A Review on Data Mining Techniques for Treatment of Cancer in Ayurveda Therapy

*Deepa Nyayadhish, **Dr.Pramod Deo
*AP, V.K.K.Menon College, Mumbai
**Asst. Professor, S.B.E.S. College, Aurangabad

Abstract-The purpose of this literature review is to evaluate and analyse the literature relevant to the topic. This paper presents an overview of the applications of data mining in Health Informatics. The current or potential applications of various data mining techniques in Health Informatics are illustrated through some published literature. Data mining techniques such as clustering, Decision tree, ANN, CART (Classification and Regression Tree) is widely used in healthcare domain. Data mining algorithms, when appropriately used, are capable of improving the quality of prediction, diagnosis and disease classification [1]. The main focus of this paper is to analyse data mining techniques required for Ayurveda therapy data especially to discover effective treatment for cancer.

Keywords-Data mining techniques, Ayurveda Therapy, Expert System, Data mining process

1. INTRODUCTION

Data Mining or "the efficient discovery of valuable, on-obvious information from a large collection of data"[2] has a goal to discover knowledge out of data and present it in a form that is easily comprehensible to humans. Knowledge detection in databases is precise process consisting of a number of distinct steps [3]. Data mining is the foundation step, which outcome in the discovery ofunknown but helpful knowledge from huge databases. A formal definition of Knowledge discovery in databases is given as follows: "Data mining, or knowledge discovery, is the computer assisted process of digging through and analysing enormous sets of data and then extracting the meaning of the data. Data mining tools predict behaviours and future trends, allowing businesses to make proactive, knowledge-drivendecisions [4].Data mining expertise provide a consumer leaning approach to new and unknown patterns in the data. The exposed knowledge can be used by the healthcare administrators to progress the superiority of service.

Ayurveda is an ancient system of health care that is native to the Indian subcontinent. The word "Ayurveda" is compound of the word āyus meaning "life" or "life principle", and the word veda, which refers to a system of "knowledge". Ayurveda is concerned with measures to protect "ayus", which includes healthy living along with therapeutic measures that relate to physical, mental, social and spiritual harmony. This traditional healthcare system is trying to prove its identity by searching newer remedies to overcome the diseases for which there is no answer in modern medical science. Medical informatics may not be able to address all the emerging areas of Ayurveda because the concepts in Ayurvedic system of medicine are different from modern system.

Ayurveda is the most suitable system of medicine in which Data mining techniques can be applied, provided both the IT experts and Ayurveda experts have very clear idea about the potentiality of both systems.

2. LITERATURE REVIEW

Many studies have been conducted previously to investigate effective and efficient Mining techniques in health care for cancer treatment. Literature reviews are secondary sources and do not report any new or original experimental work. Selected important literatures found relevant to the present study are given as follows.

K. Srinivas et al. discussed the potential use of classification based data mining techniques such as Rule Based, Decision tree, Naïve Bayes and Artificial Neural Network to the massive volume of healthcare data. Using an age, sex, blood pressure and blood sugar medical profiles it can predict the likelihood of patients getting a heart disease [5].

Shweta Kharya focused on various data mining approaches that have been utilized for breast cancer diagnosis and prognosis Decision tree is found to be the best predictor with 93.62% accuracy on benchmark dataset and also on SEER data set [6].

3. OTHER LITERATURE RELATED TO AYURVEDA THERAPY

Expert System

It is the most important branch of Artificial Intelligence .The Inference Engine of an Expert System uses the knowledge base and produces inferences. For making up an Expert System knowledge acquisition is used to obtain domain knowledge about the application. The process of building an Expert System is called knowledge engineering and is done by a knowledge engineer. Development of an Expert System continues until the performance of the system is found satisfactory by the expert.

D.S. Kalana Mendis et al. considered domain of Ayurvedic medicine as a case study for domain with tacit knowledge. A questionnaire used to classify individuals in *Ayurvedic* has been studied and found that the classification is still vague, subjective and cannot be addressed using traditional technique like Principal Component (PC) analysis. They developed an approach to model such tacit knowledge using PC and Fuzzy Logic that has been linked with Expert system technology [7].

Mendis et al. presented a multi techniques integrated expert system for diagnosis of *Prakurthi* in *Ayurvedic* medicine. The present mechanisms used for diagnosis of *Prakurthi*, which is considered, as classification of human constitutions and is inconsistent about its findings. Statistical technique and fuzzy logic have been described as multi techniques involved in the expert system [8].

Krantee Jamdaade et al. explained Expert system for diagnosing HIV/AIDS by using *Ayurveda* Therapy. This Expert System will be a standalone system which will implement the Knowledge Base and usage of Inference logic. Using this and on the basis of symptomatology and pathological findings, this system is going to infer the prescription for the HIV patient. This paper explains how to use ES that will help to formulate the base for immunization system [9].

D.S. Kalana Mendis et al. used *Vedic* psychology forms the clinical core of mental health counselling in the *Ayurvedic* medical tradition. According to *Ayurvedic* medical practises, a person is dominated on one of constitutes type known as *prakurthi parikshan.* An approach to modelling common sense knowledge in clinical psychology in *Ayurvedic* medicine. It gives three-phase an approach for modelling common sense knowledge in psychological assessment which enables holistic approach for clinical psychology. Evaluation of the system has shown 77% accuracy [10].

G. Gnana Sekari et al. explores that Expert System is a versatile tool, which can be used as multi-purpose systems such as decision support system, diagnostic system, and teaching aid etc. An expert system is design for *Siddha medical systems*. Siddha medicine, which is found in the regional language of the Tamils, is now gradually being translated into English and Hindi. It is believed that these medicines arrest the advancement of the some of the killer diseases such as Cancer and Aids, when detected at an early stage [11].

4. JOURNAL OF RESEARCH IN AYURVEDA

Premalatha Balachandrana et al.discussed about the pathology and therapeutic management of various cancers described in *Ayurveda*. Review of literature on anticancer drugs of plant origin revealed identification of newer *Ayurvedic* drugs that are not mentioned in the ancient texts. In addition, details of experimental and clinical studies conducted on single and compound A*yurvedic* preparations for their anticancer efficacy strongly emphasize A*yurvedic* therapy as a scientifically driven one and not simply unconventional [12].

Jayarajan Kodikannath elaborated descriptions in *Ayurvedic* classics on the cause, clinical features and management of tumours and cancer, with the names such as Apachi, Gulma, Granthi, and Arbuda. There are specific protocols mentioned on curative aspects of cancers that have resemblance with clinical entities of arbuda and granthi. Arbuda is considered as condition involving multiple doshas (often Tridosha), multiple dhatus and associated with multiple complications. Abuse of nature's law upsets the human system and ends up in disease like cancer [13].

Sushant Sud et al. discussed the impact of IT and the advances in information and communication technologies (ICTs) in this scientific age. *Ayurveda* needs to be restructured globally to meet the rising demands of a cyber-mobile society with the application of information and communication technology. Therefore, *Ayurveda* needs Ayur-informatics to keep pace with this modern world [14].

5. DATA MINING CONCEPTS

Data mining is defined as "a process of nontrivial extraction of implicit, previously unknown and potentially useful information from the data stored in a database" by Fayyad [18].

The data mining processes include formulating a hypothesis, collecting data, performing pre-processing, estimating the model, and interpreting the model and draw the conclusions.
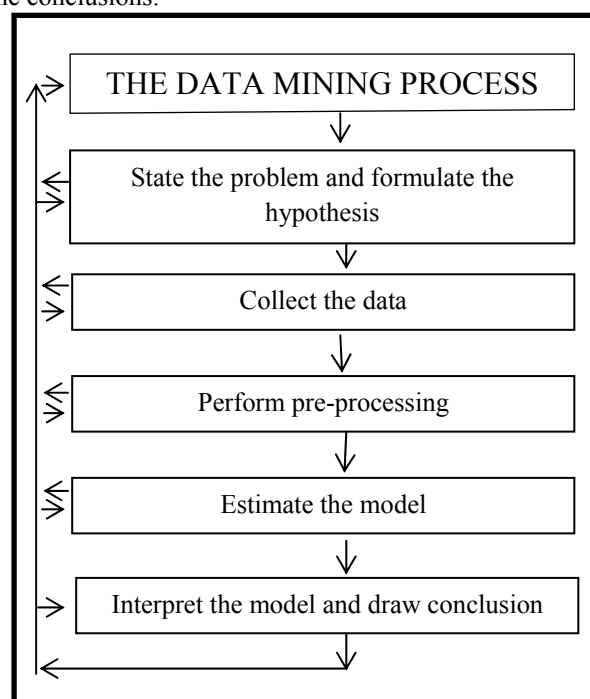

Fig.1 The Data Mining Process [22]

Healthcare databases have a huge amount of data but however, there is a lack of effective analysis tools to discover the hidden knowledge.Appropriate computer-based information and/or decision support systems can help physicians in their work. Efficient and accurate implementation of an automated system needs a comparative study of various techniques available. In this paper an overview of the current research being carried out using the DM techniques for the diagnosis and prognosis of various diseases [23].

6. DATA MINING TECHNIQUE USED FOR CLASSIFICATION OF DATA

Decision Tree:
Decision trees are an approach of representing a sequence of rules that lead to a set or value. As a result, they are used for directed data mining, mainly classification. One of the mainreward of decision trees is that the model is quite

reasonable since it takes the form of explicit rules. This allows the evaluation of results and the identification of key attributes in the process [19]. It consisting of nodes and branches organized in the form of a tree such that, every interior non-leaf node is labelled with ideals of the attributes. The branches coming out from an inner node are labelled with ideals of the attributes in that node. Each node is labelled with a rank (a worth of the goal characteristic). Tree based models which include classification and regression trees, are the common implementation of induction modelling [20]. Decision tree algorithms such as CART, ID3, C4.5, SLIQ, SPRINT. The decision tree can be built from the very small training set (Table 1). In this table each row corresponds to an enduring record. We will refer to a row as a data instance.

The data set contains three predictor attributes, namely Age, Gender, symptoms and one goal attribute, namely disease whose values to be predicted from symptoms indicates whether the corresponding enduring have a certain disease or not.

Table 1: Data set used to build decision tree of Figure.2

Age	Gender	Symptoms	Disease (Goal)
5	Female	Medium	Yes
3	Male	High	Yes
2	Female	Medium	Yes
4	Female	High	Yes
10	Female	Low	No
9	Male	Low	No
11	Female	Low	No

Decision tree can be used to classify data of the above data set given in the Table 1. The idea is to push the instance down the tree is shown in fig2; following the branches whose attribute values match the instances attribute values, until the instance reaches a leaf node, whose class label is then assigned to the instance [21]. In this example, Gender attribute is irrelevant to a particular classification task. The tree tests the intensity of symptom value in the instance. If the answer is medium; the instance is pushed down through the corresponding branch and reaches the Age node.
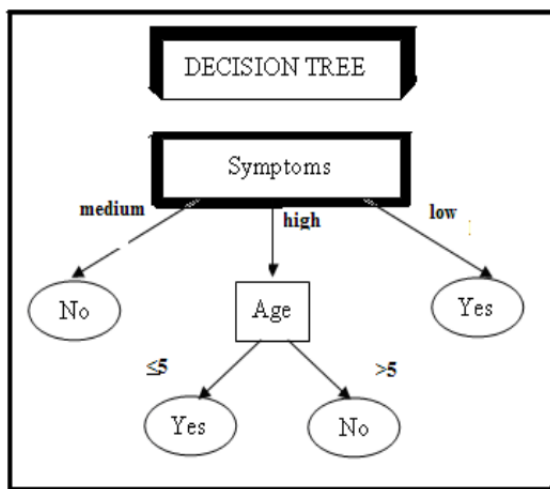

Fig. 2: A decision tree built from the data in Table 1

Then the tree tests the Age value in the instance. If the answer is 5, the instance is again pushed down through the corresponding branch. Now theInstance reaches the leaf node, where it is classified as yes [16].

ANN
An (ANN) artificial neural network, also called a neural network, is a mathematical model based on biological neural networks [24]. A neural network consists of an interconnected group of artificial neurons. Neural networks are used to model complex relationships between inputs and outputs or to find patterns in data.

Classification
The objective of the classification is to assign a class to find previously unseen records asaccurately as possible. If there is a collection of records (called as training set) and each record contains a set of attributes, then one of the attributes is class[29,30]. The motive is to find a classification model for class attributes, where a test set is used to determine the accuracy of the mode.

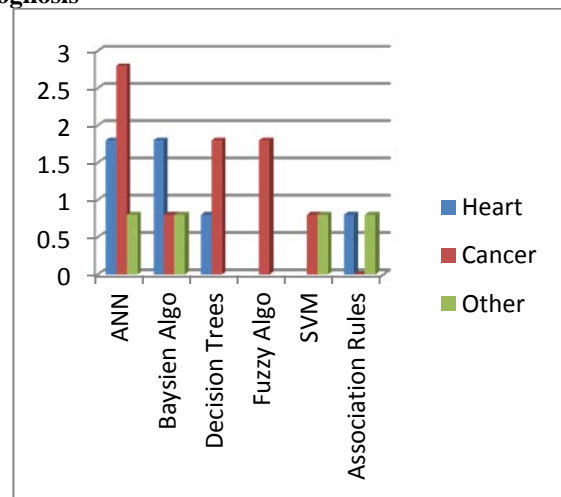**Well-performing DM algorithms used for disease prognosis**

Fig.3 Efficient Algorithms for Disease Prognosis [26, 14]

The graphs in Figure 3 shows the well-performing algorithms used for disease prognosis that can be applied in Ayurveda therapy for effective treatment of cancer. Diseases in Heart Diseases are classified (Cardiovascular disease, Heart Attack, Coronary Artery Disease, Hypertension), Cancer Diseases (Breast, Prostate, Pancreatic Cancer) and Other Diseases (Asthma, Diabetes, Hepatitis, Kidney Disease, Nerve Diseases, Chronic Disease, Skin Diseases). As we can see in Fig.3, ANNs are the well-performing in diagnosing Cancer Diseases, Bayesian Algorithms and Decision Trees in Heart Diseases, and DTS in diagnosing other diseases [15].

7. CONCLUSION
This paper presents that various data mining techniques has become a necessity for health care organizations. Data mining techniques have higher utility in medical data mining as there is voluminous data in this industry. 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.

Data mining are quite popular in allopathic field of medicine. This review identified that attempts are made to design an Expert System for Ayurveda therapy. So far, no attempt has been made to design an apply data mining techniques for cancer treatment in Ayurveda Therapy.

This review is an attempt to explore possibility of implementation of the concept of data mining to the domain of Ayurvedic treatment. This concept will definitely yield fruitful results provided it executed in a systematic manner it will surely support in reducing the time and cost of Ayurvedic medical treatment in the near future.

REFERENCES

[1] Dr. D. P. Shukla; Shamsher Bahadur Patel; Ashish Kumar Sen : A Literature Review in Health Informatics Using Data Mining Techniques, International Journal of Software & Hardware Research in Engineering Vol.2 Issue 2, February 2014 (ISSN No:2347-4890)

[2] J.P. Bigus. (1996),"Data Mining with Neural Networks", New York: McGraw- Hill.

[3] A survey of Knowledge Discovery and Data Mining process models the Knowledge Engineering Review, Vol. 21:1-68 2006, Cambridge University Press Printed in the United Kingdom.

[4] http://wwwcc.qub.ac.uk/tec/courses/datamining/stu_notes/dm_book_1.html

[5]. K. Srinivas , B. Kavitha Rani and Dr. A. Govrdhan, ―Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks International Journal on Computer Science and Engineering (2010).

[6]. Shweta Kharya, ―Using Data Mining Techniques For Diagnosis And Prognosis Of Cancer Disease, International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.2, No.2, April 2012.

[7] D.S. Kalana Mendis, Asoka S. Karunananda, U. Samarathunga:Framework for modelling of tacit knowledge -Case study using Ayurvedic domain

[8] Mendis, D.S.K. , Karunananda, A.S, Samaratunga, U.--An approach to developMulti Techniques Integrated Expert System For Diagnosis Of Human Constitutions, Information and Automation for Sustainability, 2008. ICIAFS 2008. 4th InternationalConference on12-14 Dec. 2008

[9] Mrs. Krantee Jamdaade, Mrs. Laika Satish -- Expert System For Diagnosing Hiv/Aids Using Ayurveda Therapy (Hivesat), National Conference on Recent Trends in Computer Application, 17th – 18th February 2010, Department of Management Science, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad.

[10] D.S. Kalana Mendis, Asoka S. Karunananda,Udaya Samaratunga, U. Rathnayake: Development of Commonsense Knowledge Modeling System for Psychological Assessment in Clinical Psycho, Journal of Health Science 2 (2014) 41-55

[11] G. Gnana Sekari, Manorama Srinath,G.Veluchamy : Expert System- An Innovative Tool For Managing Indian Traditional Medical Knowledge

[12] Premalatha Balachandrana, Rajgopal Govindarajan: Cancer—an ayurvedic perspective, Pharmacological Research 51 (2005) 19–30

[13] Jayarajan Kodikannath : Classical Ayurvedic approach to Cancer Care- Case Presentations, AAPNA's 8th International Conference on "Ayurveda & Cancer Care", October -2013

[14] Sushant Sud, Khyati Sud – Ayur Informatics: Need ofthe Hour, International Journal of Research in Ayurveda & Pharmacy, 4(2), Mar-Apr 2013

[15] Dimple : A Review on Data Mining Techniques Used in Healthcare Industry, International Journal in Multidisciplinary and Academic Research (SSIJMAR) Vol. 3, No. 1, February- March -2014 (ISSN 2278 – 5973)

[16] Dr M.Hemalatha S. Megala: mining techniques in health care: a survey of immunization, Journal of Theoretical and Applied Information Technology, 31 Mar 2011 Vol. 25 No 2, (ISSN: 1992-8645)

[17] M. Durairaj, V. Ranjani: Data Mining Applications in Healthcare Sector: A Study, international journal of scientific & technology research volume 2, issue 10, october 2013 ISSN 2277-8616

[18] American Medical Informatics Association, http://www.amia.org/informatics/.

[19] Elizabeth A Madign, Oliver Louis Curet, "A Data Mining Approach In Home Healthcare:Outcomes And Service Use" ,BMC Health Care Research.

[20] Curet, Jackson M, Tarar, " A Designing and Evaluating a case-based learning and resoning agent in unstructured decision making, Information Intelligence and Systems,pp2487-2492.

[21] Han,J. and M. Kamber.(2001), Data Mining concepts And Techniques. An Francisco, Morgan Kauffmann Publishers.

[22] Elma Kolçe (Çela), Neki Frasheri: A Literature Review of Data Mining Techniques Used in Healthcare Databases, CT Innovations 2012 Web Proceedings - Poster Session ISSN 1857- 7288

[23] J. C. Prather, et al., "Medical data mining: knowledge discovery in a clinical data warehouse," in AMIA Annual Fall Symposium 1997, pp. 101-105.

[24] Nidhi Bhatla, Kiran Jyoti,‖ An Analysis of Heart Disease Prediction using Different Data Mining Techniques‖ International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Vol. 1 Issue 8, October – 2012