

## Future Perception In Public Health Care Using Data Mining

Dr. Surendra Kumar Yadav

Associate professor, JECRC  
University, Jaipur

Nitesh Dugar

Research scholar in Software  
Engineering, JECRC, Jaipur

Aditi Jain

Research scholar in Software  
Engineering, JECRC, Jaipur

### Abstract

Healthcare information is diverse in scope and huge in content and its volume is so vast that traditional / routine analytical methods reveal very little of the possible conclusions. Modern data mining techniques can be applied to this data to extract otherwise hidden / unknown facets of knowledge which may be of vital importance to therapeutic, commercial and preventive aspects of healthcare. This research paper provides a survey of mining concepts in health-care, necessity of data-mining in Medicare field, algorithms used and its applications in various health care domains.

### 1. Introduction

Data mining is the exploration of large datasets to extract hidden and previously unknown patterns, relationships and knowledge that are difficult to detect with traditional statistics [1]. So it extracts helpful knowledge from huge datasets and presents it in human understandable form. Data-mining divide its task into two parts Predictive and Descriptive tasks. Predictive tasks predict the value of specific attribute and descriptive tasks drive patterns that summarize the relation between data. Classification, Regression and Deviation Deduction come under Predictive Tasks. Descriptive Tasks derive pattern that summarize the relationship between data [2]. Data mining expertise provides a consumer leaning approach to new and unknown pattern in the data [3].

With the application of data mining tools in spreadsheet of the program that analyzes data to identify patterns and relations, user profiling and development of business strategies can be started [4]. Health-Care data mining involves few steps from collecting patient's raw data and discovering new knowledge out of it.

To discover knowledge out of data there is an iterative process consists of data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation, and knowledge representation. The Fig.1 shows how Data Mining extracts Knowledge from a given Dataset.

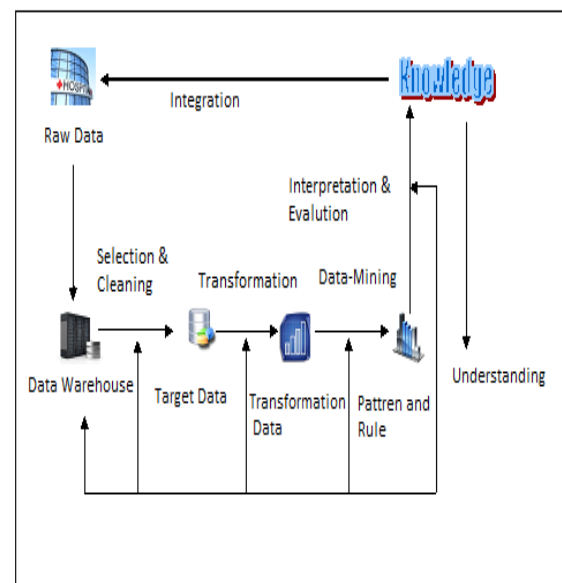


Figure 1. Knowledge discovery process

In recent years, a medical department has been frustrated by the problems of overloading, long processing time, delay patient treatment and high cost.

These problems have been caused from several internal and external factors, including patient characteristics, staffing patterns of medical department, access to health care providers, patient arrival time, management practices, and testing and treatment strategies selected by medical department [5]. The first step is to understand these issues, to improve the patient care in hospital departments. Today most of the hospitals have Hospital Information Systems (HIS) to manage patient data. These systems generate huge amount of data in the form of image, charts, texts, and numbers. The researchers in the medical field identify and predict the disease with the aid of data mining techniques [6].

## 2. Basic Mining Concepts in Health-Care

### 2.1. Definition

In medical research, data mining starts with a hypothesis and then the results are adjusted to fit the hypothesis [7]. Hence, data mining describe patterns, but do not explain trends in pattern.

### 2.2. Necessity of Data Mining in Health-Care

With the help of data mining we can improve the public health, the health care of the system users, reduce costs, save time and money [4]. For preparing Health Care Information System (HMIS) reports for usage of hospital capacities such as number of occupied and vacant beds, number of patient vs doctors and nurses etc. Combining data mining and geographical information system (GIS) we can discover disease clusters towards specific location which leads to better policy making to detect and manage disease outbreaks.

Knowledge discovery (KDD) and data mining Combined to discover fraud in credit cards and insurance claim from a huge database. Therefore data mining allows organizations and institutions to get more out of existing data at minimal extra cost [7]. Adverse drug events, some drugs and chemicals that have been approved as non-harmful to humans are

later discovered to have harmful effects after long term publicize, but data mining algorithms such as Multi-item Gamma Poisson Shrinker (MGPS) can discover drug side effects [8]. The basic objective of data mining in healthcare is decrease mortality rate, early detection of diseases, to help in diagnosis and detecting fraudulent health claims.

### 2.3. Basic Health-care Data Mining Tasks

Initially Data is extracted, transform and load from the transaction database into Data-Warehouse as shown in fig2. After extracting data into electronic machine following tasks will be performed

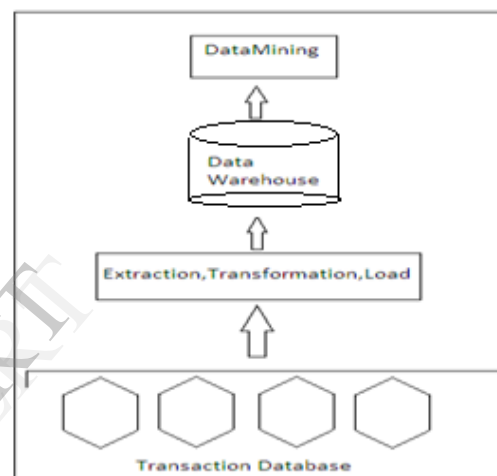


Figure 2. Basic concept of data mining

#### Health-Care Business Understanding:

This Phase concentrates on understanding health-care project objective and various requirements.

#### Medical-Data Understanding:

This phase starts with collection of health care data from Laboratories, Operation Theater, Blood Bank, drug store, Therapy Modules etc, and also focuses on understanding of patient's data to discover knowledge out of it to generate HMIS reports.

**Medical-Data Preparation:**

This phase constructs final data set to feed into modelling tools and it is iterative process. Here various database artifacts such as attribute, table, records are selected as well as transformed and cleaned for modelling tools.

**Modelling:**

This phase apply various modelling techniques such as Naive Bays, Artificial neural network, decision tree, time series algorithm, clustering algorithm, sequence clustering algorithm etc to generate optimal values.

**Evaluation:**

In this stage thorough evaluation and reviewing the model to check whether applied algorithms discovers proper hidden pattern. And this stage also checks for fast accessing on mined data.

**Deployment:**

In this phase most of the time customer carry out the deployment wizard with the help of analyst to generate HMIS reports. For generating useful knowledge out of data i.e. reports we can repeat data mining process.

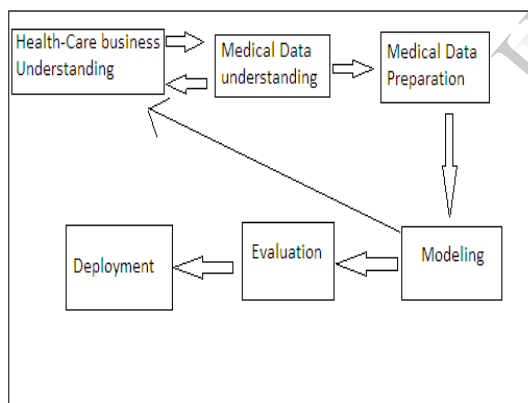


Figure 3. Health care business cycle

### 3. Common Data Mining Algorithm in Health-care Domain

#### 3.1. Decision Tree Algorithm

Decision tree is a flow chart like tree structure [9] which represents rules, where each node denotes a test on an attribute value, branch denotes test output, and leaves represent classes [10]. It recursively partitions a

data set of records using depth-first greedy approach until all the data items belong to a particular class are identified[11]. There are two phases in decision tree algorithm: tree building and tree pruning. Tree building uses top down approach for partitioning the data sets and bottom up approach is used in tree pruning to improve prediction and classification accuracy of the algorithm by minimizing the over fit[12]. It is easy to extract display rules, has smaller computation amount and could display important decision property and own higher classification precision [13]. It can be used in predictive modeling for both discrete and continuous attributes.

#### Application

- Adverse Drug Reactions [14]
- Breast Feeding [15]
- Clinical Practice In Psychiatry [16]
- Diagnosis Of Myocardial Infarction [17]
- Literature Review Of meta-analysis[18]
- Measles Detection [19]
- Orthopaedic (Fracture Data) [20]
- To Detect Patterns Across Multiple Physiologic Data Signals [21]
- To Examine Mental Health Care [22]

#### 3.2. Artificial Neural Network

Neural network is a parallel processing network which is based on biological neural network and features of neurons (computing elements), generated with simulating the image intuitive thinking of humans[23]. Neuron receives a number of input signals and performs a simple operation on this set of inputs. The output of each neuron is fanned out to the input of other neuron [24] It can be used to model complex relationships between inputs and outputs or to find patterns in data [25]. It uses the idea of non-linear mapping, parallel processing and neural network structure to express the associated knowledge of inputs and outputs [26]. Two approaches of using neural network in data mining are self organization neural network, which is learning without teachers and fuzzy neural network is used to increase output systems capacity and to make system more stable. Neural networks can be used for both supervised and

unsupervised learning applications. Neural networks have high acceptance ability for noisy data and high accuracy and are preferable in data mining [27].

#### Applications

- Anaesthesia[28]
- Analysis of side drug effects [29]
- Breast cancer [30]
- Cervical cancer [31]
- Classification of Blood Cells [32]
- Health-Care image processing [33]
- Statistical Models for Breast Cancer [34]
- To analysis ECG Waves Forms [35]
- To Classify Retina damage [36]
- Tumours [37]

### 3.3. Genetic Algorithm

Genetic algorithm performs global search based on greedy approach and performs with less time complexity. It is an adaptive heuristic search algorithm premised on the evolutionary ideas of natural selection and genetic [38]. It uses a number of artificial individuals looking through a complex search space by using functions of selection, crossover and mutation [39]. This algorithm creates a number of random solutions out of which only few solutions are optimal. Other poor solutions are discarded. Good Solutions are then hybridized and same process will be repeated again to get only best solutions. All best solutions are then combine together to obtain a universal solution.

### 3.4. Nearest Neighbor method

K nearest neighbor algorithm is used for classification in which a training data set is given. This algorithm performs search operation on training data of k nearest neighbor of the given input within a certain distance and then finds for the best fit class for new training data. It can be a costlier algorithm when training data is extremely large because whole data set is scanned to find out the best nearest neighbor. There is no learning process to create a model for this

algorithm. The data used in algorithm is itself a model [4].

## 4. References

- [1] Liangxiao J. et.al. "One Dependency Augmented Naïve Bayes", *In Proc. 1st Int'l Conf. Advance Data Mining and applications (ADMA)*, pp.186-194 (2005).
- [2] Durairaj M. et.al. "An Empirical Study on applying Data Mining Techniques for the Analysis and Prediction of Heart Disease", *Int'l conf. on Information Communication and Embedded Systems (ICICES)*, IEEE, Chennai, pp.265-270, (2012).
- [3] Hemalatha M. and Megala S., "Mining Techniques in Health Care: A Survey of Immunization", *Little Lion Scientific R&D*, **25(2)**, pp.63-70, (2012).
- [4] Milovic Boris and Milovic Milan " Prediction and Decision Making in Health Care using Data Mining", *International Journal of Public Health Science (IJPHS)*, **1(2)**, pp. 69-78, (2012).
- [5] Fromm RE et.al. "Critical care in the emergency department: a time-based study", *Crit. Care Med.*, **21(7)**, pp.970-976, (1993).
- [6] Setiawan N.A. "Rule Selection for Coronary Artery Disease Diagnosis Based on Rough Set", *International Journal of Recent Trends in Engineering*, **2(5)**, pp.198-202, (2009).
- [7] Kuo-Chung Lin and Ching-Long Yeh "Use Of Data-Mining Techniques To Detect Medical Fraud In Health Insurance" *International Journal Of Engineering And Technology Innovation*, **2(2)**, pp. 126-137, (2012).
- [8] Ruben D. and Canlas Jr. "Data Mining in Healthcare: Current Applications and Issues", (2009)
- [9] Changala R. et.al. "Classification by Decision Tree Induction Algorithm to Learn Decision Trees from the class-labelled Training Tuples", *International Journal of Advanced Research in Computer Science and Software Engineering*, **2(4)**, pp. 427-434, (2012).
- [10] Gharehchopogh F.S. et.al. "Application of Decision Tree Algorithm for Data Mining in Healthcare Operations: A Case Study", *International Journal of Computer Applications*, **52(6)**, pp.21-26, (2012).
- [11] Venkadatri M. And Lokanatha C. "A comparative study on decision tree classification algorithm in data mining", *International Journal of Computer Applications in Engineering, Technology and Science*, **2(2)**, pp. 24-29, (2010).
- [12] Andrew C. "Building Decision tree with the ID3 Algorithm", *Dr. Dobbs Journal*, (1996).

- [13] Linna Li et.al, "Study of data mining algorithm based on decision tree", *Int'l Conf. on Computer design and applications (ICCCA)*, IEEE, **vol.1**, pp. 155-158, (2010).
- [14] Jones J.K. "The role of data mining technology in the identification of signals of possible adverse drug reactions: Value and limitations, *Current Therapeutic Research Clinical and Experimental*", **62(9)**, pp. 664-672, (2001).
- [15] .Babic S.H. et.al. "Fuzzy decision trees in the support of breastfeeding", in *Proc. of 13th IEEE Symposium on Computer-Based Medical Systems CBMS*, pp. 7-11, (2000).
- [16] Dantchev N. "Therapeutic decision trees in psychiatry", *Encephale-Revue De Psychiatrie Clinique Biologique et Therapeutique*, **22(3)**, pp. 205-214, (1996).
- [17] Tsien C.L. et.al. "Using classification tree and logistic regression methods to diagnose myocardial infarction", in *Proc. of the 9th World Congress on Medical Informatics MEDINFO*, **vol.52**, pp. 493-497, (1998).
- [18] Gambhir S.S. "Decision analysis in nuclear medicine", *Journal Of Nuclear Medicine*, **40(9)**, pp. 1570-1581, (1999).
- [19] Ohno-Machado L. et.al. "Decision trees and fuzzy logic: A comparison of models for the selection of measles vaccination strategies in Brazil", *Journal of the American Medical Informatics Association:Suppl.*, pp. 625-629, (2000).
- [20] Kokol P. et.al. "The limitations of decision trees and automatic learning in real world medical decision making", in *Proc. of the 9th World Congress on Medical Informatics MEDINFO*, **vol. 52**, pp. 529-533, (1998).
- [21] Tsien, C.L. et.al. "Using classification tree and logistic regression methods to diagnose myocardial infarction", in *Proc. of the 9th World Congress on Medical Informatics MEDINFO*, **vol.52**, pp. 493-497, (1998).
- [22] Bonner G. "Decision making for health care professionals: use of decision trees within the community mental health setting", *Journal of Advanced Nursing*, **vol. 35**, pp. 349-356, (2001).
- [23] Sonalkadu. "Effective Data Mining through Neural Network", *International Journal of Advanced Research in Computer Science and Software Engineering*, **2(3)**, pp. 441-444, (2012).
- [24] Fausett and Laurene. "Fundamentals of Neural Networks: Architectures, Algorithms and Applications", Pearson First Edition, USA, (2002).
- [25] Singh Y. And Chauhan A. Singh. "Neural networks in data mining", *Journal of Theoretical and Applied Information Technology*, **5(1)**, pp. 37-42, (2009).
- [26] Xianjun Ni. "Research of Data Mining Based on Neural Networks", *World academy of science, engineering and technology (WASET)*, **vol.15**, (2008).
- [27] Nirghi S. "Potential use of Artificial Neural Network in Data Mining", *2<sup>nd</sup> Int'l Conf. on Computer and automation engineering(ICCAE)* IEEE, **vol.2**, pp. 339-343, (2010).
- [28] Sharma A. and R.J. Roy. "Design of a Recognition System to Predict Movement during Anesthesia" *IEEE Transactions*, (1997).
- [29] Domine D. et.al. "Non linear neural mapping analysis of the adverse effects of drugs", (1998).
- [30] Einstein A.J. et.al. "Fractal characterization of chromatin appearance for diagnosis in breast cytology", (1998).
- [31] Romeo M.et.al. "Infrared Microspectroscopy and Artificial Neural Networks in the Diagnosis of Cervical Cancer.", U.S. National Library of Medicine National Institutes of Health, **44(1)**, pp179-87, (1998).
- [32] Ifeachor E. and Rosen K. Eds. *International Conference on Neural Networks and Expert Systems in Medicine and Healthcare*, world scientific, pp. 248—256
- [33] Aizenberg I. et.al. "Cellular Neural Networks and Computational Intelligence in Medical Image Processing", *Image and Vision Computing*, **19(4)**, pp.177-183, (2001).
- [34] Burke H. et.al. "Comparing the Prediction Accuracy of Artificial Neural Networks and Other Statistical Models for Breast Cancer Survival", *Advances in Neural Information Processing Systems*, **Vol. 7**, pp.1063—1067, (1995).
- [35] Waltrus R.L. et.al. "Synthesize, Optimize, Analyze, Repeat (SOAR): Application of Neural Network Tools to ECG Patient Monitoring", <http://www.emsl.pnl.gov:2080/proj/neuron/workshops/WEEANN95/talks/watrous.weeann95.abs.html>
- [36] Aleynikov S. and Micheli-Tzanakou. "Classification of retinal damage by a neural network based system", (1998).
- [37] Ball G. et.al. "An Integrated Approach Utilizing Artificial Neural Networks and SELDI Mass Spectrometry for the Classification of Human Tumors and Rapid Identification of Potential Biomarkers Bioinformatics", **Vol.18**, pp 395-404, (2002).
- [38] Ghosh S. "Mining Frequent Item sets Using Genetic Algorithm", *International Journal of Artificial Intelligence & Applications (IJAI)*, **1(4)**, pp. 133-143, (2010).
- [39] Vishwakarma P. et.al. "Data Mining Using Genetic Algorithm (DMUGA)", *International Journal of Engineering Research and Development*, **5(2)**, pp. 88-94, (2012).