International Journal For Technological Research In Engineering
Volume 4, Issue 8, April-2017 *ISSN (Online): 2347 - 4718*

PERFORMING THE PRIDICTION BASED ANALYSIS FOR DISEASE USING CLUSTERING MEAN ALGORITHM AND IMAGE ANALYSIS THEOREM : A REVIEW

Deepika[1], Dr. Rakesh Joon[2]
[1]M.Tech (ECE), [2]HOD, Dept. of ECE, GITAM, Kablan

Abstract: Disease prediction is one of the most important issues in medical research. Many kind of patients trouble for his or her check up even for prognosticative illness like heart condition potentialities, internal organ injury modification and potentialities of those disease lies in prognosticative illness classes. They have not need very giant analysis if we tend to are ready to predict. However, there is an absence of effective analysis tools to search out hidden relationships and knowledge for any kind of prognosticative illness. This analysis inspire to develop a way on the premise data| through web mining that is utilized to research large volumes information. That will be regenerate useful knowledge through web mining tools. Experimental results will show that lots of of the principles facilitate at intervals the simplest prediction of disorder that even helps doctors in their diagonozing choices by victimization through A-priori and k-mean. Through this Analsis it could be easy to find the current status as well as future prediction of victim by very economical mean. Thus ,even doctors ineffective to predict illness accurately, deciding system which might predict the proper diseases with obtainable knowledge must beneficial to victim. During this analysis it tend predict the diseases by mean of the algorithms like Association rule, Predictive Rule, Priorities rule as in integrated form. Here, it include nearly more than two hundred persons informations for this analysis.

Key words: Data mining, kidney failure, heart disease, A-prior and k-mean algorithm, data set.

I. INTRODUCTION

Knowledge discovery in databases is well-defined process consisting of several distinct steps. Data mining is the core step, which results in the discovery of hidden but useful knowledge from massive databases. Quality service implies diagnosing patients correctly and administering treatments that are effective. Poor clinical decisions can lead to disastrous consequences which are therefore unacceptable. Treatment records of millions of patients can be stored and computerized and data mining techniques may help in answering several important and critical questions related to health care. The availability of integrated information via the huge patient repositories, there is a shift in the perception of clinicians, patients and payers from qualitative visualization of clinical data by demanding a more quantitative assessment of information with the supporting of all clinical and imaging data. For instance it might now be possible for the physicians to compare diagnostic information of various patients with identical conditions. The term Kidney failure and heart disease applies to a number of illnesses that affect the circulatory system, which consists of heart and blood vessels. It is intended to deal only with the condition and the factors, which lead to such condition. Acute kidney injury (also called acute renal failure) means that your kidneys have suddenly stopped working. Your kidneys remove waste products and help balance water and salt and other minerals in your blood. When your kidneys stop working, waste products, fluids, and electrolytes build up in your body. This can cause problems that can be deadly. Symptoms of Heart Disease: 1. Dizzy spell or fainting fits. 2. Discomfort following meals, especially if long continued. 3. Shortness of breath, after slight exertion. 4. Fatigue without otherwise explained origin. 5. Pain or tightness in the chest a common sign of coronary insufficiency is usually constrictive in nature and is located behind the chest bone with 6. Radiation into the arms or a sense of numbness or a severe pain in the centre of the chest. 7. Palpitation. Heart disease is a general term that means that the heart is not working not accurately. There are different kinds of heart disease like congenital heart diseases, acquired heart diseases, Coronary artery disease (CAD). Coronary artery disease (CAD) is the most frequent type of heart disease . About 80% of deaths occurred in low-and middle income countries due to heart diseases. It is predicted that if this trend continue then till 2030 around 23.6 million people will die from cardiovascular diseases (that’s heart strokes & heart attacks. It is the leading cause of death among males as well as females. This research paper analyzes how data mining techniques are used for predicting different types of diseases Symptoms of kidney failure: 1. Nausea and vomiting. 2. Passing only small amounts of urine. 3. Swelling, particularly of the ankles, and puffiness around the eyes. 4. Unpleasant taste in the mouth and urine-like odour to the breath. 5. Persistent fatigue or shortness of breath. 6. Loss of appetite.

II. BACK GROUND

[1] Ji-Jiang Yang, Jianqiang Li, Jacob Mulder d, Yongcai Wange, Shi Chen f, Hong Wug, Qing Wang b, Hui Pan proposed the main goal of this special issue and gives a brief guideline. Then, the present situation of the adoption of EMRs is reviewed. After that, the emerging information technologies are presented which have a great impact on the healthcare provision. These include health sensing for medical data collection, medical data analysis and utilization for accurate detection and prediction. Next, cloud computing is discussed, as it may provide scalable and cost-effective delivery of healthcare services. Accordingly, the current state

of academic research is documented on emerging information technologies for new paradigms of healthcare service. At last, conclusions are made. The appropriate collection and consumption of electronic health information about an individual patient or population is the bedrock of modern healthcare, where electronic medical records (EMR) serve as the main carrier. The healthcare industry is especially fastest-growing part of the economy of many countries in modern society, not only the more economically developed countries like those in Western Europe and North America, but also in areas of high growth, such as China and India. [2]Pankaj Deep Kaur, Inderveer Chana proposed the use of cloud computing for the creation and management of cloud based health care services. As a representative case study, we design a Cloud Based Intelligent Health Care Service (CBIHCS) that performs real time monitoring of user health data for diagnosis of chronic illness such as diabetes. Advance body sensor components are utilized to gather user specific health data and storein cloud based storage repositories for subsequent analysis and classification. In addition, infrastructure level mechanisms are proposed to provide dynamic resource elasticity for CBIHCS. Experimental results demonstrate that classification accuracy of 92.59% is achieved with our prototype system and the predicted patterns of CPU usage offer better opportunities for adaptive resource elasticity. The “pay for use” pricing model, on-demand computing and ubiquitous network access allow cloud ser-vices to be accessible to anyone, anytime, anywhere. The inherent benefits like fast deployment, lower costs, scalability,rapid provisioning, instant elasticity, greater resiliency, rapidre-constitution of services, low-cost disaster recovery and datastorage solutions promises the potentials of cloud computing.[3] Manjeevan Seera, Chee Peng Lim, proposed hybrid intelligent system that consists of the Fuzzy Min–Max neural network, the Classification and Regression Tree, and the Random Forest model is proposed, and its efficacy as a decision support tool for medical data classification is examined. The hybrid intelligent system aims to exploit the advantages of the constituent models and, at the same time, alleviate their limitations. It is able to learn incrementally from data samples (owing to Fuzzy Min–Max neural network), explain its predicted outputs (owing to the Classification and Regression Tree), and achieve high classification performances (owing to Random Forest). To evaluate the effectiveness of the hybrid intelligent system, three benchmark medical data sets, viz., Breast Cancer Wisconsin, Pima Indians Diabetes, and Liver Disorders from the UCI Repository of Machine Learning, are used for evaluation. A number of useful performance metrics in medical applications which include accuracy, sensitivity, specificity, as well as the area under the Receiver Operating Characteristic curve are computed. The results are analyzed and compared with those from other methods published in the literature. The experimental outcomes positively demonstrate that the hybrid intelligent system is effective in undertaking medical data classification tasks. More importantly, the hybrid intelligent system not only is able to produce good results but also to elucidate its knowledge base with a decision tree. As a result, domain users (i.e., medical practitioners) are able to comprehend the prediction given by the hybrid intelligent system; hence accepting its role as a useful medical decision support tool. [4] G. Santhosh Kumar, Lakshmi K.S proposed a new method of uncovering valid association rules from medical transcripts. The extracted rules describes association of disease with other diseases, symptoms of a particular disease, medications used for treating diseases, the most prominent age group of patients for developing a particular disease. NLP (Natural Language Processing) tools were combined with data mining algorithms (Apriori algorithm and FP-Growth algorithm) for the extraction of rules. Interesting rules were selected using the correlation measure, lift. Medical databases serve as rich knowledge sources for effective medical diagnosis. Recent advances in medical technology and extensive usage of electronic medical record systems, helps in massive production of medical text data in hospitals and other health institutions. Most of this text data that contain valuable information are just filed and not utilized to the full extent. Proper usage of medical information can bring about tremendous changes in medical field. [5] Hui Yang, Erhun Kundakcioglu proposed a broad spectrum of research and education for realizing the healthcare intelligence. They present unique perspectives, analytical methodologies, healthcare applications, and technology transfer opportunities. They also demonstrate the power of turning the Big Data generated in healthcare industry into useful knowledge for intelligent healthcare decisions. In the future, it’s anticipated that Big Data analytics and informatics research will bring a profound transformation in the arena of healthcare. It present three major areas of informatics research carried out by our group, along with our developments that intend to provide quality-assured, accessible, and patient- centered radiology best practices. We’ll briefly introduce the work here and detail it further in a bit. [6] Vishnu S. Pendyala Yi Fang JoAnne Holliday Ali Zalzala proposed our vision of automating some of the healthcare functions such as monitoring and diagnosis for mass deployment. We explain our ideas on how machines can help in this essential life supporting activity. Diagnosis part of the problem has been researched for long, so we set out working on this first, while the remaining is still in idea stage. We give insights into our work on automating medical diagnosis. There is a tremendous amount of attention being focused on improving human health these days. The World Health Organization (WHO) statistics show that disease and mortality rate greatly depend on access to proper healthcare, which is not available to a vast majority of the global population. A large section of the world population does not have access to proper healthcare. We approach this problem from two angles: prevention and cure. Prevention by way of monitoring; and diagnosing, as part of the cure. The monitoring part is still in the idea stage, while we did make some progress with respect to the diagnosis aspect. While we present our vision of machines helping with both monitoring and diagnosing, the focus of this paper is more on the diagnosis part. [7] Saba Bashir, Usman Qamar, M.Younus Javed proposed Large amount of medical data leads to the need of intelligent data mining tools in order to extract useful knowledge. Researchers have been using several statistical analysis and data mining techniques to improve the disease diagnosis accuracy in medical healthcare. Heart disease is considered as the leading cause of deaths worldwide over the past 10 years. Several researchers have introduced different data mining techniques for heart disease diagnosis. Using a single data mining technique shows an acceptable level of accuracy for disease diagnosis. Recently, more research is carried out towards hybrid models which show tremendous improvement in heart disease diagnosis accuracy. The objective of the proposed research is to predict the heart disease in a patient more accurately. The proposed framework uses majority vote based novel classifier ensemble to combine different data mining classifiers. UCI heart disease dataset is used for results and evaluation. Analysis of the results shows that the sensitivity, specificity and accuracy of the ensemble framework are higher as compared to the individual techniques. We obtained 82% accuracy, 74% sensitivity and 93% specificity for heart disease dataset. [8] Hajer Baazaoui Zghal Antonio Moreno proposed a system that allows any search engine to develop its semantic layer by applying ontology learning techniques onWeb snippets and applies it to a well-known medical digital library, PubMed. The new system (SemPubMed) automatically builds new ontology fragments related to the user's query and then it reformulates queries using the new concepts in order to improve information retrieval. Our system has endured a twofold evaluations. On the one hand, we have evaluated the quality of the modular ontologies built by the system. On the other hand, we have studied how the semantic reformulation of the queries has led to an improvement of the quality of the results given by PubMed, both in terms of precision and recall. Obtained results show that adding semantic layer to PubMed enables an improvement of query reformulation and predicted ranking score. They usually index a particular area, and thus return to the user results only from a particular field. One of their main shortcomings is that users have to make precise queries to retrieve the desired documents. Ontologies are knowledge structures allowing to represent the main concepts, relationships, instances and properties in a specific domain. Indeed, ontologies are considered as a concept modeling tool that can be used by information systems on the semantic and knowledge level; they have captured the attention of many researchers and they already play an important role in Knowledge Engineering. [9] Monali Dey, Siddharth Swarup Rautaray proposed the huge amounts of data generated by healthcare transactions are too complex and voluminous to be processed and analyzed by traditional methods. Data mining provides the methodology and technology to transform these mounds of data into useful information for decision making. The main aim of this survey is, analysis of the uniqueness of medical data mining, overview of Healthcare Decision Support Systems currently used in medicine, identification and selection of the most common data mining algorithms implemented in the modern HDSS, comparison between different algorithms in Data mining. Data mining technology provides a user oriented approach to novel and hidden information in the data. Valuable knowledge can be discovered from application of data mining techniques in healthcare system. Data mining in healthcare medicine deals with learning models to predict patients' disease. Data mining applications can greatly benefit all parties involved in the healthcare industry. For example, data mining can help healthcare insurers detect fraud and abuse, healthcare organizations make customer relationship management decisions, physicians identify effective treatments and best practices, and patients receive better and more affordable healthcare services.

III. TECHNIQUE USED

3.1 The k-means algorithm

The k-means algorithm is a simple iterative method to partition a given dataset into a specified number of clusters, k. This algorithm has been discovered by several researchers across different disciplines. The algorithm operates on a set of d-dimensional vectors, D = {xi | i = 1, . . . , N}, where xi ∈ Rd denotes the ith data point. The algorithm is initialized by picking k points in Rd as the initial k cluster representatives or "centroids". Techniques for selecting these initial seeds include sampling at random from the dataset, setting them as the solution of clustering a small subset of the data or perturbing the global mean of the data k times. The simple way to understand K-means is: Requires real-valued data.
We must select the number of clusters present in the data. Works best when the clusters in the data are of approximately equal size. Attribute significance cannot be determined. Lacks explanation capabilities.

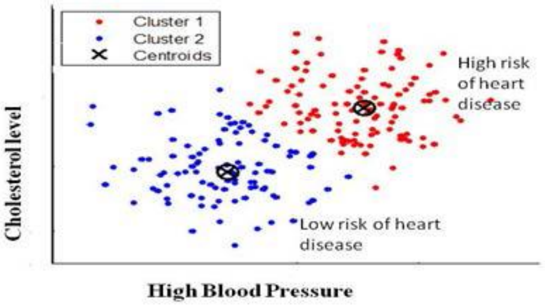

Fig3.1: K-means Clustering for Heart Disease Patients

3.2 K-means Procedure Steps

Step 1: Data Assignment. Each data point is assigned to its closest centroid, with ties broken arbitrarily. This results in a partitioning of the data.
Step 2: Relocation of "means". Each cluster representative is relocated to the center (mean) of all data points assigned to it. If the data points come with a probability measure (weights), then the relocation is to the expectations (weighted mean) of the data partitions.

3.3 The Apriori algorithm

Apriori algorithm for association is proposed by R.Agarwal., in 1994. They finds out the relationships among item sets using two inputs-support and confidence. One of the most popular data mining approaches is to find frequent itemsets from a transaction dataset and derive association rules. Finding frequent itemsets (itemsets with frequency larger than or equal to a user specified minimum support) is not trivial because of its combinatorial explosion. Once frequent itemsets are obtained, it is straightforward to generate association rules with confidence larger than or equal to a user specified minimum confidence. Apriori is a seminal algorithm for finding frequent itemsets using candidate generation. It is characterized as a level-wise complete search algorithm using anti-monotonicity of itemsets, "if an itemset is not frequent, any of its superset is never frequent". By convention, Apriori assumes that items within a transaction or item set are sorted in lexicographic order.

3.4 Apriori

Let the set of frequent item sets of size k be Fk and their candidates be Ck. Apriori first scans the database and searches for frequent item sets of size 1 by accumulating the count for each item and collecting those that satisfy the minimum support requirement. It then iterates on the following three steps and extracts all the frequent item sets.
1. Generate Ck+1, candidates of frequent itemsets of size k +1, from the frequent itemsets of size k.
2. Scan the database and calculate the support of each candidate of frequent itemsets.
3. Add those item sets that satisfies the minimum support requirement to Fk+1.

Function apriori generates Ck+1 from Fk in the following two step process:
1. Join step: Generate RK+1, the initial candidates of frequent itemsets of size k + 1 by taking the union of the two frequent itemsets of size k, Pk and Qk that have the first k−1 elements in common.
RK+1 = Pk ∪ Qk ={item1 , item2, . . . , itemk−1, itemk , itemk' }
Pk = {item1 , item2, . . . , itemk−1, itemk }
Qk = {item1 , item2, . . . , itemk−1, itemk' }
where, item1 < item2 < · · · < itemk < itemk'_ .
2. Next step: Check if all the item sets of size k in Rk+1 are frequent and generate Ck+1 by removing those that do not pass this requirement from Rk+1. This is because any subset of size k of Ck+1 that is not frequent cannot be a subset of a frequent item set of size k + 1. Function subset finds all the candidates of the frequent item sets included in transaction t. Apriori, then, calculates frequency only for those candidates generated this way by scanning the database. It is evident that Apriori scans the database at most kmax+1 times when the maximum size of frequent item sets is set at k-max.

IV. CONCLUSION

The paper proposed the integrating solution of apriori and k-mean. It will have more modified result comparing with older techniques. Individually, apriori have the simple mean of association and k mean have clustering formation of raw data. This integration technique gives result in two phases. First Phase generates the level in two phases, Higher threshold & Lower threshold. Second phase have work to do the intensity level of clustered data. Finally, combining both the stage for common purpose may give best ever result that each achieved yet in scientific world in same scenario.

V. FUTURE WORK

In future, we can expand this technique to very large size data means can be analyzed over millions of patient information. As well it could be very accurate for prediction information even by scanning technique of body.

REFERENCES

[1] Ji-Jiang Yang, Jianqiang Li, Jacob Mulder, Yongcai Wange, Shi Chen, Hong Wug, Qing Wang, Hui Pan "Emerging information technologies for enhanced healthcare" Computers in Industry xxx (2015) xxx–xxx Contents lists available at Science Direct Computers in Industry journal homepage:www.elsevier.com/locate/compind.

[2] Pankaj Deep Kaur∗, Inderveer Chana "Cloud based intelligent system for deliveringhealth care as a service" computer methods and programs in biomedicine113 (2014) 346–359 journal home page: www.intl.elsevierhealth.com/journals/cmpb. 2013 Elsevier Ltd. All rights reserved

[3] Manjeevan Seera a, Chee Peng Lim "A hybrid intelligent system for medical data classification" Expert Systems with Applications 41 (2014) 2239–2249 Contents lists available at ScienceDirect Expert Systems with Applications journal homepage: www.elsevier.com/locate/eswa. 2013 Elsevier Ltd. All rights reserved.

[4] G. Santhosh Kumar, Lakshmi K.S "Association Rule Extraction from Medical Transcripts of Diabetic Patients" 978-1-4799-2259-14/$31.00©2014 201.

[5] Hui Yang, Erhun Kundakcioglu "Healthcare Intelligence: Turning Data into Knowledge" 54 1541-1672/14/$31.00 © 2014 IEEE Ieee InTeLLIGenT SySTeMS Published by the IEEE Computer Society.

[6] Vishnu S. Pendyala Yi Fang JoAnne Holliday Ali Zalzala "A Text Mining Approach to Automated Healthcare for the Masses" 978-1-4799-7193-0/14/$31.00 ©2014 IEEE 28 IEEE 2014 Global Humanitarian Technology Conference.

[7] Saba Bashir, Usman Qamar, M.Younus Javed "An Ensemble based Decision Support Framework for Intelligent Heart Disease Diagnosis" International Conference on Information Society (i-Society 2014) 978-1-908320-38/4/$25.00©2014 IEEE 259.

[8] Hajer Baazaoui Zghal • Antonio Moreno "A system for information retrieval in a medical digital library based on modular ontologies and query reformulation" Multimed Tools Appl (2014) 72:2393–2412 DOI 10.1007/s11042-013-1527-4. H. Baazaoui Zghal Riadi-GDL Laboratory, Manouba University, Tunis, Tunisia e-mail: hajer.baazaouizghal@riadi.rnu.tn A. Moreno ITAKA Research Group, University Rovira i Virgili (URV), Tarragona, Spain e-mail: antonio.moreno@urv.cat

[9] Monali Dey, Siddharth Swarup Rautaray "Study and Analysis of Data mining Algorithms for

Healthcare Decision Support System" Monali Dey et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (1) , 2014, 470-477 www.ijcsit.com 470

[10] Dharmendra K Roy And Lokesh K Sharma "Genetic K-Mean Clustering Algorithm For Mixed Number And Categorical Data Set.". International Journal Of Intelligence & Apllications (IJAIA). Vol. 1, No. 2, April 2010.