Survey on Information Mining Procedures Utilized in Healthcare Services

S. Dhanalakshmi,
Research Scholar
Department of Computer Science
Thiruvalluvar Government Arts College
Rasipuram, Namakkal Dt, Tamilnadu

Dr. S. Sathiyabama
Assistant Professor
Department of Computer Science
Thiruvalluvar Government Arts College
Rasipuram, Namakkal Dt, Tamilnadu

Abstract: **In this paper, we have focused various techniques to resolve the issues in healthcare services. Purpose of data mining is to extract accurate information from huge set of information. In this paper contains various mining techniques and its procedures to identify and extract error free information from vast health related data like patient and their problems. We could able to identify the nature of problem and percentage of disease stage with clinical survey. Using various data mining techniques can able to analyse and produce survey of appropriate technique to find out results in healthcare services. Finally we get an idea to choose the data mining tool for making statistical survey in health care data.**

Keywords: Data Mining, Knowledge Discovery Database, In-Vitro Fertilization (IVF), Artificial Neural Network, Mining Tools

I. INTRODUCTION

The reason of data mining is to extract useful information from large databases or data warehouses. Data mining applications are used for commercial and scientific sides [1]. This study mainly talks about the Data Mining applications in the scientific area. Scientific data mining distinguishes itself in the sense that the nature of the datasets is often very different from traditional market driven data mining applications. Data mining algorithms applied in healthcare industry play a major role in guess and identification of the diseases. Large numbers of data mining applications are found in the medical related areas such as Disease identification, Medical Equipment industry, Pharmaceutical Industry and Hospital Management. Popularly data mining called knowledge discovery from the data. The knowledge discovery is an interactive process, consisting by developing an understanding of the application domain, selecting and creating a data set, preprocessing, data transformation. Data Mining has been used in a variety of applications such as marketing, customer relationship management, engineering, and medicine analysis, expert prediction, web mining and mobile and mobile computing.

Expanding the health coverage to as many people as possible, and providing financial assistance to help those with lower incomes purchase coverage [2]. Health administration or healthcare administration is the field relating to leadership, management, and administration of hospitals, hospital networks, and health care systems[1,3]. In the Healthcare sector Government spends more money.

Proposal in draft NHP 2001 is timely that State health expenditures be raised to 7% by 2015 and to 8% of State budgets thereafter .
Health spending in India at 6% of GDP is among the highest levels estimated for developing countries.
Public spending on health in India has itself declined after liberalization from 1.3% of GDP in 1990 to 0.9% in 1999. Central budget allocations for health have stagnated at 1.3%of the total Central budget. In the States it has declined from7.0% to 5.5% of the State health budget.
Data mining tools to answer the question that traditionally was a time consuming and too complex to resolve. They prepare databases for finding predictive information. Data mining tasks are Association Rule, Patterns, Classification and Prediction, Clustering. Most common modeling objectives are classification and prediction.

This paper mainly compares the data mining tools deal with the health care problems. The relative study compares the exactness level predicted by data mining applications in healthcare. Infertility is on the rise across the globe and it needs the sophisticated techniques and methodologies to predict the end results of infertility treatments particulars IVF (in-vitro fertilization) treatments, since the cost of IVF procedure is on the rise.

In this study, we have taken this issue and compare the diverse techniques of data mining applications for predicting the achievement rate of IVF treatment with the accuracy level. This comparative study could be useful for aspiring researchers in the field of data mining by knowing which data mining tool gives an accuracy level in extracting healthcare data.

II. LITERATURE REVIEW

A literature review is a text written from the current knowledge including theoretical and procedural contributions to a particular topic. Literature reviews are secondary sources and do not report any new or original experimental work.

HianChyeKoh and Gerald Tan mainly discusses data mining and its applications with major areas like Treatment effectiveness, Management of healthcare, Detection of fraud and abuse, Customer relationship management[1].

JayanthiRanjan presents how data mining discovers and extracts useful patterns of this large data to find

Special Issue - 2020

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICATCT – 2020 Conference Proceedings

observable patterns. This paper demonstrates the ability of Data mining in improving the quality of the decision making process in pharma industry. Issues in the pharma industry are adverse reactions to the drugs [2].

M. Durairaj, K. Meena illustrates a hybrid prediction system consists of Rough Set Theory (RST) and Artificial Neural Network (ANN) for dispensation medical data. The process of developing a new data mining technique and software to assist competent solutions for medical data analysis has been explained. Propose a hybrid tool that incorporates RST and ANN to make proficient data analysis and indicative predictions. The experiments onspermatological data set for predicting excellence of animal semen is carried out. The projected hybrid prediction system is applied for pre-processing of medical database and to train the ANN for production prediction. The prediction accuracy is observed by comparing observed and predicted cleavage rate[16].

K. Srinivas, B. Kavitha Rani and Dr. A. Goverdhan discusses mainly examine 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[4].

ShwetaKharya discussed 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[5].

Arvind Sharma and P.C. Gupta discussedData mining can contribute with important benefits to the blood bank sector. J48 algorithm and WEKA tool have been used for the complete research work. Classification rules performed well in the classification of blood donors, whose accuracy rate reached 89.9%[7].

III. DATA MINING CONCEPTS

Data Mining is all about discovering unsuspected/ previously unknown relationships amongst the data. Data mining is also called as Knowledge discovery, Knowledge extraction, data/pattern analysis, information harvesting, etc.

A. Definition

Data mining or knowledge discovery in database, as it is also known, is the non-trivial extraction of implicit, previously unknown and potentially useful information

This encompasses a number of technical approaches, such as clustering, data summarization, classification, finding dependency networks, analyzing changes, and detecting anomalies [8].
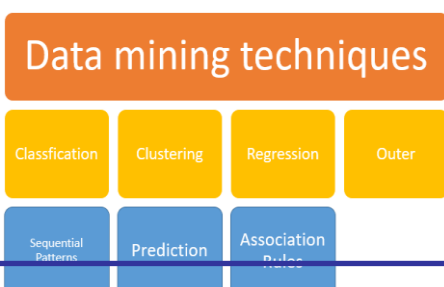
B. Development of data mining

The current evaluation of data mining functions and products is the results of influence from many disciplines, including databases, information retrieval, statistics, algorithms, and machine learning [9].
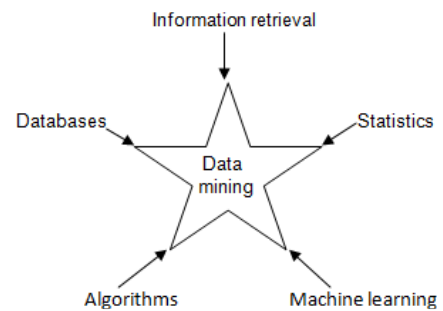

Fig. 1. Historical perspective of data mining

C. Data Mining Application Areas

The information or knowledge extracted so can be used for any of the following applications :
- Market Analysis
- Fraud Detection
- Customer Retention
- Production Control
- Science Exploration

D. Data Mining Tasks

Data mining tasks are mainly classified into two broad categories: 1. Predictive model 2. Descriptive model
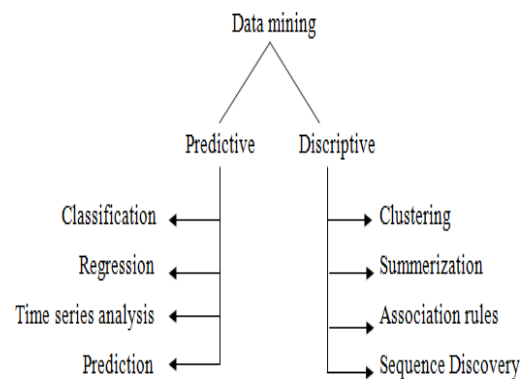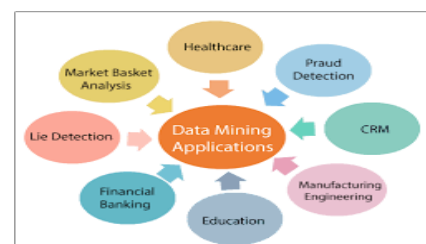

Fig 3.3 Data mining models and tasks

IV. DATA MINING APPLICATIONS

A. HEALTHCARE DIVISION

Today the healthcare industry produced large amounts of sensitive data about patients, hospital resources, disease diagnosis, electronic patient records,

medical devices etc. Data are the key resource to be processed and analyzed for knowledge mining that enables support for cost-savings and decision making. Data mining applications in healthcare can be grouped as the evaluation into broad categories[1,10],

1) Treatment efficiency & Administration

Data mining gives data for analyzing the types of treatment, causes, symptoms, and courses of treatments. Also predict which type of treatment is more suitable for particular disease.

Data mining applications developed to better identify and track chronic disease states and high-risk patients, design appropriate interventions, and reduce the number of hospital admissions and claims to aid healthcare management.

2) Customer relationship management

Customer relationship management is a main approach to managing communications between commercial organizations- typically banks and retailers- and their customers, it is no less important in a healthcare context. Customer interactions may take place through call centers, physicians' offices, billing departments, inpatient settings and emergency care settings.

3) Fraud and abuse

Detect fraud and abuses establish norms and then identify unusual or abnormal patterns of claims by physicians, clinics, or others attempt in data mining applications. Data mining applications fraud and abuse applications can highlight inappropriate prescriptions or referrals and fraudulent insurance and medical claims.

4) Medical apparatus manufacturing

Healthcare system's one significant point is medical device. Mobile communications and low-cost of wireless bio- sensors have covered the way for development of mobile healthcare applications that supply a convenient, safe and constant way of monitoring of vital signs of patients[11].

Everywhere Data Stream Mining (UDM) techniques such as light weight, one-pass data stream mining algorithms can perform real-time analysis on-board small/mobile devices while considering available resources such as battery charge and available memory.

5) Pharmaceutical Industry

The technology is used to help the pharmaceutical firms manage their inventory and to develop new product and services. A deep understanding of the information secret in the health data is very important to a firm's competitive position and organizational decision-making.

6) Hospital administration

Organizations including multispecialty hospitals are capable of generating and collecting a vast amount of data. Application which is used to maintain administration details are very sensitive. Three layers of hospital management:

- Services for hospital management
- Services for medical staff
- Services for patients

V. RESULTS OF COMPARATIVE STUDY

In this chapter gives comparative study of data mining tools and its analysis report based on usage of various tools, for particular disease.

- Diabetes Mellitus
- Kidney dialysis
- Dengue
- IVF
- Hepatitis C
- Heart Disease
- Cancer
- HIV/AIDS
- Blood
- Brain Cancer
- Tuberculosis

In the Table1, the recent healthcare issues specifically in disease side and research results have been listed. The diseases are the most serious problems in human.

To analyze the effectiveness of the data mining applications for diagnosing the disease, the traditional methods of mathematical / statistical applications are also given and compared. Eleven problems are taken for evaluation.

The following table list out the various health related issues and appropriate data mining techniques to produce statistical data on that particular disease. This table contains data mining techniques, tool, traditional method for analyzing, algorithm involved and accuracy level after using data mining application are described.

Table 1. Data mining applications in healthcare

S.No	Type of disease	Data mining tool	Technique	Algorithm	Traditional Method	Accuracy level(%) from DM application
1	Heart Disease	ODND, NCC2	Classification	Naïve	Probability	60
2	Cancer	WEKA	Classification	Rules. Decision Table		97.77
3	HIV/AIDS	WEKA 3.6	Classification, Association Rule Mining	J48	Statistics	81.8
4	Blood Bank Sector	WEKA	Classification	J48		89.9
5	Brain Cancer	K-means Clustering	Clustering	MAFIA		85
6	Tuberculosis	WEKA	Naïve Bayes Classifier	KNN	Probability, Statistics	78
7	Diabetes Mellitus	ANN	Classification	C4.5 algorithm	Neural Network	82.6
8	Kidney dialysis	RST	Classification	Decision Making	Statistics	75.97
9	Dengue	SPSS Modeler		C5.0	Statistics	80
10	IVF	ANN, RST	Classification			91
11	Hepatitis C	SNP	Information Gain	Decision rule		73.20

Bar chart created by using this table with the values of health care problems, Accuracy Level of mining tool is illustrated in Fig. 2. In this chart, the accuracy level of different data mining applications has been compared.
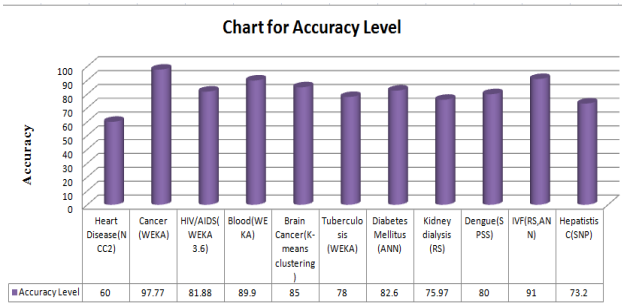

Fig. 4. Chart for Accuracy Level of using Data mining tools for diagnosis

A. Comparative Study Of Ivf Success Rate Prediction

This section deals with the comparative study of three different data mining theories for predicting the success rate of IVF treatment. The process of data mining applications, its advantages and results obtained are compared.

The detailed study of selected works gives a broad idea about the application of data mining techniques. This study mainly compares the three different data mining applications carried out on the prediction of the IVF treatment success rate.

B. Application of rough set theory for medical informatics data analysis

The research work aims to analyze the medical data by applying Rough Set Theory of data mining approach. The data reduction process has been done using rough set theory reduction algorithm. Rough set is mainly used to reduce the attributes without compromising its knowledge of the original.

To analyze the fertilization data, ROSETTA tool kit reduction algorithm is used in this work to produce the optimal result set without affecting the original knowledge. The treatment success rate is predicted and tabulated as depicted in Table 2.

		Predicted		
		SUCCESS	UN SUCCESS	
Actual	SUCCESS	17	4	0.80952
	UN SUCCESS	26	10	0.27777
		0.395349	0.714286	0.47368

The actual and desired outputs are compared with each other. It also depicts that the success rate obtained after reducing the number of attributes is 47%.

1) Artificial neural network in classification and prediction

This research work is mainly aimed to predict and classify the IVF treatment results using Artificial Neural Network (ANN).

Performance	DESIRED OUTPUT	ACTUAL NETWORK OUTPUT
MSE	0.209522132	0.212860733
NMSE	1.164459543	1.18301446
MAE	0.23114814	0.25780224
Min Abs Error	9.90854E-07	6.66044E-06
Max Abs Error	1.015785003	0.998857054
R	0.498099362	0.498099362
Percent Correct	73.07692308	75

Table 3. IVF success rate predicted by ANN

The artificial neural network is constructed with multi-layer perception and back-propagation training algorithm, and constructed network is trained, tested and validated using patients' sample IVF data. This work finally compares the success rate between desired output which is field recorded data and actual output which is predicted output of neural network. In the Table 3, the comparison between desired and actual output of the neural network is illustrated.

This work finds the actual output using patients' IVF data by applying Artificial Neural Network. By comparing success rate, desired and actual output, the result obtained has a prediction accuracy of 73%.

2) Modeling an integrated methodology of neural networks and rough sets for analyzing medical data

This work is mainly aimed to develop a combined prediction system for analyzing medical data using Artificial Neural Network and Rough Set Theory. Two kinds of rules Deterministic and Non-deterministic are effected in the application of Rough set tool. For the rough set application, the software tool Neuro solution is used to predict the result. The performance of the combined technique of Artificial neural network and rough set theory is described in the Table 4.

Performance	Unsuccess of treatment (0)	Success of treatment (1)
MSE	0.092835478	0.110601021
NMSE	0.378803726	0.451293836
MAE	0.14313612	0.191653959
Min Abs Error	0.002563409	0.005851654
Max Abs Error	1.055555499	1.055555556
R	0.789058201	0.789058201
Percent Correct	89.23076923	91.83673469

Table 4. Performance of IVF success Rate prediction using hybrid technique

The prediction accuracy of this hybrid approach of combined use of ANN and RST is around 90%. These comparison results of three different data mining applications for predicting the success rate of IVF treatments are shown in Table 5 and Fig. 5.

	Rough Set	ANN	Rough Set & ANN (Hybrid)
Percentage of Accuracy in Estimating Success	47	73	90

Table 5. Comparison analysis

Special Issue - 2020

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICATCT – 2020 Conference Proceedings

The application of combined Rough Set and Artificial Neural Network yields better result when compared with other techniques. It is observed that the hybrid technique of combined use of two or more machine learning tool yields better results than the use of a single technique for mining information from the database.
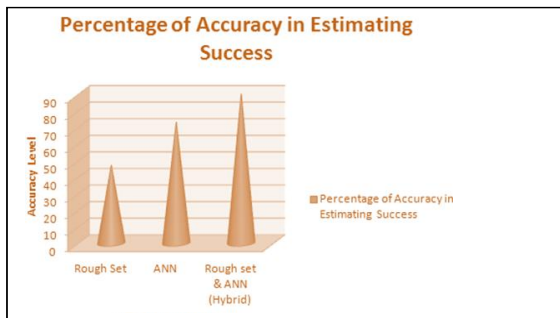

Fig. 5. The Success rate of Rough Set, ANN and Hybrid Technique

VI. CONCLUSION

This paper designed to compare the different data mining application in the healthcare division for extracting useful information. Exploring knowledge from the medical data is such a risk task as the data found are noisy, irrelevant and huge too. In this scenario, data mining tools come in practical in exploring of knowledge of the medical data and it is quite interesting. It is experimental from this study that a combination of more than one data mining techniques than a single technique for diagnosing or predicting diseases in healthcare sector could yield more promising results. The comparison study shows the interesting results that data mining techniques in all the health care applications give a more encouraging level of accuracy like 97.77% for cancer prediction and around 70% for estimating the success rate of IVF treatment.

REFERENCES

[1] HianChyeKoh and Gerald Tan―Data Mining Applications in Healthcare, journal of Healthcare Information Management– Vol19,No 2.

[2] Jayanthi Ranjan,― Applications of data mining techniques in pharmaceutical industry, Journal of Theoretical and Applied Technology, (2007).

[3] RubanD.Canlas Jr., MSIT., MBA , ― Data mining in Healthcare: Current applications and issues.

[4] 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).

[5] ShwetaKharya, ―Using Data Mining Techniques ForDiagnosis And Prognosis Of Cancer Disease, International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.2, No.2, April 2012.

[6] EliasLemuye, ― Hiv Status Predictive Modeling Using Data Mining Technology.

[7] Arvind Sharma and P.C. Gupta ― Predicting the Number of Blood Donors through their Age and Blood Group by using Data Mining Tool, International Journal of Communication and Computer Technologies Volume 01 – No.6, Issue: 02 September 2012.

[8] Arun K Punjari, ― Data Mining Techniques, Universities (India) Press Private Limited, 2006.

[9] Margaret H.Dunham, ―Data Mining Introductory and Advanced Topics, Pearson Education (Singapore) Pte.Ltd., India. 2005.

[10] PrasannaDesikan, Kuo-WeiHsu, JaideepSrivastava, ―Data Mining For Healthcare Management, 2011SIAM International Conference on Data Mining, April, 2011.

[11] ShusakuTsumoto and Shoji Hirano, ―Temporal Data Mining in Hospital Information Systems.

[12] David Page and Mark Craven, ―Biological Applications of MultiRelationalData Mining.

[13] N. AdityaSundar, P. PushpaLatha and M. Rama Chandra, ―Performance Analysis of Classification Data Mining Techniques Over Heart Disease Data Basel, International Journal of Engineering Science & Advanced Technology, (2012).

[14] HardikManiya, Mosin I. Hasan and Komal P. Patel, ―Comparative study of Naïve Bayes Classifier and KNN for Tuberculosis, International Conference on Web Services Computing (ICWSC) 2011 Proceedings published by International Journal of Computer Applications® (IJCA).

[15] B.Renuka Devi, Dr.K.NageswaraRao, Dr.S.PallamSetty and Dr.M.NagabhushanaRao, Disaster Prediction System Using IBM SPSS Data Mining Tool, International Journal of Engineering Trends and Technology (IJETT) - Volume4 Issue8- August 2013ISSN: 2231.

[16] M. Durairaj, Ranjani, Data Mining Applications In Healthcare Sector: A. Study INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 2, ISSUE 10, OCTOBER 2013.