Review on Cognitive Analysis of Data Mining Tools Application in Healthcare Services

Swapna Bhavsar[1]*, Dr. Anil Badara[2]

[1] Phd Student, University of Technology, Jaipur

[2] PhD guide, University of Technology, Jaipur

Abstract — Healthcare is a thriving sector of the economy in many countries. There is a huge volume generation of data of in healthcare which includes patient information, patient treatment plan, other plans like health insurance. This paper is a review on cognitive analysis of data mining tool in heath sector. To dig out the important and helpful data which is created in massive quantity, data mining methods are used. Data mining is for providing planned pronouncement and generates a new point of view in healthcare data. The survey gives a summary of data mining methods and offer a very good decision making for health care experts. This paper surveys the cognitive analysis of data mining tools in healthcare.

Keywords — Cognitive analysis, Data mining tools, Data Analysis, Bigdata and Healthcare.

- x -

INTRODUCTION

From large data base to get the useful information data mining is approach is used and it provides motivation of research in many area like medical. In health data set for finding the valuable and unknown information and due to the requirement of analytical methodology the technique of data mining becoming more popular. Data mining introduced many advantages in the field of healthcare like identification of medical treatment techniques, at lower cost availability of medical solution, finding of reason of diseases and detection of the fraud in health insurance. For implementing health profiles of people, for making drug recommendation systems, and for building efficient healthcare policies data mining help the scientists in the healthcare field. Cognitive analysis means using data mining is a method which is done to mimic the human thinking so that it will become automatic. This technology is used in healthcare service alot. Internet of things (IoT) is increasing with the rapid developing considerations in numerous domains particularly in personal healthcare applications. Wearable

Body Area sensor Network (WBAN) based on IoT can help the patients who are suffering from chronic diseases like heart diseases and it monitors the patient condition and obtains the opinion from the expert. Many types of wearable medical devices are ready to use in the market. The accuracy of these devives are not upto the satisfactory level but they can be used for measuring the parameters. These can help in the detection of symptoms in the initial stages. Smart health monitoring systems consist of processing and analysis of the data obtained from the smart device [2]. These systems can continuously monitor the psychological and health conditions of the patients with sensed parameters. This WBAN data were gratified by the international and local law regulation and standard which is very confidential. Healthcare [3] organizations will guarantee the safe-keeping and ease of access of the data for the healthcare expert and to the patient. Comprehensively real time individual healthcare and activity data are having the vital task in data process and analyzing hence divers efforts are considered to make immense

and demonstrative illustrations of realtime world datasets [4]. Instances of these data sets are UbiqLog [5] and CrowdSignals [6], it consists the data from the devices like a smartphone, wearable device, etc. Biomedical signal datasets are found at PhysioNet [7] that allow the free web accessibility to gather the recorded physiologic signal data. Physio-Toolkit open-source software is used to collect the dataset. This data can also be obtained from UCI data repository [8]. This health-care data obtained from data repositories are used to mine valuable information by the application of appropriate data mining algorithms.

Most significant challenge of the data mining in healthcare is "to attain the superiority and appropriate medical data. It is complicated to obtain the accurate and inclusive healthcare data. Health data is intricate and varied in nature since it is taken from a variety of basis like from the health check reports of laboratory, from the conversation with patient or from the review of doctor. The data quality should be maintained for providing good service to patients. If data ware house is faulty it will not provide effective data mining.

LITERATURE SURVEY

This papers [9] provides a brief overview of data mining methods used in healthcare services. The privacy can be maintained by keeping patient decision support system which allows service provider to diagnose patient's disease without leaking any patient's historical medical data. The system model is divided into five parties: Trusted Authority (TA), Cloud Platform (CP), Data Provider (DP), Processing Unit (PU), and Undiagnosed Patient (PA). To prevent individual historic sensitive medical data to disclose from service provider. A new aggregation technique called additive homomorphic proxy aggregation (AHPA) scheme is introduced. To securely aggregate the message to solve the collusion problem, it contains the following six algorithms: KeyGen, ReKeygen, Encrypt, Decrypt, Re-encrypt & Agg, and Re-decrypt. This algorithm (AHPA) can be applied in our Disease Risk Prediction Application to avoid the disclosure of patient's sensitive medical data without compromising the privacy of data provider. As every information is

processed in the encrypted form our prediction application can achieve patient diagnose result in privacy preserving way.

In this paper [10], the author suggests the MHN architecture and privacy preserving data aggregation scheme. Qop can achieve authentication, guarantee integrity. the paper identifies the privacy requirements from the perspective of Qop. This paper describes the schema Encrypting the data prior to uploading it with symmetric encryption. This will be used to provide the privacy through symmetric encryption. In this technique the project uses various encryption techniques for data privacy. Few of them are: Encrypting the data prior to uploading it with some symmetric encryption; Using a Trusted Execution Environments (TEE) such as OS containers. Mainly the project is focusing on the first technique. It includes two type of encryption methods Symmetric and Asymmetric key encryption. In symmetric we make use of single key for both encryption and decryption process. Both sender and receiver have a copy of same key and the algorithm used is AES. In asymmetric we make use of two keys i.e. Public and Private keys. Public key is used for encryption and private used for decryption. Usually sender will have a public key and sender will have private key.

In this paper [11], the author suggests an idea about the privacy preserving scheme for medical data. It makes use of SVM algorithm which is one of the powerful classification algorithms in terms of prediction. This paper describes a model similar to our application in which data privacy is done using encryption and decryption techniques. This paper aims at the accurate results, low computation and data privacy which is similar to our application. The encryption and decryption methods can be validated in our Disease Risk Prediction Application to avoid the disclosure of patient's sensitive medical data also without compromising the privacy of data provider. Since the identity of the users is in encrypted form, privacy is preserved and also the user can get efficient medical pre-diagnosis results.

Swapna Bhavsar[1]*, Dr. Anil Badara[2]

According to [12] V. Ranjani and M. Durairaj defined several applications of data mining in the field of healthcare. And they work for decreasing the complexity of transactions in healthcare information. Many applications, tools, and algorithms that are already exist described by them.

A short description of data mining techniques is discussed by Sonali Agarwal and Divya Tomar in [13]. They give the introduction of association, classification, regression and clustering in healthcare field and also defined their benefits and drawbacks. Future issues, applications and challenges also included in their study.

By developing the Apriori algorithm of Association technique J.Jayaprakash and R.Karthiyayini [14] analyzed several result. about chronic diseases proving the precise information for public is the main objective of this paper. On the basis of user input symptoms to predict the disease an approach is presented by Pallavi Chitte, Priyanka Vijay Pawar and Megha Sakharam Walunj [15]. To demonstrate the efficiency of these approaches they create a prototype, this prototype tells the user about his or her disease. By proving remedial solutions and suggested doctors through data mining set it can predict the probable diseases.

By the use of Apriori data mining approach that is depend on association rules for a given period of time in particular geographical location for identifying frequency of diseases M.Pounambal, Gitanjali J and C.Ranichandra proposed a model [16].

In [17] authors used Linear regression technique used for multiple linear regression models.the objective is to help to diagnose liver patients and predicts the albumin level of the patient, that level helps the doctor to find the patient disease faster.the proposed method provides 89.34% accuracy . The calculation of albumin level can help out the doctor to lessen the time and effortlessly find the patients.

In [18] Big Data Predictive Analytics Model for Disease Prediction is developed using Machine learning Technique. The objective is to classify medical disease. Medical field depend on data mining technique, which is used to find the liver disease classification and examine the medical data. This model used M-PSO for selecting input variables and Modified Artificial Neural Network algorithm for classification of disease.the proposed method alculates the accuracy, sensitivity and performance and predicts the classification of liver diseases accurately compared to other techniques.

Big data Predictive Analytics Model is created [19] using Naive Bayes Technique (BPA-NB). predicts the upcoming heart health condition of patients. It predicts disease rate up to accuracy of 97.12%

A "combination of methods for finding heart disease Hybrid recommendation system for heart disease analysis which is based on multiple kernel learning with adaptive neuro-fuzzy inference system" is discussed [20] . Deep learning method is used with Multiple Kernel Learning with Adaptive Neuro-Fuzzy Interference System (MKL with ANFIS). MKL with ANFIS used for find the heart diseases. MKL used to classify the parameters between the heart patient and normal and this method calculates the sensitity, Mean Square Error (MSE) , and Specificity. This method produced 98% sensitivity, 99% specificity and 0.01 less Mean square error.

Another combinational model for the predicting Parkinson's Disease progression using machine learning techniques is explained [21].the method is used to predict Parkinson Disease (PD) through data sets. Here UPDRS used to find PD. To forewarn Total-UPDRS and Motor-UPDRS using Incremental support vector machine. It reduces the prediction computation time. The Mean Absolute Error (MAE) result obtained for Total-UPDRS and Motor-UPDRS are respectively 0.4656 and 0.4967.

An model for diagnosis of cancer using explained which is based on machine learning and uses hidden Markov model and GM clustering [22]. DNA sequencing is important for helping biomedical and healthcare services. Array based Comparative Genomic

Hybridization (CGH) measuring the DNA copy number, which is used to find the cancer disease. Bayesian hidden Markov Model(HMM) with Gaussian Mixture (GM). Big Data analytics plays an important role in DNA sequencing method : It is used to measured accuracy and error.

A decision tree based data mining method with neural network classifiers for prediction of heart disease is explained [23]. The gini index is used for prediction purpose. To predict the heart diseases using data mining (DM)technique. The classification with decision trees are calculated using this method and find the coronary illness. It is measured 93% prediction level of coronary disease.

To predict diabetes mellitus , a model is proposed which is based on data mining [24]. The accuracy of prediction level increases. This technique is used to increase the accuracy of prediction level. Here two algorithms used, that are K- means and Logistic regression. The result get 3.04% higher accuracy of prediction compared to other researches.

Adaptive weighted fuzzy rule-based system for the risk level assessment of heart disease is discussed [25]. It finds the exact treatment of heart diseases. It will give the accuracy classification of heart disease and this method is safe and very useful for physicians.

Different Preprocessing techniques for Neural Network Based Heart Disease Prediction is explained [26] . This method used to find the heart disease network and get high accuracy for disease prediction. Multi-Layer Pi-sigma Neuron Model (MLPSNM) used to find heart disease. This method gets 94.53% classification accuracy for finding the heart diseases for particular dataset.

Machine Learning Toward Infectious Disease Treatment is explained [27].Pathogens affects the public health system. This method helps to identify infectious disease. Data mining, data pre-processing, data analysis and machine learning techniques used for process. Machine learning methods used to find accurate prognosis and treatment of infectious diseases

A model which is based on Decision Tree is used for Prediction of Heart Disease is di scussed [28]. The decision tree and artificial neural network are combined to get high performance. Here, accuracy, sensitivity and specificity. Data mining technique used to reduce the number of tests for heart diseases. And hybridization technique is used. The result of accuracy 78.14%, Sensitivity 78%   and Specificity 22.9%.

A model which is based on Decision Tree is used for Prediction of Heart Disease is di scussed [28]. The decision tree and artificial neural network are combined to get high performance. Here, accuracy, sensitivity and specificity. Data mining technique used to reduce the number of tests for heart diseases. And hybridization technique is used. The result of accuracy 78.14%, Sensitivity 78%   and Specificity 22.9%.

"An Automated Diagnostic System for Heart Disease Prediction Based on Statistical Model and Optimally Configured Deep Neural Network" is explained [29]. The authors proposed the automated diagnostic system for finding the heart diseases.the proposed method is X2 statistical model with Deep neural network (DNN). It gives accuracy 93.34%, Specificity 91.83% and sensitivity 87.80%.

A model for predicting type 1 diabetic using neural network on artificial pancreas system is explained [30]. Artificial Pancreas system (APS) used to treat diabetic patients and the separate algorithms used for controlling blood glucose level (BGL). The details of BGL data, insulin injection and food intake are collected through APS. The Model is developed with Artificial Neural Networks (ANN). NN-MPC maintains the normal level of blood glucose at 90% with the mean absolute deviation of 4.7 mg/dl.

A method for heart disease classification is explained [31]. Classification methods are considered which uses back propagation neural network and logistic regression. This method is used to predict the heart diseases for Cleveland dataset 85.074% accuracy obtained for BPNN and 92.58% accuracy obtained for LR.

Swapna Bhavsar[1]*, Dr. Anil Badara[2]

To get better accuracy of heart disease calculation, an ensemble method is considered [32]. Improving the prediction accuracy of heart disease with ensemble learning and majority voting rule. Ensemble method is used to predict the accuracy of heart diseases and it is integrated U-healthcare monitoring system that helps the cardiologists and warns patients and healthcare workers for taking decision. It gives 88.88% accuracy of prediction.

A method for Predicting fatty liver disease using machine learning algorithms is explained [33]. FLD is prognosis using machine learning method and classify high-risk patients, prevent and treatment. Random forest (RF), Naive Bayes, ANN and Logistic Regression used to predict Fatty Liver Disease. Here, the accuracy of RF 87.48%, NB 82.65%, ANN 81.85%, LR 76.96%. The classification models to forewarn fatty liver disease.

A new method for detecting hepato cellular carcinoma patients is discussed which uses machine learning [34]. This method used to calculate the performance with many databases. Here many machine learning algorithms are used and the genetic algorithm helps to select algorithms and classifier parameter optimization. Machine learning method like normalization technique used to find Hepato cellular carcinoma (HCC). The value of accuracy 0.8849 and F1-score 0.8762

DATA MINING IN HEALTHCARE

There is a huge change in data in healthcare. To get accurate and good quality service the healthcare providers should have access to the perfect and up to date data. So that they can have a good decision making. Data mining is used to give planned pronouncement and generate a new point of view in healthcare data by digging out important and helpful data which is created in massive quantity . This is how the data mining methods help the healthcare professionals.

Table 1: Comparison of Different Methods

Sl. No.	Paper	Method	Objective
1	Predicting Liver Patient Data Using Linear Regression Technique	Linear Regression	This is used diagnose liver patients and predicts the albumin level of the patient, that level helps the doctor to find the patient disease
2	Big Data Predictive Analytics Model for Disease Prediction	PSO and Artificial Neural Network	Used to find the liver disease classification and examine the medical data
3	A Model for Liver Syndrome Classification	Naive Bayes	Predicts the upcoming heart health condition of patients.
4	A combined system for heart disease diagnosis based on multiple kernel learning	Deep learning	Classify the parameters between the heart patient and normal patients
5	A combined intelligent system for predicting Parkinson's Disease progression using machine learning techniques	Machine Learning Methods	Used to predict Parkinson Disease
6	A decision tree based classifier for predicting heart disease	Neural Network and Gini Index	To predict the heart diseases

CONCLUSION

This paper analyzes the cognitive analysis data mining methods in healthcare. Data mining helps the healthcare professionals to get the most precise data and provide very good services to patients. The paper gives an introduction to data mining techniques used in healthcare and provides a literature review on it. It can be recapitulated as Data Mining applications will enormously add to the healthcare sectors. This paper analyzes the data mining techniques and give a new perspective to the health professionals in terms of decision-making processes by presenting examples about the use of Data Mining in the health sector.

REFERENCES

1. Han, J.,Kamber, M.: "Data Mining Concepts and Techniques", Morgan Kaufmann Publishers, 2006

2. M. Rodgers, V. M. Pai, and R. S. Conroy, " Recent advances in wearable sensors for health monitoring ," IEEE Sensors J., vol.15, no.6, p. p. 31193126, Jun 2015.

3. R. Wu, G. Ahn, and H. Hu, "Secure Sharing of Electronic Health Records in Cloud," in Collaborative Computing: Networking, Applica- tions and Worksharing (CollaborateCom), 2012 8th International Con- ference, 2012, p. p. 711 - 718.

Swapna Bhavsar[1]*, Dr. Anil Badara[2]

4. Rawassizadeh, R., Kotz, D.: "Datasets for mobile, wearable and IoT research" GetMobile: Mob. Comput. Commun. 20(4), 57 (2017). https://doi.org/10.1145/3081016.3081018.

5. Rawassizadeh, R., Momeni, E., Dobbins, C., Mirza-Babaei, P., Rah- namoun, R "Lesson learned from collecting quantified self information via mobile and wearable devices." J. Sens. Actuator Netw. 4(4), 315335 (2015). https://doi.org/10.3390/jsan4040315

6. Welbourne, E., Tapia, E.M.: "CrowdSignals: a call to crowdfund the communitys largest mobile dataset." In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, UbiComp 2014 Adjunct, p. p. 873877. ACM, New York (2014). https://doi.org/10.1145/2638728.2641309 http://www.physionet.org.

7. Dua, D. and Graff, C. (2019). UCI "Machine Learning Repository [h]ttp://archive.ics.uci.edu/ml]." Irvine, CA: University of California, School of Information and Computer Science.

8. J.Kim, J.Lee, and Y.Lee, "Data - mining - based coronary heart disease risk prediction model using fuzzylogic and decision tree," HealthcareInform. Res., vol.21, no.3, p. p.167174, Jul.2015, doi:10.4258/hir.2015.21.3.167.

9. Ximeng Liu, Rongxing Lu, Jianfeng Ma, "Privacy Preserving Patient Centric Clinical Decision Support System on Naive Bayesian Classification", 2015

10. Kuan Zhang, Kan Yang, Xiaohui Liang, Zhou Su, "Security and Privacy for Mobile Healthcare Networks: From a Quality of Protection Perspective", 2015.

11. Yonglin Ren, Richard Werner Nelem Pazzi, And Azzedine Boukerche, "Monitoring Patients Via a Secure and Mobile Healthcare System ", 2010

12. International Journal Of Scientific & Technology Research Volume 2, Issue 10, OCTOBER 2013. "Data Mining Applications In Healthcare Sector: A Study"

13. Indian Institute of Information Technology, Allahabad, India. "A survey on Data Mining approaches for Healthcare"

14. Assistant Professor, Department of Computer Applications, Anna University, Trichy, India P.G. Student, Department of Computer Applications, Anna University. "Association Technique on Prediction of Chronic Diseases Using Apriori Algorithm"

15. Department Of Computer Engineering, Ramrao Adik Institute of Technology Nerul,Navi Mumbai. "Estimation based on Data Mining Approach for Health Analysis"

16. School of Information Technology and Engineering, VIT UNIVERSITY, Vellore-632014, Tamil Nadu, India. "APRIORI algorithm based medical data mining for frequent disease identification"

17. Garg, Deepankar, and Akhilesh Kumar Sharma. "Prediction and Analysis of Liver Patient Data Using Linear Regression Technique" Advances in Machine Learning and Data Science. Springer, Singapore, 2018. 71 - 80.

18. Venkatesh, R., C. Balasubramanian, and M. Kaliappan Journal of medical systems 43.8 (2019): 272.

19. Anand, L., and SP Syed Ibrahim "HANN: A Hybrid Model for Liver Syndrome Classification by Feature Assortment Optimization" Journal of medical systems 42.11 (2018): 211.

20. Manogaran, Gunasekaran, R. Varatharajan, and M. K. Priyan. "Hybrid recommendation system for heart disease diagnosis based on multiple kernel learning with adaptive neuro - fuzzy inference system" Multimedia tools and applications 77.4 (2018): 4379 - 4399.

21. Nilashi, Mehrbakhsh, et al. "A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques" Bio-

Swapna Bhavsar[1]*, Dr. Anil Badara[2]

cybernetics & Bio-medical Engg., 2018, 01 - 015.

22. Manogara, Gunasekara, et al. "Machine learning based big data processing framework for cancer diagnosis using hidden Markov model and GM clustering" Wireless personal communications 2018: 02099 - 02116.

23. Manthan, et al. "A novel Gini index decision tree data mining method with neural network classifiers for prediction of heart disease" Design Automation for Embedded Systems 2018: 0225 - 0242.

24. Wu, Han, et al. "Type 2 diabetes mellitus prediction model based on data mining, Informatics in Medicine Unlocked" 010 (2018): 0100 - 0107.

25. Paul, Animesh Kumar, et al. "Adaptive weighted fuzzy rule - based system for the risk level assessment of heart disease" Applied Intelligence 48.7 (2018): 1739 - 1756.

26. Burse, Kavita, et al. "Various Preprocessing Methods for Neural Network Based Heart Disease Prediction Smart Innovations in Communication & Computational Sciences" Springer, 2019. 055 - 065.

27. Bhardwaj, Tulika, and Pallavi Somvanshi. "Machine Learning Toward Infectious Disease Treatment" Machine Intelligence and Signal Analysis. Springer, Singapore, 2019. 683 - 693.

28. Maji, Srabanti, and Srishti Arora. "Decision Tree Algorithms for Prediction of Heart Disease" Information and Communication Technology for Competitive Strategies. Springer, Singapore, 2019. 447 - 454.

29. Ali, Liaqat, et al. "An Automated Diagnostic System for Heart Disease Prediction Based on Statistical Model and Optimally Configured Deep Neural Network" IEEEaccess 07 (2O19): 034938 - 034945.

30. Bahremand, Saeid, et al. "Neural network - based model predictive control for type 1 diabetic rats on artificial pancreas system" Medical & Biological Engineering & Computing 57.1 (2019): 177 - 191.

31. Desai, Shrinivas D., et al. "Back - propagation neural network versus logistic regression in heart disease classification" Advanced Computing and Communication Technologies. Springer, Singapore, 2019. 133 - 144.

32. Raza, Khalid. "Improving the prediction accuracy of heart disease with ensemble learning and majority voting rule" U - Healthcare Monitoring Systems. Academic Press, 2019. 179 - 196.

33. Wu, Chieh - Chen, et al. "Prediction of fatty liver disease using machine learning algorithms-Computer methods and programs in bio-medicine" 1.70 (2019): 023 - 029.

34. Książek, Wojciech, et al. "A novel machine learning approach for early detection of hepatocellular carcinoma patients" Cognitive Systems Research 54 (2019): 116 - 127

Corresponding Author

Swapna Bhavsar*

Phd Student, University of Technology, Jaipur