



Data Mining Techniques in E-Health Systems: An Analysis

A. MAHAR, S. RAJPER

Department of Computer Science, Shah Abdul Latif University, Khairpur

Received 16<sup>th</sup> June 2016 and Revised 2<sup>nd</sup> March 2017

**Abstract:** Data mining implies a lot of information to extricate valuable information, most imperative and well known data mining techniques are classified, clustering, sequential patterns, regression, association, and prediction. In Health concern organizations, data mining has a vital part, in early forecast of diseases There are numerous approaches to analyze the malady; one of these strategies is Data mining technique. This research to survey and analyze that which Data mining techniques are being used for diagnosis of various diseases to contribute researchers and scholars. The objective of this study is to conduct a systematic analysis of data-mining techniques in the area of e- health systems. This analysis will be helpful to compare robustness of various Data Mining techniques in this area.

**Keywords:** Data Mining; E-Health; classified, Naïve Bayes; Decision Tree.

1. **INTRODUCTION**

The Information systems capture large data in data bases for medical analysis and research. The customary manual information examination has gotten to be wasteful and techniques for productive PC based investigation are vital this point, numerous ways to deal with modernized information investigation have been considered and inspected. Information mining speaks to a noteworthy development in the sort of systematic devices as of now accessible. Data mining software finds and analyzes relationships in large data and usually, the relationships, i.e., classes, clusters, Associations, Sequential Patterns are sought. Various data mining techniques are being used to diagnose/find related patterns from huge data of health concern organizations. Data Mining gives advantages, for example, discovery of the extortion in health care coverage, accessibility of therapeutic solutions for the patients at lower cost, early detection of reasons for illnesses and can identify the treatment techniques. It moreover helps the human services specialists for making effective social insurance strategies, building drug proposal frameworks, creating health care profiles of people (Koh and Tan 2011) (Alafaireet *et al.* 2012; Tomar and Agarwal 2013). In present time different human services organizations are creating gigantic measures of data. Data Mining instruments for investigation and deciphering the helpful data from this information. This data is exceptionally important for medicinal services authority to comprehend the reasons for ailments and identify treatment to patients (Silver, Sakata *et al.* 2001; Bellazzi and Zupan 2008). the Data Mining techniques used for health care. Data mining is a

fundamental stride of knowledge discovery. In recent years it has pulled in extraordinary arrangement of enthusiasm for information concerned businesses. Knowledge discovery process comprises of an iterative arrangement of the following steps:

- Data Cleaning,
- Data Integration,
- Data Selection,
- Data Mining Pattern/Relationship Recognition
- Knowledge Presentation (Kaur and Wasan 2006) (Fayyad, Piatetsky-Shapiro *et al.* 1996).

(Fig. 1) is used to show the Data Mining process life-cycle to extract and analyze treasured knowledge.

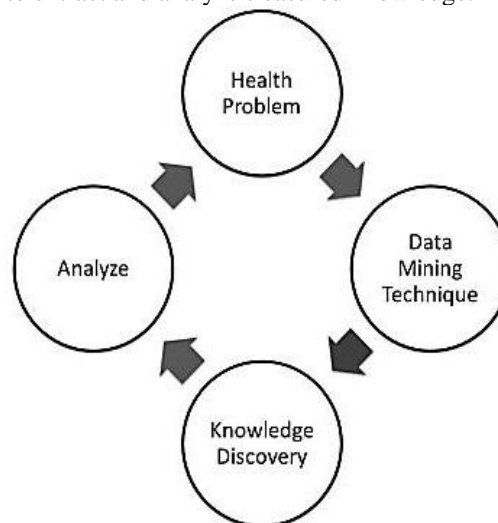


Fig.1 Data Mining Process

++Corresponding Author's mail:samina.rajper@gmail.com

## 2. MATERIALS AND METHODS

To conduct this research survey, research studies from 2001-2016 were studied in which data mining techniques were used for various health care issues and their solutions. The study was conducted by planning, shepherding and reporting the research studies in the mentioned field. (Fig. 2). Shows the flow of undertaken study

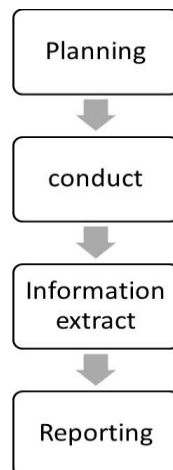


Fig.2. Flow of undertaken study

## 3. RESULTS AND DISCUSSION

The outcomes of this research survey are reported using the technique used in various health care studies :

### 3.1 Classification and Clustering

(Nagarajan and Chandrasekaran 2015) used classification and clustering techniques of data mining for detection of diabetes, digestion and kidney diseases dataset and also compare the results with decision tree technique results for the same dataset. (Yasin, *et al.* 2011) used classification technique for Hepatitis-C patients. The yield indicates whether patients with hepatitis are alive or dead. (Fathima, 2011) used classification techniques for separating the Dengue and Chikungunya patients. (Rajkumar and Reena 2010; Chandna 2014) used Naïve Bayesian Algorithm for Diagnosis of Heart Disease But got 52.33% accuracy in results. (Karegowda, and Jayaram 2012) Categorization of Diabetic Patients using Cascading K-means Clustering and K-Nearest Neighbor (KNN) Classifier were conducted. (Srinivas, and Rani 2010) used Naïve Bayes extended algorithm to discover patterns from huge medical data to discover relationship between different factors causing heart disease.

### 3.2 Neural Network and Association Rule

(Kharya 2012) conducted the study for diagnosis of Breast Cancer Disease using data mining techniques. Author notified that large ratio of deaths reported in women specially in developed countries. Therefore, DM is the most effective way to detect early breast cancer. This study performed case-control study on dataset collected by the Department of Genetics Medical Sciences Faculty, Universidad Nova de Lisboa. Authors performed experiments for tumor detection in digital mammography. This research used various data mining techniques, i.e., Association Rules and neural networks. (Fathima, 2011) studied that how to find infection of Dengue/ Chikungunya. The most essential segment of this framework will likewise allow to separate whether the sickness is Dengue or Chikungunya. The goal of these forecasts is to relegate patients to either a "Dengue" or a "Chikungunya" or "some other contamination" and to handle perplexing cases for the viral sickness. Data mining methods, i.e., classification (Bayesian Classifier, Neural network, Support Vector Machines SVM) were used. (Boughorbel, *et al.* 2016) compared the performance of various prediction models for breast cancer prognosis, based on AU-ROC performance (Area Under ROC). They compared eight techniques, i.e., Generalized Linear Model (GLM), Partial Least Square (PLS), GLM-Net, Random Forests (RF), Neural Networks, k-Nearest Neighbors (k-NN) Support Vector Machines (SVM), and Boosted Trees. (Er, Yumusak 2010) used Artificial Neural Network (ANN) to analyze chest infections/diseases and performed a comparative analysis of chest diseases/infections using generalized regression, multilayer, probabilistic neural networks.

### 3.3 Decision Tree

(Chang and Chen 2009) used Decision Tree (DT) for diagnosis of skin diseases. (Smit, Smit *et al.* 2013) used DT for Skin Auto fluorescence (SAF) for detection of impaired Glucose Tolerance and Diabetes. (Song, Choi *et al.* 2016) used DT and apriori algorithm for finding the likenesses between the amino acid sequences between viruses. (Devi and Devi 2016) focused on investigating the automated diagnosis of breast cancer. The proposed approach has three steps of a process. They used Farthest First clustering algorithm, then ODA (Outlier Detection Algorithm) and finally, J48 DT. (Table 1). Shows the summary of the research techniques used by various research studies to detect and diagnose various diseases.

**Table 1. Summary of DM Techniques**

Study	Techniques used
(Nagarajan and Chandrasekaran 2015), (Yasin, Jilani <i>et al.</i> 2011), (Fathima, Manimegalai <i>et al.</i> 2011), (Rajkumar and Reena 2010; Chandna 2014)	Classification and clustering techniques
(Kharya 2012), (Fathima, Manimegalai <i>et al.</i> 2011),	Bayesian Classifier, Neural network, Support Vector Machines SVM
(Chang and Chen 2009), (Song, Choi <i>et al.</i> 2016)	Decision Tree
(Boughorbel, Al-Ali <i>et al.</i> 2016)	Generalized Linear Model (GLM), Partial Least Square (PLS), Random Forests (RF), k-Nearest Neighbors (k-NN) and Boosted Trees GLM-Net, Support Vector Machines (SVM), Neural Networks.
(Devi and Devi 2016)	clustering algorithm, ODA (Outlier Detection Algorithm) and DT.
(Er, Yumusak <i>et al.</i> 2010)	Artificial Neural Network (ANN)

#### 4. **CONCLUSION**

Data mining techniques are found efficient during the undertaken study. Many Data Mining techniques were compared on the same datasets of patients to check the accuracy of the technique. Mostly classification, clustering, Neural Networks are used in various studies. The undertaken study concluded that the accuracy, relevance and quality of datasets is important to acquire any data mining technique for accuracy in results.

#### **REFERENCES:**

Chandna, D. (2014). "Diagnosis of heart disease using data mining algorithm." *IJCSIT International Journal of Computer Science and Information Technologies* 5(2): 1678-1680.

Chang, C. L. and C. H. Chen (2009). "Applying decision tree and neural network to increase quality of dermatologic diagnosis." *Expert Systems with Applications* 36(2): 4035-4041.

Devi, R. D. H. and M. I. Devi (2016). "Outlier Detection Algorithm Combined With Decision Tree Classifier For Early Diagnosis Of Breast Cancer." *Int J Adv Engg Tech/Vol. VII/Issue II/April-June* 93: 98.

Er, O., N. Yumusak, (2010). "Chest diseases diagnosis using artificial neural networks." *Expert Systems with Applications* 37(12): 7648-7655.

Fathima, A. S., D. Manimegalai, (2011). "Data mining classification techniques applied for diagnosis and prognosis of the arbovirus-dengue." *IJCSI International Journal of Computer Science* 8(6). 27-34.

Fayyad, U., and G. Piatetsky-Shapiro, (1996). "From data mining to knowledge discovery in databases." *AI magazine* 17(3): 37-40.

Kharya, S. (2012). "Using data mining techniques for diagnosis and prognosis of cancer disease." *arXiv preprint arXiv:1205.1923*.

Nagarajan, S. and R. Chandrasekaran (2015). "Design and implementation of expert clinical system for diagnosing diabetes using data mining techniques." *Indian Journal of Science and Technology* 8(8): 771-776.

Rajkumar, A. and G. S. Reena (2010). "Diagnosis of heart disease using datamining algorithm." *Global Journal of Computer Science and Technology* 10(10): 38-43.

Song, S. H., Y. Choi, (2016). Comparison of episodes of mosquito borne disease: Dengue, yellow fever, west nile, and filariasis with decision tree, apriori algorithm. 18<sup>th</sup> International Conference on Advanced Communication Technology (ICACT), IEEE.

Srinivas, K., and B. K. Rani, (2010). "Applications of data mining techniques in healthcare and prediction of heart attacks." *International Journal on Computer Science and Engineering (IJCSE)* 2(02): 250-255.

Yasin, H., and T. A. Jilani, (2011). "Hepatitis-C classification using data mining techniques." *International Journal of Computer Applications* 24(3): 1-6.