**NOVATEUR PUBLICATIONS**
**International Journal of Research Publications in Engineering and Technology [IJRPET]**
**ISSN: 2454-7875**
**VOLUME 3, ISSUE 10, Oct. -2017**

# DATA MINING TECHNIQUES IN MEDICAL FIELD FOR HEALTH CARE SYSTEM

PROF .DHOBALE MANOJ RAMCHANDRA
(VVPPS, Lecturer) manoj.dhobale@yahoo.co.in

PROF .GHOLASGAON KIRAN CHANDRAKANT
(VVPPS, Lecturer) kiran_cg15@yahoo.co.in

PROF. THORAT PAWAN DATTATRYA
(VVPPS, Lecturer) pavan.thorat14feb@gmail.com

**ABSTRACT:**

**Due to the vast improvement in IT sectors ,the popularity health care organization conserve there data elctoncally. The mining of this well-informed data, in series from the mass data which will be very useful for huge progress in medical field but it is insufficient. There is need of proficient analysis tools to resolve covered exact data. Data mining can represent new biomedical and healthcare details for medical preference. The relationship comes together from the data of patients composed in database aid in the resultant progression. This assessment investigates the worth of a variety of Data Mining techniques such as classification, regression, and clustering and association health domain. Now a days this documentation of presents a brief introduction of the techniques that are currently used in medical field followed by various qualities and demerit of the existing techniques. This investigation also focusing on applications, challenges and future issues of Data Mining in healthcare. At the end , this work provides the recommendations for exact selection of available Data Mining techniques.**

**KEYWORDS: Data Mining, Association, Classification, Clustering Decision making and Healthcare.**

## INTRODUCTION:

Data mining in the Medical field has been an enormous latent process for exploring veiled patterns in data sets of medical sphere. In healthcare, despite the fact that data mining is not broadly used, its reputation now more and more accepted in the medical datasets for its earlier discovery overall development. Data mining can enliven Decision-making by discovering patterns and trends in large amounts of all-around data. There are two main goals of data mining- prediction and portrayal. Prediction involves some variables or fields in the data set to visualize mysterious or future values of other variables of curiosity. On the other hand narrative focuses on verdict patterns recitation of the data that can be interpreted by humans. The data generated by the health organizations is very enormous and versatile due to which it is difficult to investigate the data, in order to make significant announcement concerning patient's health. This data contains details regarding hospitals, patients, medical assert, cure cost etc. So, there is a must to create a commanding tool for scrutinizing and extracting significant information from this complicated data. The analysis of health data improves the healthcare by enhancing the concert of patient organization tasks. The results created by Data Mining technologies improve the progression of predicting the comparable patients and clustering them under a challenging group based on illness or fitness issues, so that healthcare involvement offers them powerful treatments. It can also be constructive for forecasting the span of life of patients in hospital, for medical investigation and constructing plan for effectual information system management. Current technologies are used in medical field to enhance the medical facilities in cost valuable manner. Data Mining techniques are also used to investigate the many no's of reasons that are accountable for diseases such as nature or type of food, different working condition, culture and knowledge level, livelihood conditions, availability of pure water, health care services, artistic, environmental and farming factors. The rest of this paper is organized as follows: Section 2 - 1.reviews the related work on data mining techniques on Parkinson's disease, 2. discusses the survey on Diabetes Disease Diagnosis using data mining technique, 3 - presents the finding from the survey of Heart disease Diagnosis.

## DATA MINING FOR DISEASE DIAGNOSIS:

Data mining is not a substitution to doctors as an alternative, it is a tribute to them to predict the diseases in advance stages because ,With the fast increase in population, there is a substantial quantity of augmentation in the health related diseases. Numerous diseases are strongly associated with a symptom which makes it complicated for the doctors to forecast the precise diseases on one go. This is where data mining

IJRPET

appears into backing; it helps in forecasting the disease which is almost perfect. Even-though the forecasting is not extremely accurate, it gives the doctor a concise idea what the disease might be.

**PARKINSON'S DISEASE**:

In the Parkinson's illness is a neurodegenerative ailment, the malady impinges on by brain cells (neurons) in human brain. The neurons make a chief chemical called dopamine. The dopamine send signal to the fraction of brain that pedals travel. The small signals can help those parts of the brain work enhanced. The decrease of dopamine in the brain makes the person motionless or immobile. The four types of symptoms of Parkinson's bug are: tremor, rigidity, Bradykinesia and postural volatility.

When any patient suffering from that quiver then, it pulse by hands, arms, legs or jaws. The sign of determination makes limbs and trunk rigid. Bradykinesia is a sign which leads to sluggish travels. Postural instability causes gloominess and bad changes. The basic indicator affects almost 80-90% of citizens with Parkinson's bug. The Parkinson's disease of remote tracking used by eleven techniques.

The can refers previous research work in developed the voice dimensions of disease primarily spotlights the speech signals. The Parkinson dataset was a series of biomedical voice measurements from 31 people, 23 trait features in Parkinson's disease. The error rate of confusion matrix of 2*2 matrix is the output. The major objective is to get the lowest amount of error rate with the minimum characteristic of Parkinson's dataset. The work was presented in the speech of vocal sound examination for the Parkinson's disease patients to be evaluated by the health control (HC) people. The speech was assessed for four features like NHR, SPLD, RFPC and F0 SD. The classification accuracy is upto 80-85%.It was developed to be evaluated for the performance of Artificial Neural Networks (ANN) and Support Vector Machines (SVM). The ANN and SVM are designed by metrics like accurate, truly positive, false positive, positive predictive value and negative predictive value

**DIABETES DISEASE DIAGNOSIS USING DATA MINING TECHNIQUE:**

A good number of researches have been reported in literature on diagnosis of diabetes disease diagnosis. This section deals with discussion of some related work in diabetes disease diagnosis using data mining technique. The proposed work is based on neuropathy diabetics which is a kind of nerve disorder due to diabetic mellitus. This is mainly affected by the long term diabetics. This paper deals with the symptoms and risk factors of neuropathy taken into the consideration for deployment of fuzzy based relation equation. It is linked with the Multilayer perceptron in composition of binary relation using fuzzy inference model.

The authors of devised an automatic retinal diabetic detection using a multilevel perceptron neural network. To evaluate the best worldwide threshold to minimize the pixel classification errors, the network is trained by this algorithm. The performance of the proposed work improves by the uncovering and enough index based on neuro fuzzy subsystem. The fuzzy set and linguistic variable are used into diagnose diabetes. This thing used the maximum and minimum relationship to deal with uncertainty availed in the dataset. Data sets of forty patients were collected to produce this relationship.

In this presented a binary categorization based neural network techniques used diabetes diagnosis problem. The review is done based on three benchmark data sets obtained from UCI machine learning repository which is the standard one.

In our proposed work fuzzy membership function is used in connection with fuzzy neural network to detect the diabetes in early stages. This work uses two experimental examinations for medical data. In this technique developed a alternate pruning technique based on the Minimal Description Length principle. It can be viewed as substitutions between theory complexity and data prediction accuracy. This work proposed a greedy search algorithm to prune the fuzzy ARTMAP categories one by one. The results proved that fuzzy ARTMAP pruned with the MDL principle gave improved recital with less categories shaped compared to original fuzzy ARTMAP and other machine learning systems.

**SURVEY ON HEART DISEASE DIAGNOSIS USING MINING TECHNIQUES:**

This piece presents the literature survey associated to heart bug dataset forecast by means of data mining techniques for judgment & achieved dissimilar probabilities for diverse methods as discussed below.

- A gifted Heart bug prophecy System (IHDPS) urbanized by via data mining techniques. Immature Bayes, Neural Network, and conclusion Trees was projected in the vocation. The effort has its own might to get suitable fallout. To erect this system veiled patterns and affiliation were used. It is web-

IJRPET

based, user friendly & stretchy.

- To widen the multi-parametric trait with linear and nonlinear (HRV - Heart Rate Variability) a narrative practice was projected by authors To attain this, they have used classifiers like Bayesian Classifiers, Classification based on Multiple Association Rules, and Support Vector Machine.
- The guess of Heart bug, Blood heaviness and Sugar with the support of neural networks was projected by the novelist of .The dataset contains records with 13 attributes in each trace. The supervised networks i.e. Neural Network with back broadcast algorithm is used for preparation and difficult data.
- The crisis of identifying unnatural organization regulations for heart bug guess was deliberate by Carlos Ordonez. The ensuing dataset contains records of patients having heart ailment. Three constraints were introduced to diminish the number of patterns

**SUMMARY:**

The above review reveals that there are lots of present techniques are available for detection of disease diagnosis by performing classification or clustering techniques. But still there is a low awareness on the quality of the dataset, when it is incomplete. There are very few papers which work on data preprocessing, feature selection and reduction

**STATEMENT OF THE PROBLEM**

The present techniques used for disease diagnosis are still facing problem for false detection rate, due to the unrefined(raw) nature of dataset. The unsatisfactory and incomplete dataset may lead to false alarms in diagnosis of accurate results. The preprocessing of dataset is less concentrated in the existing work, which leads to major setback in the overall process. The future work of this survey will start with a data preprocessing technique to produce a complete dataset instead of using raw dataset.

**CONCLUSION:**

Over the past few decades, the automated problem- solving tools have been intended to assist the physician with a clear sense of medical data. In healthcare, data mining is becoming ever more and more important . From the above study it is examined that the support vector machine due to its regularization parameter is often recommended by a lot of the researches to evade over-fitting. With the help of its kernel trick it can box to build an expert inference system. In clustering models k-means based forecasting and variants of it also produce promising results. In evolutionary based models, particle swarm optimization with its variants contributed more innovatively in many disease diagnosis works. It also assumes the real number code, and it is determined honestly by the solution. The selection of data mining approaches depends on the nature of the dataset. If the dataset consists of the labeled features then the classification techniques can be suggested for best prediction. If the dataset is with unlabelled features then the clustering techniques are most excellently suited for pattern.

**REFERENCES:**

1) H. C. Koh and G. Tan, "*Data Mining Application in Healthcare*", Journal of Healthcare Information Management, vol. 19, no. 2, (2005)
2) R. Naveen Kumar [1], M. Anand Kumar [2]"*Medical Data Mining Techniques for Health Care Systems*" International Journal of Engineering Science and Computing, April 2016
3) Ming Li, Member, IEEE, Shucheng Yu, Member, IEEE, Yao Zheng, Student Member, IEEE, Kui Ren, Senior Member, IEEE, and Wenjing Lou, Senior Member, IEEE "*Scalable and Secure Sharing of Personal Health Records in Cloud Computing Using Attribute-Based Encryption",* Ieee Transactions On Parallel And Distributed Systems, Vol. 24, No. 1, January 2013
4) Rahul Isola, Student Member, IEEE Rebeck Carvalho, Student Member, IEEE, Amiya Kumar Tripathy, Member, IEEE *Knowledge Discovery in Medical Systems Using Differential Diagnosis, LAMSTAR, and k-NN*
5) Dr. R.Geetha Ramani, G.Sivagami, Shomona Gracia Jacob " *Feature Relevance Analysis and Classification of Parkinson's Disease TeleMonitoring data Through Data Mining"* , International Journal of Advanced Research in Computer Science and Software Engineering,vol-2,Issue 3, March 2012.
6) Peyman Mohammadi, Abdolreza Hatamlou and Mohammed Msdaris "*A Comparative Study on Remote Tracking of Parkinson's Disease Progression Using Data Mining Methods"* , International Journal in Foundations of Computer Science and Technology(IJFCST),vol- 3,No.6, Nov 2013.
7) Dr. R.Geetha Ramani and G.Sivagami "*Parkinson Disease Classification using Data Mining Algorithms*", International Journal of Computer Applications (IJCA),Vol-32,No.9, October 2011
8) Shanghais Wu, Jiannjong Guo "*A Data Mining Analysis of the Parkinson's Disease*", Scientific Research, iBusiness 2011, 3, 71-75.
9) J.Rusz, R. Cmejla, H. Ruzickova, J.Klempir et al.,

IJRPET

"*Acoustic Analysis of Voice and Speech Characteristics in Early Untreated Parkinson's Disease".*

10) David Gil A, Maguns Johnson B, "*Diagnosing Parkinson by Using Artificial Neural Networks and Support Vector Machines",* Global Journal of Computer Science and Technology, page 63-71.
11) Farhad Soleimanian Gharehehopogh, Peymen Mohammadi, "*A Case Study of Parkinson's Disease Diagnosis Using Artifical Neural Networks*" , International Journal of Computer Applications, Vol-73,No.19, July 2013.
12) Tripti Kapoor, R.K.Sharma, "*Parkinson's Disease Diagnosis Using Mel-Frequency Cepstral Coefficients and Vector Quantization",* International Journal of Computer Applications, Vol-4, No.3, Jan 2011.
13) Mehmet can "*Boosting committee Machines to Detect the Parkinson's disease by Neural Networks".*
14) SellappanPalaniappan, RafiahAwang, "*Intelligent Heart Disease Prediction System Using Data Mining Techniques*", IJCSNS International Journal of Computer Science and Network Security, Vol.8 No.8, August 2008
15) Niti Guru, Anil Dahiya, NavinRajpal, "*Decision Support System for Heart Disease Diagnosis Using Neural Network*", Delhi Business Review, Vol. 8, No. 1 (January - June 2007).
16) HeonGyu Lee, Ki Yong Noh, KeunHoRyu, "*Mining Biosignal Data: Coronary Artery Disease Diagnosis using Linear and Nonlinear Features of HRV,"* LNAI 4819: Emerging Technologies in Knowledge Discovery and Data Mining, pp. 56-66, May 2007.
17) Carlos Ordonez, "*Improving Heart Disease Prediction Using Constrained Association Rules*," Seminar Presentation at University of Tokyo, 2004
18) Franck Le Duff, Cristian Munteanb, Marc Cuggiaa, Philippe Mabob, "*Predicting Survival Causes After Out of Hospital Cardiac Arrest using Data Mining Method",* Studies in health technology and informatics, Vol. 107, No. Pt 2, pp. 1256-9, 2004.
19) Shantakumar B. Patil,Y. S. kumaraswamy "*Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network"*. ISSN 1450-216X Vol.31 No.4 (2009), pp.642-656.
20) Sarvestan Soltani A. , Saf`avi A. A., Parandeh M. N. and Salehi M., *"Predicting Breast Cancer Survivability using data mining techniques*," Software Technology and Engineering (ICSTE), 2nd International Conference, 2010, vol.2, pp.227-231.
21) Anunciacao Orlando, Gomes C. Bruno, Vinga Susana, Gaspar Jorge, Oliveira L. Arlindo and Rueff Jose, "*A Data Mining approach for detection of high-risk Breast Cancer groups*," Advances in Soft Computing, vol. 74, pp. 43-51, 2010

IJRPET