# Ubiquitous Data Access For Medical Services

**M.Kanimozhi[1], Dr.D.J.Evanjaline[2]**
[1, 2] Department of Computer Science
[1, 2] Rajah Serfoji Government college, Thanjavur

***Abstract-*** *Data mining is defined as sifting through very large amounts of data for useful information. Some of the most important and popular data mining techniques are association rules, classification, clustering, prediction and sequential patterns. Data mining techniques are used for variety of applications. Data mining holds great potential for the healthcare industry to enable health systems to systematically use data and analytics to identify inefficiencies and best practices that improve care and reduce costs and also data mining plays an important role for predicting diseases. For detecting a disease number of tests should be required from the patient. But using data mining technique the number of test should be reduced. This reduced test plays an important role in time and performance. This technique has an advantages and disadvantages. This research paper analyzes how data mining techniques are used for predicting different types of diseases. This paper reviewed the research papers which mainly concentrated on predicting heart disease, Diabetes and so on.*

## I. INTRODUCTION

Data mining is one of the most important steps of KDD process and it is the process of extracting hidden information from large database and transform it into understandable format by considering different perspectives.[4]

- It is the computer-assisted process of digging through and analysing enormous sets of data and then extracting the meaning of the data.
- Data mining explores data and analyses large observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner.
- It uses information from past data to analyse the outcome of a particular problem or situation that may occur in future.

We are concentrating on Healthcare databases, which have a huge amount of data but however, there is a lack of effective analysis tools to discover the hidden knowledge. In this survey we present an overview of the current research being carried out using the DM techniques for the diagnosis and prognosis of various diseases, highlighting critical issues and summarizing the approaches in a set of learned lessons. Data Mining used in the field of medical application can exploit the hidden patterns present in voluminous medical data which otherwise is left undiscovered. Data mining techniques which are applied to medical data include association rule mining for finding frequent patterns, prediction and classification. Traditionally data mining techniques were used in various domains. However, it is introduced relatively late into the Healthcare domain.[1]

## II. DATA MINING TECHNIQUES

Data Mining Tasks can be classified into :

- ✓ Supervised learning
- ✓ Unsupervised learning

The distinction is based on how learner classifies the data, if the classification takes under supervision then it is supervised learning else if the classes are not known then they are unsupervised learning. In supervised learning the classes are predetermined, the inputs are either assumed or known at the beginning. In unsupervised learning the classes are not predetermined, the inputs are not known at the beginning and it is not carried out under supervision. The model is not provided with input data ,so it can be used to form clusters based on the statistical properties or similarities among the values Supervised models: Classification, etc. Unsupervised Models: Clustering, etc[11].

**Major Issues in Medical Data Mining**

- ✓ Heterogeneity of medical data
  - Volume and complexity
  - Physician's interpretation
  - Poor mathematical categorization
  - Canonical Form
  - Solution: Standard vocabularies, interfaces between different sources of data integrations, design of electronic patient records
- ✓ Ethical, Legal and Social Issues
  - Data Ownership
  - Lawsuits
  - Privacy and Security of Human Data
  - Expected benefits
  - Administrative Issues

The most frequently used Data Mining techniques are specified below

**Classification learning:-** The learning algorithm takes a set of classified examples (training set) and use it for training the algorithms. With the trained algorithms, classification of the test data takes place based on the patterns and rules extracted from the training set.

**Numeric predication:-** This is a variant of classification learning with the exception that instead of predicting the discrete class the outcome is a numeric value.

**Association rule mining:-** The association and patterns between the various attributes are extracted and from these attributes rules are created. The rules and patterns are used predicting the categories or classification of the test data.

**Clustering: -** The grouping of similar instances in to clusters takes place. The challenges or drawbacks considering this type of machine learning is that we have to first identify clusters and assign a new instance to these clusters[8].

Out of these four types of learning methods we need to identify the algorithm which performs better. The application of data mining techniques depends on the types of data which is fitted to be used in the techniques, and solving data mining problems depend on the types of data to be used and the selection of data mining technique which is most suitable for the data.

Mining association rules : The purpose of mining association rules is to find out the associations between items from a huge transactions. It is usually applied to transactional data forms. But it is also suitable for data stored in relational data forms.

Classification: The basic purpose is to find out the classification principle from a pre-classified data set (training data). This principle can be used to classify the newly coming data. ID3, CART,C4.5, and neural network are all popular classification methods.

There are numerous data mining tools and methods available today. In this survey, we will define the new approach of using hybrid Data Mining system for medical database[11].

**Data Integrity**

Data integrity refers to the overall completeness, accuracy and consistency of data. This can be indicated by the absence of alteration between two instances or between two updates of a data record, meaning data is intact and unchanged. Data integrity is usually imposed during the database design phase through the use of standard procedures and rules. Data integrity can be maintained through the use of various error checking methods and validation procedures.

**Random Sampling**

A random “sampling” checking greatly reduces the workload of services. Thus, a probabilistic automatic on sampling checking is preferable to realize the secret key manner, as well as to rationally allocate resources and non-repeat keywords. An efficient algorithm is used to since the single sampling checking may overlook a small number of data abnormalities.

**Entropy**

The entropy of a random variable is a function which attempts to characterize the “unpredictability” of a random variable. Consider a random variable X representing the number that comes up on a roulette wheel and a random variable Y representing the number that comes up on a fair 6-sided die. The entropy of X is greater than the entropy of Y. In addition to the numbers 1 through 6, the values on the roulette wheel can take on the values 7 through 36. In some sense, it is less predictable. But entropy is not just about the number of possible outcomes.

**Infrequent Data**

The outlier detection in this type of data set is more challenging since there is no inherent measurement of data mining for discovering novel or rare events. The infrequent data is especially challenging because of the difficulty of defining a rare data for categorical data or combination of relevant and irrelevant data. Outlier detection can be implemented as a preprocessing step prior to the application of an identifying the physical significance of an object.

**Mutual Information**

The **mutual information** or formerly of two random variables is a measure of the mutual dependence of the two random variables combine entropy and total correlation with attributes.

The mutual information disorder of a data set and the total correlation measures the attribute relationship between questions and answers. Based on this concept, we build a formal model of outlier detection and propose a criterion for

estimating the "goodness" of a subset of objects as potential outlier online treatments.

**Real Data Sets**

The results indicate that our proposed factor for watermarking better reflects the intuitive understanding of the data set is not changed. Specifically possible to adapt ITB-SS and ITB-SP to continuous attributes either through data sets. These results are evidence of the importance of large-scale categorical data sets. The effectiveness of our algorithms results from a new concept of weighted greedy approaches that considers both the data distribution and attribute correlation for real data sets.

## III. CONCLUSION

Data mining can be beneficial in the field of medical domain .However privacy, security and misuse of information are the big problems if they are not addressed and resolved properly. This survey describes about the proposal of hybrid data mining model to extract classification knowledge for aid of various disease in clinical decision system and presents a framework of the tool various tools used for analysis .It describes about the advantages and various issues faced by data mining technique and a description of various algorithms that has been applied in the field of medical diagnosis. The accuracy of the algorithms can be enhanced by hybridizing or combining algorithms or their most prevalent features, as a single algorithm may not be accurate for weakly classified sets of data

## REFERENCES

[1] X. D. Wu, M. Q. Ye, D. H. Hu, G. Q. Wu, X. G. Hu, and H. Wang, "Pervasive medical information management and services: Key techniques and challenges," Chin. J. Comput., vol. 35, no. 5, pp. 827–845, May 2012.

[2] R. L. Richesson and J. Krischer, "Data standards in clinical research: Gaps, overlaps, challenges and future directions," J Amer. Med. Informat. Assoc., vol. 14, no. 6, pp. 687–696, 2007.

[3] L. Wang, G.-Z. Yang, J. Huang, J. Zhang, L. Yu, Z. Nie et al., "A wireless biomedical signal interface system-on-chip for body sensor networks," IEEE Trans. Biomed. Circuits Syst., vol. 4, no. 2, pp. 112–117, Apr. 2010.

[4] R. Agarwal and S. Sonkusale, "Input-feature correlated asynchronous analog to information converter for ECG monitoring,"IEEETrans.Biomed. Circuits Syst., vol. 5, no. 5, pp. 459–468, Oct. 2011.

[5] Dohr, R. Modre-Osprian, M. Drobics, D. Hayn, and G. Schreier, "The Internet of things for ambient assisted living," in Proc. 7th Int. Conf. Inf. Technol., New Gener., 2010, pp. 804–809.

[6] O. S. Adewale, "An internet-based telemedicine system in Nigeria," Int. J. Inf. Manag., vol. 24, no. 3, pp. 221–234, Jun. 2004.

[7] R. S. H. Istepanaian and Y.-T. Zhang, "Guest editorial introduction to the special section: 4 G health—The long-term evolution of m-health," IEEE Trans. Inf. Tech. Biomed., vol. 16, no. 1, pp. 1–5, Jan. 2012.

[8] R. Kyusakov, J. Eliasson, J. Delsing, J. V. Deventer, and J. Gustafsson, "Integration of wireless sensor and actuator nodes with IT infrastructure using service-oriented architecture," IEEE Trans. Ind. Informat., vol. 9, no. 1, pp. 43–51, Feb. 2013.

[9] K. Wang, X. Bai, J. Li, and C. Ding, "A service-based framework for pharmacogenomics data integration," Enterp. Inf. Syst., vol. 4, no. 3, pp. 225–245, 2010.

[10] Pereira,B.Andersson,andE.Tovar,"WiDom:Adominance protocolfor wireless medium access," IEEE Trans. Ind. Informat., vol. 3, no. 2, pp. 120–130, May 2007.

[11] Parvathi I, Siddharth Rautaray ,"Survey on Data Mining Techniques for the Diagnosis of Diseases in Medical Domain", International Journal of Computer Science and Information Technologies, Vol. 5 (1) , 2014, 838-846,ISSN 0975-9646.