Review Article

International Journal of Advanced Technology and Engineering Exploration, Vol 3(15)
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
http://dx.doi.org/10.19101/ IJATEE.2016.315003

A hybrid framework for heart disease prediction: review and analysis

Ankita Shrivastava[1*] and Shivkumar Singh Tomar [2]
Research Scholar, Department of Computer Science, TIT, Bhopal[1]
Assistant Professor, Department of Computer Science, TIT, Bhopal[2]

Abstract

According to WHO the mortality rates are higher in case of heart diseases in the world and it is increasing continuously. The main reasons suggested in the several research work are smoking, alcohol, obesity, diet and hereditary. In this paper we are analysing several methods of heart disease classification and prediction so that we can detect it in the earlier stages. As it can be cured if it is detected in the early stage. For this we have examined a few systems displayed till now and in view of the study introduced before recommending some future bits of knowledge which may be a superior structure for finding. Factors, Prevention and early discovery are additionally talked about here. The main objective of this study is to discuss on hybrid techniques specially data mining and optimization. So that better decision making and prediction framework will be designed.

Keywords

Heart Diseases, Risk Factor, Mining, Optimization.

1.Introduction

The intend of data mining is to abstract acquaintance data of hefty amount of Text. Data mining is an interdisciplinary approach, whose debased is at the standpoint of machinery way of life, statistics, and databases. In this paper we have studied and analysed the use of data mining to emphasize to discover knowledge that is not only accurate, but also comprehensible for the user [1], [2], [3]. According to Yang Jianxiong et al. [4] comprehensibility is imperative at different point found learning will be utilized for supporting a human decision and choice. All things considered, if found information is not understandable for a client, it won't be conceivable to decipher and accept the learning.

In terms of health disease prediction and factor extraction data mining can be used as a nontrivial extraction of implicit, earlier hidden, and relevant information from data. It includes a series of different computational approaches, such as clustering, data visualization, learning, classification, associated patterns, and pattern detection [5][6]. The found data and learning are valuable for different applications, including business sector investigation, choice bolster, misrepresentation identification, and business administration.

Numerous methodologies have been proposed to concentrate data, and mining therapeutic information is an imperative field [7-9]. In [10] and [11] authors have suggested the used of data mining along with the evolutionary techniques as it is efficient in finding the optimal solutions so that available solutions can be found. It can strength the other techniques to make it predict on the basis of threshold values so it can be milestone in heart disease prediction. An efficient way is to insert the nearby boundaries into the system of Population-based procedures, including developmental calculations, subterranean state routines, and so forth[12]. In [13] by utilizing subterranean insect settlement calculation, we number thing sets as indicated by the pheromone focus. In addition, utilizing the gathering and volatilization normal for pheromones, we mimic the change of recurrence of Item set. So in our abstract we need to join streamlining for better coronary illness elements expectation. There are several hybridization is possible through the use of several data mining and optimization techniques as suggested in [14-17].

2.Related work

In 2011, Hnin Wint Khaing [18] exhibited a proficient methodology for the forecast of heart assault danger levels from the coronary illness database. Firstly, the coronary illness database is bunched utilizing the K-means grouping calculation,

*Author for correspondence

which will separate the information significant to heart assault from the database. This methodology permits mastering the quantity of parts through its k parameter. In this way the regular examples are mined from the removed information, significant to coronary illness, utilizing the MAFIA (Maximal Frequent Itemset Algorithm) calculation. The machine learning calculation is prepared with the chose huge examples for the successful expectation 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 per the creator that the outlined forecast framework is equipped for foreseeing the heart assault adequately.

In 2012, Peter, T.J. et al. [19] proposes utilization of example acknowledgment and information mining systems into danger forecast models in the clinical area of cardiovascular drug is proposed. The information is to be demonstrated and grouped by utilizing order information mining procedure. A restrictions' percentage of the ordinary medicinal scoring frameworks are that there is a vicinity of inherent direct blends of variables in the data set and henceforth they are not capable at displaying nonlinear complex associations in restorative areas. This restriction is taken care of in this exploration by utilization of characterization models which can verifiably identify complex nonlinear connections in the middle of indigent and autonomous variables and also the capacity to distinguish every single conceivable cooperation between indicator variables.

In 2012, Muhammed et al. [20] exhibit and examine the trial that was executed with gullible bayes strategy keeping in mind the end goal to construct prescient model as a manufactured analyze for coronary illness taking into account information set which contains set of parameters that were measured for people already. At that point they contrast the outcomes and different methods as indicated by utilizing the same information that were given from UCI archive information.

In 2012, Debabrata Pal et al. [21] recommend that Coronary supply route illness (CAD) influences a large number of individuals everywhere throughout the world incorporating a noteworthy bit in India consistently. Albeit much advance has been done in restorative science, however the early location of this malady is still a test for counteractive action. The target of their paper is to portray creating of a screening master framework that will help to recognize CAD at an early stage. They concentrates on tenet association utilizing the idea of modules, meta-principle base, guideline address stockpiling in tree representation and standard consistency checking for proficient inquiry of substantial number of guidelines in tenet base. Their created framework prompts 95.85% affectability and 83.33% specificity in CAD hazard calculation.

In 2013, Muhammad Usman et al. [22] recommend that incorporation of information mining strategies with information warehousing is picking up fame because of the way that both controls supplement one another in extricating learning from expansive datasets. They propose a philosophy that chooses a subset of instructive measurement and actuality variables from a starting arrangement of hopefuls. Their test results led on three true datasets taken from the UCI machine learning storehouse demonstrate that the learning found from the diagram that we created was more different and educational 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. [23] propose that the greater part of the present calculations for mining affiliation guidelines expect that two item sub portrayals are comparable when they are precisely equivalent, however in numerous genuine issues some other comparability capacities are utilized. Usually these calculations are isolated in two stages: Frequent example mining and era of fascinating affiliation rules from incessant examples. Furthermore, the GenRules Algorithm is adjusted to create intriguing affiliation rules from regular comparable examples. Trial results demonstrate that their calculations are more viable and acquire preferred quality examples over the current ones.

In 2013, Jesmin Nahar et al. [24] examines the wiped out and sound elements which add to coronary illness for guys and females. Affiliation standard mining, a computational insight methodology, is utilized to distinguish these components and the UCI Cleveland dataset, a natural database, is considered alongside the three guideline era calculations – Apriori, Predictive Apriori and Tertius. Breaking down the data accessible on wiped out and solid people and taking certainty as a marker, females are seen to have less risk of coronary illness then guys. Be that as it may, resting ECG being either ordinary or hyper and slant being level are potential high hazard variables for ladies just. For men, then again, just a solitary standard communicating resting ECG being hyper was appeared to be a critical component. This implies, for ladies, resting ECG status is a key unmistakable element for coronary illness forecast. Looking at the sound status of men and ladies, incline being up, number of hued vessels being zero, and old peak being not exactly or equivalent to 0.56 demonstrate a solid status for both sexual orientation.

In 2014, Sonawane et al. [25] present an expectation framework for coronary illness utilizing multilayer perceptron neural system. The neural system in this framework acknowledges 13 clinical elements as info and it is prepared utilizing back-proliferation calculation to foresee that there is a vicinity or nonappearance of coronary illness in the patient with most noteworthy exactness of 98% relative to different frameworks. The exactness along these lines got with this framework demonstrates that it is preferred and effective over different frameworks.

In 2014, Jabbar et al. [26] suggested a decision support system which can be able to predict the risk score of a patient. This can be helpful in taking precautionary steps like balanced diet and medication which will in turn increase life time of a patient. They have proposed a lazy associative classification for prediction of heart disease in Andhra Pradesh and present some experimental results which will help physicians to take accurate decisions.

In 2014, Krishnaiah et al. [27] recommended to evacuate instability of unstructured information, an endeavor was made by presenting fluffiness in the deliberate data. A enrollment capacity was planned and consolidated with the deliberate worth to uproot vulnerability and fuzzified information was utilized to anticipate the coronary illness patients. They have characterized the patients in light of the characteristics gathered from therapeutic field. Least Euclidean separation Fuzzy K-NN classifier was intended to order the preparation and testing information fitting in with diverse classes. It was found that Fuzzy K-NN classifier suits well as contrasted and different classifiers of parametric strategies.

In 2014, Sonawane et al. [28] exhibited a forecast framework for coronary illness utilizing Learning vector Quantization neural system calculation the neural system in this framework acknowledges 13 clinical elements as data and predicts that there is a vicinity or nonappearance of coronary illness in the patient, alongside distinctive execution measures.

3.Analysis

This dataset we have considered from UCI repository which concerning the data from heart disease diagnosis. The attributes is in the form of number. The data sources are following [29]:

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)

Every database has the same case position. While the databases have 76 crude properties, just 14 of them are really utilized. In this way I've taken the freedom of making 2 duplicates of every database: one with every one of the properties also, 1 with the 14 qualities really utilized as a part of past trials. The attributes are:

- #3 (age)
- #4 (sex)
- #9 (cp)
- #10 (trestbps)
- #12 (chol)
- #16 (fbs)
- #19 (restecg)
- #32 (thalach)
- #38 (exang)
- #40 (oldpeak)
- #41 (slope)
- #44 (ca)
- #51 (thal)
- #58 (num) (the predicted attribute)

In [31] they have used some other like which can increase the high risk are following:

- Family history
- Smoking
- Poor diet
- High blood pressure
- High blood cholesterol
- Obesity
- Physical inactivity
- Hyper tension

According to the current report presented by the WHO the death rates are very high in case of heart diseases as shown in *figure 1*.

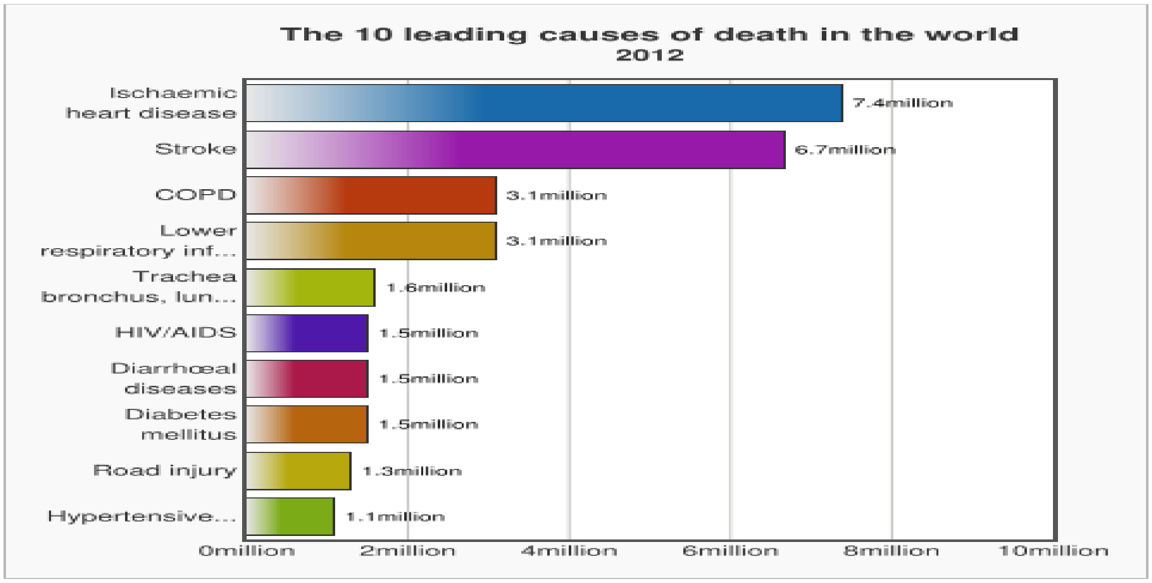

Figure 1 Leading causes of death in the world source: WHO

According to Waghulde et al. [30] the death rate brought about by coronary illness has been changing so there is requirement for advancement of routines for coronary illness expectation is of prompt investigative and useful premium. There are a few calculations which have been as of now created for hazard stratification and indicative models for heart ailment forecast.

Data mining has used in the intelligent medical systems [32-33]. The connections of disarranges and the genuine reasons for the issue and the impacts of side effects that are suddenly found in patients can be assessed by the clients through the built programming effectively. Vast databases can be connected as the info information to the product by utilizing the product's extendibility. The impacts of connections that have not been assessed enough have been investigated and the connections of shrouded learning laid among the extensive restorative databases have been looked in this study by method for discovering incessant things utilizing hopeful era. The arrangements of disorders at the same time found in the medicinal databases can be decreased by utilizing our no hopeful methodology. Same like the evolutionary algorithms are trending as it is efficient in finding the optimal solution. The result analyses from different approaches are shown below:

Table 1 Result analysis

S.No	Reference	Methods and Tools	Results	Objective
1	[34]	Radial Basis Function	The system has analysed 125 sample data and is achieved 97% accuracy in the clinical assessment.	The diagnosis performances diagnosis, and operating parameters improvements are the main objective.
2	[35]	Genetic Algorithm	BP algorithm which is the best classifier of Artificial Neural Network which uses the updating technique of weights by propagating the errors backward can be used.	To add to a model which can focus and concentrate obscure learning related with coronary illness from a past coronary illness database record.
3	[36]	Associative Classification and Hybrid Feature Subset Selection	Better results have been found in terms of cross over rate, mutation and selection.	To develop a prediction system for heart disease.
4	[26]	Lazy Associative Classification	Their proposed system has achieved 90.2 % average	To predict the risk score a decision support system is needed.

S.No	Reference	Methods and Tools	Results	Objective
			accuracy.	
5	[27]	Fuzzy K-NN method	Their proposed system has achieved 97 % average accuracy.	To remove uncertainty in unstructured data.
6	[37]	Naïve Bayes	The result showed that Naïve Bayes tress has highest prediction accuracy than clustering Algorithm.	To check the hybrid techniques used for the diagnosis.
7	[25]	Multilayer Perceptron Neural Network	Their proposed system has achieved 98 % average accuracy.	To develop a prediction system for heart disease.
8	[28]	Learning Vector Quantization Algorithm	Their proposed system has achieved 85.55 % average accuracy.	To improve the performance of the prediction system.
9	[38]	Naïve Bayes	Their proposed system has achieved 90.74 % average accuracy.	To analysis of HDP using Different DM Techniques
10	[39]	Fuzzy Logic	Their proposed system has achieved 83.85 % average accuracy.	To compare on of various classification techniques.

4.Gap identification

1. Method hybridization is missing as it is important to predict the symptom level wise.
2. Data Mining and Evolutionary algorithms combined more affect instead of single trends.
3. Parameters used are also important factors which will be helpful in the design of prediction system.
4. The combined effect of clustering and classification can be a better decision support system.
5. Efficient way of unwanted data removal can be efficient in the final results.
6. To reduce the number of attributes and also determine the attribute which contribute towards the diagnosis of disease is an important factor.

5.Conclusions and future direction

In the above perspective study and exchange there are a few systems and methodology are introduced till now. A systems' portion are go in some so of parameters yet general precision is not up to the imprint there are degree in this course so that expectation method can be moved forward. So in future work a framework with improved mining can be a better way in early identification. There are a several research work are heading in this field and the degree is in the bearing to satisfy it in the prior stage. The symptoms of heart diseases are diverse thus it is dealt with contrastingly additionally so that the chances are more positive. So our bearing of exploration is in the distinguishing proof in the early stages with the help of data mining and evolutionary algorithms.

Acknowledgment

None.

Conflicts of interest

The authors have no conflicts of interest to declare.

References

[1] Dorigo M, Di Caro G, Gambardella LM. Ant algorithms for discrete optimization. Artificial Life. 1999; 5(2):137-72.

[2] Fayyad U, Piatetsky-Shapiro G, Smyth P. From data mining to knowledge discovery in databases. AI Magazine. 1996; 17(3):37-54.

[3] Watada J, Aoki K, Kawano M, Hitam MS. Dual scaling in data mining from text databases. JACIII. 2006; 10(4):451-7.

[4] Jianxiong Y, Watada J. Wise mining method through ant colony optimization. In international conference on systems, man and cybernetics 2009 (pp. 1833-39). IEEE.

[5] Fayyad U, 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.

[6] 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 (IJACR). 2012; 2(2):23-8.

[7] Vaska JS, Sowjanya AM. Clustering diabetics data Using M-CFICA. International Journal of Advanced Computer Research. 2015; 5(20):327-33.

[8] 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 (IJATEE). 2014; 1(1):1-6.

[9] Chen YL, Tony CK, Hui-Ling H. A general framework for discovering sequential patterns based on fuzzy concept. Nanya Academic Journal. 2008; 25: 45-58

[10] Yadav BS, Rai Y, Kushwaha S. Iterative K-Means (IKM) approach for Wisconsin breast cancer data prediction. International Journal of Advanced Technology and Engineering Exploration (IJATEE).2016;3(14):1-7.

[11] Shokhan MH, Khitam AM. Biometric identification system by lip shape. International Journal of Advanced Computer Research. 2015; 5(18):19-24.

[12] Yadav BS, Rai Y, Kushwaha S. Data mining based breast cancer analysis: a review. International Journal of Advanced Technology and Engineering Exploration (IJATEE). 2015; 2(12):157-162.

[13] ShengBing C, XiaoFeng W, XiaoFang W. Mining dynamical frequent itemsets based on ant colony algorithm. In international conference on computer science and automation engineering (CSAE) 2011 (pp. 252-255). IEEE.

[14] 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.

[15] Li K, Cui L. A kernel fuzzy clustering algorithm with generalized entropy based on weighted sample. International Journal of Advanced Computer Research. 2014; 4(2):596-600.

[16] Kushwah J, Singh D. Classification of cancer gene selection using random forest and neural network based ensemble classifier. International Journal of Advanced Computer Research. 2013; 3(2):30-4.

[17] Sadh AS, Shukla N. Apriori and ant colony optimization of association rules. International Journal of Advanced Computer Research. 2013; 3(2):35-42.

[18] Khaing HW. Data mining based fragmentation and prediction of medical data. In international conference on computer research and development (ICCRD) 2011(pp. 480-5). IEEE.

[19] 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. 514-8). IEEE.

[20] 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.

[21] 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.

[22] Usman M, Pears R, Fong AC. Discovering diverse association rules from multidimensional schema. Expert Systems with Applications. 2013; 40(15):5975-96.

[23] 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.

[24] 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.

[25] Sonawane JS, Patil DR. Prediction of heart disease using multilayer perceptron neural network. In international conference on information communication and embedded systems (ICICES) 2014 (pp. 1-6). IEEE.

[26] Jabbar MA, Deekshatulu BL, Chandra P. Heart disease prediction using lazy associative classification. In international multi-conference on automation, computing, communication, control and compressed sensing (iMac4s) 2013 (pp. 40-46). IEEE.

[27] 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.

[28] Sonawane JS, Patil DR. Prediction of heart disease using learning vector quantization algorithm. In conference on IT in business, industry and government (CSIBIG) 2014 (pp. 1-5). IEEE.

[29] https://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/heart-disease.names. Accessed 15 June 2015.

[30] Waghulde NP, Patil NP. Genetic neural approach for heart disease prediction. International Journal of Advanced Computer Research. 2014; 4(3):778-84.

[31] Xing Y, Wang J, Zhao Z, Gao Y. Combination data mining methods with new medical data to predicting outcome of coronary heart disease. In international conference on convergence information technology 2007 (pp. 868-72). IEEE.

[32] Aflori C, Craus M. Grid implementation of the Apriori algorithm. Advances in Engineering Software. 2007; 38(5):295-300.

[33] 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.

[34] Hannan SA, Mane AV, Manza RR, Ramteke RJ. Prediction of heart disease medical prescription using radial basis function. In international conference on computational intelligence and computing research (ICCIC) 2010 (pp.1-6). IEEE.

[35] Dewan A, Sharma M. Prediction of heart disease using a hybrid technique in data mining classification. In international conference on computing for sustainable global development (INDIACom) 2015 (pp. 704-6). IEEE.

[36] Jabbar MA, Chandra P, Deekshatulu BL. Prediction of risk score for heart disease using associative classification and hybrid feature subset selection. In international conference on intelligent systems design and applications (ISDA) 2012 (pp. 628-34). IEEE.

[37] Sivagowry S, Durairaj M, Persia A. An empirical study on applying data mining techniques for the analysis and prediction of heart disease. In international conference on information communication and embedded systems (ICICES) 2013 (pp. 265-70). IEEE.

[38] Bhatla N, Jyoti K. An analysis of heart disease prediction using different data mining techniques. International Journal of Engineering Research and Technology. 2012; 1(8):1-4.

[39] Rahman RM, Afroz F. Comparison of various classification techniques using different data mining tools for diabetes diagnosis. Journal of Software Engineering and Applications. 2013; 6(3):85-97.