International Journal of Engineering & Technology, 7 (2.32) (2018) 97-99

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper



Analysis Of Single And Hybrid Data Mining Techniques For Prediction Of Heart Disease Using Real Time Dataset

[1]syed Ahmed Yasin, [2]Dr.P.V.R.D.Prasad Rao

[1,2]Department of CSE, K L E F, Guntur, Andhra Pradesh, India – 522502
**Corresponding author E-mail: syedahmed.cse@gmail.com*

Abstract

Data mining and healthcare has strong relations as data mining is a process where we can analyze enormous set of data and after extraction meaning of data can be understood. As medical data is in bulk which leads to the need of data analytics tool for extraction of useful information or knowledge .Disease Prediction is one of the applications where many researchers are working so that they can answer whether to apply single or compound data mining technique. Heart related disease is the major cause of deaths in past decades. Use of single technique in treatment of heart disease is showing variable accuracy. Current leading research study is all about to check the effect of compound data mining technique will show enhanced results in the prognosis of heart related disease or not, but the major issue arises of identification a suitable treatment for heart disease patient. This survey paper notify gap in research findings on heart disease prognosis and prescription data and proposes a model which will fill the gap and shows enhanced accuracy.

Keywords:*Compound Data Mining, Data Mining, Disease Prognosis, Data Analytics, Heart Disease, Single Data Mining Technique..*

1. Introduction

Information mining is a procedure to extricate learning from past, covered up, already obscure examples, relations and learning that are hard to recognize with conventional factual strategy. Information mining hast tremendous application region, for example, examination of money related foresight, regardless of whether and medicinal services. [1] Prognosis of any malady is a standout amongst the most requesting zones where information mining method can give 360 subtle elements of therapeutic dataset. Learning removed can give collaborator to human services association to settle on quality choices, specialists to settle on choices in their training treatment, arranging hazard analysis,diagnoisis,predication like early disclosure of any disese,predecting future conduct of patient knowing current indications ,diminish unfavorable medication impact i.e. restorative information mining causes doctors to set up a few techniques for conclusion ,anticipation and treatment Data digging gives various approaches to guess like regulated learning, unsupervised learning group learning, and compound characterization. Gathering learning consolidates homogeneous machine approaches utilizing different blending technique its will probably get change in precision over uniclassifier demonstrate. [2] Compound systems join heterogeneous machine learning approaches. There are two kinds of learning philosophy directed and unsupervised learning. Both compound and outfit strategy demonstrates upgraded exactness over single grouping model. Late examination demonstrates some downside of outfit technique. For example, 1. Ensemble method does not enhance exactness dependably of model but rather tends to build error of every individual base classifier.2. Utilization of various classifier make yield complex which is hard for analysis.3. It's unrealistic for all outfits to enhance exactness of every classifier for all dataset.

2. Need

The significant issues looked by health care are

1. How to Improve work efficiency of the physician
2. How to provide quality service at moderate cost
3. Accuracy of results with less number of test.

Medical Services are regularly in light of doctor solution and experience as opposed to the shrouded information which contain heft of data in the database. Such practice causes, therapeutic mistakes and unnecessary costs which influences the nature of administration given to patients. It's discovered that coordination of therapeutic DSS with PC based patient information records could lessen restorative errors, tolerant security change, diminish undesirable practice variety, and enhance persistent result comes about. When we observe traditional method of diagnosis it's found that availability of experts is not possible everywhere and every time, Dependency of diagnosis is on doctor's experience and talent, difficult to establish multivariable relationship. We are getting intelligent diagnosis by doctors but above mention complication are major hurdles in Quality of service for health care. From writing overview it's seen that, one average errand in information mining and KDD is the choice of best classifier for specific dataset in light of the fact that every classifier display has certain preferred standpoint and disservice in predicating the class, and from review it's prominent that compound model performs over outfit and single model. Diverse kind of heart issue may have comparative indications. In India and lion's share of world most extreme deaths are because of Heart related issues. The foundation of pharmaceutical announced that around 98000 individuals kick the bucket in doctor's facilities consistently because of medicinal mistakes and around 29$ billion is spent each year I settling that.Motiviated by above history of death rate of coronary

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

illness persistent every year and nearness of substantial measure of recorded information from which to remove helpful data, we can utilize information mining approaches predication which can help as second assessment for human services proficient in finding.

This proposed architecture will be useful for giving elective choices specialists in conclusion and giving fitting treatment to coronary illness analysis and treatment [3].Information mining strategies, for example, Neural Network, naive bays, decision tree, kernel density, support vector machine indicates distinctive level of exactness. Applying prescient system in conclusion is useful yet less work is done in distinguishing treatment get ready for quiet particularly for heart related malady. Investigation demonstrates that healing facilities don't serve same nature of administration despite the fact that they giving same kind of administration. Concentrate likewise propose that applying information mining strategy in distinguishing proof of compelling treatment for patient can enhance specialists execution, Many researchers utilizes single as well as multiple information mining procedure in finding of coronary illness for recognizable proof of dependable precision [7].Investigation additionally propose that there is no past work that distinguishes exactness in recognizing suitable treatment for heart related sickness. To decrease treatment time and exactness change it has turned out to be requesting issue to plan a dependable and effective medicinal finding framework to help the social insurance framework .i.e. moving from treatment based practice to result based practices. Human services determination is confused and troublesome errand and ought to be done proficiently. The majority of the analysis is completed in view of doctor understanding and exhortation as opposed to concealed learning in database. This strategy for training may cause abundance restorative cost and blunders which can influence trust of patient on specialists and doctor's facilities and furthermore influence nature of therapeutic administration gave to understanding thus information mining techniques can enhance the nature of medicinal sickness altogether. Past Study demonstrates that overall human services has 3 noteworthy test cost, quality and reach.

1. At Operative cost how to do treatment?
2. The most effective method to enhance the quality i.e. enhance analysis, better result from quiet treatment
3. The most effective method to broaden the scope of social insurance. What is coronary illness and why treatment is essential?

3. Motivation

1. To anticipate whither single information mining calculation will demonstrate expected exactness in distinguishing proof of guess medicines for HD patients?

2. To anticipate which specific single information mining calculation will demonstrate most noteworthy precision in distinguishing proof best medicines for HD patients?

3. To foresee whither compound information mining calculation demonstrate most elevated exactness than single information mining calculation distinguishing proof of guess medicines for HD patients?

4. To anticipate which specific compound information mining calculation will indicate most astounding exactness in distinguishing proof of best medicines for HD patients?

Coronary artery disease is one of the type of cardiovascular disease in which detection of atherosclerotic plaques in coronary arties leads to heart attack or sudden death. For diagnosis positive signs of heart disease and to assess the level of damage of heart muscles, certain test may be prescribed by a medical practitioner including nuclear scan, angiography, ECG, exercise stress testing, ECG is a noninvasive technique used to identify CAD cases, though it could leads to undamaged symptoms of CAD. This limitations leads to angiography which is an invasive diagnosis to confirm CAD cases and is considered as gold standard for disease detection and severity analysis however it may cost more and need high level of technical expertise ,Researchers are therefore seeking less expensive and effective alternatives say using data mining for predicating CAD cases.ECG and echo cardiogram are specialized process conducted by trained practioners,sometimes ECG is not able to confirm CAD cases this process is complex ,costly and involves lots of time and effort. These methods are non-invasive, low complexity level, low cost, reporucdicible and objective diagnosis can do automated detection of disease and can be used for large number of patient based on medical data can be obtained from hospitals.

Here the proposed model woks for data mining technique for heart disease treatment data can provide reliable performance as that achieved in diagnosis heart disease patient. They overall objective of this paper is to make a literature review data mining technique used for disease predication and analyze the performance of compound model over single classification model, ensemble model on different medical data set. Also survey of different single and compound data mining techniques used for heart disease prediction and level of accuracy shown by the method. Table 2 summarizes techniques and data set used by different authors and researchers and their corresponding results.

4. Existing System Study

Various techniques are used for predication of healthcare issues, different authors has used single, multiple and ensemble techniques which shows variable accuracy. Online repository as well as use of real hospital dataset is done. Table no 2 shows summary of literature review regarding techniques or approach used for prediction and tools used. Table also shows the summary of results obtained and accuracy. Many Researchers have opted for single data mining technique and some has utilized hybrid or compound techniques. Few researchers has used ensemble method. Each technique shows individual results and predictive modeling can be done with machine learning techniques.

Paper No	Approach/Techniques, Data Sets used	Results &Accuracy
1	Use of KEEL TOOL & Cleveland	Performance improvement is possible by mixing a mining technique [1]
2	Use Of privacy preserving techniques is used	Performance measurement is done, Results are better. [2]
3	Use of Radial bias function Gaussian Function	Highest accuracy shown by RBFN [3]
4	Applicability of RBF is done. Testing of 75 samples are done	RBF showing satisfactory results as compared to other technique. [4]
5	DMS-PSO technique is used Diagnosis idea is given by selection of critical attribute	Accuracy can be increased by combining of DMS-PSO and fuzzy logic [5]
6	Use of Apache Hadoop clusters, Google Cloud Storage. Map reduce based Classification via clustering, C 4.5 Decision tree, Analysis is done by WEKA tool.	Accuracy of 98.5% is achieved Tool Used : WEKA [6]
7	Multilayer & Multinomial perceptron with logistic regression model, Use of 1. FURIA 2.C4.5	Advantage by this models it needs few data for training and testing. Improvement in accuracy is possible by use of more data instances. [7]
8	Following techniques are used1. Neural Network 2.Decision Tree3. Naive Bayes, 4.Support Vector	NB and NN shows highest accuracy.Back propagation is used for nonlinear relationship ,Use of Genetic algorithm which is based on mutation and crossover. [8]
9	Use of, PSO, F score, LDA, rough sets , Simulated annealing	Improvement in accuracy is observed [9]
10	MO, C4.5, MLP and FFBP are used.	PSO algorithm shows accuracy of 91.94%,Cost reduction is possible [10]
11	Naïve Bayes, Decision tree and Support vector machine data Set of 13 attributes are used	Multi disease predication is possible, level of accuracy is high [11]

5. Proposed Method

Following architecture shows how single and hybrid data mining techniques can be utilized for predication of heart disease diagnosis

and suitable treatment based on given input symtoms.The proposed idea is firstly try for single data mining algorithm on disease data and check for technique accuracy, then apply same data mining technique on disease treatment data to test its accuracy.Similairly We can apply Hybrid or compound data mining technique on disease diagnosis data to test its accuracy and then applying same hybrid data mining technique on treatment data to test its accuracy. From this analysis we can find out which data mining technique is best for predication as well as which medicine or treatment is suitable form real time data set.

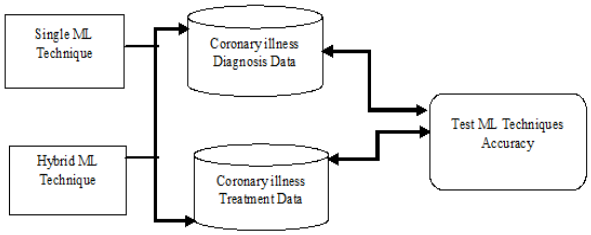


Fig1:. Proposed Architecture showing how single and hybrid machine learning technique can useful for knowledge extraction

6. Conclusions

Utilization of more than one mining techniques can show accurate results for comparing the performance in the prognosis of heart related disease.

1. Training data set is used From UCI (University of California Irvine) Machine learning dataset or other online data set.

2. Need of Real time data, because validity of algorithm implementation is based on validity of collected data set.
This survey paper depicts the significant advances in health sector alongside dialog about well-known methods, and database, alongside subtle elements of single and compound information mining procedures utilized. As of late. The primary target of this paper is to think about different methods utilized for predication of coronary illness and contrast the procedures with discover best strategies from them.

This review paper describes the major technologies in healthcare along with discussion about popular techniques, tools and database, along with details of single and compound data mining techniques used. In recent years. The main objective of this paper is to study various techniques used for predication of heart disease and compare the techniques to find out best techniques from them.

Acknowledgment

First of all, I would like to thank my Research Supervisor Dr.P.V.R.D Prasad Rao, Professor, Department of Computer Science and Engineering (CSE), KLEF for his continuous guidance and support without which this paper writing would not have been possible. I would like to thank Dr.V.Kirshna Reddy, Chairman of the Doctorial Committee and Dr.P Siva Kumar and Dr.P. Raja Rajeswari, Members, Doctorial committee, KLEF for their valuable guidance and constant support throughout my work.

References

[1] Purushottam, K. Saxena and R. Sharma, "Efficient heart disease prediction system using decision tree," International Conference on Computing, Communication & Automation, Noida, pp. 72-77,2015.
[2] PVRD Prasad Rao and E.Rudrani "Privacy Preserving for Hospital Data Using Earth Mover's Distance "Published in Scopus indexed International Journal of Applied Engineering Research in May 2014 No 15, PP. 2799-2808.
[3] Shaikh Abdul Hannan, AV Mane, RR Manza, RJ Ramteke," Prediction of heart disease medical prescription using radial basis function" IEEE International Conference on Computational Intelligence and Computing Research,pp 1-6,2010.
[4] Masilaslsam, Dr.Saad Amin, "Diagnosis of Heart disease by using radial basics function network classification on patient medical record". RF and Wireless Technologies for Biomedical and Healthcare Applications (IMWS-Bio), 2014.
[5] M.A.JabbarB.LDeekshatuluPriti Chandra, "Computational Intelligence Technique for early Diagnosis of Heart disease" IEEE International Conference on Engineering and Technology (ICETECH), 2015.
[6] M.A.Jabbar B.L Deekshatulu Priti Chandra "Alternating decision trees for early diagnosis of heart disease" International Conference on Circuits, Communication, Control and Computing (I4C), 2014.
[7] Shivani Batra,ShellySachdevan ,"Organizing standardized electronic Healthcare records data for mining" Elsevier, Volume 5, Issue 3, Pages 226-242,September 2016.
[8] FurqanAlama, Rashid Mehmoodb,Iyad Katiba, Aiiad Albeshria "Analysis of Eight Data Mining Algorithms for Smarter Internet of Things (IoT)" Elsevier, Pages 437-442,Volume 98, 2016
[9] Saba Bashir, Usman Qamar, M.Younus "An Ensemble based Decision Support Framework for Intelligent Heart Disease Diagnosis" International Conference on Information Society (i-Society), 2014.
[10] Animesh Kumar Paul, Pintu Chanda Shill, Md. Rafiqul Islam Rabin "Fuzzy Membership Function Generation using DMSPSO for the Diagnosis of Heart Diseases",18th International Conference on Computer and Information Technology (ICCIT), 2015.
[11] Majid Ghonji Feshki, OmidSojoodi Shijani "Improving the Heart Disease Diagnosis by Evolutionary Algorithm of PSO and Feed Forward Neural Network" International Conference On Artificial Intelligence and Robotics (IRANOPEN), 2016.
[12] Kulwinder Singh,Navjot Kaur "Cloud Deployable Health Data Mining using Framework for CDSS" International Conference and Workshop on Computing and Communication (IEMCON), 2015.
[13] Luxmi Verma,Sangeet S,P.C Negi "A Hybrid Data Mining Model to Predict Coronary Artery Disease cases using Non Invasive Clinical Data "International Journal of Journal of Medical Systems,2016.
[14] Ankita Dewan ,Meghana Sharma Fatemeh Nematik,Hedieh "Prediction of Heart Disease using a Compound Technique in Data Mining Classification compound data mining model of feature selection algorithm and ensemble learning classifier for credit scoring" 2nd International Conference on Computing for Sustainable Global Development (INDIA Com), 2015