A SURVEY ON MEDICAL DATA MINING TECHNIQUES

MRS. KRUPALI DHAWALE
Assistant Professor ,DYPIMCA, Akurdi,
Email id:krupali.acem@gmail.com

MRS. DIPTINASHINE
Assistant Professor ,DYPIMCA, Akurdi,
Email id:dipti.nashine@gmail.com

ABSTRACT
In data mining, we are extracting data from historical knowledge and create data sets. Data mining is similar to discover data knowledge from data warehouse and converts it into understandable format. In India, furthersocieties are listening aboutthe medical fitness and health opinion problems. In China, more than 200 thousand people die due to medication errors and 100 thousand in USA .More than 45% medication errors are caused by physicians as specialists write the prescription giving to their practices which are quite limited. In India, at every moment to take care of health and maintain medicine records is very tedious task in pharmaceutical industry, medical ministry and health industry. So, data mining techniques provide necessary information and awarenessabout medical history and help doctors to classify various medicines correctly to reduce medication error effectively. This paper reviews the results of different classification techniques such as Support vector machine, Naive Bayes, Decision tree, neural networks to diagnosis various diseases that helps to offer many advantages over the traditional healthcare. This paper shows that the medical data mining techniques related to the patients are rising on increasing and extended towards storage on cloud. So now to extract important information from this huge amount of data securely, requires classification, so, in future C4.5 algorithm is used for classification and it is the latest version of ID3 algorithm effective to classify the data.

KEY WORDS: Data Mining, Health, Aggregation, C4.5 algorithm.

I:INTRODUCTION
Data mining is the system of miningdata form responsible hidden patterns which can be converted into understandable format. Data mining essential is one of a number of analytical tools for applied to data. It allows users to analyze data from many different databases and categorize it, and summarize the relationships identified.Generally, data mining is the process of searching patterns or correlations among relational databases repository of fields in large databases where data sources can include database, data warehouse, web, information repositories etc. Data mining turns a huge collection of data into knowledge for global challenges. It is nothing but 'Knowledge mining data" which is somewhat long and similar to knowledge extraction, data/pattern analysis, data dreading.In India, further societies are listening about the medical fitness and health opinion problems. In China, more than 200 thousand people die due to medication errors and 100 thousand in USA .More than 45% medication errors are caused by physicians as specialists write the prescription giving to their practices which are limited.

In India, at every moment to take care of health and maintain medicine records is very tedious task in pharmacological manufacturing, medicinal department and health industry.So, data mining techniques provide necessary information and awareness about medical history and help doctors to classify various medicines correctly to reduce medication error effectively. C4.5 decision tree algorithm is improved and implemented to classify the instances. Datasets are analyzed on WEKA (Waikato Environment for Knowledge Analysis).

Data Mining Methods are useful for medical data including if else condition for find out frequent patterns, clustering and classification. Now a days, data mining techniques introduced in in Healthcare Domain. Lot of Research found and various data mining techniques are useful in predicating heart disease, lung cancer, and breast cancer, uterusfibroid diagnosis and so on.

II .LITERATURE SURVEY
SrinivasaDurga Prasad G[1] ,Siva SankarUppala[2] Proposed system helps thepatients to find an alternative to the medicine which is prescribed by the doctor. This isnot only helps the patients to find an alternative but also to find an optimum price of the drug required. Doctors are benefited by studying about the different drug composition in a medicine and the effects and side effects of the medicine. Chemist can refer for finding the list of the medicines which are in demand. Pharmaceutical companies can find their competitorsand the price of the medicine they are selling at, which will help them in advertising for their respective medicines. The Insurance Companies can find out which area is more prone to diseases and expand their sales out there. The use of data mining not only reduces the human efforts for findinginformation but also provides the information in fraction of seconds.GirijaD.K[1],M.S. Shashidhara[2] Suggested data mining technique identifies the risk factors regarding with the effect of fibroids in uterus and its application in area where women's are affected by uterine fibroids. VivekBarot[1], Prof.NikuBrahmbhatt[2] explains This paper reviews the results of different classification techniques such as Support vector machine, Naive Bayes, Decision tree, neural networks to diagnosis breast cancer.Trupti .S.Mokati[1], Prof. Vijay B. Gadicha [2] shows that as the clinical data mainly related to the patients are growing on increasing and reached towards storage on cloud. So now to extract important information from this huge amount of data securely, requires Homomorphic encryption techniques and SVM classification techniques that results in generation of proper Clinical Decision Support System. Shaker H. El-Sappagh[1], Samir El-Masri[2]explains proposed

framework will take advantages of EHR, datamining techniques, clinical databases, domain expert'sknowledge bases, and available technologies and standards to provide decision making support for the healthcare personnel. Christos A. Frantzidis[1], CharalamposBratsas[2]Proposed method forthe efficient discrimination of negative and positive emotions of people and monitor elderly or chronically ill people. JankiNaik[1],Sagar Patel[2] proposed method improves the efficiency than the traditional image mining methods. Here, results which we get are compared with Naive Bayesian classification algorithm.R. Castaldo[1], L. Pecchia[2] and C. James[3] proposed the best performing machine learningmethod was the C4.5 tree algorithm, which discriminated between stress and rest with sensitivity, specificity and accuracy rate of 78%, 80% and 79% respectively.Christos A. Frantzidis[1], Manousos A. Klados[2], Chrysa D. Lithari, Ana B. Vivas, Christos L. Papadelis[3], EleniKaldoudi[4] the proposed approach holds promise for the efficient discrimination of negative and positive emotions, and it is hereby discussed how future developments may be steered to serve for affective healthcare applications, such as the monitoring of the elderly or chronically ill people.S.PitchumaniAngayarkanni[1], Dr.NadiraBanu Kamal[2],Dr.V.Thavavel[3]presents the mathematical morphological and rough set based approach in detection and classification of cancerous masses in MRI ammo gram images. Breast cancer is one of the most common forms of cancer in women. In order to reduce the death rate, early detection of cancerous regions in mammogram images is needed. R .GeetaRamani[1], K.Sivaselvi[2] proposed Thus, this study has attempted to analyses the capability of various supervised data mining techniques in classifying the brain MR images. Initially, the images are pre-processed and the volumetric features are extracted. Then, these are fed into feature selection techniques viz. Principal Component Analysis, Runs, Fisher filtering and Relief feature selection to determine relevant features. The selected features are utilized for the supervised data mining techniques viz. Naive Bayes, Support Vector Machine, Random Tree and C4.5 to identify the abnormal images of brain. Among them, SVM has achieved highest accuracy of 71.33% with the features extracted through Relief feature selection with Leave-One-Out cross validation. Random Tree achieved accuracy of 82% with Runs filtered features.Mohammed Abdul Khaleel[1],Sateesh Kumar Pradham[2] ,G.N. Dash[3]proposed to analyze data mining techniques required for medical data mining especially to discover locally frequent diseases such as heart ailments, lung cancer, breast cancer and so on. They evaluate the data mining techniques for finding locally frequent patterns in terms of cost, performance, speed and accuracy. They, also compare data mining techniques with conventional methods.MeghanaNagori[1] , ShivajiMutkule[2] and PrafulSonarkar[3] Proposed system stores the metabolite values in dataset instead of storing fMRI images so reduces the image processing tasks and memory requirements.

III: SUMMARY OF TECHNIQUES FOR MEDICAL DATA MINING

Table 1:Different DataMiningTechniques used in Healthcare

References.	Techniques	Diseases
1	Naive Bayes, Support Vector Machine, Random Tree and C4.5	To find out complexities of abnormalities presence and judgment of images.
2	(quadratic discriminate analysis (QDA), linear discriminate analysis (LDA), C4.5, random forests, and nearest neighbors (KNN)and Machine Learning Algorithm	To Detect and report sedentary young men.
3	Graph scanning method is used to extract the values from graph. Graph scanning algorithm decision tree algorithm used.	Brain tumor
4	Machine learning method was the C4.5 tree algorithm, which discriminated between stress and rest with sensitivity, specificity and accuracyrate of 78%, 80% and 79% respectively.	Mental stress detectionvia short-term Heart Rate Variability (HRV)
5	The healthcare domain by providing information to different bodies like Doctors, Patients, Chemists, Pharmaceutical and theInsurance Companies.	Alternative solution prescribed by the doctor
6	Decision Tree algorithm used for discoveringmeaningful new correlation, patterns and trends by siftingthrough a large amount of data stored in repositories.	Uterine Fibroids
7	clustering, classification, regression, association rule mining, CART (Classification and Regression Tree)	Heart ailments, lung cancer, breast cancer
8	Metabolites like NAA, Creative, Choline and Cr2 are used to detect the brain tumor ,Decision Tree Algorithm used	Brain tumor
9	Machine Learning Techniquesto build a modelthat makes the knowledge discovered from the data explicit and communicable to domain experts.	Choose better data for Clinical Use.
10	Ten-fold cross-validation was used to evaluate the performance ofvarious classification algorithms trained on predicting the survivalstatus	Bone marrow stem cell transplant (BMT
11	Decision trees, Naive Bayes classifiers, K-nearest neighbor classification (KNN), support vector machine (SVM), and artificial neural networks technique used.	Heart disease diagnosis
12	MFS had the most notable effect on Support Vector Machine (SMO) and Decision Tree (J48) for the dataset	Cardiac Stress Test

IV. PERFORMANCE ANALYSIS OF DIFFERENT CLASSIFIERS

Different data mining techniques have been used to help healthcare professionals in the diagnosis of

variousdiseases. Table 1 shows a various data mining techniques used in the diagnosis of different diseases.

V CONCLUSION AND FUTURE SCOPE

In this survey paper, we have studied different data mining techniques for diagnosis different disease. From this survey, wegot the information about how to apply various data mining technique to diagnosis of various diseases. In future a Decision tree can be developed by using C4.5 algorithm which gives better accuracy and decision tree to interpret. This paper shows that the medical data mining techniques related to the patients are rising on increasing and extended towards storage on cloud. So now to extract important information from this huge amount of data securely, requires classification, so, C4.5 algorithm is used for classification and it is latest version of ID3 algorithm.

REFERENCES

1)R. Castaldo, Student Member, IEEE ,W. Xu, P. Melillo, Member, IEEE, L. Pecchia, Member, IEEE, L.Santamaria, Student Member, IEEE, C. James, Member, IEEE."Detection of Mental Stress due to Oral Academic Examination viaUltra-short-term HRV Analysis".

2)Meghana Nagori[1], Shivaji Mutkule[2], Praful Sonarkar[3],"Survey on Data Mining Techniques for Diagnosis and Prognosis of Breast Cancer",International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 4, January 2013.

3)Babak Taati[1], Jasper Snoek[2], Dionne Aleman[3], and Ardeshir Ghavamzadeh[4], "Data Mining in Bone Marrow Transplant Records to Identify Patients With High Odds of Survival", Ieee Journal Of Biomedical And Health Informatics, Vol. 18, No. 1, January 2014.

4)Kittipol Wisaeng [1] "An Empirical Comparison of Data Mining Techniques in Medical Databases",
International Journal of Computer Applications (0975 – 8887) Volume 77– No.7, September 2013

5) Srinivasa Durga Prasad G[1]., sivaSankarUppala Knowledge Discovery[2], "Analysis And Prediction In Healthcare Using Data Mining And Analytics".

6)Girija D.K[1] ,M.S. Shashidhara[2], "Classification of Women Health Disease (Fibroid)Using Decision Tree algorithm",International Journal of Computer Applications in Engineering Sciences.

7)Mohammed Abdul Khaleel[1],Sateesh Kumar Pradhan G.N. Dash[2],"A Survey of Data Mining Techniques on Medical Data for Finding Locally Frequent Diseases", International Journal of Advanced Research in Computer Science and Software Engineering.

8)MeghanaNagori[1],ShivajiMutkule[2],PrafulSonarkar[3] ,"Detection of Brain Tumor by Mining fMRI Images",International Journal of Advanced Research in Computer and Communication EngineeringVol. 2, Issue 4, January 2013.

9)Paula de Toledo[1], Pablo M. Rios[2], AgapitoLedezma [3], "Predicting the Outcome of PatientsWith Subarachnoid Hemorrhage Using MachineLearning Techniques".

10) BabakTaati[1], Jasper Snoek[2], Dionne Aleman[3], and ArdeshirGhavamzadeh[4], "Data Mining in Bone Marrow Transplant Recordsto Identify Patients With High Odds of Survival", IEEE Journal Of Biomedical And Health Informatics, Vol. 18, No. 1, January 2014

11) Salha M. Alzahani[1], AfnanAlthopity[2], AshwagAlghamdi[3], BoushraAlshehri[4], and SuheerAljuaid[5]"An Overview of Data Mining Techniques Applied for Heart Disease Diagnosis and Prediction" ,Lecture Notes on Information Theory Vol. 2, No. 4, December 2014.

12) JesminNahar, Tasadduq Imam, Kevin S. Tickle, Debora Garcia- Alonso, "Medical knowledge based data mining for cardiac stress test".