

Data Mining Technique Based Critical Disease Prediction in Medical Field

Preetha J ^{a,1}, Raju S ^b, Abhishek Kumar ^c, Sayyad Samee ^d, and Vengatesan R^e

^a*Dept of CSE, Muthayammal Engineering College, Rasipuram, India*

^b*Dept of IT, Mahendra Engineering College, Mallasamudram*

^c*Dept of CSE, Banaras Hindu University, Varanasi, India*

^d*AlSadara, Al Ain, United Arab Emirates*

^e*Dept of CSE, Sanjivani College of Engineering, Kopergaon*

Abstract. In the present days' deaths because of some critical disease has become a significant issue in the medical field. Data mining is one of the significant territories of research that is famous in wellbeing associations. Data mining has a functioning job for finding new patterns and examples in the healthcare association which is valuable for every one of the gatherings related to this field. The medical dataset has heterogeneous data as numbers, content, and pictures that can be mined to convey an assortment of helpful data for the physicians. The examples picked up from the medical data can be helpful for the physicians to find diseases, foresee the survivability of the patients after disease, the seriousness of diseases and so forth. The focal point of this paper is to break down the utilization of data mining in medical space and a portion of the systems utilized in critical disease prediction. We have completely reviewed many research papers of data mining identified with some critical disease prediction.

Keywords. Data mining, critical disease, medical dataset.

1. Introduction

Healthcare informatics data incorporates clinic subtleties, patient's subtleties, disease subtleties and treatment cost. These immense data are produced from various sources and configuration. It can have unessential qualities and missing data. Applying data mining methods is a key way to deal with extricate information from huge disease data. Data mining has different techniques to extricate information from tremendous disease data set. Data mining methods like characterization, grouping and rule mining can be utilized to break down data and concentrate important data. A portion of the significant momentum utilizations of data mining in health care incorporates anticipating the future results of diseases dependent on past data gathered from comparative diseases, analysis of disease dependent on patient data, breaking down treatment expenses and request of assets, pre-processing of boisterous, missing data and limiting an opportunity to hang tight for the disease finding. Data mining tools like Weka, Rapid digger and Orange are utilized to investigate and foresee better outcome for health care data. New and flow data mining tools and advancements are utilized in disease finding and health care informatics to improve the health care benefits in practical way and limiting the ideal opportunity for disease determination.

¹Preetha J, Dept of CSE, Muthayammal Engineering College, Rasipuram, India.

Email : unpraju_cse@acoee.edu.in

2. Data Mining Algorithms

Different calculations plan because of various research jobs at data mining. These methods are clearly used for creating structures or to discover critical derivations and ends from the came about a dataset. Different surely understood strategies are K-means, ANN, Naïve Bayes, SVM and so on are examined.

2.1. Classification

It is a machine learning-based data mining strategy. It is used to describe each data in a lot of data into one of pre-defined set of crowds or classes. It settles on utilize numerical methods, for example, choice trees, direct programming, neural network and insights to order the data into various gatherings. Present-day classification procedures give increasingly keen techniques to effective prediction of diseases. Various kinds of taxonomy systems incorporate linear and non-linear relapse, SVM, decision tree, discriminant examination and naive based.

2.2. Clustering

Clustering is a data mining system that causes a cluster of articles that to have comparable trademark utilizing programmed strategy. Clustering method characterizes the classes and put protests in them where the class isn't predefined. Various sorts of group methods incorporate Fuzzy Cmeans(FCM), K-means,Rough-Fuzzy Cmeans (RFCM), Rough C-means (RCM), Robust RFCM (rRFCM), Gaussian mixture and hierarchical.

2.3. Association rule mining

Association rule learning is a mainstream and very much explored technique for finding fascinating relations between various data with regards to huge databases. It is proposed to distinguish well-manufactured rules found in databases utilizing various strategies of significance dependent on input data set.

Decision Support: This tool uses a tree like model or chart of choices with the goal that potential outcomes which incorporate possibility occasion results and utility. This tool is one of the normal way to deal with show calculations. These are commonly used for explore tasks increasingly explicit in investigation of choices to help and assess a method that can accomplish the goal.

2.4. Naïve Bayes Algorithm

An easiest system for building classifiers and a probabilistic classifier relies upon Bayes' hypothesis. Bayes hypothesis is determined as follows: $P(C|X) = P(X|C) * P(C)/P(X)$, where C is the class to such an extent that P(X) is consistent for all classes and X is data tuple.

2.5. Support Vector Machine

SVM is utilized for getting to data and example for relapse investigation and classification. The goal is to discover most exact classification capacity to perceive people from the two classes in the preparation dataset. It relies upon numerical capacities and used in model complex, and genuine issues. It functions admirably on data sets that have numerous traits and guarantees that the best such capacity is found by augmenting the edge between the two classes. It separates data via looking through the best hyper plane that partitions all data purposes of one class from those of the different class.

3. Data Mining Tools

Tools give prepared applications to be utilized for mining calculations. They have a simple to utilize interface and scientists can undoubtedly utilize them on account of free open source programming. Mainstream data mining tool is ORANGE, WEKA, and MATLAB etc.

3.1. Critical disease with DM in the medical field

Medical science is another field where huge measure of data is created utilizing distinctive clinical reports and other patient indications. Disease prediction assumes a significant job in data mining. Finding of a disease requires the presentation of various tests on the patient. Be that as it may, utilization of data mining methods, can lessen the quantity of tests. This decreased test set assumes critical job in execution and time. Health care data mining is a significant assignment since it enables specialists to see which traits are progressively significant for finding, for example, age, weight, indications and so forth. This will enable the specialists to analyze the disease all the more effectively. Picture mining strategies improved the disease prediction assignment and settles on health care basic leadership simpler.

3.2. Asthma Disease

There is a constant need to distinguish factors related with asthma in light of the fact that the predominance of asthma is rising. Different elements, for example, prosperity, stationary way of life, ecological tobacco smoke (ETS), youth viral contaminations and air contamination, have been proposed to be significant in the pathogenesis of asthma. PeymanRezaei and Hachesu (2017) tests comprise of 600 alluded patient with asthma disease. Data were gathered dependent on the examination's incorporation criteria. Preprocessing was performed and different calculations incorporate DT, SVM, KNN and Naïve Bayesian was surveyed.

3.3. Heart Disease

Cardiovascular diseases (CVDs) are disarranges of the heart and veins and incorporate cerebrovascular disease, coronary heart disease, rheumatic heart disease and different conditions. Four out of five CVD passings are because of heart attacks and strokes. People in danger of CVD may exhibit raised pulse, glucose, and lipids just as overweight and stoutness. Chaitrali S. Dangare et.al (2012) has dissected prediction

frameworks for Heart disease utilizing progressively number of info traits. The data mining classification procedures, to be specific DT, NB, and NN are examined on Heart disease database. The exhibitions of these methods are analyzed, in light of exactness.

4. Literature review

AvinashGolande et.al, (2019) suggested that the heart disease has become a significant issue around one individual bites the dust every moment because of heart disease. This is thinking about both male and female class and this proportion may shift as indicated by the locale likewise this proportion is considered for the individuals of age bunch 2569. This doesn't demonstrate that the individuals with other age gathering won't be influenced by heart diseases. This issue may begin in early age bunch additionally and anticipate the reason and disease is a significant test these days. Here in this paper, talked about different calculations and tools utilized for prediction of heart diseases.

Table 1. Data mining techniques used for diagnosis of some critical diseases

Disease	Author	Year	Technology
Heart Disease	MontherTarawneh et al	2019	Naïve bayes, SVM
Heart Disease	T.Nagamani et al	2019	Mapreduce
Heart Disease, Breast Cancer, PimaIndian Diabetes	M.Akhil et al.	2012	GA & Associative Classification
breast cancer, heart disease	Mohammad et al	2012	C5.0 C4.5
Diabetes, heart diseases	Humar et al.	2008	FNN, Back propagation, Classification,

Koh& Tan, et.al, (2011) expressed that the healthcare business creates a lot of complex data on patients, emergency clinic assets, and finding of diseases, electronic patient records and medical gadgets. Increasingly bounteous measures of data are a fundamental asset for data mining. There is a huge potential in healthcare data mining applications, and the absolute most critical applications in healthcare data mining are prediction and finding, treatment effectiveness, healthcare the executives, extortion and misuse, client relationship the board, and the medical gadget industry. Cinetha and Maheswari (2014) proposed a Decision Support System which predicts the probability of heart disease danger of patients for the following ten years utilizing Fuzzy Logic and DT. This model predicts with 97.67% evaluated exactness. Devi and Anto (2014) proposed a transformative fluffy master framework for the finding of coronary supply route disease dependent on a dataset with a sum of 303 records and 14 properties. In [12] executed a correlation of heart disease analysis with the assistance of Decision Tree and Naive Bayes. The outcomes show that the precision of NB and the DT is 85.03 % and 84.01 %.

5. Conclusion

Data mining systems helps in finding the shrouded information in a gathering of disease data that can be utilized to dissect and anticipate the future conduct of diseases. Classification is one of the data mining methods that assigned a class name to a lot of unclassified cases. The primary goal of this paper is to look at the data mining tools based on their classification precision. In the future, more disease datasets can be utilized for classification methods and other data mining systems can be utilized to survey and examine the presentation of different data mining tools.

References

- [1] V.D.Ambeth Kumar, G.Gokul, S.Malathi, K.Vengatesan, D.Elangovan, B.Chitra, "Implementation Of The Pulse Rhythmic Rate For The Efficient Diagonising Of The Heart Beat", ", Healthcare Technology Letters (IET) 2019 Apr 17;6(2):48-52.
- [2] Bhatla, N., & Jyoti, K. (2012). An analysis of heart disease prediction using different data mining techniques. *International Journal of Engineering*, 1(8), 1-4.
- [3] Chaitrali S. Dangare, Sulabha S. Apte, "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques", *International Journal of Computer Applications* (0975 –888), Volume 47– No.10, June 2012, page no 44-48
- [4] Koh, H. C., & Tan, G. (2011). Data mining applications in healthcare. *Journal of healthcare information management*, 19(2), 64-72.
- [5] PeymanRezaei-Hachesu, TahaSamad-Soltani, RuhollahKhara, Mehdi Gheibi, NazilaMoftian, Prediction Of Asthma Control Levels Using Data Mining Methods: An Evidence-Based Approach, 5th International Society for Evidence-Based Healthcare Congress, Kish Island, IranOral., February 2017.
- [6] Ratmakar, S., Rajeswari, K., & Jacob, R. (2013). Prediction of heart disease using genetic algorithm for selection of optimal reduced set of attributes. *International Journal of Advanced Computational Engineering and Networking*, 1(2), 51-55.
- [7] Ambeth Kumar, V. D., Malathi, S., Venkatesan, R., Ramalakshmi, K., Vengatesan, K., Ding, W., & Kumar, A. (2019). Exploration of an innovative geometric parameter based on performance enhancement for foot print recognition. *Journal of Intelligent & Fuzzy Systems*, 1–16. <https://doi.org/10.3233/jifs-190982>
- [8] Singh, V. K., Singhal, A., Rai, K. N., Kumar, A., & Dwivedi, A. N. D. (2019). Randomized key-based gmo-bcs image encryption for securing medical image. *International Journal of Recent Technology and Engineering*. <https://doi.org/10.35940/ijrte.C4453.098319>
- [9] Rai, A. K., Agarwal, S., & Kumar, A. (2018). A novel approach for agile software development methodology selection using fuzzy inference system. *Proceedings of the International Conference on Smart Systems and Inventive Technology, ICSSIT 2018*. <https://doi.org/10.1109/ICSSIT.2018.8748767>
- [10] Kesavan, S., Kumar, E. S., Kumar, A., & Vengatesan, K. (2019). An investigation on adaptive HTTP media streaming Quality-of-Experience (QoE) and agility using cloud media services. *International Journal of Computers and Applications*. <https://doi.org/10.1080/1206212X.2019.1575034>
- [11] Kumar, A., Vengatesan, K., Rajesh, R., Parthibhan, M., & Singhal, A. (2018). Review of gene subset selection using modified k-nearest neighbor clustering algorithm. *Proceedings of the International Conference on Smart Systems and Inventive Technology, ICSSIT 2018*. <https://doi.org/10.1109/ICSSIT.2018.8748667>
- [12] Vengatesan, K., Kumar, A., Naik, R., & Verma, D. K. (2019). Anomaly based novel intrusion detection system for network traffic reduction. *Proceedings of the International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), I-SMAC 2018*. <https://doi.org/10.1109/I-SMAC.2018.8653735>
- [13] S.Hema Kumar, J.Uday Kiran, V.D.Ambeth Kumar, G.Saranya, Ramalakshmi V, "Effective Online Medical Appointment System", *International Journal of Scientific & Technology Research*, Volume 8, Issue 09, September 2019, Pages 803 – 805.
- [14] SellappanPalaniappan, RafiahAwang .Intelligent Heart Disease Prediction System Using Data Mining Techniques .IJCSNS *International Journal of Computer Science and Network Security*, aug 2008 ; Vol.8 No.8.