more attention towards health care domain data and tries to establish the relationship between health care data and data analytics approaches and mine the significant information to built clinical decision support system which can treated as supportive tools for experts. Predictive data analytics in medicine involves the development of computational models which are capable of analyzing large amounts of data in order to predict health care outcomes for individual patient. Experts use predictive analysis in health care mainly to determine which patients are at high risk of developing certain circumstances, like diabetes, cancer, heart disease, and other lifetime illnesses. Due to the above advantages, medical research has a strong inclination towards predictive data mining which motivated to work in the field of predictive data mining. The objective of this paper is to summaries various reviews and technical articles on diagnosis and prognosis of medical illness. It gives an overview of the current research being carried out on various medical fields using the predictive data mining techniques to enhance the disease diagnosis and prognosis.

Keywords: Data Analytics, Predictive Modeling, Clinical Informatics, Decisin Trees, Support vector machine, Neural Networks, Bayesian Belief Networks.

1. Introduction

Today data analytics methods are being used in an extensive range of medical applications including detecting and classifying disease outcomes. Data Analytics draws from statistics and probability, but it is deeply more commanding because it allows decisions to be made that is not possible by using conventional statistical methodologies. Given the budding importance [1] of predictive medicine and the growing reliance on data analytics to make predictions, it would be of great interest to conduct a detailed review of published studies employing predictive data analytics methods in disease prediction and prognosis. If analytics approaches experiment is properly deliberate, correctly implemented and the

results robustly validated, then one usually has a superior chance of success. Interestingly, almost all analytical algorithms used in disease prediction and prognosis utilize supervised learning. Predicting the result of a disease is one of the most interesting and exigent 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 between large numbers of variables and made them able to predict the effect of a disease using the historical cases stored within datasets.

1.1 Clinical informatics

Clinical informatics is the study of information technology and how it can be applied to the health care field. It includes the study and practice of an information-based approach to health care delivery in which data must be structured in a certain way to be effectively retrieved and used in a report or evaluation. Clinical informatics can be applied in a range of health care settings including hospital, physician's practice.

1.2 Predictive data analytics

Predictive data mining is becoming an essential instrument for researchers and clinical practitioners in medicine. Understanding the main issues underlying these methods and the application of agreed and uniform procedures is mandatory for their deployment and the broadcasting of results. Predictive models can be used to conjecture explicit values, based on patterns determined from known results. This technique is becoming an essential appliance for researchers and clinical practitioners in medical. These methods may be applied to the edifice of decision models for procedures such as prognosis, diagnosis and treatment planning, which once evaluated and verified may be embedded within clinical information systems. Classification and prediction are major predictive data mining ta

2. Literature survey on predictive data analytics approaches in clinical informatics

Data Analytics is an assortment of effective tool set to perform the *statistical analysis* on an immense dataset and to retrieve the valuable information from the dataset. In this work we have carried out an analytical survey on *predictive data mining approaches on clinical* dataset.

2.1 Decision Trees

Decision Trees are a straightforward approach to guide one's path to a decision. It is a classification scheme which generates a tree and a set of rules representing the model of different classes from a given set of data. The goal is to create a model that predicts the value of a class variable by learning simple

decision rules extracted from the data features. The tree formation is used in classifying unidentified data records. The main purpose of decision tree is to expose structural information that is contained in the data. Decision trees are commonly used to answer relatively simple binary decisions to help identify a strategy most likely to reach a goal. A number of algorithms have been anticipated for decision tree induction. Popular Decision tree algorithms are ID3, C4.5 ,CART,CHAID.

Some advantages of decision trees are:

- Simple to understand and to interpret by domain expert as Trees can be easily visualized.
- Requires little data ground work.
- Numerical and Categorical data can be tackled easily.
- Possible to validate a model using statistical tests which makes it possible to account for the reliability of the model.
- Human readability can be improved as trained trees can be represented as a set of if-then rules.

Disadvantages of decision trees are:

- Decision-tree does not generalize the data in good manner and thereby creating over-complex tree which is known as over fitting.
- Tree building is very hectic task and computationally exhaustive as the training data set is traversed over and over again.
- Un stability in Decision trees due to minute variations in the data yields fully new trees.

In [2] authors analyzed the prediction of survivability rate of breast cancer patients using data mining

techniques. The data used is the SEER Public-Use Data. The preprocessed data set consists of 151,886 records, which have all the available 16 fields from the SEER database. Authors investigated three data mining techniques: the Naïve Bayes, the back-propagation neural network, and the C4.5 decision tree algorithms. They found out that C4.5 algorithm has a much better performance than the other two techniques.

In [3], authors 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 [4] Decision tree has been successfully applied for feature selection to analyze the performance of classifier. The analysis of CART Classifier is performed with and without feature selection in terms of accuracy, time to build a model and size of the tree on various Breast Cancer Datasets. The results show that a particular feature selection using CART has enhanced the classification accuracy of a particular dataset.

In this paper [5] the authors accordingly developed a new knowledge-based system for classification of breast cancer disease using clustering, noise removal, and classification techniques. Expectation Maximization (EM) is used as a clustering method to cluster the data in similar groups. Then Classification and Regression Trees (CART) is used to generate the fuzzy rules to be used for the classification of breast cancer disease in the knowledge-based system of fuzzy rule based reasoning method. To overcome the multi- collinearity issue, authors incorporate Principal Component Analysis (PCA) in the proposed knowledge-based system. Experimental results on Wisconsin Diagnostic Breast Cancer and Mammographic mass datasets show that proposed methods remarkably improves the prediction accuracy of breast cancer. The proposed knowledge-based system can be used as a clinical decision support system to assist medical practitioners in the health care practice. By the combination of EM, PCA, CART and Fuzzy Rule-Based methods, a hybrid intelligent system is developed to increase the predictive accuracy of breast cancer disease classification.

2.2 Artificial Neural Networks

Artificial Neural Networks are learning algorithms that are inspired by biological neural networks and they consist of a set of interconnected input/output units in which each connection has an associated weight. During the learning phase, neural network learns by adjusting their weights in order to predict the class label of the input records. Neural Network must be trained for long periods of time before they can be applied to solving problems such as classification or prediction. Artificial Neural Network performs well with classifying data which have unseen by model previously.

Advantages:

- Data generalization is used to handle noisy and new situations automatically.
- Prediction of future indicator values based on previous data and trend recognition.
- computerized real-time analysis and diagnosis.
- Allows rapid identification and classification of input data.
- Reduces error associated with human fatigue and habituation.

Disadvantages:

- Due to over fitting ,it give rise to new spurious patterns.
- They belong to black-box methods, i.e., the relation between a neural network and the problem it represents is not easy to understand.

In this paper [6] authors proposed that network ensembles have successfully utilized to assist in medical diagnosis. Neural network ensemble scan significantly improve the generalization ability of learning systems through training a finite number of neural networks and then combining their results. The significant study is to investigate the performance of various classifiers in an ensemble classifier and solo classifiers. Use of various evaluation criteria like confusion matrix, accuracy, sensitivity, specificity are made to examine the performance of these classifiers with real-life UCI datasets of Wisconsin Diagnostic Breast cancer, Indian liver patient dataset, Vertebral column data set, Heart disease data sets for the purpose of disease prediction. Finally the use of statistical testing is done to evaluate the significance of the difference in performance among the classifiers. The statistical testing results indicate that an ensemble classifier performs better than an individual classifier involved within an ensemble.

Authors in this [7] research paper proposed that neural networks are currently a burning research area in medical science, especially in the areas of cardiology, radiology, oncology, urology and etc. An Artificial Neural Network usually known as a neural network is basically a mathematical model motivated by biological nervous systems like brain processes information. A neural network comprises an interconnected group of simulated neurons and it uses connectionist approach to process information for computation. In this paper survey of various neural network technologies for classification of cancer is done. The main aim of this study in medical diagnostics is to guide researchers to develop most cost effective and user friendly systems, processes and approaches for clinicians. Although neural network techniques provide good classification rate, but their training time is very high. Several researchers thus hybridize neural network techniques with optimization algorithms like PSO for further enhancement of accuracy. These optimization algorithms are used for dimensionality reduction, they suppresses search space and therefore, reduces the training time of neural network.

In this paper [8] authors predict trends in the patient conditions and their behaviors, which is accomplished by data analysis from different perspectives. The methodology used is Genetic algorithm, Artificial Neural Network and Nearest Neighborhood algorithm. **Artificial neural networks** are analytical techniques that are formed on the basis of superior learning processes in the human brain.

Genetic algorithms are based on the principle of genetic modification, mutation and natural selection. Genetic algorithms are used in data mining to formulate hypotheses about the dependencies between variables in the form of association rules or other internal formalism.

Nearest neighborhood method is a technique that is also used for data classification. The application of nearest neighbor method on benchmark data set to detect efficiency in the diagnosis of heart diseases, produced the results that application of this method achieved an accuracy of 97.4% which is a higher percentage than any other published study on the same set of data. Authors concluded that Data mining technology provides customer oriented approach towards new and hidden patterns in data, from which the knowledge is being generated, the knowledge that can help in providing of medical and other services to the patients. Healthcare institutions that use data mining applications have the possibility to predict future requests, needs, desires, and conditions of the patients and to make adequate and optimal decisions about their treatments.

In [9] authors devised an Artificial Neural network for the prognosis of cancer, the aim of this work is to help doctors to decide whether a biopsy is necessary. Total of 121 patients data which are definitively diagnosed with cancer after biopsy are used. Their result revealed an Area Under Curve of 94.44% which exhaustively suggest that an ANN system can help doctors to make quick and reliable diagnoses.

2.3 Support Vector Machine

Support vector machines (SVM) are today's most powerful classification algorithm in terms of predictive accuracy [10]. The SVM algorithm as a relatively new classification or prediction method, has been developed in the [11] as a result of the collaboration between the statistical and the machine learning research community. The methods are based on strong mathematical foundations and statistical learning theory [12]. Basically the method is to find a hyper plane that separates the examples of different outcomes. Being primarily designed for two-class problems, SVMs find a hyper plane with a maximum distance to the closest point of the two classes; such a hyper plane is called the optimal hyper plane. A set of instances that is closest to the optimal hyper plane is called a support vector. Support vector machines are becoming increasingly popular in Health mining sector. It uses a formalism that is often unsuitable for interpretation by human experts. If the concerned is only accuracy prediction, support vector machines are a strong contender to artificial neural networks, especially since their performance may be more robust and less depends on the specific selection of method's parameters.

Advantages:

- Support vector machines are powerful pattern recognizers.
- The development of SVMs involved sound theory first and then implementation and experiments assumed to belong to the same family of methods.
- The solution to an SVM is global and unique .
- Two more advantages of SVMs are that they have a simple geometric interpretation and give a sparse solution.
- SVMs use structural risk minimization.

Disadvantage:

- They yield black-box models.

In [13] authors have used neural networks Support Vector Machine method for diagnosis of breast cancer. SVMs can only be used for classification, not for function approximation. The Support Vector Machine (SVM) is implemented using the kernel Adatron algorithm. Support Vector Machine method was used on the set of 683 samples of actual data. Additional set of data of 117 samples is generated using Neural Network. Using Support Vector Machine ANN, the prediction of diagnosis and detection of breast cancer is comparably accurate than the human being. The efficiency of manual detection of breast cancer is 85% and the efficiency of the Support Vector machine recognition obtained is nearly 97%. This high rate of accuracy can be utilized to support the Doctor's decision to avoid Biopsy.

In [14] authors proposed a Support Vector Machines (SVMs) based classifier in comparison with Bayesian classifiers and Artificial Neural Networks for the prognosis and diagnosis of breast cancer disease. The paper provides the implementation details along with the corresponding results for all the assessed classifiers. A SVM model is implemented for the breast cancer diagnosis and prognosis problem using the Wisconsin Diagnostic Breast Cancer (WDBC) as well as the Wisconsin Prognostic Breast Cancer (WPBC) datasets. The optimized SVM algorithm performed excellently, exhibiting high values of accuracy (up to 96.91%), specificity (up 97,67%) and sensitivity (up to 97,84%).

Authors in [15] this paper explored that detecting the cancer and differentiating between the diagnosis that affirm whether a patient has breast cancer or not has been considered as a big challenge. In order to have an accurate diagnosis, Support Vector Machine (SVM) and Artificial Neural Network (ANN) have been selected in many research papers to solve this problem with high classification accuracy. In this paper the breast cancer diagnosis is addressed using SVM and ANN combined with feature selection. The feature selection is based on the correlation coefficient of each feature against the target class where different feature subsets are used. The model is tested on the popular Wisconsin Diagnosis Breast Cancer

(WDBC) dataset to conduct the experiments. 10- Fold Cross validation has been used for data partitioning while developing the model and the outcome indicates better classification accuracy. As for comparison between SVM and ANN, empirical studies outcome indicated that SVM outperformed ANN with classification accuracy of 97.14 and 96.71 respectively.

2.4 Bayesian Belief Networks

A Bayesian Network is a directed acyclic graph consisting of (i) Nodes (circles), which represent random variables; arcs (arrows), which represent probabilistic relationships among these variables and (ii) for each node and a local probability distribution attached to it, which depends on the state of its parents [16]. A Bayesian Network is a graphical model that represents relationships of probabilistic nature among the input variables of interest. Such networks consist of a qualitative part (structural model), which provides a visual representation of the interactions amid variables, and a quantitative part (set of local probability distributions), which allows probabilistic inference and numerically measures the impact of a variable or sets of variables on others. A unique joint probability distribution over the variables in a specific problem is determined by both the qualitative and quantitative parts. In other words, many of the early Bayesian networks developed for real-life applications in biomedicine and health-care have been constructed by hand, i.e., they are based on medical background knowledge. Manual constructions of a Bayesian network requires access to the knowledge of human experts and, are quite time consuming. With the increasing availability of clinical and biologic data, machine learning is clearly the more feasible alternative for developing a Bayesian network. Integrating background knowledge and evidence derived from past data is also supported by Bayesian networks. Missing data can be handled both in the construction process and in using a Bayesian network model

Advantages:

- The Bayesian network is a powerful tool to describe the uncertainty and complexity of many problems in the real world having rigorously justified mathematical basis.
- They deal in a natural way with uncertainty modeled as a joint probability distribution.
- They are easy to understand because of their graphical representation.
- BNs can be used both as predictive and descriptive models.
- In prediction they constitute an efficient tool for solving different inference tasks (posterior probability, additive or diagnostic reasoning, relevance analysis, classification). As a descriptive tool they possess the ability to efficiently represent the dependence/independence relationships among the random variables.
- That compose the problem domain we desire to model.
- It is possible to represent a large instance in a Bayesian network using little space, and it is possible to perform probabilistic inference among the features in an acceptable amount of time.

- The graphical nature of Bayesian network gives us a much better intuitive grasp of the relationships among the features.

Suitability in Medical Prediction:

Medical Prediction is special Domain where symptoms have certain degree of dependency among them. For example if the Obesity will increase it will also increase Blood Pressure, blood sugar level etc. In BN's the probability values for links between nodes in Bayesian networks reflect degrees of dependence between variables. This makes it possible for the structure of these networks to be examined by human experts to uncover relationships between the variables, hence, enabling the assessment of the reasonableness of the decision process. The other prediction Model discussed so far does not consider the dependency among the input parameter rather the model is built considering the independency of input parameter and this is an important factor in their gaining acceptance in the field of medical diagnosis.

Characteristics of BBN:

- BBN provides an approach for capturing the prior knowledge of a particular domain using a graphical model.
- The network also be used to encode casual dependencies among the variables. BBN are well suited to dealing with incomplete data.
- Because the data is combined probabilistically with prior knowledge, the method is quite robust to model over fitting [17].

The objective of this [18] paper is to explore the implementation of a Bayesian Belief Network for an automated breast cancer detection support tool. It is intuitive that Bayesian networks are employed as one viable option for computer-aided detection by representing the relationships between diagnoses, physical findings, laboratory test results and imaging study findings. It was explained that, by exploiting conditional independencies entailed by influence chains, it is possible to represent a large instance in a Bayesian network using little space, and it is often possible to perform probabilistic inference among the features in an acceptable amount of time. For the application of computer-aided detection in mammography, the researchers intend to design an interface between the project's Bayesian network learning algorithm and the radiologists, so that the radiologists can have interaction with the system by labeling only a small number of informative images presented by the active learning algorithm.

Authors in this paper [19] convey an impression of the current state of art of the use of Bayesian networks in biomedicine and health care. A detailed formalism of Bayesian Network is introduced and methods for their construction like manual construction and by learning are discussed. Manual construction involves series of steps like.

- 1) Selection of relevant variables.
- 2) Identification of relationships among the variables.
- 3) Identification of qualitative probabilistic and logical constraints.
- 4) Assessment of probabilities.
- 5) Sensitivity analysis and evaluation.

And learning a Bayesian network from data involves the task of structure learning i.e identifying the graphical structure of the network and parameter learning i.e estimating the conditional probability distribution to be associated with the network's diagram. Algorithms for learning a Bayesian network from data listed from survey are K2, Minimum description length by using genetic algorithm, Dependence analysis approach to learning Bayesian network from data, and also great deal of interest in estimating probability distributions using maximum likelihood estimation. Expectation maximization algorithm can be used by many researchers for this purpose. And elaborates the biomedical problem involving uncertainty for which Bayesian networks can be employed like Diagnostic reasoning, Prognostic reasoning, and Treatment selection and Discovering functional interactions.

In the paper [20] authors presented Bayesian networks as a highly practical framework for working with a kind of classification problem. Authors intended to demonstrate how the BayesiaLab software can extremely quickly, and relatively simply, create Bayesian network models that achieve the performance of the best custom-developed models, while only requiring a fraction of the development time. Author's wishes to demonstrate how Bayesian networks can help researchers and practitioners generate a deeper understanding of the underlying problem domain. BayesiaLab's speed of model building, its excellent classification performance, plus the ease of interpretation provide researchers with a powerful new tool. Bayesian networks and BayesiaLab has thus become a driver in accelerating research.

DATASETS used: Wisconsin Breast Cancer Database

In this paper authors had used Wisconsin breast cancer Database consisting of 699 instances of patient consisting of two classes: 458 benign cases (65.5%) and 241 malignant cases (34.5%). The Wisconsin Breast Cancer Database is available to any interested researcher from the UC Irvine Machine Learning Repository. The following eleven attributes² are included in the database:

1. Sample code number
2. Clump Thickness (1 - 10)
3. Uniformity of Cell Size (1 - 10)
4. Uniformity of Cell Shape (1 - 10)
5. Marginal Adhesion (1 - 10)

6. Single Epithelial Cell Size (1 - 10)
7. Bare Nuclei (1 - 10)
8. Bland Chromatin (1 - 10)
9. Normal Nucleoli (1 - 10)
10. Mitoses (1 - 10)
11. Class (benign/malignant)

In this paper a tool known as BayesiaLab is used to generate a Bayesian network. By using Bayesian belief networks as the framework and BayesiaLab as the tool, authors have shown a practical new modeling and analysis approach based on the widely studied Wisconsin Breast Cancer Database. BayesiaLab can rapidly machine-learn reliable models, even without prior domain knowledge and without hypothesis. The classification performance of the BayesiaLab generated Bayesian network models are on equivalence with all research based on this topic. Beyond the predictive performance, BayesiaLab enables a range of analysis and interpretation functions, which can help the researcher gain deeper domain knowledge and perform inference more efficiently.

Authors in this paper [21] have developed and evaluate a new Bayesian network based for patient survivorship prediction method. The aim of the method is to predict patient survivorship, while having the capability to handle high-dimensional data and being able to incorporate it into a clinical decision support system. They developed EBMC_Survivorship which is based on newly developed EBMC which is a Bayesian Network algorithm as it was designed to handle high dimensional data and further it can be incorporated into complete CDSS. Author evaluated the system using a breast cancer dataset known as METABRIC (Molecular Taxonomy of Breast Cancer International Consortium, and compared the prediction performance of EBMC_Survivorship to the standard the Cox proportional hazard model and RSF method using same dataset and by using 5-fold cross validation models are evaluated over 5,10,15 years into the future and concluded that EBMC_S outperformed others.

In this paper [22] authors focused on predicting the graft survival for kidney transplantation and proved that the strength of any predictive model depends on the selection of proper predictors and potential solution for prediction is done by machine learning approaches. In machine learning methods probabilistic graphical models plays a major role in modeling the complex non linear relationships among different variables and reasoning under uncertainty. Bayesian belief networks have become popular among PGMs because they can handle previously unknown but potentially useful information embedded in networks.

This study offers novel methodological solution to Prediction problem by

- a) Analyzing database including >31000 US patients.

- b) Introducing comprehensive feature selection framework like medical literature, data analytics methods and elastic net regression.
- c) Using sensitivity analyses and information fusion to evaluate and combine features from machine learning approaches like SVM, ANN, BF.
- d) Using best performing sets in Bayesian belief models and statistical models in addition to expert knowledge.

In summary BBN models could help the health care providers in examining the interactions among the different variables to develop clinically relevant protocols that could improve clinical outcomes in transplant populations.

3. CONCLUSIONS

It is vital to move research beyond the experimental stage onto clinical validation stage, a step towards achieving patient benefit via improved predictions of outcomes. Hence Predictive analytics are promising approaches to build efficient & accurate Clinical Decision support system for improved outcomes. This paper provides a comprehensive review of the state of the art of predictive data analytics field. From the exhaustive review of work carried out during last 15 years, ANN is found to be most widely used predictive technique in medical prediction as compare to traditional methods like Decision Tree, Regression Tree etc. Owing to the fact that ANN technique provides robust solution to real time prediction problem till date they have invaded almost all the realm of medical prediction process. SVM has been mainly used in the computational biology such as microarray data analysis, protein secondary structure prediction, prediction of human signal peptide cleavage sites, translational initiation site recognition in DNA, protein fold recognition, prediction of protein-protein interactions, prediction of protein sub cellular localization, and peptide identification from mass spectrometry derived data. Like ANN, SVM is also Black Box method have not yet seen widespread adoption in the communities working with very large datasets due to the high computational in the training phase. Bayesian network is recently developed technique suitable to perform prediction under uncertainty with incomplete data. It offers a very attractive formalism for representing uncertain knowledge, have so far only been used in medicine essentially to assist in the diagnosis of disorders and to predict the natural course of disease after treatment (prognosis). The graphical nature of Bayesian networks gives a much better intuitive grasp of the relationships among the features. Bayesian Belief Networks is found to be best technique for medical prediction. When taken together, the findings of study can offer useful tools for medical practitioners that they can be incorporate into decision making process which potentially could affect the overall survival in Medical field. This paper may act as a base map for selecting an appropriate data mining approach for diversified Clinical systems.

REFERENCES

- [1] Kharya, S., Dubey, D., &Soni, S. (2013). Predictive Machine Learning Techniques for Breast Cancer Detection. *International Journal of Computer Science and Information Technologies*, 4(6), 1023-1028.
- [2] Bellaachia, A., &Guven, E. (2006). Predicting breast cancer survivability using data mining techniques. *Age*, 58(13), 10-110.
- [3] Anunciaçao, O., Gomes, B. C.,Vinga, S., Gaspar, J., Oliveira, A. L., &Rueff, J. (2010). A data mining approach for the detection of high-risk breast cancer groups. In *Advances in bioinformatics* (pp. 43-51). Springer, Berlin, Heidelberg.
- [4] .Lavanya, D. ,Dr.K.Usha Rani (2011), ANALYSIS OF FEATURE SELECTION WITH CLASSIFICATION: BREAST CANCER DATASETS, *Indian Journal of Computer Science and Engineering (IJCSE)* ISSN : 0976-5166 Vol. 2 No. 5, pp 756-763.
- [5] Nilashi, M., Ibrahim, O., Ahmadi, H., &Shahmoradi, L. (2017). A knowledge-based system for breast cancer classification using fuzzy logic method. *Telematics and Informatics*, 34(4), 133-144.
- [6] Weng, C. H., Huang, T. C. K., & Han, R. P. (2016). Disease prediction with different types of neural network classifiers. *Telematics and Informatics*, 33(2), 277-292.
- [7] .Agrawal, S., & Agrawal, J. (2015). Neural network techniques for cancer prediction: A survey. *Procedia Computer Science*, 60, 769-774.
- [8] Milovic , B. (2012). Prediction and decision making in health care using data mining. *Kuwait chapter of arabian journal of business and management review*, 1(12), 126-136.
- [9] Saritas, I., Ozkan, I. A., &Sert, I. U. (2010). Prognosis of prostate cancer by artificial neural networks. *Expert Systems with Applications*, 37(9), 6646-6650.
- [10] Delen, D. (2009). Analysis of cancer data: a data mining approach. *Expert Systems*, 26(1), 100-112.
- [11] Vapnik, V. N. (1998). Adaptive and Learning Systems for Signal Processing Communications, and control. *Statistical learning theory*.
- [12] Sun, Y., Wong, A. K., & Wang, Y. (2006, June). An Overview of Associative Classifiers. In *DMIN* (pp. 138-143).
- [13] Sawarkar, S. D., Ghatol, A. A., &Pande, A. P. (2006, June). Neural network aided breast cancer detection and diagnosis using support vector machine. In *Proceedings of the 7th WSEAS International Conference on Neural Networks* (pp. 158-163).
- [14] Maglogiannis, I., Zafiropoulos, E., &Anagnostopoulos, I. (2009). An intelligent system for automated breast cancer diagnosis and prognosis using SVM based classifiers. *Applied intelligence*, 30(1), 24-36.
- [15] Alyami, R., Alhajjaj, J., Alnajrani, B., Elaalami, I., Alqahtani, A., Aldhaffer, N., ...&Olatunji, S. O. (2017, February). Investigating the effect of Correlation based Feature Selection on breast cancer diagnosis using Artificial Neural Network and Support Vector Machines. In *Informatics, Health & Technology (ICIHT), International Conference on* (pp. 1-7). IEEE.
- [16] Cooper, G. (1999). An overview of the representation and discovery of causal relationships using Bayesian networks. *Computation, causation, and discovery*, 4-62.
- [17] .Kharya, S.,Soni,S.(2013).Using Bayesian Belief Network of breast cancer detection.,All India Conference -2013,paper no-50.

- [18] Gadewadikar, J., Kuljaca, O., Agyepong, K., Sarigul, E., Zheng, Y., & Zhang, P. (2010). Exploring Bayesian networks for medical decision support in breast cancer detection. *African Journal of Mathematics and Computer Science Research*, 3(10), 225-231.
- [19] Lucas, P. J., Van der Gaag, L. C., & Abu-Hanna, A. (2004). Bayesian networks in biomedicine and health-care. *Artificial intelligence in medicine*, 30(3), 201-214.
- [20] Conrady, S., &Jouffe, L. (2011). Breast cancer diagnostics with Bayesian networks. *Conrady Applied Science LLC and Bayesia SAS: USA*
- [21] Jiang, X., Xue, D., Brufsky, A., Khan, S., & Neapolitan, R. (2014). A new method for predicting patient survivorship using efficient Bayesian network learning. *Cancer informatics*, 13, CIN-S13053.
- [22] Topuz, K., Zengul, F. D., Dag, A., Almehti, A., &Yildirim, M. B. (2017). Predicting graft survival among kidney transplant recipients: A Bayesian decision support model decision support systems. *Decision Support Systems*.