International Journal on Emerging Technologies (Special Issue NCETST-2017) **8**(1): 124-127(2017)
(Published by Research Trend, Website: www.researchtrend.net*)*

ISSN No. (Print) : 0975-8364
ISSN No. (Online) : 2249-3255

Analytical Study of Applied Data Mining in Healthcare

Anurag Bhatt[1] and Manish Joshi[2]
[1]Assistant Professor, CSE Department, Amrapali Institute of Technology and Sciences, Haldwani, Uttarakhand, India
[2]M. Tech. Scholar, CSE Department, Amity University Uttar Pradesh, Lucknow Campus,Uttar Pradesh, India

ABSTRACT: Enormous amount of data is being generated every year, adding a huge bulk of unprocessed data that have potential to outrun various health issues with the help of technology. Data mining in healthcare is a boon to our society by extracting various hidden patterns that helps in understanding hidden patterns and automated monitoring of the various disease. Electronic medical devices generate data that are analyzed later and various predictive models are then built upon them. This paper provides a review of the role of data mining in healthcare in order to provide the better understanding of the subject. Data Mining Algorithms and Techniques are being used to detect and predict Sudden Cardiac Arrest, Diabetes etc and other health issues using Decision Trees, Naïve Bayes classifiers etc. We have analyzed major Data Mining Techniques being deployed now a days.

Keywords: Data mining algorithms, Diabetes, Sudden Cardiac Arrest, Decision Tree, Niave Bayes, Healthcare

I. INTRODUCTION

In the past few years, it has been analyzed that increasing volume of data and its various undefined resources has taken the shape of revolution in the field of knowledge extraction. Data Mining is quite the filed that incorporates the process of combining multiple data sources in large volumes and then processing of that data to get optimized results in the data. Data mining as the new field has played a vital role in various fields like Industrial data analysis, Market research as well as healthcare informatics.

Healthcare industry has always been the honeywell for the betterment of the society as there is unstoppable growth in terms of data collected through electronic medical devices. Patient data is then analyzed to compute the analysis of the continuous data generated from devices through various data mining techniques. Healthcare industry has provided immense services to their clients in terms of better treatment as well as understanding the overall analytics behind the huge amount of data generated .In this paper, we have presented the analytical survey of the increasing role of data mining techniques for the better analysis and prediction of data mining in healthcare informatics.

II. RELATED WORK

Healthcare industry has always been buzzword for years and researchers has given various theories in subject of applying data mining techniques in healthcare industry for better insight into healthcare data.

There are various previous works discussed in this paper regarding the role of data mining in prediction and analysis of data collected from various data resources.

Researcher [1] has discussed the importance of applyingdata mining in clinical data collected of diabetic patients and enormous amount of data generated from electronic medical devices. Researcher has also showed that data mining techniques are widely accepted in the field of diabetes for better insight into the matter. Researcher [2] has contributed a novel approach in analysis and prediction of heart disease through data mining techniques like KNN, MAFIA (Maximal Frequent Itemset Algorithm) based algorithms. Data is first preprocessed through attribute selection and then it is clustered and the classified to find the patterns in the processed data.

Researcher [3] have has presented a technique to detect arrhythmia in ECG signal generated form patients heart in the form of medical records. These recorded signal are then analyzed to get irregular patterns of having arrhythmia in the patient data. Various studies have been conducted in the past few years by researchers gaining deep insights into the matter.

Following is the table showing various tools and techniques used in various studies and the databases used to gain insight into the matter.

Table 1. Table showing various tools and data mining techniques used.

S.no	Databases	Tools & Techniques	Data Mining Algorithms
[1]	Cleveland Database	Feed Forward Neural Network	Neural Network Approach
[2]	Cleveland Heart Disease Database	DMX Query Language	ID3 and Decision Tree algorithm is used
[3]	South African Medical Practitioners Database	J48, Bayes Network	Simple Cart and REPTREE Algorithm
[4]	UCI Machine Learning Repository	WekaTool is used	Naïve Bayes and Decision Tree algorithm
[5]	Trace of ECG Signal	HRV(Heart Rate Variability) is used	Wigner Ville Transformation Technique
[6]	UCI Breast Cancer Database	Classification, Clustering and Rule Mining is used	Association Rule Mining method
[7]	Blood-Glucose Homo Monitoring database	Association and classification Techniques	Association and classification Techniques
[8]	Physio-Bank Database is used	Attribute Selection and EM Clustering	clustering techniques

III. DATA MINING TECHNIQUES

Healthcare industry has moved from traditional way of generating patient data to the automated and well analyzed way of collecting data from electronic devices. With the advent of recent technology, it has been found that huge amount of voluminous data has been generated in medical institutions and healthcare industry. This data is required to be analyzed through data mining techniques like classification, clustering etc. These analysis and predictions on the present data is done through finding the hidden patterns in this huge voluminous data through various effective data mining algorithms.

Researcher [4] has unfolded the analysis of data collected of dengue disease through the machine learning tool i.e. weka. In this paper, it has been discussed that various data mining algorithmic approaches has been used to predict the dengue disease in patients using various patient input data. Raw data collected has been first preprocessed and then classified through data mining algorithms like J48 and Naïve Bayes etc. with maximum accuracy.

There are various data mining techniques available, for better prediction and analysis of data collected from numerous sources.

Association: This data mining techniques is quite literal and straightforward, focusing on the fact of finding associations among various items. In this technique, two or more than two items of the same type are associated with each other, to find the hidden patterns behind the data and correlate them. Suppose, in an airport, it has been analyzed from the buying pattern of men is that, while buying beer from the store, men are likely to buy baby products. In this way, the hidden pattern is all about buying two different items of the same store at the same time. Therefore, finding patterns through association will lead to generate business suggestion to keep both baby products and men's accessories close to each other. This buying habits is sometimes generates more profitable market research of buying patterns.

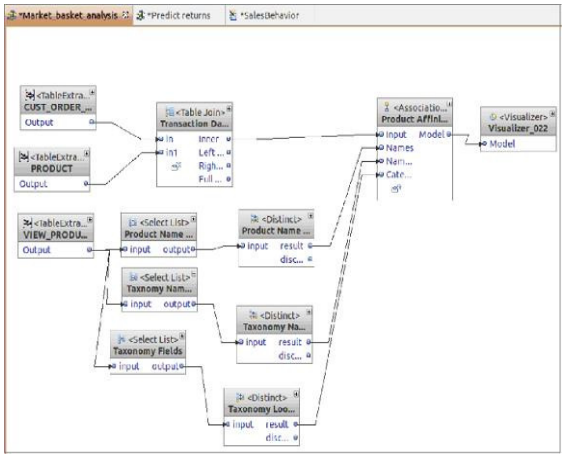

Fig 1. Figure showing knowledge flow in association [https://www.ibm.com/developerworks/library/ba-data-mining-techniques/].

Classification: Classification techniques under data mining is concerned all about differentiating an entity based on its class labels. Classification techniques involve labelling with class label and then classify the objects based on their features with their appropriate class labels. Classification techniques like SVM (Support Vector Machine) etc. Researcher [5] have

used decision tree and neural network as classification techniques to classify heart disease database and to generate intelligent heart disease prediction system. Classification techniques are all based on the method of identifying different attributes i.e. differentiating two cars i.e. one of sedan and other of Volkswagen , can be done on the basis of their attributes i.e. car size, features etc.

Clustering: The process of clustering in data mining techniques is concerned with the process of differentiating data objects according to the characteristics of features of an object. In the process of clustering, it is highly likely that objects in the same clusters shares the same features as other attributes i.e. similar to nearest neighbor identity.

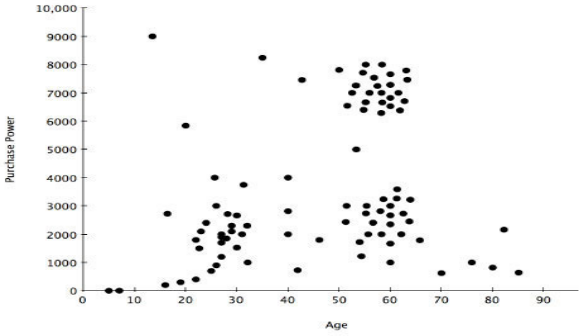

Fig 2. Figure showing clustering technique[https://www.ibm.com/developerworks/library/ba-data-mining-techniques/].

Sudden Cardiac Arrest prediction using data mining techniques in the field of heart disease is one of the hottest research topic due to the seriousness and lack of appropriate solutions to this problem. Researcher [6] have presented a review done on this topic for further research. Various techniques like HRV, T-Wave Alternans etc. are used to predict SCD and in this review statistical techniques are found much reliable as compared to various other classification techniques to successfully predict the Sudden Cardiac Arrest(SCD) in patients.

Predictive Analytics is another aspect of data mining that has been extensively used in healthcare industry, in orderto provide the efficient analysis and predictions based on the data collected. In the paper [7] researcher has used diabetes data for the prediction of the treatment given to the different age group of diabetes patients. This paper, on the basis of data mining predictions discloses the eye opening fact about the different treatment strategies for different age group of patients. Researcher[8] have used classification techniques like neural network and back propagation algorithm to recognize dengue fever using blood report as input and extracted features and signals are then analyzed to visualize and predict the results.

IV. DATA MINING IN HEALTHCARE

Predictive data mining is a part of data mining that helps in improving the efficiency as well as produces better results for future enquiries. Predictive data mining helps in generating various predictive models based on the current data on applying data mining algorithms. Models based on predictive data mining helps in creating framework as well as in order to produce automated mechanism that relies on various machine learning algorithms for its functionality. Researcher [9] have provided a survey revealing about the use of data mining techniques specially predictive data mining techniques for knowledge discovery in building heart disease prediction system where decision tree and Bayesian outperforms all others while neural network based classifications are not working well.

Extensive use of data mining in healthcare doesn't limit data mining to heart disease, diabetes and any other similar diseases. Data mining techniques has been widely used in detecting pattern of disease in various states of our country by finding various similarity patterns between diseases in throughout the country, which will be a benefit to the medical as well as IT professionals. [10]. Heart disease prediction is considered as another complex task to gain insight into the hidden patterns of the voluminous data collected from various medical institutions and hospitals. Researcher [11] have presented a review stating about the fact of using various data mining techniques like classification, clustering, naïve Bayes, decision tree, neural networks etc. in order to predict the occurrence of fatal heart disease prediction.

Sudden cardiac Arrest prediction is one of the fatal situation found in patients having medical history of heart disease. SCD is considered to be as the sudden stopping of heart leading to death. Researcher[12] have proposed a system by creating a Code blue alert system, that enables rapid action team to start responding to save the patient from sudden cardiac arrest problem. Electronic records of patients have heart disease are constantly monitored and any recent update in the records are considered to be the upcoming heart attack problem, that might lead a patient to death.In general, heart disease patients are flagged from the medical team or ranked based on the data gathered from their electronic records. Medical Fraud Detection [13] is being considered as one of the application of applying data mining in healthcare industry. Medical professionals as well as medicine providers are detected to analyze the hidden patterns behind data and in this way original database is compared with faked database to generate effective results. Supervised machine learning algorithms like Decision tree, Naïve Bayes and Neural network are used to generate heart disease prediction system [14]. In paper [15], breast cancer is analyzed using association and classification techniques to generate results. Association rule mining and

classification techniques are combined under associative classification technique, and it helps to predict the severity level of the problem in the patient.Researcher [16] have reviewed the diagnosis of Arbovirus- dengue in the patients using classification techniques under data mining. This research is focused in finding the difference or hidden patterns using classification techniques to differentiate chikungunya and dengue viral fever.

V. CONCLUSION

In a nutshell, we can conclude that the applications of data mining in healthcare industry is vast and yet to discover. The process of extracting knowledge from the huge volume of data I quite tedious but it enables to find the hidden patterns in the disease that enables us to analyze various eye opening facts about the data. Classification, clustering, neural networks etc. have their different approach to find the correlation among different data sets. Results generated through this database enables us to make predictions and analyze the current scenario. Various predictive models gets generated from the datasets for better functioning of the model.

REFERENCES

[1] Miroslav Marinov, M.S., Abu Saleh Mohammad Mosa, M.S.,Illhoi Yoo, Ph.D.,and Suzanne Austin Boren, Ph.D., MHA, "Data-Mining Technologies for Diabetes: A Systematic Review", *Journal of Diabetes Science and Technology,* Vol **5**, Issue 6, November 2011, pp. 1555.

[2] G.Karthiga, C.Preethi, R.Delshi Howsalya Devi, "Heart Disease Analysis System Using Data Mining Techniques", *International Journal of Innovative Research in Science, Engineering and Technology*, Volume **3**, Special Issue 3, March 2014, pp. 3101, 3105.

[3] Neerakj Kumar S Sathawane& Pravin K Shirsagar, "Prediction and Analysis of ECG Signal Behaviour using Soft Computing", *International Journal of Research inEngineering & Technology*, Vol. **2**, Issue 5, May 2014, 199-206, pp. 205.

[4] Kashish Ara Shakil, Shadma Anis and Mansaf Alam, "Dengue Disease Prediction using Weka Data Mining Tool", pp. 1, 23.

[5] Sellappan Palaniappan, Rafiah Awang, "Intelligent Heart Disease Prediction System Using Data Mining Techniques", IEEE, 2008, pp. 115

[6] L Murukesan , M Murugappan , M Iqbal, "Sudden Cardiac Death Prediction using ECG Signal Derivative (Heart Rate Variability): AReview", *IEEE 9th International Colloquium on Signal Processing and its Applications*, 2013, pp. 269, 272.

[7] Abdullah A. Aljumah, Mohammed Gulam Ahamad, Mohammad Khubeb Siddiqui, "Application of data mining: Diabetes health care in youngand old patients", *Journal of King Saud University – Computer and Information Sciences* (2013), pp. 127

[8] N.Subitha, Dr.A.Padmapriya, "Diagnosis for Dengue Fever Using Spatial DataMining", *International Journal of Computer Trends and Technology (IJCTT)* – volume **4** Issue 8–August 2013, pp.2651

[9] Jyoti Soni, Ujma Ansari, Dipesh Sharma, Sunita Soni, "Predictive Data Mining for Medical Diagnosis: An Overviewof Heart Disease Prediction", *International Journal of Computer Applications*, Volume **17**– No.8, March 2011, pp.43

[10] Hemlata Sharma, Pallavi Sharma, "Application of Data Mining in Detecting pattern of Disease Spread in Various States of India", *International Journal of Advanced Research in Computer Science and Software Engineering,* Volume **4**, Issue 6, June 2014, pp.291.

[11] K. Sudhakar, Dr. M. Manimekalai, "Study of Heart Disease Prediction using Data Mining, *International Journal of Advanced Research in Computer Science and Software Engineering,* Volume **4**, Issue 1, January 2014, pp. 1157

[12] Sriram Somanchi, Samrachana Adhikari, Allen Lin, Elena Eneva, Rayid Ghani, "Early Prediction of Cardiac Arrest (Code Blue) usingElectronic Medical Records", ACM, 2015, pp. 2119, 2124

[13] Jacqulin Margret J, Shrijina Sreenivasan, " Implementation of Data Mining in Medical Fraud Detection" *International Journal of Computer Applications,* Volume **69**– No.5, May 2013, pp. 1

[14] Abhishek Taneja, "Heart Disease Prediction System Using Data Mining Techniques", *Oriental Journal of Computer Science and Technology,* December 2013, Vol. **6**, pp. 457

[15] A.Kavipriya, B.Gomathy, "Data Mining Applications in Medical Image Mining: An Anlaysis of Breast Cancer using weighted rule mining and classifiers", *IOSR Journal of Computer Engineering*, Volume **8**, Issue 4, pp.18,23

[16] A.Shameem Fathima, D.Manimegalai and Nisar Hundewale, " A Review of Data Mining Classification Techniques applied for diagonosis and prognosis of the Arbovirus-Dengue" *IJCSI*, Vol. **8**, Issue 6, No. 3, 2011, pp. 322.