International Conference on Latest Advancements & Future Trends In Engineering, Science & Management
Sambhram Institute of Technology, Bangalore (India)
03rd - 04th May 2019 www.conferenceworld.in ISBN : 978-93-87793-86-6

A review on an efficient disease prediction system to predict the heart disease with better accuracy utilizing different data mining classification techniques and domain knowledge of experts in the field at the lowest cost

DR.Mt Gopalkrishna1
Department of CSE,VTU
Professor
Bengaluru

Megha J2
Department of CSE,VTU
Assistant professor,KSSEM
Bengaluru

Vishruth B G3
Department of ISE,VTU
Assistant professor,SaIT
Bengaluru

Abstract: *A popular saying goes that we are living in an "information age". Terabytes of data are produced every day. Data mining is the process which turns a collection of data into knowledge. The health care industry generates a huge amount of data daily. This huge repository of information contains wealth of knowledge. The hidden patterns and relationships in the data is mostly overlooked. Diagnosing cardio vascular diseases in patients is a difficult task and doctors who can accurately predict such diseases are few in number. However, most of it is not effectively used. Efficient tools to extract knowledge from these databases for clinical detection of diseases or other purposes are not much prevalent. Analysis of these issues at beginning period helps the doctors in treating it at starting stage and to enhance the patient's wellbeing, The aim of this paper is to summarize some of the current research on predicting heart diseases using data mining techniques, analyse the various combinations of mining algorithms used and conclude which techniques are effective and efficient. Also, some future directions on prediction systems have been addressed.*

Keywords: Data Mining, Machine Learning, Decision Tree, Heart Disease.

I. INTRODUCTION

In the modern life style health diseases are increasing tremendously. Our life style had a great impact on our health causing heart diseases and other health problems. Taking a survey of present population it is seen that about sixty percentages are suffering from heart diseases[2]. Early detection of heart diseases can prevent the death rate, people are not aware about the detection of heartdisease earlier due to lack of knowledge. Health care industries are aiming to diagnose the disease at early stages. In most cases it is noticed at the final stages of disease or after death. The cost of treatment for heart disease is very expensive. The treatment cost is not affordable for everyone. Therefore people are reluctant to do proper treatment at early stages of disease.

Hospitals have to reduce the charge of medical tests. They can attain these consequences by employing suitable decision support systems. Health care data is enormous[19]. It consists of patient centric data, resource organization data and altered data. Medical care organizations must have capability to explore data. Treatment

records of millions of patients can be hoarded and data mining techniques will aid in answering numerous essential and decisive questions Interrelated to health care.

Data mining techniques has been performed in healthcare domain. This realization is in the arouse of explosion of difficult medical data. Medicinal data mining can utilize the veiled patterns present in huge medical data which otherwise is left undiscovered[20]. Data mining techniques which are useful to medical data include association rule mining for finding frequent patterns, prediction, classification and clustering. Data mining techniques are more useful in predicting heart diseases, breast cancer lung cancer, diabetes and etc.

By using data mining technique it is possible to detect disease at early stage and can completely cure the disease by proper diagnosis[14]. Health care industry collect huge amount of data, which are not mined to discover hidden information. Remedy of this problem is data mining technique[24]. Data Mining is the way towards extracting interesting patterns and knowledge from huge amount of information. The Data Mining process is a combination of choosing, analyzing, planning, interpreting and evaluating the outcomes[5].

Data mining techniques are:

1. Association
2. Classification
3. Clustering

Association rule mining is method for discovery of interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using some measurements. The Apriori algorithm and MAFIAalgorithm are used for generating association rules. MAFIAalgorithm generate maximal frequent item set before finding all frequent item set and once we find maximal frequent itemset, we can generate all frequent item set in single scan. These are the main advantage of MAFIA algorithm than Apriori algorithm.

Classifiation model use to extract a model describing important classes. Classification techniques are decision treealgorithm and Naïve Bayes algorithm [4]. Decision tree arevery flexible, easy to understand and easy to debug. It takescare of various issues like missing value, outlier andidentifying significant dimensions. Naïve Bayes is supervisedalgorithm.it assume underlying probabilistic model. It is assumption, so loss accuracy and if no occurrence of attributeor class label then probability estimate will be zero. Sodecision tree is better than naïve Bayes.

Clustering is the process of grouping same characteristic data into classes or cluster. K-mean clusteringalgorithm used for clustering. K-mean algorithm is faster than other clustering algorithm and Works great if clusters arespherical. K-means becomes a great solution for pre-clustering; reducing the space into disjoint smaller sub-spaceswhere other clustering algorithms can be applied.

HOW DM IS USEFUL FOR THE MEDICAL FIELD: Due to the vast use of computers in the hospitals by doctors who practice, a large amount of information is gathered. Huge set of data consist of relevant information of the patient along with lot of other information which is the noise. The entire set of data may be used by the practitioners but the data miners have to extract only specific concerned information know as knowledge. Today as the research is emerging in a very rapid pace it is a major requirement to use the technology available to be helpful for the society globally. With the available mining tools it is possible to

design a model which can be helpful for the health care industry. The DM tools can provide us with accurate and time to time report needed for the practitioners so that the patient is benefited.

1.1 DATA MINING ALGORITHMS

Research on data mining has led to the formulation of several data mining algorithms.These algorithms can be directly used on a dataset for creating some models or to draw vital conclusions and inferences from that dataset. Some popular data mining algorithms are Decision tree, Naïve Bayes, k-means, artificial neural network etc. They are discussed in the follows section.

1. Decision Tree

A Decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences including chance event outcomes and utility. It is one of the ways to display an algorithm. Decision trees are commonly used in operations research, specifically in decision analysis to help and identify a strategy that will most likely reach the goal. It is also a popular tool in machine learning. A Decision tree can easily be transformed to a set of rules by mapping from the root node to the leaf nodes one by one. Finally by following these rules, appropriate conclusions can be reached.

2. C4.5

It is a classifier in the form of a Decision tree. It is a supervised learning method which uses information gain and pruning for improved results. It is quite fast, popular and the output is easily interpretable.

3. K-means Algorithm

K-means creates k groups from a set of given objects so that the members of a groupare more similar. Other than specifying the number of clusters, k-means also “learns”the clusters on its own without any information about which cluster a particular observation should belong to. That’s why k-means can be called as semi-supervised learning method. K-means is specially effective over large datasets.

4. ID3 Algorithm

The ID3 algorithm (Quinlan86) is a Decision tree building algorithm which determines the classification of objects by testing the values of the properties. It builds the tree in a top down fashion, starting from a set of objects and the specification of properties. At each node of the tree, a property is tested and the results used to partition the object at that point are set. This process is recursively continued till the set in a given sub tree is homogeneous with respect to the classification criteria. Then it becomes a leaf node. At each node, information gain is maximized and entropy is minimized. In simpler words, that property is tested which divides the candidate set in the most homogeneous subsets.

5. Support Vector Machine(SVM)

It is a supervised learning method which classifies data into two classes over a hyper plane. Support vector machine performs a similar task like C4.5 except that it doesn’t use Decision trees at all. Support vector machine attempts to maximize the margin (distance between the hyper plane and the two closest data points from each respective class) to decrease any chance of misclassification. Some popular implementations of support vector machine are scikit-learn, MATLAB and of LIBSVM.

ISBN : 978-93-87793-86-6

6. Naive Bayes(NB)

It is a simple technique for constructing classifiers. It is a probabilistic classifier based on Bayes' theorem. All Naive Bayes classifiers assume that the value of any particular feature is independent of the value of any other feature, given the class variable. Bayes theorem is given as follows: $P(C|X) = P(X|C) * P(C)/P(X)$,

where X is the data tuple and C is the class such that P(X) is constant for all classes. Though it assumes an unrealistic condition that attribute values are conditionally independent, it performs surprisingly well on large datasets where this condition is assumed and holds.

7. Artificial Neural Network (ANN)

An artificial neural network (ANN) is a computational model based on the structure and functions of biological neural networks. Information which flows through the network affects the structure of the artificial neural network because a neural network changes or learns in a sense-based on input and output, for that particular stage and consequently for each stage. ANN's are considered nonlinear statistical data modelling tools where the complex relationships between inputs and outputs are modelled or patterns are found .ANNs have layers that are interconnected. Artificial neural networks are fairly simple mathematical models to enhance existing data analysis technologies.

8. CART

CART stands for Classification and Regression Trees methodology. In classification trees the target variable is categorical and the tree is used to identify the "class" within which a target variable would likely fall into.In regression trees, the target variable is continuous and a tree is used to predict its value. The CART algorithm is structured as a sequence of questions, the answers to which determine what will be the next question if there should be any questions. The result of these questions look like a tree structure where the ends are terminal nodes which represent that there are no more queries.

9. Random Forest

Random Forests are an ensemble learning method (also thought of as a form of nearest neighbor predictor) for classification and regression techniques. It constructs a number of Decision trees at training time and outputs the class that is the mode of the classes output by individual trees. It also tries to minimize the problems of high variance and high bias by averaging to find a natural balance between the two extremes.. Both R and Python have robust packages to implement this algorithm.

10. Regression

Regression is a statistical concept which is used to determine the weight of relationship between one dependent variable (usually denoted by Y) and a series of other changing variables (known as independent variables). Two basic types ofregression are linear regression and multiple linear regression. Also, there are several non-linear regression methods that are used for more complicated data analysis.

11. J48

J48 is a Decision tree that is an implementation of ID3 (Iterative Dichtomiser 3) developed by the WEKA project team. R language also has a package to implement this. J48 does not require discretization of numeric attributes.

ISBN : 978-93-87793-86-6

12. A-Priori Algorithms

It is an algorithm for frequent item set mining and association rule learning. A-priori uses breadth-first search algorithm and a hash structure to count candidate item sets efficiently. It generates candidate item sets of length k from item sets of length k-1 .Then it prunes the candidates which have an infrequent sub pattern.

13. Fuzzy Logic

It is a form of many-valued logic in which the truth values of variables may be any real number between 0 and 1. Fuzzy logic is applicable in many fields from control theory to artificial intelligence. Fuzzy logic is mainly employed to handle the concept of partial truth where the truth value may range between completely true and completely false. Among various combinations of methodologies in soft computing, fuzzy logic and neuro computing are very practical and popular techniques leading to development of neuro-fuzzy systems.

14. Association Rules

Association rules are basically if/then statements which help us to find out the relationships between apparently unrelated data in an information warehouse. It has two parts, an antecedent (if) and a consequent (then).Association rules are created by analyzing a data set for frequent if/then patterns. Using the criteria support and confidence, it identifies the most important relationships. Support indicates that how frequently the items appear in the database while confidence shows the number of times the if/then statements have been found tobe true. In data mining, association rules are very useful for analyzing and predicting customer behaviour. Programmers use association rules to build programs capableof machine learning.

1.2 DATA MINING TOOLS

Data mining tools provide ready to use implementation of the mining algorithms. Most of them are free open source software's so that researchers can easily use them. They have an easy to use interface. Some of the popular data mining tools are WEKA, RapidMiner, TANAGRA, MATLAB etc. Some of them are discussed as follows.

1. WEKA

It stands for Waikato Environment for Knowledge Learning. It is a computer program that was developed at the University of Waikato in New Zealand for the purpose of identifying information from raw data. WEKA supports different standard data mining tasks such as data pre-processing, classification, clustering, regression, visualization and feature selection. The basic premise of this application is to utilizecomputer application that can be trained to perform machine learning capabilities and derive useful information in the form of trends and patterns. Originally written in C, the WEKA application was then completely rewritten in Java and is now compatible with almost every computing platform. Its user friendly graphical interface allows for quick set up and operation.

2. RAPIDMINER

Formerly called as YALE (Yet Another Learning Environment), is an environment for providing data mining and machine learning procedures including data loading and transformation (ETL), data preprocessing and visualization, modeling, evaluation and deployment. Rapid Miner is written in the Java programming language. Also, it can be used for text mining, multimedia mining, feature engineering, data stream mining etc.

3. TANAGRA

It is a free data mining software designed for academic and research purposes. It proposes several data mining methods such as exploratory data analysis, statistical learning and machine learning. TANAGRA comprises some paradigms and algorithms such as clustering, association rule, parametric and nonparametric statistics, factorial analysis, feature selection and construction algorithms.

4. APACHE MAHOUT

It is a project of the Apache Software Foundation designed for free implementations of distributed or otherwise scalable machine learning algorithms that focus primarily in the areas of collaborative filtering, clustering and classification. Apache Hadoop isanother open source, Java-based programming framework which supports the processing and storage of extremely large data sets in a distributed computing environment. It is a part of the Apache project which is sponsored by the Apache Software Foundation.

5. MATLAB

It is the short form for matrix laboratory. It supports a multi-paradigm numerical computing environment. It is a fourth-generation programming language. MATLAB provides matrix manipulations, plotting of functions and data, algorithm implementations, creation of user interfaces and interfacing with programs written in other languages including C, C++,C#, Java, Fortran and Python [41].

6. Java

Java is a high level programming language developed by Sun Microsystems and now owned by Oracle Inc. It is widely used for developing and delivering content on the web. Java has numerous object oriented programming features much like C++, but is simplified to eliminate language features that cause common programming errors. Java language is well suited for use on the World Wide Web. Java applets (small Java applications) can be downloaded from a web server and run on a computer by a Java-compatible web browser.

7. C

C was developed by Dennis M. Ritchie at Bell Labs for the Unix Operating System in the early 1970s.It was originally intended for writing system softwares. C is a high-level, general-purpose programming language which is ideal for developing firmware and portable applications.

8. Orange

It is a toolkit for data visualization, machine learning and data mining. It is interactive and can be used as a Python library.

II. LITERATURE REVIEW

Over the years, numerous works have been done related to heart disease prediction system using different data mining algorithms by different authors. They tried to achieve efficient methods and accuracies in finding out diseases related to heart by their work including datasets and different algorithms along with the experimental results and future work that can be done on the system to achieve more efficient results. Researchers have been investigating the use of data mining techniques to help practitioners in accurate diagnosis. Works done by various researchers in heart disease diagnosis using different data mining techniques with various performance measures are discussed below:

A. John Peter T et al., (2012) used classification technique for the heart disease prediction. Dataset of heart disease which is in the ARFF format uses 14 attributes which has large quantity of intrinsic linear combination of variables. The limitations of medical scoring systems are handled and the data is classified by using classification models that assign data in a collectio3n of target classes. The classifierswere implemented on the reduced data. Accuracy of the classifiers was 83.70% for Naïve Bayes, 76.66% for decision tree, 75.18% for K-Nearest Neighbour and 78.148% for Neural Network[1].

B. Chaitrali S et al., (2012) used classification methods such as Neural Network, Decision Tree and Naive Bayes for classifying 13 common attributes like age, gender, blood pressure, cholesterol, etc. for the prediction of heart disease. Two more attributes called smoking and obesity were also added. Confusion matrix was obtained for 3 classification methods using 13 attribute datasets and 15 attribute datasets.Accuracy obtained by these techniques were 100%, 99.62% and 90.74% respectively[2].

E. Hlaudi Daniel Masethe et al., (2014) used data mining algorithms like J48, NB and REP TREE for predicting heart attacks. The medical database was collected from the doctors in South Africa. The various attributes considered were Gender, age, CPT, ECG, RBP, Thalach, serum cholesterol, alcohol, obesity(diet) and smoking. WEKA - Waikato Environment for Knowledge Analysis tool was used for discovering, analysing and predicting patterns for heart disease. The accuracy obtained were 99.0741, 99.222, 98.148 for J48, REPTREE and NB respectively[4].

F. Venkatalakshmi B et al., (2014) aims to predict heart disease using predictive mining. 13 factors from UCI Repository is taken as the source data, to compare the performance of DT and NB where the accuracies such as 84.01% and 85.03% was obtained respectively[6].

Lokanath Sarangi et al. (2015) designed a cost efficient model by using Genetic Algorithm optimizer technique. The weights were optimized and fed as an input to the given network. The accuracy achieved was 90% by using the hybrid technique of GA and neural networks[7]

Kalaiselvi C (2016) proposed a new website called average K-nearest neighbour algorithm which is to improve classification accuracy and efficiency. The advantage in AKNN is that grouping the samples based on super classes reduces the number of samples used for training, thus making the KNN the faster algorithm. It uses 13 attributes and the accuracy obtained is 96.5% and with 12 attributes the accuracy obtained is 97%. This works well if the data are well segregated and does not work if the data are noisy [8].

VivekanandanTet al., (2017) proposed the challenging tasks of selecting critical features from the enormous set of available features and diagnosing heart disease. DE (Modified Differential Evolution) algorithm is used to perform feature selection. Prediction of heart disease was carried out using Fussy AHP and Feed-Forward neural network. Using 9 attributes an accuracy of 98% was achieved[11].

K. Sharmila S et al., (2017) proposed a method to improve Naïve Bayes performance. It takes only two values for the prediction. It classifies the data into two classes 0-Absent and 1-Present. The proposed algorithm is used for identification of the values ranging from 0 to 1. It uses 14 attributes available in UCI machine repository which contains 303 records. It takes only two values for the prediction with accuracy of 97%[12].

Chaitanya Suvarna Abhishek Sali Sakina Salmani focused on developing prediction algorithm with the help of data mining and prediction algorithm using the Particle Swarm Optimization technique which is an inherently distributed algorithm where the solution for a problem emerges from the interactions between many simple individual agents called particles.The data source we have used for experimental testing are commonly used and considered as a de facto standard for heart disease prediction reliability ranking. We will also be using a slightly modified version of PSO with constriction factor called Constricted PSO.The results obtained show that Particle Swarm Data Mining Algorithms are competitive, not only with other evolutionary techniques, but also with industry standard algorithms, and can be successfully applied to heart disease prediction. [13].

Cincy Raju, Philipsy E, Siji Chacko, L Padma Suresh, Deepa Rajan S proposed a method to improve accuracy by considering 13 factors . Various steps are taken to apply pertinent techniques in the disease prediction. Support vector machines, Association Rule, K- nearest neighbor classification are used to diagnosis the heart diseases. Among these algorithms Support Vector Machine (SVM) gives best result Support Vector Machine (SVM) technique is an efficient method for predicting heart disease[15].

Bandarage Shehani Sanketha Rathnayakc ; Gamage Upeksha Ganegoda a survey about different data mining and neural network classification technologies used in predicting the risk of occurring heart diseases based on risk factors. The risk level of a person is classifying using techniques like K-Nearest Neighbor Algorithm, Decision Trees, Genetic algorithm, Naïve Bayes etc. and the accuracy is high when using more attributes and combinations of above techniques.sujjested to use the combination of algorithms in order to achieve high accuracy instead of one particular problem and one subject area[16].

III. OBJECTIVES OF THE RESEARCH WORK

The main objectives of this research work are:

1. The main objective is to identify the key patterns and features from the medical data of the patient by combining data mining techniques to predict the heart disease before it causes to help the medical practitioners.
2. To reduce the data sets and increase the accuracy of prediction model.
3. To use identical set of input data for measuring the performance if it is used for comparison.
4. To use a combination of algorithms to gain highest accuracy.

IV. METHODOLOGY OF THE RESEARCH WORK

From the literature survey so many authors have discussed many algorithms such as Decission Tree, Id3 algorithm, Support Vector Machine, Navie Bias, Artificial Neural Network, Random Forest, Regression , Fuzzy Logic and Association Rules. One of the major drawbacks of these works is that the main focus has been on the application of classification techniques for heart disease prediction, rather than studying various data cleaning and pruning techniques that prepare and make a dataset suitable for mining. It has been observed that a properly

cleaned and pruned dataset provides much better accuracy than an unclean one with missing values. Selection of suitable techniques for data cleaning along with proper classification algorithms will lead to the development of prediction systems that give enhanced accuracy.

V. CONCLUSION

Data mining algorithms can be effectively used to 'mine' relevant information from the huge amounts of data generated by the healthcare industry. These works show that rather than applying a single mining technique on a data set, results are far better if a collection of mining techniques are used. Java is chosen in most of the research work for practical execution of the project. WEKA, Tanagra, Matlab etc. are some of the other popular tools used for data analysis. Careful selection of the combination of mining techniques and accurate implementation of those techniques on the data set yields a fast and effective implementation of a system for heart disease management. The required dataset is divided into two parts, one is used for mining and the smaller partition is used for verifying. Most of the time, 10 fold cross validation technique is used. Some of the works are about the comparison of different classification techniques on a dataset to correctly classify if a given patient has any probability of a heart disease or not. Others have worked on 'mining' the causes that lead to heart diseases from a given dataset. Commonly used classification techniques are Decision tree, Naïve Bayes, artificial neural network, association rule mining and fuzzy logic.

VI. REFERENCES

[1] An empirical study on prediction of heart disease using classification data mining techniques, Peter, T.J, . Somasundaram, K., 2012 International Conference on Advances in Engineering, Science and Management (ICAESM).

[2] Chaitrali S. Dangare Sulabha S Apte, "Improve study of Heart Disease prediction system using Data Mining Classification techniques", International journal of computer application, 2012.

[3] Syed Umar Amin1, Kavita Agarwal2, Dr. Rizwan Beg ,"Genetic Neural Network based Data mining in prediction of Heart disease using risk factors", Proceedings of 2013 IEEE Conference on Information and Communication Technologies (ICT 2013).

[4] Shamsher Bahadur Patel, Pramod Kumar Yadav and Dr. D.P. Shukla, "Predict the Diagnosis of Heart Disease Patients using classification Mining Techniques", IOSR Journal of Agriculture and Veterinary Science (IOSR-JAVS), 2013.

[5] K.Sudhakar, and Dr. M. Manimekalai, January 2014, "Study of Heart Disease Prediction using Data Mining", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 4, Issue 1,pp. 1157-1160 .

[6] Hlaudi Daniel Masethe, Mosima Anna Masethe, "Prediction of Heart Disease using classification Algorithm", Proceedings of the World Congress on Engineering & Computer Science 2014, WCECS.

[7] Rishi Dubey, and Santosh Chandrakar, Aug. 2015,"Review on Hybrid Data Mining Techniques for The Diagnosis of Heart Diseases in Medical Ground" ,Vol. 5, Issue 8, ISSN: 2249-555X, pp. 715-718.

[8] Sairabi H. Mujawar, and P. R. Devale, October 2015,"Prediction of Heart Disease using Modified k-means and by using Naive Bayes", International Journal of Innovative Research in Computer and Communication Engineering(An ISO 3297: 2007 Certified Organization) Vol. 3, Issue 10, pp. 10265-10273.

[9] Sumitra Sangwan, and Tazeem Ahmad Khan, Mar. 2015,"Review Paper Automatic Console for Disease Prediction using Integrated Module of A-priori and k-mean through ECG Signal",International Journal For Technological Research In Engineering, Vol. 2, Issue 7, ISSN(Online): 2347-4718, pp. 1368-1372.

[10] Lokanath Sarangi, Mihir Narayan Mohanty, Srikanta Pattnaik, "An Intelligent Decision Support System for Cardiac Disease Detection", IJCTA, International Press 2015.

[11] G. Purusothaman, and P. Krishnakumari, June 2015,"A Survey of Data Mining Techniques on Risk Prediction: Heart Disease", Indian Journal of Science and Technology, Vol. 8(12), DOI:10.17485/ijst/2015/v8i12/58385, pp. 1-5.

[12] Gunsai Pooja Dineshgar, and Mrs. Lolita Singh, February 2016, "A Review on Data Mining for Heart Disease Prediction",International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE), Vol. 5, Issue 2, pp. 462-466.

[13] B.Venkatalakshmi, M.V Shivsankar, Heart Disease Diagnosis Using Predictive Data mining; International Journal of Innovative Research in Science, Engineering and Technology Volume 3, Special Issue 3, March 2016

[14] Ashwini Shetty A, and Chandra Naik, May 2016,"Different Data Mining Approaches for Predicting Heart Disease", International Journal of Innovative Research in Science,Engineering and Technology(An ISO 3297: 2007 Certified Organization), Vol. 5, Special Issue 9, pp. 277-281.

[15] Serdar AYDIN, Meysam Ahanpanjeh,and Sogol Mohabbatiyan,February 2016, "Comparison And Evaluation of Data Mining Techniques in the Diagnosis of Heart Disease",International Journal on Computational Science & Applications (IJCSA), Vol. 6,No.1, pp. 1-15

[16] C.Kalaiselvi,,"Diagonising of heart disease using k-neighbour algorithm of data mining" , International Conference on Computing for Sustainable Global Development (INDIACom) 2016.

[17] S.Radhimeenakshi "Classification and prediction of heart disease risk using data mining techniques of support vector machine and artificial neural network", International Conference on Computing for Sustainable Global Development (INDIACom) 2016.

[18] M.A.Jabbar ,Shirina samreen," Heart disease prediction system based on hidden naïve base classifier ", Page(s):1 – 5 ,2016.

[19] Manpreet Singh, Levi Monteiro Martins, Patrick Joanis, and Vijay K. Mago,2016, "Building a Cardiovascular Disease Predictive Model using Structural Equation Model & Fuzzy Cognitive Map",IEEE International Conference on Fuzzy Systems (FUZZ),pp. 1377-1382.

[20] Prajakta Ghadge,Vrushali Girme, Kajal Kokane, and Prajakta Deshmukh, 2016,"Intelligent Heart Attack Prediction System Using Big Data", International Journal of Recent Research in Mathematics Computer Science and Information Technology,Vol. 2, Issue 2, pp.73-77, October 2015–March.

[21] Vivekanandan T et al., "Optimal feature selection using a modified differential evolution algorithm and its effectiveness in human heart disease prediction", www.elsevier.com/locate/compbiomedhttps://doi.org/10.1016/j.compbiomed, Pages: 125-136 (2017).

[22] Sharmila S et al., "Analysis of Heart Disease Prediction using Data Mining Techniques", International Journal of Advanced Networking & Applications (IJANA), Volume: 08, Issue: 05,Pages: 93-95 (2017).

[23] Chaitanya Suvarna Abhishek Sali Sakina Salmani "Efficient heart disease prediction system using optimization technique", Proceedings of the IEEE 2017 International Conference on Computing Methodologies and Communication (ICCMC).

[24] M. Deepika Dr. K. Kalaiselvi "empirical study on Disease Diagnosis using Data Mining Techniques", sep 30 2018.

[25] Cincy Raju, Philipsy E, Siji Chacko, L Padma Suresh, Deepa Rajan S ,"A Survey on Predicting Heart Disease using Data Mining Techniques", Proc. IEEE Conference on Emerging Devices and Smart Systems (ICEDSS 2018) 2-3 March 2018.

[26] Bandarage Shehani Sanketha Rathnayakc ; Gamage Upeksha Ganegoda "Heart diseases prediction with Data Mining and Neural Network Techniques", 018 3rd International Conference for Convergence in Technology (I2CT) The Gateway Hotel, XION Complex, Wakad Road, Pune, India. Apr 06-08, 2018.