# Survey: Techniques Of Data Mining For Clinical Decision Support System

*Manodnya Shitole* [1], *M.A.Wakchaure* [2]

[1] *Student, Computer Department, Amruthvahini College Of Engineering, Maharashtra, India.*
[2] *Assistance Professor, Computer Department, Amruthvahini College Of Engineering, Maharashtra, India.*

**ABSTRACT**

*Clinical decision support system, which uses advanced data mining techniques access as well store data on server. The advantages of clinical decision support system include not only providing diagnosis accuracy but also minimize diagnosis time .Typically, with large amounts of clinical data generated every day, naive Bayesian classification can be utilized to formed valuable information to improve clinical decision support system. The Clinical Decision support system is very flourishing but it also having some critical problems. I propose a new privacy-preserving patient-centric clinical decision support system, which helps clinician complementary to diagnose the risk of patients' disease in a privacy-preserving way. Also , the past patients' historical data are stored in cloud and can be used to train the naive Bayesian classifier without leaking any individual patient medical data, and then the trained classifier can be applied to determine the disease risk for new coming patients and top-k disease names are also extracted from their according to the own preferences, which is provided for protecting the privacy of past patients' historical data, a new cryptographic tool called additive homomorphism proxy aggregation scheme is designed. Moreover, to leverage the leakage of na¨ive Bayesian classifier, we introduce a privacy-preserving top-k disease names retrieval protocol in our system. The privacy analysis gives security t the patient information and will not be leaked out at the time of disease diagnosis phase. This can be concluding that our system can efficiently calculate patient's disease risk with high accuracy in a privacy-preserving way.*

**Keyword :** - *Privacy-preserving, cryptography, patient-centric, clinical decision support system, etc.*

## 1.Introduction

In 1950 the first research article dealing with medicine and computers appeared in late 1950s . After that an experimental prototype appeared in the early 1960. At that time limited capabilities of computer did not allow it to be a part of medical domain. In 1970s the three advisory systems: de Dombal's system for diagnosis of abdominal pain and helps system for medical alerts delivery .1990s gives a large scale shift from administrative systems to clinical decision support systems. Healthcare Industry, no appropriate technique is developed to find great potential economic values from big healthcare data, these data might not only become meaningless but also requires a large amount of space to store and manage. Over the past two decades, the miraculous evolution of data mining technique has imposed a major impact on the revolution of human's lifestyle by predicting behaviors and future trends on everything which can convert stored data into meaningful information. These techniques are well suitable for providing decision support in the healthcare setting. To speed up the diagnosis time and improve the diagnosis accuracy, a new system in healthcare industry should be workable to provide a much cheaper and faster way for diagnosis.[1] Clinical Decision Support System (CDSS), with various data mining techniques being applied to assist physicians in diagnosing patient diseases with similar symptoms, has received a great attention recently. Naive Bayesian classifier, one of the popular machine learning tools, has been widely used recently to predict various diseases in CDSS . Despite its simplicity, it is more appropriate for medical diagnosis in healthcare than some sophisticated techniques.[4]

## 2 Literature Survey:

Data mining techniques have been widely used in clinical decision support systems for prediction and diagnosis of various diseases with better accuracy. The techniques have been very effective in developing clinical support systems because they are able to detect hidden patterns and relationships in medical data. There are several classification techniques which can be used for clinical decision support system. The aim of classification is to predict the target class for each case in the data accurately. Classification is important when a data repository contains samples that can be used as the basis for future decision making.

### 2.1. Bayesian Belief Network

The Bayesian network is a knowledge based graphical representation that shows a set of variables and their probabilistic relationships between diseases and symptoms. Bayesian network is utilized to find the probability of the presence of possible diseases given their symptoms. Its advantage is that it requires the knowledge and conclusions of experts in the form of probabilities. It is very important for the physician who has no computer expertise to understand about the Bayesian network. Which is also gives as a clinician reference with a searchable database of diseases and clinical manifestations. [5]Another example is SimulConsult, by focusing ecialty by specialty which uses Bayesian systems to input data in a scalable way and determine probabilities, accomplishing it. It applies a statistical pattern-matching approach that considers the age of onset and offset of the findings in each disease.[6]

### 2.2 Neural Network

Neural Networks is allows the systems to learn from existing knowledge and experiences. : Input, Output and Hidden layer are the three main layers of Neural Network. Neural Network is made of nodes called neurons. And there is weighted connection between nodes of different layers, which is used to transfer signals between the nodes. Neural Network is able to continue with incomplete data that gives educated guesses about missing data and get improved with every use due to its adaptive system learning. Mr. P. A. Kharat et al 2011 proposed clinical decision support system based on Jordan/Elman neural network for the diagnosis of epilepsy and they got relatively high overall accuracy for training data 99.83% and for cross-validation data and testing data 99.92% [7] R.R.Janghel et al 2009 developed a CDSS using artificial neural network to predict the fetal delivery to be done normal or by surgical procedure. In that system, they used three different training algorithms to train the neural network, which are Back Propagation algorithm, Radial Basis function and Learning vector quantization Network and they were able to gain accuracy of 93.75%, 99% and 87.5% respectively. [9] Mrudula Gudadhe and Kapil Wankhade et al 2010 designed decision support system based on neural network for heart diseases classification and they classified the data into 5 categories of heart disease with 97.5% accuracy by using multilayer perceptron with back propagation training algorithm. [8]

### 2.3.Decision Tree

Decision tree is the most often used techniques of data analysis. It is applied to classify records to a proper class. In medical field decision trees determine the sequence of attributes. First it makes a set of solved cases. Then the whole set is divided into training set and testing set. A training set is used for the induction of a decision tree. A testing set is used to find the accuracy of an obtained solution. AY AI-Hyari et al 2013 developed a CDSS for diagnosing patients with Chronic Renal Failure using various classification methods like neural network, naïve bays and decision tree. They proved that there is (92.2%) accuracy of Decision tree algorithm as compared to all other algorithms/implementations involved in their study. [10] They applied supervised decision tree classifier C4.5 to classify image samples with sensitivity of 98.1% and specificity of 99.6%. [11].

### 2.4. Naïve Bays

Naïve Bays uses the kernel estimator for numeric attributes rather than a normal distribution and utilized Supervised Discretization while converting numeric attributes to normal ones. We got an Output in text form of Naïve Bayes classifier. Abeer Y. Al-Hyari et al 2013 designed a CDSS for prediction and diagnosis of Chronic Renal Failure (CRF) using naïve bays. The implemented CDSS can be used to observe the progression stage of the disease in patient. They were able to achieve accuracy of 88.2% using naïve bays algorithm. [12]Mrs.G.Subbalakshmi et al 2011 developed a CDSS for heart disease prediction. Their system extracts hidden knowledge from a historical heart

disease database. They claimed that it is the most effective model to predict patients with heart disease.[13] Advantages of Naïve bays are that it is simple and efficient and it gives better classification performance.

**2.5. Support Vector Machine**

Support vector machine (SVM) has become more and more popular tool for machine learning tasks involving classification, regression etc. SVM separate the data into two categories of performing classification and constructing an N-dimensional hyper plane. SVM is supervised learning model that is applied for classification. SVM serves as the linear separator between two data points to identify two different classes in the multidimensional environment. SVM algorithms are in the binary formate; so in the case of multi-class problem one must reduce the problem to a set of multiple binary classification problems. They applied rough set for feature selection and SVM for classification. They attained very high classification accuracy of 99.41% for 50–50% of training-test partition, 100% for 70–30% of training-test partition, and 100% for 80–20% of training-test partition. And they were also able to discover a combination of five informative features, which can be important to the physicians for breast diagnosis. [14]

**2. 6.Fuzzy Set Approach**

Fuzzy set gives operations for combining fuzzy measurements. Fuzzy Logic is a type of multi-valued logic derived from fuzzy set theory to deal with approximate reasoning.Aniele C. Ribeiro et al 2014 proposed fuzzy breast cancer system to map two controlled and two non-controlled input variable into the risk of breast cancer occurrence. In this the Fuzzy Logic Rule based classifier is very effective in high degree of positive predictive value and diagnostic accuracy. It can provide health support to predict measurement of developing breast cancer to thefemale population and the health authorities, to reduce both the outcomes and mortality rate. [15]Chang-Shing Lee et al 2011 presented a novel five-layer fuzzy ontology to model the domain knowledge withuncertainty and extend the fuzzy ontology to thediabetesdomain. They proved that the proposed method works more effectively for diabetes application than previously developed ones. [16]Markos G. Tsipouras et al 2007 presented a methodology for the automated development of fuzzyexpert systems (FES).

**2.7.Genetic Algorithm**

Genetic Algorithm is proposed in the 1940s at the Massachusetts Institute of Technology based on Darwin's evolutionary theories that dealt with the survival of the fittest In this algorithm, the system goes through an iterative process to get an optimal solution. Because of its lacks of transparency in the reasoning involved for the decision support systems, it is unwanted by physicians. But still it is being used in clinical decision support with other algorithm for improved result. Similar to neural networks, the genetic algorithms derive their information from patient data.S.U. Amin et al 2013 presented a hybrid system of genetic algorithm and neural network for prediction of heart disease using major risk factors. They used the global optimization benefit of genetic algorithm to initialize the weight of neural network. [17] AZ Shabgahiet al 2011 proposed cancer detection on Global Cancer Map dataset by creating fuzzy rule with genetic algorithm. [18]

**2.8. Rough Set Approach**

Rough set theory provides mathematical tools to determine hidden patterns in data that can be used in data mining, it is determined by a lower and upper bound of a set. The lower and upper bound is chosen based on selection of attributes. Therefore it may not be applicable for some application. It does not need any preliminary or extra information corning data. [19]

**2.9. K-Nearest Neighbor**

K-Nearest neighbor classifies item based on nearest training data in the feature space. Which is the type of instance base learning or lazy learning. It is very simple but its accuracy can be affected by noisy or irrelevant features.[19]

## 3. CONCLUSIONS

The main Aim of this survey was to describe the most common classification algorithms of data mining which used in Clinical Decision Support Systems. The Survey gives us the result that of different classifiers is depend on various factor like dataset, number of attributes in dataset, number of class variable, number of missing values in

dataset. The selection of classifier for clinical decision support is application dependent .And the Naïve Bayesian Classifier is most fruitfully use for clinical system.

## 4. REFERENCES

[1] Wyatt JC, Liu JL. J Epidemiol Community Health 2002; 56(11): 808- 812, "Basic concepts in medical informatics"

[2] Characteristics of Clinical Decision Support System,en.wikipedia.org/wiki/Clinical_decision_support_system

[3]David Crockett, Ryan Johnson, and Brian Eliason, Analytics in health care, " What is Data mining in HealthCare?"

[4]Decision Support System, Iliad,www.openclinical.org/aisp_iliad.html

[5]Barnett GO, Cimino JJ, Hupp JA, Hoffer EP, JAMA. July 3, 1987. "DXplain. An evolving diagnostic decision-support system".

[6]A simultaneous consult on your patient's diagnosis, Simulconsult, www.simulconsult.com/

[7]Mr. P. A. Kharat, Dr. S. V. Dudul, IEEE 2011, Clinical Decision Support based on Jordan/Elman Network"

[8]Mrudula Gudadhe, Kapil Wankhade and Snehlata Dongre, IEEE 2010, Decision Support System for heart disease using support vector machine and artificial neural network"

[9]R. R. Janghel, Anupam Shukla and Ritu Tiwari, IEEE 2009, "Clinical Decision Support System for Fetal Delivery using artificial neural network"

[10]AY AI-Hyari, A. M. Al-Taee and M. A. Al-Taee, IEEE 2013, "Clinical Decision Support for diagnosis and management of Chronic Renal Failure

[12]Abeer Y. Al-Hyari, Ahmad M. Al-Taee and Majid A. Al Taee,IEEE 2013, "Clinical Decision Support System for Diagnosis and Management of Chronic Renal Failure"

[13]G.Subbalakshmi, K. Ramesh and M. Chinna Rao,IJCSE 2011, "Decision Support in Heart Disease Prediction System using Naive Bayes"

[14]Hui-Ling Chen, Bo Yang, Jie Liu and Da-You Liu, Springer 2011, "A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis"

[15]Aniele C. Ribeiro, Deborha P. Silva and Ernesto Araujo,IEEE 2014, "Fuzzy Breast Cancer Risk Assessment"

[16]Chang-Shing Lee and Mei-Hui Wang, IEEE 2011 "A Fuzzy Expert System for Diabetes Decision Support Application"

[17]S. U. Amin, K. Agarwal and R. Beg, IEEE 2013, "Genetic neural network based data mining in prediction of heart disease using risk factors"

[18]AZ Shabgahi and MS Abadeh, IEEE 2011, "Cancer Tumor Detection by Gene Expression Data Exploration using genetic Fuzzy System"

[19] 1.Chaitali Vaghela,2.Nikita Bhatt 3.Darshana Mistry "A Survey on Various Classification Techniques for Clinical Decision Support System." *International Journal of Computer Applications (0975 – 8887) Volume 116 – No. 23, April 2015*