

Study and Analysis of Medical Data Mining Techniques in Healthcare for Heart Disease Using Hybrid Approach

Khodke harish Eknath^{a,1}, Yadav S K ^b, and Kyatanavar D N ^b

^{a,b} Dept of CSE, Shri Jagdishprasad Jhabarmal Tibrewala University, Rajasthan-333001,India

Abstract. Information mining frameworks are exhaustively used in coronary affliction for affirmation and figure. As heart condition is that the essential clarification for death for individuals, recognizing confirmation . The work proposed is inductive type and needs deep analysis of the data to ensure the right predictions on the data sets provided. A sample dataset of patients for heart disease will be collected from repository. It involves the steps and procedure. The proposed research work can be carried out step by step to conclude it with the accurate results.

Keywords. Data mining; hybrid method; base classifier; classification accuracy; sensitivity; specificity

1. Introduction

Availability of data mining techniques has now made the medical field to formulate more and more easy steps and predict the diseases like heart disease, diabetes, cancer, malaria, measles much before they start affecting the health of people etc. The classification methods gradually help to predict these problems for future analysis by building the classification models. In this proposed research work, classification techniques are used for the prediction of disease from medical data warehouse[1]. An integration of association rule mining and classification, which improves the blueprint accuracy. Approach is a sensible report that is being used by data disclosure and decision help which direction rule introduction structures and requesting for need. classifiers are used especially to fit to applications where most basic precision may be a perfect essential to show up for figure. There are various spaces, for instance, appear, clinical, etc where the most wonderful exactness of the using models is diligently major [2]. It selects a little set of top quality rules and uses these rules for prediction. . The proposed classification algorithms will be predict the by-diseases from the data sets of the patients who suffer from heart disease. Data sets are collected from variety of repositories. This work of research will tested using various performance measures applicable in data mining. Heart disease remains the quantity one clarification for death all through the world for whatever length of time that decades. the world Health Organization (WHO) has quantifiable that seventeen.7 million passages have happened by and large because of heart ailments. If we will envision mess and give alerted ahead of time, a few passages zone unit for the most part hindered. the machine of data mining carries a substitution estimation to mess

¹Khodke harish Eknath, Dept of CSE, Shri Jagdishprasad Jhabarmal Tibrewala University, Rajasthan; E.mail. hekhodke@gmail.com

conjecture. different dealing with frameworks domain unit used for specific and expelling obliging data from the clinical dataset with base customer wellsprings of information and try. The potency of knowledge mining for the most part varies on the techniques used and also the options elect. The medical datasets within the tending trade area unit redundant and inconsistent[3]. it's more durable to use processing techniques while not previous and applicable preparations. the guts malady information can collect from UCI machine learning repository or kaggle. Among the attribute options utilized in cardiopathy prediction. when choosing the options, the models can created with the foremost in style classification techniques in processing. The performance of the classification models can live victimization 3 performance measures: accuracy, f-measure, and exactitude. Accuracy is that the degree of fittingly foreseen cases among all events. F-measure is that the weighted mean of exactitude and recall. exactitude is that the proportion of correct predictions for the positive category. Process of discovering the patterns hidden and therefore the extraction of those patterns from the varied sources so on analyze the info then come to deciding are often mentioned as data processing[12] . Generally, there are two types of learning methods in data mining called supervised learning and therefore the unattended learning. supported these 2 learning ways we tend to do perform the assorted processing tasks like classification, cluster and pattern recognition, etc. In supervised learning, the class is preset and it's acknowledged at intervals information the info the information set or coaching data set. By this, we'll explore for the output pattern whereas at intervals the unattended learning there is no any there is not any construct of sophistication attribute and thus the design for the pattern isn't restricted to 1 attribute[20]. The disease analysis of the info sets has been a decent approach and higher example of mining. the info at intervals the medical space is sometimes huge and contains the information that relates to completely different diseases. Here, regardless of the data which will belong to a heart condition , it are often wont to detect the opposite diseases from the prevailing problems of the patients supported the principles formed from the classification. There are different data processing techniques that are wont to determine the lon ger term diseases or the issues that shall arise. Here, to predict the by-diseases and therefore the symptoms of those which can be occurring heart condition are going to be predicted. The classification mining can be used in an efficient way when compared to other techniques like association rule mining, decision tree, clustering and classification etc[21-22].

2. Dataset

SPECT cardiopathy dataset is utilized all through this paper is out there on the college of CA, Irvine (UCI) AI dataset store. The dataset is accommodated work the analyze of viscus Single nucleon Emission electronic pivotal tomography (SPECT) pictures abuse AI calculations. SPECT, or less ordinarily, SPET, could likewise be a medications tomographic imaging system abuse gamma beams.

3. Literature Review

a) Mohammad Shafenoor Amin, Yin Kia Chiam, Kasturi Dewi Varathan (2019) examined huge highlights and information preparing systems which will improve the exactness of foreseeing issue . Forecast models were created utilizing diverse blend of highlights, and 7 grouping methods: k-NN, Decision Tree, Naive Bayes, Logistic Regression (LR), Support Vector Machine (SVM), Neural Network etc[1].

- b) T. Vivekanandan, N. Ch Sriman Narayana Iyengar (2017) studied the testing assignments of picking fundamental features from the immense game plan of practical features and diagnosing heart condition are overseen . Feature assurance is one among the overwhelming commonly used pre-taking care of steps in gathering issues. A changed differential headway (DE) figuring is used to perform incorporate decision for disarray and streamlining of picked features. Of the ten open procedures for the average DE computation, the seventh technique, which is addressed by DE/rand/2/exp, is considered for relative assessment.
- c) Ankita Dewan,Meghna Sharma(2015) examined productive hereditary calculation mixture with the back engendering procedure approach for heart condition expectation. To build up a model which may decide and remove obscure information (examples and relations) related with heart condition from a past heart condition database record. It can explain entangled questions for recognizing heart condition and subsequently help clinical experts to shape savvy clinical choices which customary choice emotionally supportive networks weren't prepared to . By giving productive medications, it can assist with downsizing expenses of treatment[20]
- d) H. Benjamin Fredrick David and S. Antony Belcy(2018) considered distinguish the most straightforward characterization calculation appropriate for giving greatest precision when arrangement of typical and irregular individual is regulated . Along these lines avoidance of the loss of lives at a prior stage is possible . The trial arrangement has been made for the assessment of the exhibition of calculations with the help of heart condition benchmark dataset recovered from UCI AI vault. it's discovered that Arbitrary Woodland calculation performs best with 81% accuracy in contrast with different calculations for heart condition prediction[21].
- e) R.Kavitha,E.Kannan(2016) explored order of the guts illness informational collection a high dimensional informational collection is utilized inside the preprocessing phase of information mining process. This crude dataset contains excess and conflicting information accordingly expanding the pursuit space and capacity of the data . to understand the grouping precision we'd prefer to dispose of the repetitive and in this way the unimportant information present. The dimensionality decrease procedure is utilized to pack the high dimensional information to bring down dimensional information with certain imperatives. A system is coordinated for the clear expectation of the guts sickness. The structure is made by utilizing the foremost part investigation (PCA) to separate the highlights and numerical model is processed to pick the applicable highlights utilizing the significant constraint[22].
- f) Theresa Princy. R,J. Thomas(2016) predicated heart condition chance rate utilizing information dealing with structures. Assorted information arranging procedures and classifiers are talked about in different evaluations which are utilized for gainful and convincing heart condition finding. According to the appraisal mode,it is seen that a gigantic measure of creators utilize different movements and specific number of attributes for their assessment. Hence, various progressions give unquestionable exactness endless supply of characteristics considered. Utilizing KNN and ID3 calculation the danger pace of heart condition was perceived and exactness level likewise fit particular number of qualities [20].
- g) Priyanka N,Dr.Pushpa RaviKumar(2017) studied accuracy of the results obtained through two algorithms Naïve Bayes & Decision tree. albeit both of them were ok in predicting the guts disease using various parameters, a choice tree was found to be the simplest . It gave the foremost accurate result whether the patient had the likelihood of heart condition . this technique also can be utilized in future projects to detect the precise sort of

heart condition especially . Thereby the diagnosis & management of heart condition are often made simpler[21].

h) Marjia Sultana, Afrin Haider and Mohammad ShorifUddin(2016) investigated heart condition prediction using KStar, J48, SMO, Bayes Net, and Multilayer Perceptron through Weka software. The introduction of those information preparing techniques is assessed by joining the results of farsighted precision, ROC turn and AUC respect utilizing a standard instructive rundown besides as an aggregated edifying combination. bolstered execution factor SMO and Bayes Net techniques show impeccable introductions Net techniques show optimum performances than the performances of KStar, Multilayer Perceptron and J48 techniques[22].

i) Saba Bashir, Zain Sikander Khan, Farhan Hassan Khan, Aitzaz Anjum, Khurram Bashir(2019) improved heart condition prediction and studied different feature selection approach. Data science is one among the more important things in early prediction and solves large data problems nowadays. This author's main ambition is to enhance accuracy within the prediction of heart condition by using feature selection techniques[23, 18-19].

4. Performance Evaluation Measures

Execution evaluation is required unmitigated robotized contamination attestation structures and is guided during this assessment to measure the power of base classifiers also as proposed crossbreed bundle model for anticipating believability of heart condition in patients upheld SPECT pictures. Disregarding the way that precision and review are powerfully normal by and large information preparing undertakings, in clinical space, specialists like better to outline what degree sensitive and express their proposed model is and in this manner the standard examination measures are affectability and expressness. Really, in clinical setting, an unquestionably fragile model is immaculate considering the way that the expense of expelling a positive model is fabulously high and a progressively unequivocal model is flawless considering the way that the expense of enrolling a model as positive for the models that aren't the objective of testing is inconceivably high [20].

$$\text{Sensitivity} = \frac{tp}{tp + fn}$$

$$\text{Specificity} = \frac{tn}{tn + fp}$$

The request precision is in addition considered as appraisal measure during this examination since it empowers assessment of the eventual outcomes of present assessment with other top tier models. The gathering precision, CA, depends upon the proportion of tests viably described (certifiable encouraging points notwithstanding veritable negatives) and is surveyed by the formula:

$$CA = \frac{t}{n} \cdot 100$$

where t is the number of sample cases correctly classified, and n is the total number of sample cases.

5. Method

In order to conduct and continue the research in an effective manner research methodology is should and must. A systematic plan is to be derived and defined for the usage of data and observation of the data.

The sample data will be collected and this will follow the statistical forecasting methods to achieve the desired goal of the research. Here the research methodology of quantitative analysis will be used so as to conclude with the results in effective manner.

The work proposed is inductive type and needs deep analysis of the data to ensure the right predictions on the data sets provided.

A sample dataset of patients for heart disease will be collected from UCI machine learning repository or kaggle. It involves the following steps and procedure. The proposed research work can be carried out step by step to conclude it with the accurate results.

- 1) Select the data sets from repository.
- 2) Preprocess the data selected from repository.
- 3) Feature Selection.
- 4) Classification Modeling. .
- 5) Analyze the results using various metrics for future analysis.
- 6) Predict the heart disease.

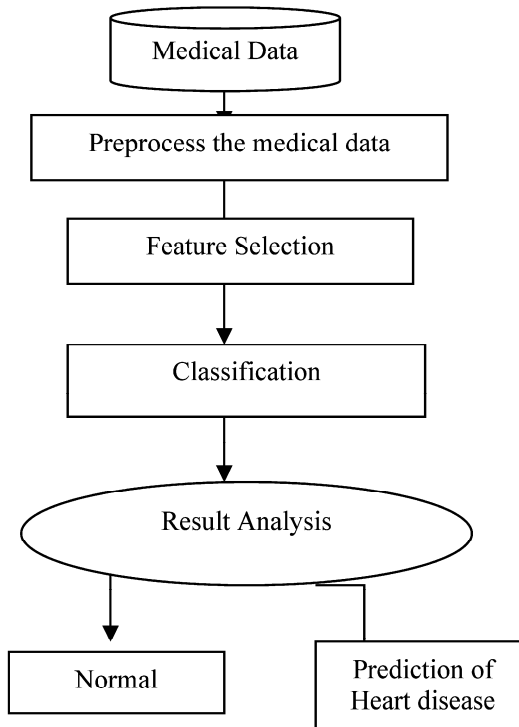
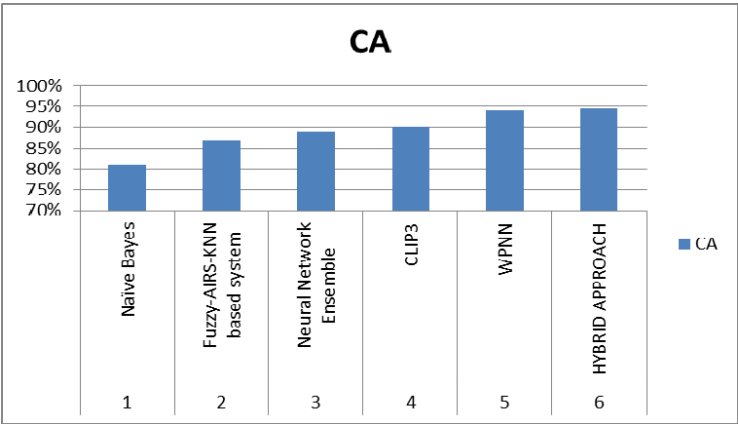


Figure 1. Proposed System

6. Result

Table 1. The results indicate that the best base classifiers

S.NO	METHOD	CA
1	Naïve Bayes	81%
2	Fuzzy AIRS KNN based systems	87%
3	Neural Network Ensemble	89%
4	CLIP 3	90%
5	WPNN	94%
6	Hybrid Approach	94.06%



Graph 1. The results indicate that the best base classifiers

7. Conculsion and Future Scope

This appraisal investigated completely outstanding fuser classifiers and thought of an association between's basic outfit and flavor pack additionally by virtue of the evaluation between mix troupe and base classifiers. The outcomes obtained from completely various of the model show that the sorted out model is also an extra solid framework which can giving extra dependable information. In this paper, we used hybrid approach towards heart disease and using this approach result is improved. This research is used real life problem.

References

[1] Mohammad Shafenoor Amin, Yin Kia Chiam, Kasturi Dewi Varathan .Identification of significant features and datamining techniques in predicting heart disease .ELSEVIER, Telematics and Informatics 36,2019.

[2] T. Vivekanandan, N. Ch Sriman Narayana Iyengar .Optimal feature selection using a modified differential evolution algorithm and its effectiveness for prediction of heart disease.ESLEVIER, computers in Biology and Medicine 90,2017.

- [3] Randa El-Bialy, Mostafa A. Salamay, Omar H. Karam and M.Essam Khalifa.Feature Analysis of Coronary Artery Heart Disease Data Sets.ELSEVIER,Procedia Computer Science65 (2015)459–468.
- [4] Rashmi G Saboji, Prem Kumar Ramesh .A Scalable Solution for Heart Disease Prediction using Classification Mining Technique .IEEE,International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS-2017).
- [5] Sarath Babu, Vivek EM, Famina KP, Fida K, Aswathi P, Shanid M, Hena M .Heart Disease Diagnosis Using Data Mining Technique.IEEE,International Conference on Electronics, Communication and Aerospace Technology ICECA 2017.
- [6] Prabakaran.N and Kannadasan.R .Prediction of Cardiac Disease Based on Patient's Symptoms .IEEE,Proceedings of the 2nd International Conference on Inventive Communication and Computational Technologies (ICICCT 2018)
- [7] Cincy Raju, Philippsy E, Siji Chacko, L Padma Suresh, Deepa Rajan S .A Survey on Predicting Heart Disease using Data Mining Techniques .IEEE,Proc. IEEE Conference on Emerging Devices and Smart Systems (ICEDSS 2018).2-3 March 2018, Mahendra Engineering College, Tamilnadu, India.
- [8] Gaurav Meena,Pradeep Singh Chauhan,Ravi Raj Choudhary .Empirical Study on Classification of Heart Disease Dataset-its Prediction and Mining .IEEE,International Conference on Current Trends in Computer, Electrical, Electronics and Communication (ICCTCEEC-2017). ISBN:978-1-5386-3243-7
- [9] M.A.Jabbar,Shirina samreen .Heart disease prediction system based on hidden naïve bayes classifier .IEEE,2015, ISSN:2278-9359.
- [10] Monika Gandhi,Dr. Shailendra Narayan Singh .Predictions in Heart Disease Using Techniques of Data Mining",IEEE,2015 1st International Conference on Futuristic trend in Computational Analysis and Knowledge Management (ABLAZE-2015), ISBN:978-1-4799-8433-6.
- [11] Salma Banu N.K,Suma Swamy .Prediction of Heart Disease at early stage using Data Mining and Big Data Analytics: A Survey .IEEE,2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECOT), ISBN:978-1-5090-4697-3.
- [12] Minas A. Karaolis, Member, IEEE, Joseph A. Moutiris, Demetra Hadjipanayi, and Constantinos S. Pattichis, Senior Member, IEEE .Assessment of the Risk Factors of Coronary Heart Events Based on Data Mining With Decision Trees",IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE, VOL. 14, NO. 3, MAY 2010, ISSN:1089-7771
- [13] Carlos Ordóñez .Association Rule Discovery With the Train and Test Approach for Heart Disease Prediction .IEEE,IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE, VOL. 10, NO. 2, APRIL 2006, ISSN:1089-7771.
- [14] R.Kavitha,E.Kannan .An Efficient Framework for Heart Disease Classification using Feature Extraction and Feature Selection Technique in Data Mining . International journal of Advanced Research in Computer Science and Software Engineering, Vol4, Issues 1, ISSN:2277i28x, pp 1157-1160, January (2014).
- [15] Anchana Khemphila, Veera Boonjing .Heart disease Classification using Neural Network and Feature Selection .IEEE,2011, 21st International Conference on Systems Engineering, ISBN:978-0-7695-4495-3
- [16] V Krishnaiah, M Srinivas .Diagnosis of Heart Disease Patients Using Fuzzy Classification Technique .vol.107, No. Pt2, pp. 1256-9, 2004.
- [17] Lamia Abed Noor Muhammed .Using Data Mining technique to diagnosis heart disease .machine Learning, Kluwer Academic Publishers, vol. 57, pp. 233-269, 2004
- [18] Theresa Princy. R,J. Thomas .Human Heart Disease Prediction System using Data Mining Techniques.IEEE,2016 International Conference on Circuit, Power and Computing Technologies [ICCPCT], ISBN:978-1-5090-1277-0.
- [19] Marjia Sultana, Afrin Haider and Mohammad Shorif Uddin,"Analysis of Data Mining Techniques for Heart Disease Prediction",IEEE,iCEEICT 2016, ISBN:978-1-5090-2906-8
- [20] Saba Bashir, Zain Sikander Khan, Farhan Hassan Khan, Aitzaz Anjum, Khurram Bashir,"Improving Heart Disease Prediction Using Feature Selection Approaches .IEEE,2019,978-1-5386-7729-2.
- [21] H.E.Khodke . Pattern Matching and analysis of vehicle number plate images for number extraction using template matching and Neural Network As Hybrid Method .IJCESR,2017
- [22] H.E.Khodke . Pattern Matching and analysis of vehicle number plate's Recognition system using template matching International Journal for Research in Applied Science & Engineering Technology,2016.
- [23] H.E.Khodke . Evaluation for Pattern Matching and Analysis of Car's Number Plate Recognition Using Template Matching .international journal,2016.
- [24] 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.