



## A Comparative Study of Data Analysis Techniques in the domain of Medicative care for Disease Predication

Prof. Samiksha H. Zaveri  
Ph. D. Research Scholar  
Parul University,  
Vadodara, India

Dr. Narayan Joshi  
Director-MCA, Dean-FITCS  
Parul University,  
Vadodara, India

**Abstract :** In healthcare sector every day we collect a huge data including clinical examination, vital parameters, investigation reports, treatment follow up and drug decisions etc. But it is not analysed and mined in an appropriate way. There are several fatal diseases which cause more deaths in the world. How to avoid the dangerous complications of these types of diseases? If we come to know in advance, we can protect ourself. There are different techniques to predict disease in well advance. One of the techniques is Data Mining. In past, research works have been done on it using mathematical equations. In this paper we try to study such type of diseases like TB, Cancer, heart disease, diabetes etc. We briefly examined that; the Data Mining could help in the identification or the Predication of disease. We also present a comparative study of different data mining application and techniques applied for extracting knowledge in healthcare industry.

**Keywords:** Data Mining, Diseases-TB, Cancer, Diabetes, Heart Disease

### I. INTRODUCTION

Today the buzz word is “Health Care” all over the world. Early Prediction of diseases can reduce the fatal rate of human. There are very large and enormous data available in hospitals and medical related institutions. Information technology plays a vital role in Health Care. Data mining is a process to extract useful information from large database. It involves computational process, machine learning, statistical techniques, classification, clustering and discovering patterns. Data mining techniques have proved for early prediction of disease with higher accuracy in order to save human life and reduce the treatment cost.

This Paper is organized as follows. In the next section Role of Data Mining in health care is discussed. Section 3 gives the objectives of the paper. Review of literature is given in section 4. Observations on study are presented in section 5 and last section is about Conclusion.

### II. DATA MINING IN HEALTHCARE

In present era various public and private healthcare institutes are producing enormous amounts of data which are difficult to handle. So, there is a need of powerful automated Data Mining tools for analysis and interpreting the useful information from this data. This information is very valuable for healthcare specialist to understand the cause of diseases and for providing better and cost effective treatment to patients. Data Mining offers novel information regarding healthcare which in turn is helpful for making administrative as well as medical decision such as estimation of medical staff, decision regarding health insurance policy, selection of treatments, disease prediction *etc.*, [8-11]. Several studies identified with primary focus on various challenges and issues of data mining in healthcare [12, 13]. Data mining is also used for both analysis and prediction of various diseases. Some research work proposed an enhancement in available Data Mining methodology in order to improve the result and

some studies develop new methodology and framework for healthcare system.

This paper explores various Data mining techniques such as classification, clustering and association are used by healthcare organization to increase their capability for making decision regarding patient health. There are ample of research resources available regarding Data Mining tasks which are presented in subsequent sections with their advantages and disadvantages.

### III. OBJECTIVES

The objective of this paper is to summarize the different algorithms of data mining for the major life threatening diseases in the field of medical predication. The main focus is to use different algorithm and combination of several targets attributes for different types of disease identification or predication using data mining techniques. We also present a comparative study of different data mining application and techniques applied for extracting knowledge in healthcare industry.

### IV. LITERATURE REVIEW

Manish Shukla and Sonali Agarwal examined application of classification technique and Data clustering a machine learning approach. They present an approach for centroid selection in k-mean algorithm for health datasets which gives better clustering results in comparison to traditional k-mean algorithm [1]. Kavitha K V et al., proposed Immune responses protect against infection by microbes like viruses, bacteria, fungi and other parasites and against entry of non-microbial foreign matter [2].

Aicha Boutorh & Ahmed Guessoum used, a hybrid intelligent technique based on Association Rule Mining (ARM) and Neural Networks (NN) which uses an Evolutionary Algorithm (EA) to deal with the dimensionality problem for the

diagnosis of breast cancer. ARM optimized by Grammatical Evolution (GE) is used to select the most informative features and reduce the dimensionality by extracting associations between SNPs, while NN is used for efficient classification[3]. Muhammad Sufyan Bin Mohd Azmi et al., developed a system that can classify tumor using neural network with Feed forward Back propagation Algorithm from a symptom that causes the breast cancer disease[4]. Juan F. De Paz et al., presented a Case-based reasoning (CBR) system for automatic classification of leukemia patients from microarray Data. The system incorporates novel algorithms that allow to filter and classify as well as extraction of knowledge[6]. Farhad Soleimanian Gharehchopoghi et al., proposed, to successfully managing heart diseases FinaBy, the case study is modeBed by Neural Network [7].

Rita Samikannu et al., examined Classification analysis to support medical diagnosis, improving quality of patient care. They propose a feature selection approach for finding an optimum feature subset that enhances the classification accuracy of Naive Bayes classifier [8]. K.Prasanna Lakshmi & C.R.K.Reddy enlightened, Associative Classification approach which combines associative rule mining and classification Associative classifiers are useful for application where maximum predictive accuracy is desired. They proposed an efficient technique for heart disease prediction. They build a classifier with prediction rules of high interestingness values and results show that this work helps doctors in their diagnosis decisions [9]. Ayush Anand and Divya Shakti examined, Diabetes Mellitus or Diabetes has been portrayed as worse than Cancer and HIV. Initially, a Chi-Squared Test of Independence was performed followed by application of the CART (Classification and Regression Trees) machine learning algorithm on the data and finally using Cross-Validation, the bias in the results was removed [10].

Madiha Guftar et al., suggested K-means clustering algorithm for classification of dataset. Syncope is classified as one of such disease. They presents novel framework for predicting possible causes of syncope disease. Results are compared by applying K-means fast, K medoids and X means algorithms. Empirical results prove that proposed framework improve the predication accuracy through novel clustering approach for possible syncope causes [11]. A. Soltani Sarvestani et al., reveals, appropriate and efficient networks for breast cancer knowledge discovery from clinically collected data sets. The results, help in choosing a reasonable treatment of the patient. Several neural network structures are evaluated for this investigation. The performance of the statistical neural network structures, self organizing map(SOM), radial basis function network (RBF),general regression neural network (GRNN) and probabilistic neural network (PNN) are tested. To overcome the problem of high dimension of the data set and realizing the correlated nature of the data, principal component techniques are used to reduce the dimension of data and find appropriate networks. The results are quite satisfactory while presenting a comparison of effectiveness each proposed network for such problems [12].

Archana L. Rane designed model using data mining techniques such as neural network, decision tree, statistical method, Naive Bayes, classifier and clustering pattern analysis for improving human life survivability[13]. Ranganatha S. et

al., stored medical information of patients who come for hospitalization for heart disease and algorithms are run on that information and result will be provided in the form of user understandable words and graph. When very large data sets are present, data mining algorithms are used. ID3 outputs the result in the form of decision tree which can be easily understood. Naïve Bayesian predicts the chances of heart disease based on conditions given [14]. Yao Liu and Yuk Ying Chung examined, data classification algorithms, discrete particle swarm optimization (DPSO), a technique based on standard PSO has proved to be competitive in predicting breast cancer[16]. K. Rameshkumar, et al., suggested Association rule mining (ARM) algorithm called as PVARM (Partition based Validation for Association Rule Mining). The performance analysis attempted with Apriori, most frequent rule mining algorithm and non redundant rule mining algorithm to study the efficiency of proposed PVARM. They reduce large number of irrelevant rules and produces new set of rules with high confidence [19].

Monika Gandhi & Shailendra Narayan Singh used, different data mining techniques for prediction of heart disease. In this, data mining methods namely, Naive Bayes, Neural network, Decision tree algorithm are analysed on medical data sets using algorithms [22]. Eiko Kai et al., used association rule technique to find common set of rules in order to build a clinical decision support system. They also showed examples of the meaningful information from the analysed data to build a better clinical decision support.

Nishchol Mishra1, Sanjay Silakari proposed, Predictive analytics uses data-mining techniques in order to make predictions about future events, and make recommendations based on these predictions. A model can be created to predict using Predictive Analytics modelling techniques. Classification & Regression are the two main objectives of predictive analytics [6]. Aqueel Ahmed et al., proposed to find out the heart diseases through data mining, Support Vec tor Machine, Genetic Algorithm, rough set theory, association rules and Neural Networks. They, examined Decision tree and SVM are most effective for the heart disease [11]. Aastha Joshi & Rajneet Kaur compares six types of clustering techniques- k-Means Clustering, Hierarchical Clustering, DBSCAN clustering, OPTICS , STING[15].

## V. OBSERVATION ON STUDY

This survey presents a systematic review of the application of Data Mining methods in healthcare domain, with a focus on the application and the techniques used which will optimize the results. In this literature survey we present an overview of the current research being carried out using the data mining techniques for the diagnosis and prognosis of various diseases. Analysis shows that it is very difficult to name a single data mining algorithm as the most suitable for the diagnosis and/or prognosis of diseases. At times some algorithms perform better than others, but there are cases when a combination of the best properties of some of the algorithms together results more effective.

## VI. CONCLUSION

In this Paper the problem of summarizing the different algorithms of Data Mining for the major life threatening disease are used in the field of medical Predication are discussed. The main focus is on using different algorithm and combination of several targets attributes for different types of disease prediction using data mining. A foremost class of problems in medical science absorbs the diagnosis of disease, based upon an assortment of tests carried out upon the patient. Due to the rapid growth of medical data, it has become indispensable to use data mining techniques to help decision support and predication systems in the field of Healthcare. In order to achieve better accuracy in the predication of disease, improving survivability rate regarding serious death related problems etc. various data mining techniques must be used in combination.

## VII. REFERENCES

- [1] Manish Shukla and Sonali Agarwal, "Hybrid Approach for Tuberculosis Data Classification Using Optimal Centroid Selection Based Clustering", IEEE, 2014
- [2] Kavitha K V Saritha R Vinod Chandra S S, "Computational Prediction of Continuous B-Cell Epitopes Using Random Forest Classifier", 4th ICCCNT July 4 - 6, 2013
- [3] Aicha Boutorh & Ahmed Guessoum, "Classification of SNPs for Breast Cancer Diagnosis using Neural-Network-based Association Rules"
- [4] Muhammad Sufyan Bin Mohd Azmi & Zaihisma Che Cob, "Breast cancer prediction based on backpropagation algorithm", Proceedings of 2010 IEEE Student Conference on Research and Development (SCORED 2010), 13 - 14 Dec 2010
- [5] Juan F. De Paz, Sara Rodríguez, Javier Bajo and Juan M. Corchado, "CBR System for Diagnosis of Patients", Eighth International Conference on Hybrid Intelligent Systems IEEE, 2008
- [6] Nishchol Mishra1, Dr.Sanjay Silakari, "Predictive Analytics: A Survey, Trends, Applications, Opportunities & Challenges", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (3), 2012
- [7] Farhad Soleimani Gharehchopoghi, Zeynab Abbasi Khalifelu, "Neural Network Application in Diagnosis of Patient: A Case Study", IEEE, 2011
- [8] Rita Samikannu, Nickolas Savarimuthu, Ramaraj Narayanaswamy, Sarojini Balakrishnan, "SVM Ranking with Backward Search for Feature Selection in Type II Diabetes Databases", IEEE, 2008
- [9] K.Prasanna Lakshmi, Dr. C.R.K.Reddy, "Fast Rule-Based Heart Disease Prediction using Associative Classification Mining", IEEE International Conference on Computer, Communication and Control (IC4-2015).
- [10] Ayush Anand & Divya Shakti, "PREDICTION OF DIABETES BASED ON PERSONAL LIFESTYLE INDICATORS", 1st International Conference on Next Generation Computing Technologies (NGCT-2015) IEEE, 2015
- [11] Madiha Guftar, Ammar Asjad Raja, Syed Hasnain Ali, Usman Qamar, "A Novel Framework for Classification of Syncope Disease using K-Means Clustering Algorithm", IEEE-SAI Intelligent Systems Conference 2015 November 10-11, 2015
- [12] A. Soltani Sarvestani, A. A. Safavi, N.M. Parandeh, M.Salehi, "Predicting Breast Cancer Survivability Using Data Mining Techniques", 2nd International Conference on Software Technology and Engineering (ICSTE) IEEE, 2010
- [13] Archana L. Rane, "Clinical Decision Support Model for Prevailing Diseases to Improve Human Life Survivability", International Conference on Pervasive Computing (ICPC) IEEE, 2015
- [14] Ranganatha S., Pooja Raj H.R., Anusha C., Vinay S.K., "MEDICAL DATA MINING AND ANALYSIS FOR HEART DISEASE DATASET USING CLASSIFICATION TECHNIQUES", IEEE
- [15] Aastha Joshi & Rajneet Kaur, "A Review: Comparative Study of Various Clustering Techniques in Data Mining", International Journal of Advanced Research in Computer Science and Software Engineering, 2013
- [16] Yao Liu and Yuk Ying Chung, "Mining Cancer data with Discrete Particle Swarm Optimization and Rule Pruning", IEEE, 2011
- [17] K. Rameshkumar, M.Sambath, S.Ravi, "Relevant Association Rule Mining from Medical Dataset Using New Irrelevant Rule Elimination Technique", IEEE
- [18] Monika Gandhi, Dr. Shailendra Narayan Singh, "Predictions in Heart Disease Using Techniques of Data Mining", 1st International Conference on Futuristic trend in Computational Analysis and Knowledge Management (ABLAZE-2015) IEEE, 2015
- [19] Eiko Kail, Andrew Rebeiro-Hargrave, Sozo Inoue, Yasunobu Nohara, Rafiqul Islam Maruf, Naoki Nakashima and Ashir Ahmed, "Empowering the healthcare worker using the Portable Health Clinic", IEEE 28th International Conference on Advanced Information Networking and Applications, 2014
- [20] Aqueel Ahmed, Shaikh Abdul Hannan, "Data Mining Techniques to Find Out Heart Diseases: An Overview", International Journal of Innovative Technology and Exploring Engineering (IJITEE), 2012