International Journal of Advanced Technology and Engineering Exploration, Vol 3(19)
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
http://dx.doi.org/10.19101/IJATEE.2016.319003

Review Article

Review and computational meta-analysis for heart disease prediction

Megha Manohar Darne[1*], Shiv K. Sahu[2] and Amit Mishra[2]
M.Tech Scholar, TIT, Bhopal[1]
Assistant Professor, TIT, Bhopal[2]

Abstract
World is regular to eyewitness a change in deaths apropos in case of heart diseases. Ancient age ascertaining may count the ruin due to the heart diseases. In this paper a detail analysis and discussion on the related technique have been presented for heart disease prediction so that we can prevent it in the earlier stages. For fulfilling the above objective data mining and optimization techniques have been considered for review. The data mining and optimization techniques can help in preparing an intelligent framework for heart disease prediction.

Keywords
Data mining, Heart disease factors, Optimization techniques, Early stage detection.

1.Introduction
The goal of data mining is to extract the intend of data mining is to abstract acquaintance newcomer disabuse of hefty amount of Text [1]. Data mining is an interdisciplinary district, whose debased is at the standpoint of machinery way of life, statistics, and databases [2-8].

To explain and analyse the use data mining to emphasize to discover knowledge that is not only accurate, but also comprehensible for the user [9]. In [10] Comprehensibility is important whenever discovered knowledge will be used for supporting a human decision. After all, if discovered knowledge is not comprehensible for a user, it will not be possible to interpret and validate the knowledge.

Data mining is the nontrivial extraction of implicit, previously unknown, and potentially useful information from data. This encompasses a number of different technical approaches, such as clustering, data summarization, learning classification rules, finding dependency networks, analysing changes, and detecting anomalies [11]. The found data and learning are helpful for different applications, including market investigation, choice bolster, misrepresentation location, and business administration. Numerous methodologies have been proposed to concentrate data, and mining medicinal information is an essential field [12].

*Author for correspondence

Streamlining has leeway that we can accomplish great results or improved results regardless of the fact that the mathematical statement is not accessible. Then again, hybridization of enhancement systems has been proposed to exploit qualities of diverse strategies. There are diverse types of hybridization. One structure is incorporating segments from diverse strategies. Numerous reconciliation plans have been proposed. A well-known plan is to insert the nearby hunt into the system of Population-based strategies, including transformative calculations, insect settlement techniques, and so on [13]. In addition, utilizing the gathering and volatilization normal for pheromones, it can reproduce the change of recurrence of Item set. Some other related techniques have been discussed in [14] and [15]. Positive and negative association rules mining procedures have been discussed in [16] and [17]. It shows the impact and power of data mining. It is also discussed in [18] and [19]. So in this paper combined techniques for better heart disease factors prediction have been analysed.

2.Literature review
In 2008, Palaniappan, Set al. [20] proposes about social insurance industry which gathers tremendous measures of human services information which, lamentably, are not "; mined"; to find shrouded data for compelling choice making. Revelation of concealed examples and connections frequently goes unexploited. Propelled information mining methods can cure this circumstance. They have added to a model Intelligent Heart Disease Prediction System

(IHDPS) utilizing information mining strategies, to be specific, Decision Trees, Naive Bayes and Neural Network. Results demonstrate that every system has its one of a kind quality in understanding the targets of the characterized mining objectives. IHDPS can answer complex "; consider the possibility that"; inquiries which conventional choice emotionally supportive networks can't. Utilizing therapeutic profiles, for example, age, sex, circulatory strain and glucose it can foresee the probability of patients getting a coronary illness. It empowers huge information, e.g. designs, connections between medicinal elements identified with coronary illness, to be set up. IHDPS is Web-based, easy to understand, versatile, solid and expandable. They actualized on the .NET stage.

In 2010, Srinivas, K. et al. [21] tell that Heart infection (HD) is a noteworthy reason for dreariness and mortality in the advanced society. Restorative finding is critical yet confounded errand that ought to be performed precisely and proficiently. As per their study dissects the Behavioral Risk Factor Surveillance System, study to test whether self-reported cardiovascular sickness rates are higher in Singareni coal mining areas in Andhra Pradesh state, India, contrasted with different locales after control for different dangers. Subordinate variables incorporate self-reported measures of being determined to have cardiovascular illness (CVD) or with a particular type of CVD including (1) mid-section torment (2) stroke and (3) heart assault. As per the creators heart care study indicates 15 credits to foresee the dismalness. Alongside standard characteristics other general qualities BMI (Body Mass Index), doctor supply, age, ethnicity, training, wage, and others are utilized for forecast. A mechanized framework for medicinal finding would improve therapeutic care and lessen costs. They apply information mining procedures to be specific; Decision Trees, Naïve Bayes and Neural Network are utilized for expectation of coronary illness.

In 2011, Hnin Wint Khaing [22] exhibited an effective methodology for the expectation of heart assault hazard levels from the coronary illness database. Firstly, the coronary illness database is bunched utilizing the K-implies grouping calculation, which will extricate the information pertinent to heart assault from the database. This methodology permits mastering the quantity of parts through its k parameter. Along these lines the successive examples are mined from the extricated information, pertinent to coronary illness, utilizing the MAFIA (Maximal Frequent Item set Algorithm) calculation. The machine learning calculation is prepared with the chose critical examples for the powerful forecast of heart assault. They have utilized the ID3 calculation as the preparation calculation to show level of heart assault with the choice tree. The outcomes as indicated by the creator that the composed forecast framework is equipped for foreseeing the heart assault successfully.

In 2012, Muhammed et al. [23] display and talk about the investigation that was executed with gullible Bayes method keeping in mind the end goal to manufacture prescient model as a fake analyse for coronary illness in light of information set which contains set of parameters that were measured for people already. At that point they contrast the outcomes and different systems as indicated by utilizing the same information that were given from UCI store information.

In 2012, Debabrata Pal et al. [24] propose that Coronary artery disease (CAD) influences a great many individuals everywhere throughout the world incorporating a noteworthy segment in India consistently. Albeit much advance has been done in therapeutic science, however the early identification of this malady is still a test for counteractive action. The goal of their paper is to portray creating of a screening master framework that will recognize CAD at an early stage. Principles were defined from the specialists and fluffy master framework methodology was brought to adapt to instability present in restorative space. This work portrays the danger variables in charge of CAD, learning procurement and information representation strategies, strategy for standard association, fuzzification of clinical parameters and defuzzification of fluffy yield to fresh esteem. The framework usage is done utilizing object situated investigation and outline. Their proposed philosophy is created to help the restorative professionals in foreseeing the patient's danger status of CAD from guidelines gave by medicinal specialists. They concentrates on guideline association utilizing the idea of modules, meta-principle base, standard location stockpiling in tree representation and standard consistency checking for productive hunt of vast number of tenets in standard base. Their created framework prompts 95.85% affectability and 83.33% specificity in CAD hazard calculation.

In 2013, Muhammad Usman et al. [25] propose that coordination of information mining systems with

information warehousing is picking up fame because of the way that both controls supplement one another in extricating learning from extensive datasets. On the other hand, the greater part of methodologies spotlight on applying information mining as a front end innovation to mine information distribution centers. While routines, for example, information grouping connected on multidimensional information have been appeared to upgrade the learning disclosure prepare, various essential issues stay uncertain as for the outline of multidimensional construction. These identify with robotized support for the determination of enlightening measurement and actuality variables in high dimensional and information serious situations, an action which might challenge the capacities of human fashioners because of the sheer size of information volume and variables included. They have proposed a procedure that chooses a subset of useful measurement and certainty variables from a starting arrangement of hopefuls. Their trial results led on three genuine datasets taken from the UCI machine learning store demonstrate that the information found from the mapping that we produced was more differing and instructive than the standard methodology of mining the first information without the utilization of our multidimensional structure forced on it.

In 2013, Ansel Y. Rodríguez-González et al. [26] propose that the greater part of the present calculations for mining affiliation decides expect that two article sub portrayals are comparable when they are precisely equivalent, yet in numerous genuine issues some other comparability capacities are utilized. Ordinarily these calculations are separated in two stages: Frequent example mining and era of intriguing affiliation rules from successive examples. Creators exhibited two calculations for mining successive comparable examples utilizing comparability capacities unique in relation to the balance they are proposed. Furthermore, the Gen Rules Algorithm is adjusted to create intriguing affiliation rules from successive comparable examples. Test results demonstrate that their calculations are more viable and get preferred quality examples over the current ones.

In 2013, Jesmin Nahar et al. [27] explore the debilitated and sound variables which add to coronary illness for guys and females. Affiliation standard mining, a computational insight methodology, is utilized to recognize these variables and the UCI Cleveland dataset, a natural database, is considered alongside the three principle era calculations – Apriori, Predictive Apriori and Tertius. Dissecting the data accessible on wiped out and sound people and taking certainty as a pointer, females are seen to have less risk of coronary illness then guys. Additionally, the traits showing sound and wiped out conditions were distinguished. It is seen that components, for example, mid-section agony being asymptomatic and the vicinity of activity impelled angina demonstrate the conceivable presence of coronary illness for both men and ladies. On the other hand, resting ECG being either ordinary or hyper and slant being level are potential high hazard elements for ladies just. For men, then again, just a solitary tenet communicating resting ECG being hyper was appeared to be a huge variable. This implies, for ladies, resting ECG status is a key particular variable for coronary illness forecast. Looking at the solid status of men and ladies, slant being up, number of shaded vessels being zero, and old crest being not exactly or equivalent to 0.56 shows a sound status for both genders.

In 2015 Dubey et al. [19] suggest the early detection may help in providing better treatment and can cure. They also suggest that the framework based on data mining and optimization may predict diseases properly.

In 2014, Krishnaiah et al. [28] designed a membership function to remove uncertainty of unstructured data, by the use of fuzziness in the measured data. A membership function was designed and incorporated with the measured value to remove uncertainty and fuzzified data was used to predict the heart disease patients. Their results of Fuzzy K-NN classifier suits to other classifiers of parametric techniques.

3.Problem domain

After studying several research papers we observe that there is lots of work in the area of heart diseases detection. But there is still the need of betterment. So our direction of research is to the improvement in terms of Heart disease detection. Some of the gaps identified are following:

1. Neural network and Fuzzy based technique to train data set for finding better classification and accuracy.
2. Optimization technique like Ant Colony Optimization to optimize the classification for improving the detection.
3. Machine learning environment or Support Vector machine is also an insight for better detection.

4. Data homogeneity based algorithm to find over fitting and overgeneralization Characteristics can be applied by clustering algorithm like K-Means.
5. Hybrid platform is missing where we can cluster, classify and optimize the heart disease dataset to better predicting the heart disease symptoms.

4.Dataset analysis

The dataset discussed here are Cleveland Heart Disease data Base [29] and Statlog Heart disease database [30]. Both of the dataset can be found from UCI repository.

Cleveland Heart Disease data Base:
This data was collected from the four following locations:

1. Cleveland Clinic Foundation (cleveland.data)
2. Hungarian Institute of Cardiology, Budapest (hungarian.data)
3. V.A. Medical Center, Long Beach, CA (long-beach-va.data)
4. University Hospital, Zurich, Switzerland (switzerland.data)

This database has 76 raw attributes and 14 of them are actually used. The number of instances as per the UCI repository are:

Cleveland: 303
Hungarian: 294
Switzerland: 123
Long Beach VA: 200

The main 14 attributes are following:
1.Age
2.Sex
3.CP
4.trestbps
5.chol
6.FBS
7.restecg
8. thalach
9.exang
10.oldpeak
11.slope
12.ca
13.thal
14.num

Statlog Heart disease database: It is a collection of 13 attributes from a set of 75.

The main 13 attributes are following:
1.Age
2.Sex
3.chest pain type
4.resting blood pressure
5.serum cholestoral in mg/dl
6.fasting blood sugar > 120 mg/dl
7.resting electrocardiographic results (values 0,1,2)
8. maximum heart rate achieved
9.exercise induced angina
10.oldpeak = ST depression induced by exercise relative to rest
11.the slope of the peak exercise ST segment
12.number of major vessels (0-3) colored by flourosopy
13.thal: 3 = normal; 6 = fixed defect; 7 = reversable defect

Cleveland Heart Disease database consists of 303 records and Statlog Heart disease database consists of 270 records. Combined data bases of 550 records.

5. Conclusion

In this paper terrific status of heart diseases and the methodologies for finding it in the initial stages. There are several methodologies which are performing well, but there is a need of improvement. From the analysis it is presumed that, information mining assumes a significant part in coronary illness order. The order precision can be made strides by decrease in peculiarities and using optimization techniques. In future hybrid platform is suggested where cluster, classification and optimization can be used on heart disease dataset to better predicting the heart disease symptoms.

Acknowledgment

None.

Conflicts of interest

The authors have no conflicts of interest to declare.

References

[1] Fayyad UM, Piatetsky-Shapiro G, Smyth P, Uthurusamy R. Advances in knowledge discovery and data mining. AAAI/MIT Press; 1996, p. 37-58.
[2] Berkhin P. A survey of clustering data mining techniques. In grouping multidimensional data 2006 (pp. 25-71). Springer Berlin Heidelberg.
[3] Peter TJ, Somasundaram K. An empirical study on prediction of heart disease using classification data mining techniques. In international conference on advances in engineering, science and management (ICAESM) 2012 (pp. 514-18). IEEE.
[4] Dubey AK, Dubey AK, Agarwal V, Khandagre Y. Knowledge discovery with a subset-superset approach for mining heterogeneous data with dynamic support. In CSI sixth international conference on software engineering (CONSEG) 2012 (pp. 1-6). IEEE.

[5] Dubey AK, Shandilya SK. A novel J2ME service for mining incremental patterns in mobile computing. In information and communication technologies 2010 (pp. 157-64). Springer Berlin Heidelberg.
[6] Gupta R, Satsangi CS. An efficient range partitioning method for finding frequent patterns from huge database. International Journal of Advanced Computer Research. 2012; 2(2):62-9.
[7] Singh S, Yadav M, Gupta H. Finding the chances and prediction of cancer through Apriori algorithm with transaction reduction. International Journal of Advanced Computer Research. 2012; 2(2):23-8.
[8] Purohit MN, Purohit MS, Purohit MR. Data mining, applications and knowledge discovery. International Journal of Advanced Computer Research (IJACR). 2012; 2(6):458-62.
[9] Fayyad UM, Piatetsky-Shapiro G, Smyth P. The KDD process for extracting useful knowledge from volumes of data. Communications of the ACM.1996; 39(11):27-34.
[10] Zhao J, Wang T. A general framework for medical data mining. In international conference on future information technology and management engineering (FITME) 2010 (pp. 163-5). IEEE.
[11] Hart WE. Adaptive global optimization with local search (Doctoral dissertation, University of California, San Diego).
[12] ShengBing C, XiaoFeng W, XiaoFang W. Mining dynamical frequent itemsets based on ant colony algorithm. In IEEE international conference on computer science and automation engineering (CSAE) 2011(pp. 252-5). IEEE.
[13] Dorigo M, Di Caro G, Gambardella LM. Ant algorithms for discrete optimization. Artificial life. 1999; 5(2):137-72.
[14] Vaska JS, Sowjanya AM. Clustering diabetics' data using M-CFICA. International Journal of Advanced Computer Research. 2015; 5(20):327-33.
[15] Mansour AM, Obaidat MA, Hawashin B. Elderly people health monitoring system using fuzzy rule based approach. International Journal of Advanced Computer Research. 2014; 4(17):904-14.
[16] Jain N, Sharma V, Malviya M. Reduction of negative and positive association rule mining and maintain superiority of rule using modified genetic algorithm. International Journal of Advanced Computer Research (IJACR). 2012; 2(6):31-6.
[17] Sadh AS, Shukla N. Association rules Optimization: A survey. International Journal of Advanced Computer Research (IJACR). 2013; 3(9):111-5.
[18] Cheng J, Ke Y, Ng W. A survey on algorithms for mining frequent itemsets over data streams. Knowledge and Information Systems. 2008; 16(1):1-27.
[19] Dubey A, Patel R, Choure K. An efficient data mining and ant colony optimization technique (DMACO) for heart disease prediction. International Journal of Advanced Technology and Engineering Exploration. 2014; 1(1):1-6.
[20] Palaniappan S, Awang R. Intelligent heart disease prediction system using data mining techniques. In IEEE/ACS international conference on computer systems and applications 2008 (pp. 108-15). IEEE.
[21] Srinivas K, Rao GR, Govardhan A. Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques. In international conference on computer science and education (ICCSE) 2010 (pp. 1344-49). IEEE.
[22] Khaing HW. Data mining based fragmentation and prediction of medical data. In international conference on computer research and development (ICCRD) 2011 (pp. 480-85). IEEE.
[23] Muhammed LN. Using data mining technique to diagnosis heart disease. In international conference on statistics in science, business, and engineering (ICSSBE) 2012 (pp. 1-3). IEEE.
[24] Pal D, Mandana KM, Pal S, Sarkar D, Chakraborty C. Fuzzy expert system approach for coronary artery disease screening using clinical parameters. Knowledge-Based Systems. 2012; 36: 162-74.
[25] Usman M, Pears R, Fong AC. Discovering diverse association rules from multidimensional schema. Expert systems with applications. 2013; 40(15):5975-96.
[26] Rodríguez-González AY, Martínez-Trinidad JF, Carrasco-Ochoa JA, Ruiz-Shulcloper J. Mining frequent patterns and association rules using similarities. Expert Systems with Applications. 2013; 40(17):6823-36.
[27] Nahar J, Imam T, Tickle KS, Chen YP. Association rule mining to detect factors which contribute to heart disease in males and females. Expert Systems with Applications. 2013; 40(4):1086-93.
[28] Krishnaiah V, Srinivas M, Narsimha G, Chandra NS. Diagnosis of heart disease patients using fuzzy classification technique. In international conference on computer and communications technologies (iccct) 2014 (pp. 1-7). IEEE.
[29] Cleveland database: http://archive.ics.uci.edu/ml/datasets/Heart+Disease. Accessed 26 March 2016.
[30] Statlog database: http://archive.ics.uci.edu/ml/machine-learningatabases/statlog/heart/. Accessed 26 March 2016.